

Article

UAV Remote Sensing for Drought-Adaptive Sesame Breeding: Flight-Altitude Benchmarking, Predictive Modelling, and Composite Stress Tolerance Indexing

Christos Petsoulas ¹, Alexandros Tsitouras ^{1,2} , Eleftherios Evangelou ¹ , Anastasia Kargiotidou ¹ ,
Chrysanthi I. Pankou ¹  and Dimitrios N. Vlachostergios ^{1,*} 

¹ Institute of Industrial and Forage Crops, Hellenic Agricultural Organization-Dimitra, 41335 Larissa, Greece; petsoulaschristos@elgo.gr (C.P.); tsitalex@elgo.gr (A.T.); evangelou@elgo.gr (E.E.); kargiotidou@elgo.gr (A.K.); cpankou@elgo.gr (C.I.P.)

² Laboratory of Precision Agriculture, Department of Agrotechnology, University of Thessaly, Gaiopolis, 41110 Larissa, Greece

* Correspondence: vlachostergios@elgo.gr

Highlights

What are the main findings?

- UAV flights at 40 m AGL best matched ground plant height ($R^2 \approx 0.83$ under leave-one-genotype-out cross-validation; $R^2 \approx 0.55$ under the stricter leave-one-date-out scheme—the figure relevant to a new phenological stage).
- Spectral-index repeatability was up to ~2-fold higher under water stress; a UAV composite Stress Tolerance Index captured 10 of the breeder's 48 selections in its top 30 (4.1-fold enrichment; $p < 10^{-4}$).

What are the implications of the main findings?

- Forty m AGL should be the operational standard for canopy height phenotyping in sesame breeding, especially under drought where sparse canopies amplify altitude sensitivity.
- The UAV composite STI provides a scalable, non-destructive framework for drought ranking in large unreplicated mutagenesis populations, with meaningful concordance to breeder selection.

Abstract

Early-generation sesame (*Sesamum indicum* L.) breeding requires high-throughput phenotyping of large unreplicated populations across contrasting environments. A DJI Phantom 4 Multispectral UAV was flown at 40, 80, and 120 m above ground level (AGL) over 588 M₂ genotypes under full irrigation (ENV1) and terminal drought (ENV2; irrigation withheld from reproductive onset) on four dates (July–September 2025). Structure-from-motion canopy height models were compared with ground measurements, and four spectral reflectance indices—Normalised Difference Vegetation Index (NDVI), Normalised Difference Red Edge (NDRE), Green Normalised Difference Vegetation Index (GNDVI), and Leaf Chlorophyll Index (LCI)—were derived from 40 m imagery. Ordinary least squares (OLS), Random Forest, and Gradient Boosting were evaluated under leave-one-genotype-out (LOGO), leave-one-environment-out (LOEO), and leave-one-date-out (LODO) cross-validation; genotypic repeatability was quantified by intraclass correlation (ICC), and drought performance was ranked by a composite Stress Tolerance Index (STI) validated against an independent breeder assessment. The 40 m altitude gave the highest height accuracy ($R^2 = 0.812$ in ENV1; 0.663 in ENV2). LOGO accuracy ($R^2 \approx 0.83$) fell to $R^2 \approx 0.55$



Academic Editors: Michal Aibin and Ryan Rad

Received: 29 April 2026

Revised: 9 June 2026

Accepted: 18 June 2026

Published: 4 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.

under LODO—the operationally relevant figure for a new phenological stage—and the full structural–spectral OLS model collapsed ($R^2 = -0.203$) where tree ensembles remained stable. Spectral-index repeatability was up to ~2-fold higher under stress ($\text{ICC}(3,4) > 0.84$). The composite STI flagged 38 elite genotypes (7.6% of 498); 10 of its top 30 were confirmed in the breeder’s 48-best selection from all 588 rows—a 4.1-fold enrichment over chance (hypergeometric $p = 4.5 \times 10^{-5}$).

Keywords: sesame; UAV remote sensing; high-throughput phenotyping; spectral reflectance indices; canopy height; drought tolerance; stress tolerance index; mutagenesis breeding

1. Introduction

Sesame (*Sesamum indicum* L.; family Pedaliaceae) is one of the earliest domesticated oilseed crops, cultivated for more than five millennia [1]. Its seeds contain 45–63% oil with a favourable fatty acid profile and high concentrations of antioxidant lignans, conferring nutritional and functional value [2–6]. Global production reached ~6.74 million tonnes in 2022, with about 79% concentrated in ten countries led by Sudan and India [7].

Greece is the highest per capita sesame consumer in Europe—the crop being central to tahini, halva, and pasteli—yet European production is negligible and the continent depends almost entirely on imports [8]. Scientific work on sesame in Greece is limited [9,10], and the narrow genetic base of available Greek germplasm constrains selection, making the introduction of new genetic variation a prerequisite for breeding progress.

Sesame breeding globally is constrained by the absence of high-yielding, locally adapted varieties, susceptibility to capsule shattering, biotic and abiotic stresses, and limited modern pre- and post-harvest technologies [11]. Capsule shattering alone causes seed losses of up to 50% at harvest, rendering the crop unsuitable for mechanised production [12]. Key breeding targets—determinate growth, high harvest index, non-shattering capsules, disease resistance—are rarely combined in cultivated material [13,14], and susceptibility to drought at the capsule filling stage further constrains yield stability in Mediterranean environments [15,16].

Mutagenesis broadens the genetic base of self-pollinating crops with narrow diversity. In sesame, gamma irradiation introduces novel allelic variation for capsule indehiscence, plant architecture, and stress tolerance, generating broad phenotypic variation in M_2 populations [17,18]; adequate coverage of the mutation spectrum requires hundreds to thousands of progeny rows [17], rendering ground-based phenotyping across environments and dates operationally impractical [19,20]. Direct yield assessment at M_2 is further precluded by the absence of within-genotype replication, sub-commercial sowing densities, and the low broad-sense heritability of seed yield in sesame ($H^2 \approx 14\%$), contrasting with $H^2 = 62\text{--}72\%$ reported for plant height and phenology [21,22]. Indirect selection through higher-heritability secondary traits correlated with yield—canopy height, vigour, spectral indices—offers greater expected response per cycle [9,23], motivating UAV-based remote sensing as a high-throughput phenotyping strategy for early-generation populations.

UAV canopy height modelling from structure-from-motion (SfM) photogrammetry of overlapping RGB or multispectral imagery has been validated against ground reference in barley [24], wheat [25,26], and sorghum [27,28], with sufficient precision to support genotypic selection. Flight altitude is a critical determinant of accuracy: lower altitudes yield finer ground sampling and denser point clouds, improving digital surface model fidelity, while higher altitudes reduce resolution and introduce sub-canopy mixing artefacts [27,29]. In sesame, UAV remote sensing has been applied to spectral

screening [9], hyperspectral yield estimation [30], and temporal phenotyping [31], but systematic benchmarking of flight altitude on canopy height accuracy under contrasting water-availability has not been reported. As plant height shows substantially higher heritability than seed yield in sesame [21] and correlates with biomass and stress response, accurate UAV-based estimation across environments and growth stages is directly relevant to early-generation screening.

Compact integrated multispectral UAV platforms—combining an RGB camera with narrow-band sensors spanning the blue, green, red, red-edge, and near-infrared regions—enable concurrent acquisition of canopy structural and spectral data in a single flight [32–34], a dual capability that supports repeated assessment across growth stages and environments in breeding applications [31].

Multispectral indices sensitive to chlorophyll concentration, photosynthetic activity, and canopy water status—Normalised Difference Vegetation Index (NDVI) [35], Green Normalised Difference Vegetation Index (GNDVI) [36], Normalised Difference Red Edge (NDRE) [37], and Leaf Chlorophyll Index (LCI) [38]—track canopy senescence dynamics, respond to drought-induced defoliation, and discriminate genotypic variation [39–41]. Their utility in breeding derives from sensitivity to chlorophyll degradation and greenness loss under stress, physiologically linked to remobilisation efficiency and terminal drought tolerance [42,43]. Petsoulas et al. [9] showed these indices effectively discriminate genetic variation in canopy spectral properties in sesame, with recent work extending to hyperspectral yield estimation [30] and multi-sensor precision phenotyping [31].

Translating UAV canopy metrics into ground-equivalent height requires calibration against direct field measurement, typically by ordinary least squares regression, with performance varying by altitude, growth stage, and irrigation [26,27,29]. Combining canopy height with spectral reflectance indices as joint predictors improves accuracy by capturing complementary vertical and foliar dimensions [44,45]. A frequently overlooked requirement for operational deployment is model transferability: calibrations that perform well under leave-one-genotype-out cross-validation but fail across contrasting irrigation regimes or phenologically distinct stages offer limited utility for early-generation selection [31,41,46].

Selection for drought tolerance in multi-environment evaluations commonly employs the Stress Tolerance Index (STI) of Fernandez [47], which quantifies joint performance under stress and non-stress conditions relative to the population mean—identifying accessions combining high productivity with resilience rather than those merely tolerating stress through yield suppression. Related indices [48,49] have been applied across species and traits to rank germplasm under contrasting water regimes [50,51], including in sesame drought evaluation [15,16].

Despite the expanding application of UAV remote sensing in field crop phenotyping, systematic evaluation within a mutagenesis-based sesame breeding programme—spanning multiple altitudes, contrasting irrigation environments, repeated seasonal campaigns, and a large unreplicated M₂ population—has not been reported. The present study differs from prior sesame UAV work—spectral screening [9], hyperspectral yield estimation [30], genomic-temporal phenotyping [31] in adapted germplasm—by jointly benchmarking flight altitude, testing model transferability across environments and phenological stages, and validating a composite drought index against breeder selection in a segregating mutagenesis population, consistent with the recent emphasis on rigorous, deployment-oriented phenotyping pipelines [52,53]. Four specific objectives were addressed: (i) to benchmark UAV canopy height estimation across three altitudes (40, 80, 120 m AGL) against ground-measured height in two irrigation environments (ENV1: full irrigation; ENV2: irrigation withheld from reproductive onset, imposing a 135 mm cumulative deficit) across four campaigns; (ii) to evaluate four spectral reflectance indices (NDVI, NDRE, GNDVI, LCI) as

additional predictors of ground height under leave-one-genotype-out (LOGO), leave-one-environment-out (LOEO), and leave-one-date-out (LODO) cross-validation, comparing OLS with Random Forest and Gradient Boosting; (iii) to quantify within-season genotypic repeatability of spectral indices via intraclass correlation coefficients across the two irrigation regimes; and (iv) to rank the 588 M₂ genotypes for drought performance using a composite Stress Tolerance Index of five UAV-derived traits, validated against an independent breeder visual assessment by a hypergeometric concordance test.

2. Materials and Methods

2.1. Plant Material, Experimental Design and Conditions

The 588 entries originated from 24 M₂ families derived from γ -irradiation of 24 selected sesame lines [9]. Each family contributed a variable number of progeny rows (denoted family–progeny, e.g., 1-1, 1-2, . . . , 22-50) plus one control row representing the unmutagenised parent line (1C through 24C) for within-trial comparison. The identical layout—genotype allocation, row order, and spatial arrangement—was replicated across both environments, so that each genotype was grown once under full irrigation and once under water deficit. This corresponds to an unreplicated single-row design, with progeny rows evaluated once per environment.

Field trials were established during the 2025 growing season at the Institute of Industrial and Forage Crops (IIFC) of the Hellenic Agricultural Organisation—DIMITRA (ELGO-DIMITRA) in Larissa, central Greece (Figure 1). Sowing was carried out on 22 May 2025; the 588 M₂ progeny rows were arranged as two spatially separated field blocks corresponding to ENV1 (normal irrigation) and ENV2 (water stress), each comprising 21 rows $\times$ 28 plots = 588 single-row plots. Each plot was 1 m long (one row per plot), with 1.5 m between adjacent rows. Each block occupied approximately 882 m², separated by a 9 m buffer strip to prevent cross-irrigation. Total trial area, including buffer, was approximately 0.20 ha.

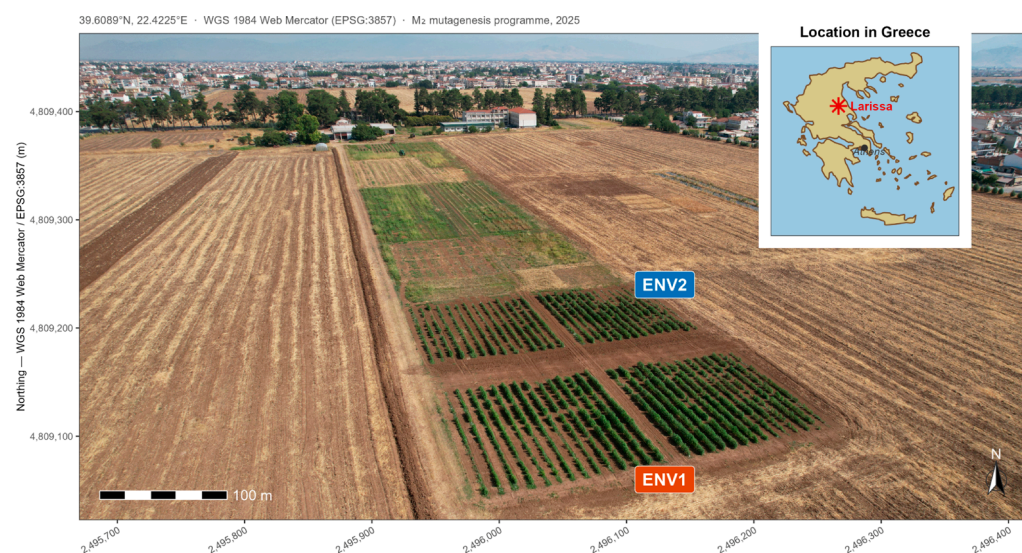


Figure 1. Location of the experimental site at IIFC/ELGO-DIMITRA, Larissa, central Greece, and ground-level photograph of the sesame M₂ trial showing the two adjacent field blocks (ENV1, full irrigation; ENV2, water stress) during the 2025 growing season. Inset map: red marker denotes the trial site. Meteorological conditions and the water application regime during the trial period (22 May–30 September 2025) are presented in Figure 2. Irrigation to ENV2 was terminated on 7 July 2025 (46 days after sowing), coinciding with the onset of flowering in sesame under central-

Mediterranean conditions [54]. ENV2 received no further water input—neither supplementary irrigation nor substantial rainfall—for the subsequent ~85 days, spanning flowering, capsule filling, and maturation. Total precipitation over the trial period was 62.2 mm, of which the largest single event (19.6 mm on 31 July) was the only rainfall exceeding 10 mm during the imposed-stress phase. The cumulative water deficit in ENV2 relative to ENV1 reached 135 mm by the end of the season. Complete irrigation cessation at reproductive onset, low summer precipitation, and the high thermal regime of the central Greek plain during July–September together imposed severe terminal drought stress on a crop sensitive to water deficit at the capsule filling stage [55,56].

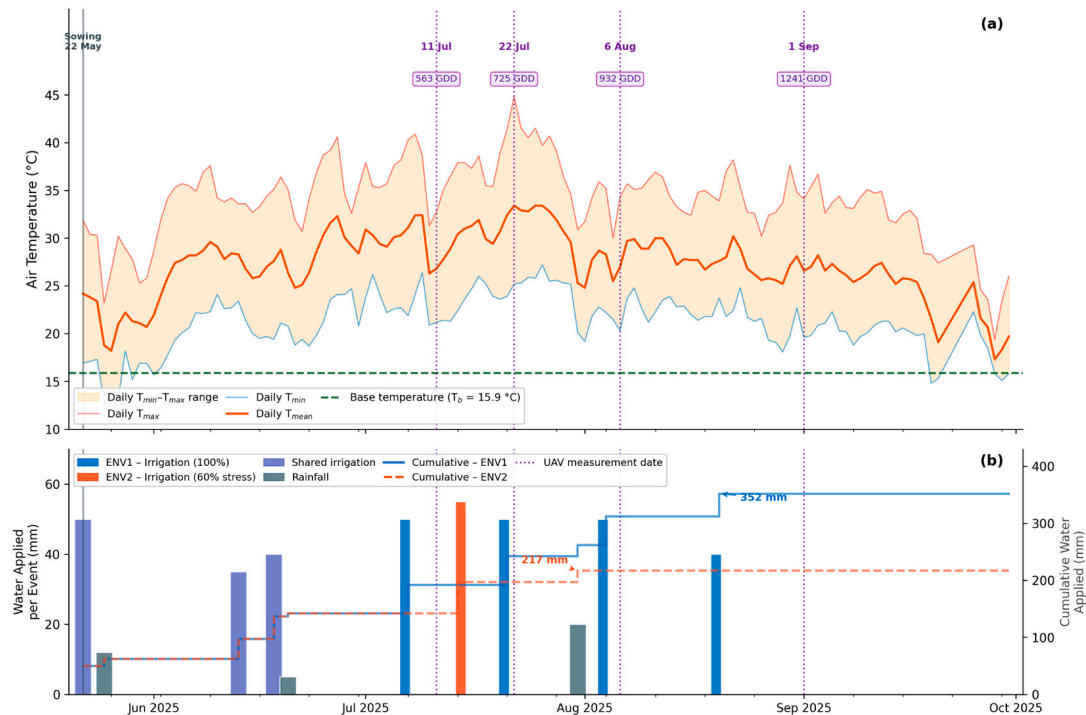


Figure 2. Meteorological conditions and water input during the sesame M₂ trial (Larissa, 22 May–30 September 2025). (a) Daily maximum, minimum, and mean air temperatures; shaded band, daily Tmax–Tmin range; dashed line, sesame base temperature ($T_b = 15.9\text{ }^{\circ}\text{C}$); vertical dotted lines, the four UAV campaigns (11 July, 22 July, 6 August, 1 September) with cumulative growing degree days (GDD) annotated. (b) Daily water input events (mm): ENV1-exclusive irrigation (blue); ENV2 events (orange)—identical to ENV1 until 7 July (Days After Sowing (DAS) 46), after which irrigation was withheld for the remainder of the season, yielding a 135 mm cumulative deficit relative to ENV1; shared irrigation events (indigo); natural precipitation (grey). Step lines (right axis), cumulative seasonal water totals for ENV1 (solid) and ENV2 (dashed).

2.2. UAV Data Acquisition and Plant Height Estimation

Ground plant height was measured manually on four dates—11 July (days after sowing, DAS 40; growing degree days, GDD 563), 22 July (DAS 51; GDD 725), 6 August (DAS 66; GDD 932), and 1 September (DAS 92; GDD 1241)—using a graduated rule on the tallest fully expanded apex of each row, recorded to the nearest centimetre. GDD was calculated following Petsoulas et al. [9].

UAV data were acquired on the same four dates under clear, calm conditions with stable illumination and high solar elevation angles to minimise shadow effects. A bare-soil reference flight was conducted at DAS 0, before canopy emergence, using the same platform, flight plan, and processing pipeline as the canopy flights. All missions were planned perpendicular to the sowing lines using DJI GS Pro (version 2.0.18 (10674); DJI, Shenzhen, China) to reduce anisotropic row-orientation effects and improve photogrammetric reconstruction, at three altitudes above ground level (40, 80, and 120 m AGL) with

80% forward and side overlap and a flight speed of 5 m s^{-1} . Due to strong winds during the DAS 92 flight at 40 m, the corresponding DSM was excluded from height analyses; multispectral imagery from that date was retained.

The UAV platform was the DJI Phantom 4 Multispectral (DJI, Shenzhen, China), carrying one RGB camera and five multispectral sensors (blue $450 \pm 16 \text{ nm}$, green $560 \pm 16 \text{ nm}$, red $650 \pm 16 \text{ nm}$, red-edge $730 \pm 16 \text{ nm}$, near-infrared $840 \pm 26 \text{ nm}$). Ground sampling distances at 40, 80, and 120 m AGL were 2.1, 4.2, and 6.4 cm pixel^{-1} , respectively. Positional accuracy was ensured by the onboard real-time kinematic (RTK) receiver, with real-time Radio Technical Commission for Maritime Services (RTCM) corrections delivered via Networked Transport of RTCM via Internet Protocol (NTRIP) from the JGC-net network, providing centimetre-level horizontal and vertical georeferencing across all missions. Image processing—including photogrammetric reconstruction, RGB and multispectral orthomosaics, vegetation indices, and DSM extraction—was performed in DJI Terra Agriculture v5.1.1 [57] (SZ DJI Technology Co., Ltd., Shenzhen, China).

Plot-level canopy height was derived by canopy height model (CHM) differencing between each canopy-date DSM and the DAS 0 bare-soil reference DSM in ArcGIS Pro (version 3.5.3; Esri, Redlands, CA, USA) [58,59]. A grid-based polygon layer was overlaid on the trial layout, with each cell corresponding to one genotype. Ground elevation per plot was extracted from the bare-soil DSM as the cell-wise mean, providing a stable baseline less sensitive to micro-topographic variability [58]. Canopy elevation per plot was extracted from each canopy-date DSM as the cell-wise 95th percentile (P95), which approximates the upper canopy boundary more reliably than the mean—which underestimates canopy height—or the maximum—which is sensitive to reconstruction outliers, particularly at 120 m AGL [60,61]. Plant height per plot was computed as P95 canopy elevation – mean ground elevation using the Calculate Field tool in ArcGIS Pro.

2.3. Spectral Reflectance Indices

Spectral reflectance data were extracted exclusively from 40 m AGL flights. For each progeny row, the cell-wise mean reflectance was computed from the multispectral orthomosaic and combined into four spectral reflectance indices (SRIs): NDVI, NDRE, LCI, and GNDVI:

$$\text{NDVI} = (R_{\text{NIR}} - R_{\text{RED}}) / (R_{\text{NIR}} + R_{\text{RED}}) \quad (1)$$

$$\text{NDRE} = (R_{\text{NIR}} - R_{\text{RED EDGE}}) / (R_{\text{NIR}} + R_{\text{RED EDGE}}) \quad (2)$$

$$\text{LCI} = (R_{\text{NIR}} - R_{\text{RED EDGE}}) / (R_{\text{NIR}} + R_{\text{RED}}) \quad (3)$$

$$\text{GNDVI} = (R_{\text{NIR}} - R_{\text{GREEN}}) / (R_{\text{NIR}} + R_{\text{GREEN}}) \quad (4)$$

where R_{λ} denotes the mean plot-level reflectance in the spectral band λ .

2.4. Visual Field Evaluation for Drought Ranking

At physiological maturity, an independent visual assessment was conducted exclusively in ENV2 by the breeder, who walked all 588 M_2 progeny rows and selected the 48 top-performing genotypes as the most promising candidates for drought tolerance. These 48 genotypes were subsequently scored on two independently rated traits: drought stress tolerance and agronomic productivity potential, each on a discrete 1–8 scale (8 = highest). Stress tolerance scores reflected canopy condition, leaf wilting, senescence progression, and capacity to maintain green biomass under water deficit; productivity scores captured stem vigour, branching, capsule load, and overall plant architecture at maturity. The maturity class of each genotype (early, medium, late) was recorded relative to the phenological standard of the respective family control row. All evaluations were

conducted in a single pass on the same day, with the breeder blinded to UAV composite ranking results to ensure independence between the two selection procedures.

2.5. Statistical Analyses

2.5.1. Descriptive Statistics and Correlation

Descriptive statistics (n , mean, standard deviation, minimum, maximum, median, interquartile range, skewness, excess kurtosis) were computed per trait, environment, and measurement date. Pearson product-moment correlations were computed between ground height, UAV canopy height at the three altitudes, and the four SRIs, pooled across environments and dates.

2.5.2. UAV-Derived vs. Ground-Measured Plant Height

Agreement between UAV-derived and ground-measured plant height was assessed by OLS linear regression (UAV height as dependent variable) performed separately for each flight altitude and irrigation environment, pooling observations across all measurement dates. Model fit was characterised by Pearson r , R^2 , and RMSE (m). Regressions were additionally computed by measurement date to examine temporal stability.

2.5.3. Predictive Modelling with Multi-SRI Data

Ground-measured plant height was predicted from UAV canopy height and SRIs using three algorithms: ordinary least squares (OLS), Random Forest (RF), and Gradient Boosting (GB). Two predictor sets were compared: (i) UAV canopy height at 40 m AGL only; (ii) the full structural-spectral set comprising canopy heights at all three altitudes jointly with the four SRIs. Irrigation environment was retained as a categorical covariate throughout; measurement date was handled differently under each cross-validation scheme (CVS), as described below.

Analyses used plot-level (cell-wise) aggregates: the P95 canopy height per altitude and the mean reflectance per band, from which the four SRIs were computed (Equations (1)–(4)). Each genotype contributed one record per environment and date. The 40 m DSM from the wind-affected DAS 92 flight was excluded, leaving three usable 40 m height dates; records with any missing predictor were dropped, and the composite STI used the 498 genotypes with complete data across all five traits and both environments. The predictor matrix comprised the three altitude canopy heights, the four 40 m SRIs, and one-hot-encoded environment and date covariates; no further derived or engineered features were constructed. Continuous predictors were standardised (zero mean, unit variance) for OLS, with scaling parameters estimated on each training fold only to avoid leakage; the tree ensembles were trained on the raw feature scale, being invariant to monotone feature transformations.

Three CVSs of progressively different generalisation targets were applied. Under LOGO (leave-one-genotype-out), all observations of one genotype were held out per fold; measurement date and irrigation environment were retained as categorical covariates since both levels are represented in training and test partitions. Under LOEO (leave-one-environment-out), the model was fitted on one irrigation environment and applied to the other, iterated in both directions, with measurement date retained as a covariate. Under LODO (leave-one-date-out), all observations from one measurement date were held out and the model trained on the remaining dates (three usable 40 m AGL dates: 11 July, 22 July, 6 August); measurement date could not be used as a training covariate since the held-out date's level is unobserved, enforcing strict temporal generalisation. LOGO, LOEO, and LODO therefore test three complementary generalisation tasks: to a new genotype at a known date and environment, to a new irrigation environment, and to a new phenological stage, respectively.

RF and GB hyperparameters were selected by nested cross-validation to prevent information leakage from the test partition. For LOGO, the outer loop was 5-fold GroupKFold by genotype; for LOEO and LODO, the outer fold was defined by the held-out environment or date, respectively. The inner loop was 3-fold GroupKFold on training genotypes, minimising RMSE. The RF grid covered $\text{max_depth} \in \{\text{None}, 15\}$, $\text{min_samples_leaf} \in \{1, 5\}$, and $\text{max_features} \in \{\text{sqrt}, 0.5\}$ with 300 estimators; the GB grid covered $\text{max_depth} \in \{3, 5\}$, learning rate = 0.05, and subsample $\in \{0.8, 1.0\}$ with 300 estimators. The best-scoring configuration in each inner loop was refit on the full outer-training partition before generating out-of-sample predictions, and all estimators used a fixed random seed ($\text{random_state} = 42$) for reproducibility.

Out-of-sample predictions were pooled across folds and summarised by five metrics: Pearson r , R^2 , RMSE, mean absolute error (MAE), and bias ($\text{mean}(\hat{y}_i - y_i)$; positive = overprediction). Paired two-sided Wilcoxon signed-rank tests on absolute residuals were used to compare OLS with RF and GB within each CVS [62].

2.5.4. Genotypic Repeatability of Spectral Reflectance Indices

The consistency of genotype rankings across four measurement campaigns was quantified by two-way ANOVA without replication, applied separately to each SRI and irrigation environment, with genotype and measurement date as classification factors. The proportion of total variance attributable to genotype (η^2) was computed as the genotype sum of squares divided by the total sum of squares. Additivity of the two-way model was verified by Tukey's non-additivity test prior to ANOVA; all index–environment combinations satisfied the assumption ($p > 0.05$).

The intraclass correlation coefficient (ICC) was computed under the two-way mixed-effects, consistency, single-measures model—ICC(3,1) as follows:

$$\text{ICC}(3,1) = (\text{MS}_G - \text{MS}_E) / (\text{MS}_G + (k - 1) \cdot \text{MS}_E)$$

where MS_G is the mean square for genotype, MS_E the residual mean square from the two-way ANOVA (i.e., the date effect partitioned out), and $k = 4$ measurement dates. The average-measure ICC(3,4) was derived via the Spearman–Brown formula:

$$\text{ICC}(3,4) = k \cdot \text{ICC}(3,1) / (1 + (k - 1) \cdot \text{ICC}(3,1)).$$

Ninety-five percent confidence intervals for ICC(3,1) were obtained using the large-sample asymptotic variance method [63]. Repeatability (R) equals ICC(3,1) and quantifies the within-season multitemporal ranking consistency of the UAV measurement protocol. Because the augmented design provides no within-genotype replication, the residual mean square confounds genotype $\times$ date interaction, measurement error, and micro-environmental variation, so R is neither broad-sense heritability nor an upper bound thereof; it estimates the consistency with which the four-date mean orders genotypes within a single growing season.

2.5.5. UAV-Based Composite Drought Ranking

Drought performance was quantified using the Stress Tolerance Index (STI) of Fernandez [47], computed independently for five UAV-derived traits at 40 m AGL: canopy height, NDVI, NDRE, GNDVI, and LCI. For each genotype and trait, mean values under ENV1 ($\bar{Y}_p$) and ENV2 ($\bar{Y}_s$) were computed by averaging across all four measurement dates. The STI was calculated as follows:

$$\text{STI} = (\bar{Y}_p \times \bar{Y}_s) / (\bar{Y}_p)^2$$

where $\bar{Y}_p$ is the population mean under ENV1. $STI > 1$ indicates above-average performance under both conditions, reflecting combined yield potential and stress tolerance. Only genotypes with complete data across all five traits and both environments were included ($n = 498$). Individual trait ranks were assigned in descending STI order (rank 1 = highest); a composite rank was computed as the arithmetic mean of the five trait ranks, integrating canopy structural response (height-based STI) with spectral indicators of chlorophyll content and photosynthetic activity (SRI-based STIs). Concordance between the UAV composite ranking and the breeder's independent selection was quantified by a one-sided hypergeometric test for enrichment [64], analogous to gene-set enrichment analysis for ranked-list overlap [65]. Sensitivity to the UAV top-k cut-off was examined at $k \in \{10, 20, 30\}$. The five traits constitute the complete set of canopy variables resolved by the UAV at 40 m—one structural (canopy height) and the four spectral indices—and were combined with equal weights, as no a priori or yield-based justification for differential weighting was available at the M_2 stage. The robustness of the equal-weight composite was tested against three alternative schemes (structural-emphasis, spectral-emphasis, and ICC-weighting) by Spearman rank correlation and top 30 overlap with the published ranking (Table S5).

All analyses were conducted in Python 3.14 [66] (NumPy v2.4.4 [67], pandas v3.0.2 [68,69], SciPy v1.17.1 [70], scikit-learn v1.9.0 [71], matplotlib v3.10.8 [72], seaborn v0.13.2 [73]). The complete analytical workflow is summarised in Figure S2.

3. Results

3.1. Ground Height Variation and Temporal Trajectories

Ground-measured plant height increased progressively across the four phenotyping campaigns in both environments (Figure 3). In ENV1 (normal irrigation), mean height rose from 0.49 ± 0.15 m on 11 July to 1.28 ± 0.25 m on 1 September (+161% over 52 days). In ENV2 (water stress), the corresponding increase was from 0.44 ± 0.15 m to 1.04 ± 0.24 m (+136%), indicating that water deficit progressively constrained vegetative elongation rather than suppressing early-season establishment—consistent with the 21.8% height reduction under water stress reported by Yemata and Bekele [74]. The divergence between environments was modest at the first two dates (11% and 4% difference in means on 11 July and 22 July, respectively) but became pronounced by 1 September (ENV1 exceeding ENV2 by ~23%; 1.28 vs. 1.04 m), coinciding with the largest accumulated irrigation differential between treatments. The coefficient of variation (CV) was consistently higher in ENV2 than in ENV1 across all dates (23.1–34.1% vs. 19.5–30.6%), suggesting that water stress amplified genotypic differentiation in height—a pattern favourable for drought tolerance screening. CV declined over the season in both environments, consistent with convergence as slower-developing genotypes approached their asymptotic height. Height distributions remained approximately unimodal and symmetric at all dates, with a slight rightward skew emerging by 1 September in ENV1, indicative of a subset of high-vigour genotypes exceeding the population mean under unrestricted water supply (Figure 3).

3.2. UAV-Based Estimation of Plant Height

Scatter plots of UAV-estimated canopy height versus ground-measured plant height are presented in Figure 4 for all three flight altitudes and both environments, with regression statistics summarised in Table S1. UAV-derived height was positively and significantly correlated with ground measurements across all altitude–environment combinations, but agreement varied markedly with altitude and was consistently stronger under normal irrigation (ENV1) than under water stress (ENV2).

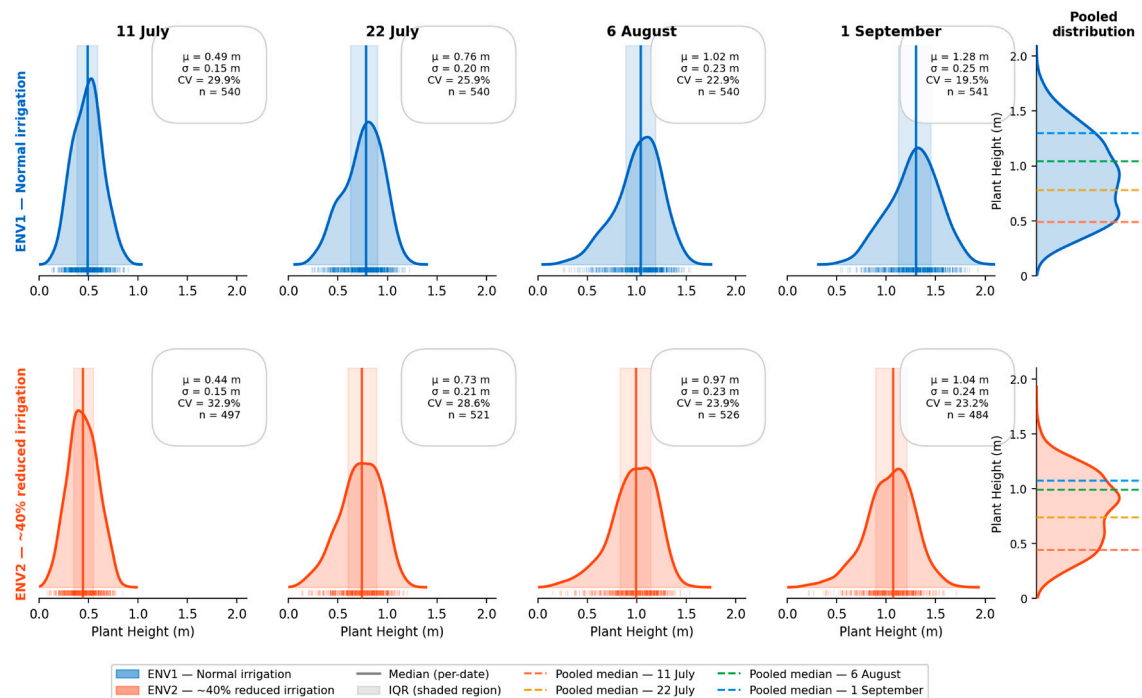


Figure 3. Distribution of ground-measured plant height (m) across four dates and two irrigation environments. Kernel density estimates (KDEs; Silverman bandwidth) for ENV1 (full irrigation, blue) and ENV2 (water stress, orange); vertical line, per-date median; shaded band, interquartile range; rug marks, individual observations. Annotations: μ = mean; σ = SD; CV = coefficient of variation; n = observations. The rightmost strip in each row shows the pooled KDE across all four dates.

At 40 m AGL, the strongest agreement was observed. In ENV1, $r = 0.901$ ($R^2 = 0.812$, RMSE = 0.112 m, $n = 1620$); in ENV2, $r = 0.814$ ($R^2 = 0.663$, RMSE = 0.148 m, $n = 1544$). Point clouds at this altitude were tight and near-linear along the 1:1 reference in both environments, with limited systematic bias.

At 80 m AGL, agreement was moderate. In ENV1, $r = 0.702$ ($R^2 = 0.493$, RMSE = 0.140 m); in ENV2, $r = 0.457$ ($R^2 = 0.209$, RMSE = 0.107 m). The pronounced R^2 drop under water stress likely reflects reduced canopy structural differentiation among genotypes combined with increased point-cloud sensitivity to sparse drought-stressed canopies.

At 120 m AGL, the weakest agreement was recorded. In ENV1, $r = 0.677$ ($R^2 = 0.459$, RMSE = 0.308 m, $n = 2161$); in ENV2, $r = 0.618$ ($R^2 = 0.382$, RMSE = 0.283 m, $n = 2028$). The elevated RMSE relative to lower altitudes is consistent with reduced point-cloud spatial resolution at greater flight heights, introducing reconstruction uncertainty that is particularly acute for short-statured sesame plants; broader dispersion and pronounced underestimation of tall plants were evident in the scatter plots.

Across all altitudes, predictive performance declined under water stress, consistent with drought-induced compression of the height distribution and reduced dynamic range available for regression. The 40 m altitude provided the best representation of ground-measured plant height across both environments and is adopted as the operational standard for subsequent analyses.

3.3. Temporal Dynamics and Genotypic Variation in UAV-Derived Spectral Reflectance Indices

Temporal trajectories of NDVI, NDRE, GNDVI, and LCI are presented in Figure S1, with descriptive statistics in Table S2 and the corresponding spatial index maps in Figure 5. All four indices exhibited a unimodal seasonal pattern under ENV1, increasing from early July to a peak on 6 August and declining thereafter, consistent with the progression from

vegetative growth to maturity in sesame. Under ENV2, trajectories diverged markedly from ENV1 at the final campaign (1 September), with all indices showing a sharp decline indicative of accelerated drought-induced senescence.

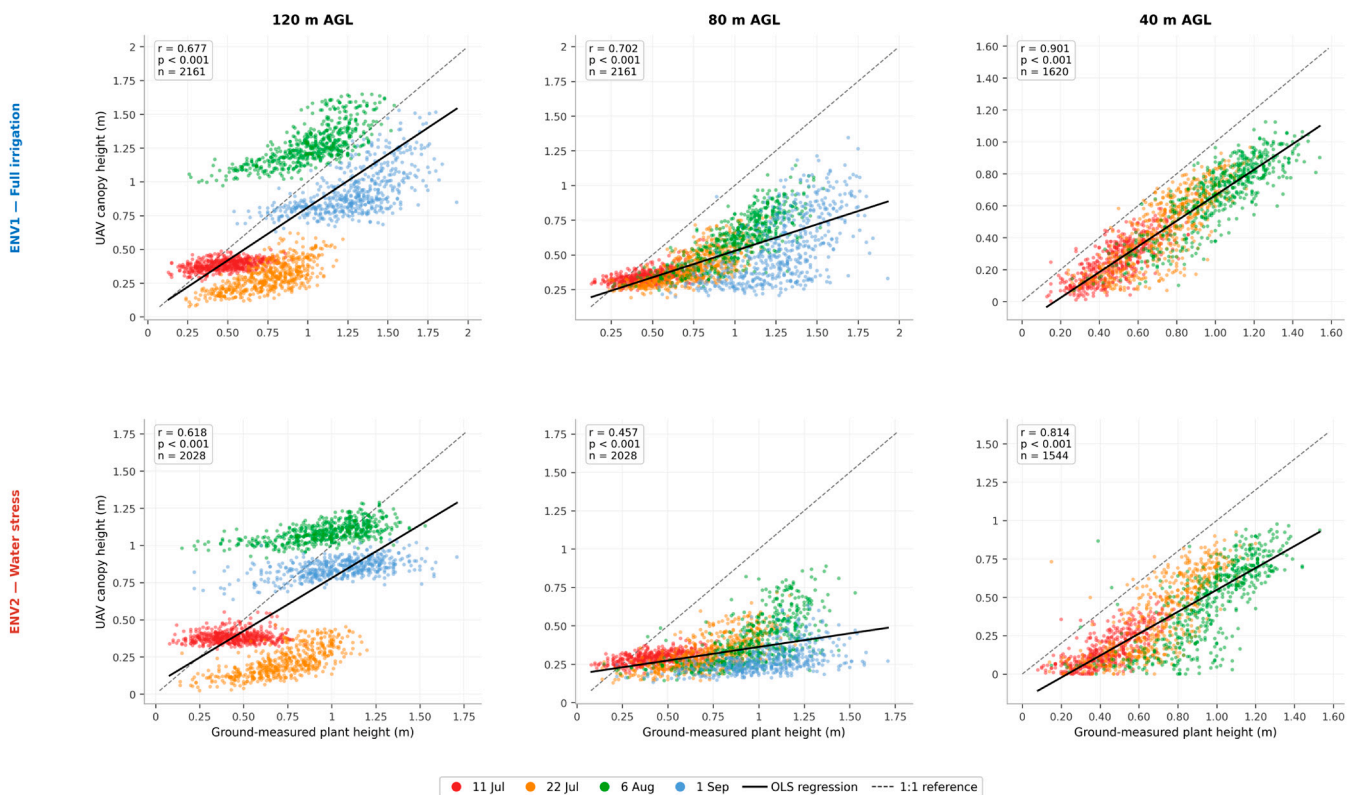


Figure 4. UAV-derived canopy height (y) vs. ground-measured plant height (x) at three flight altitudes (120, 80, 40 m AGL), faceted by environment (ENV1: full irrigation, left; ENV2: water stress, right). Points coloured by measurement date (11 July, red; 22 July, orange; 6 August, green; 1 September, blue). Solid line, OLS fit pooled across dates; dashed line, 1:1 reference. Pearson r , two-tailed p -value, and n annotated per panel.

NDVI was the highest-magnitude index throughout the season, peaking at 0.744 ± 0.063 in ENV1 and 0.626 ± 0.107 in ENV2 on 6 August. The environment divergence was greatest by 1 September (ENV1: 0.606 ± 0.078 ; ENV2: 0.387 ± 0.084 ; $\Delta \approx 0.22$). GNDVI followed the same pattern (peak: ENV1 0.561 ± 0.042 , ENV2 0.511 ± 0.069 ; 1 September: ENV1 0.482 ± 0.053 , ENV2 0.356 ± 0.047), confirming that the green-band formulation captured the same stress response as the standard red-band index. NDRE and LCI, sensitive to red-edge chlorophyll content, showed lower absolute values but the same unimodal trend. NDRE peaked at 0.151 ± 0.022 (ENV1) and 0.138 ± 0.032 (ENV2) on 6 August, declining steeply under stress by 1 September (ENV2: 0.051 ± 0.021 vs. ENV1: 0.118 ± 0.027). LCI followed an analogous trajectory, peaking at 0.230 ± 0.034 (ENV1) and 0.201 ± 0.049 (ENV2), before falling to 0.171 ± 0.042 and 0.070 ± 0.031 , respectively, on 1 September. The larger standard deviations in ENV2 across all indices reflect greater inter-genotypic variability in stress response, which is favourable for drought-tolerant genotype identification.

Notably, ENV2 showed comparable or slightly higher index values than ENV1 on 22 July across all four indices, indicating that early vegetative growth was not substantially impaired at that stage and that the stress effect became apparent only from 6 August onward—consistent with progressive water depletion during flowering and capsule filling, which in sesame occur between mid-July and late August under central Greek conditions.

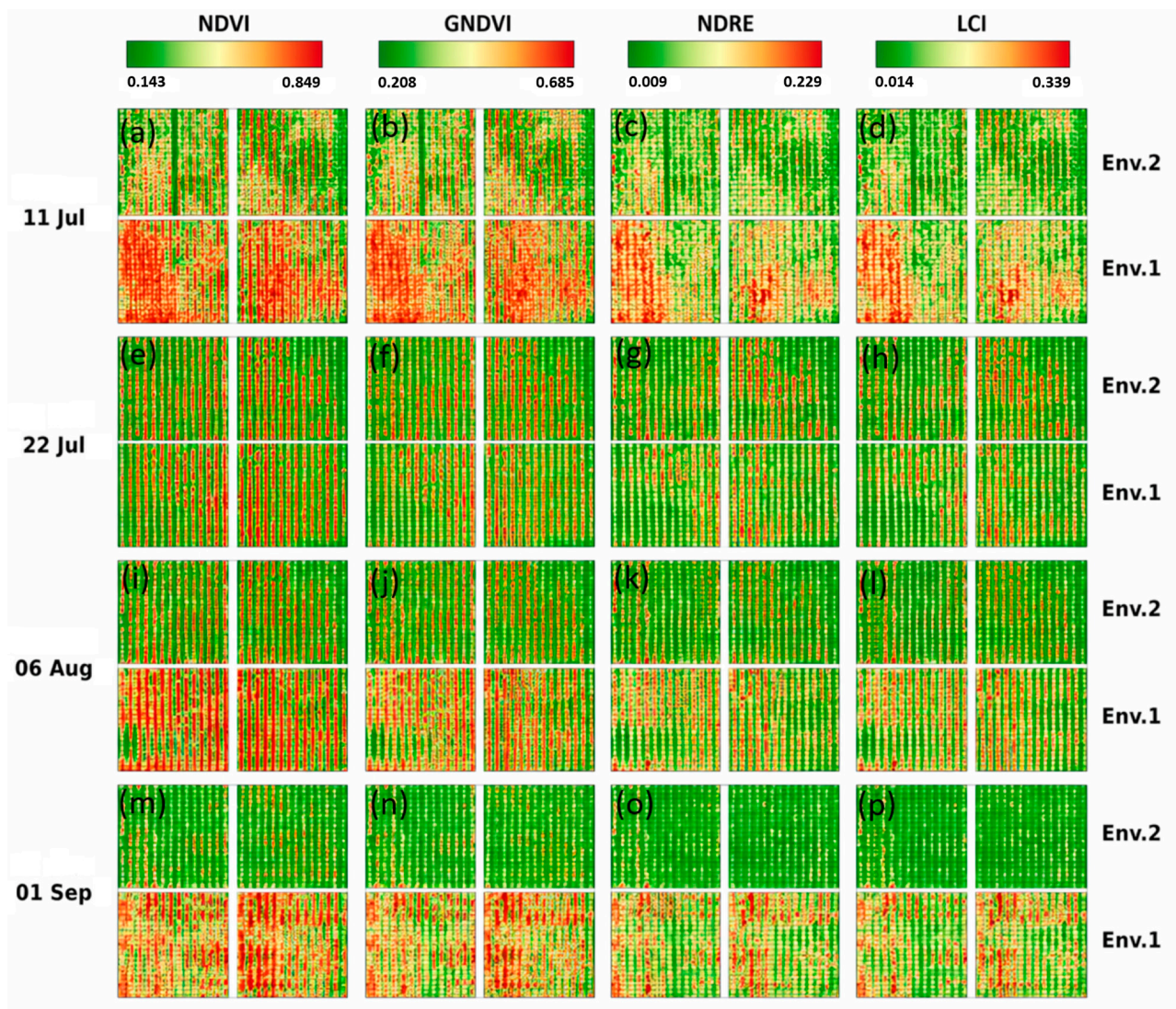


Figure 5. Spatial distribution of vegetation indices derived from UAV multispectral imagery acquired at 40 m AGL across four dates and two irrigation environments. Columns correspond to spectral reflectance indices (NDVI, GNDVI, NDRE, LCI); rows correspond to acquisition dates (11 July, 22 July, 6 August, 1 September). Within each subplot, the upper block shows ENV2 (water stress) and the lower block shows ENV1 (full irrigation). Subplot labels: (a–d) 11 July; (e–h) 22 July; (i–l) 6 August; (m–p) 1 September; within each date group, subplots progress left to right as NDVI, GNDVI, NDRE, and LCI. Colour scale bars (green–red, index-specific range) are shown at the top of each column.

Highly significant genotypic variation ($p < 0.001$) was detected for all four indices in both environments (Table 1), confirming substantial inter-genotypic diversity in canopy spectral response throughout the season. NDVI showed the highest single-measure ICC in both environments: $\text{ICC}(3,1) = 0.485$ (95% CI: 0.441–0.528) in ENV1 and 0.710 (95% CI: 0.679–0.742) in ENV2, corresponding to $\text{ICC}(3,4)$ of 0.790 and 0.907, respectively. GNDVI followed a similar pattern ($\text{ICC}(3,1) = 0.401$ in ENV1, 0.672 in ENV2); NDRE and LCI showed moderate but significant consistency ($\text{ICC}(3,1) = 0.300$ – 0.333 in ENV1; 0.573–0.592 in ENV2). Across all four indices, $\text{ICC}(3,1)$ was up to ~2-fold higher in ENV2 than in ENV1, indicating that drought amplified the expression of genotypic differences in canopy spectral properties and improved the discriminatory power of UAV-derived indices. Average-measure $\text{ICC}(3,4)$ exceeded 0.84 for all indices in ENV2, confirming that the season mean of repeated UAV observations provides a highly reliable estimate of genotypic performance under water-limited conditions.

Table 1. Two-way ANOVA without replication. η^2 , proportion of total variance attributable to genotype; ICC(3,1), single-measure intraclass correlation (consistency model); ICC(3,4), average-measure ICC across the four dates (Spearman–Brown); 95% CI via asymptotic. Repeatability (R) is the proportion of phenotypic variance attributable to genotypic differences, interpreted here as the within-season multitemporal ranking consistency of the UAV protocol; in the absence of within-genotype replication, it is neither broad-sense heritability nor an upper bound thereof. Sig.: *** $p < 0.001$. ENV1, full irrigation; ENV2, water stress.

Index	Environment	<i>n</i>	F	<i>p</i> -Value	η^2	ICC(3,1)	95% CI ICC(3,1)	ICC(3,4)	Repeatability (R)	Sig.
NDVI	ENV1	548	4.76	<0.001	0.429	0.485	[0.441–0.528]	0.79	0.485	***
NDVI	ENV2	530	10.81	<0.001	0.442	0.71	[0.679–0.742]	0.907	0.71	***
NDRE	ENV1	547	2.72	<0.001	0.285	0.3	[0.255–0.346]	0.632	0.3	***
NDRE	ENV2	529	6.38	<0.001	0.279	0.573	[0.533–0.614]	0.843	0.573	***
GNDVI	ENV1	547	3.68	<0.001	0.376	0.401	[0.356–0.446]	0.728	0.401	***
GNDVI	ENV2	529	9.2	<0.001	0.402	0.672	[0.638–0.707]	0.891	0.672	***
LCI	ENV1	547	2.99	<0.001	0.301	0.333	[0.287–0.378]	0.666	0.333	***
LCI	ENV2	529	6.8	<0.001	0.29	0.592	[0.553–0.631]	0.853	0.592	***

3.4. Predictive Modelling of Plant Height

Predictive performance was evaluated under LOGO, LOEO, and LODO cross-validation, targeting generalisation to a new genotype (at a known date and environment), a new irrigation environment, and a new phenological stage, respectively. Full metrics for OLS, RF, and GB under both predictor sets and all three schemes are reported in Table S3; Figure 6 shows predicted-vs-observed scatter plots for the full structural–spectral predictor set under LOGO and LODO for all three algorithms.

Under LOGO, the 40 m only OLS model achieved $R^2 = 0.827$, RMSE = 0.121 m ($n = 3158$), with negligible bias. Adding the two higher-altitude heights and four SRIs marginally improved OLS to $R^2 = 0.833$, RMSE = 0.119 m. RF and GB produced a small but statistically significant improvement over OLS for the full predictor set ($R^2 = 0.840$ for both; RMSE = 0.116 m; Wilcoxon $p = 0.007$ and 0.001 , respectively).

Under LOEO, OLS outperformed both ensembles: pooled OLS $R^2 = 0.779$ (40 m only) and 0.751 (full set), versus RF $R^2 = 0.741$ and 0.711, and GB $R^2 = 0.744$ and 0.722 (all ensemble–OLS gaps: $p < 10^{-6}$). The directional asymmetry persisted (ENV2→ENV1: $R^2 = 0.812$ vs. ENV1→ENV2: $R^2 = 0.743$ for 40 m only OLS), though with only two CV folds available under LOEO, this directionality is descriptive of the present trial rather than a generalisable property of the calibration scheme.

Under LODO—training on two dates and predicting the third, with the held-out date’s phenological stage unrepresented during training—accuracy dropped substantially relative to LOGO (Figure 6). For the 40 m only predictor set, pooled LODO R^2 was 0.557 (OLS), 0.496 (RF), and 0.512 (GB), representing a drop of 0.27–0.32 R^2 units from LOGO and an RMSE increase from ~0.12 m to ~0.20 m. For the full structural–spectral set, LODO revealed a qualitative divergence between algorithms: OLS collapsed ($R^2 = -0.203$; RMSE = 0.319 m), while RF and GB remained stable ($R^2 = 0.527$ and 0.486; RMSE ≈ 0.20 m; Wilcoxon $p < 10^{-150}$ vs. OLS). The OLS collapse was concentrated on the date-3 fold (6 August, seasonal peak): trained on the two vegetative-phase dates (11 and 22 July), OLS extrapolated beyond the SRI training range (per-fold $R^2 = -3.408$). The 40 m only structural model is robust to this failure mode across all three algorithms; the full model is robust only when a non-linear algorithm constrains predictions to the training-set support.

The LODO–LOGO gap clarifies the headline LOGO figure. Because LOGO retains measurement date as a training covariate, the model borrows population-level seasonal trajectory information from ~1000 training observations per date, placing held-out genotype

predictions within the correct seasonal band before the UAV predictors are consulted. LODO severs this channel, measuring the harder, deployment-relevant task of predicting at a new phenological stage. The resulting gap— $\sim 0.27 R^2$ units for the 40 m only OLS—quantifies the date-leakage component of LOGO performance, consistent with the general finding that within-level CV schemes overstate generalisation in spatio-temporal prediction problems [66,67]. LOGO $R^2 = 0.827$ should be treated as an upper bound on genotype-level generalisation within a season; LODO $R^2 = 0.557$ is the operationally relevant figure for deployment to a new phenological stage, and the 40 m only structural model is the safer operational choice.

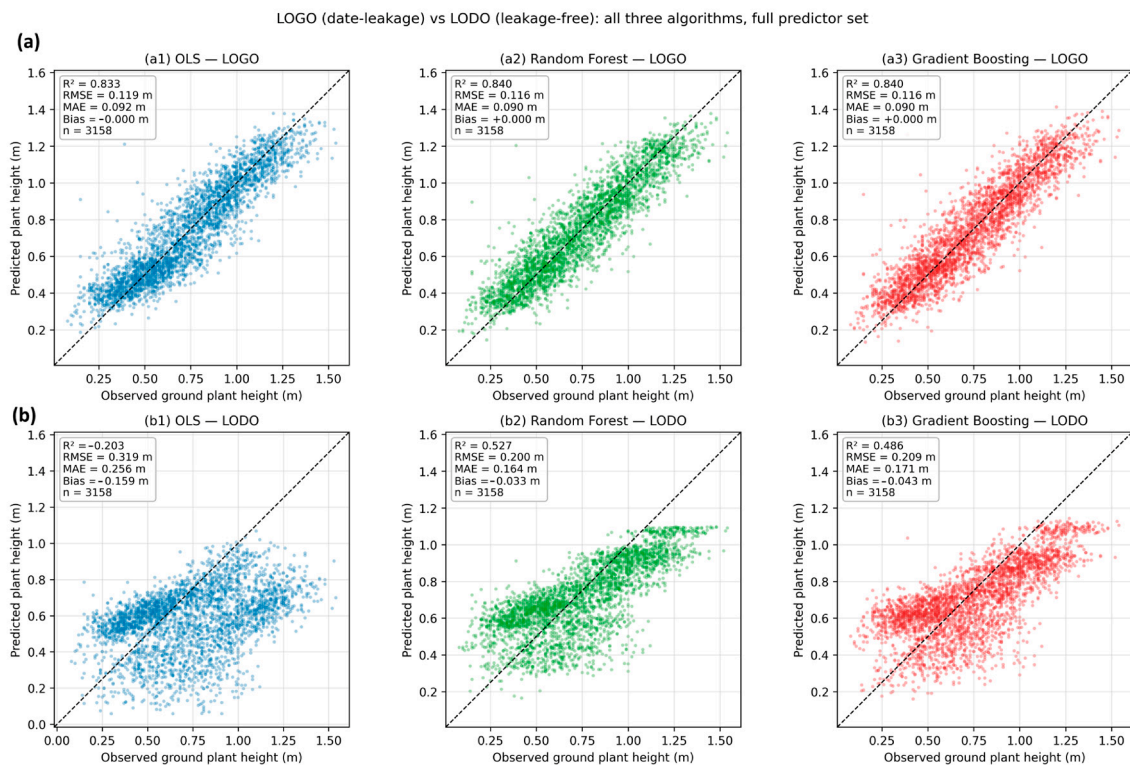


Figure 6. Predicted vs. observed ground plant height (m) for OLS (a1,b1), Random Forest (a2,b2), and Gradient Boosting (a3,b3), fitted to the full structural–spectral predictor set. Top row (a), LOGO (leave-one-genotype-out); bottom row (b), LODO (leave-one-date-out) pooled across three hold-out dates. Dashed line, 1:1 reference. Point colours identify the algorithm: blue, OLS; green, Random Forest; red, Gradient Boosting.

3.5. Composite Stress Tolerance Indexing

3.5.1. UAV-Based Composite Drought Ranking

The composite STI ranking integrating canopy height and four SRIs (NDVI, NDRE, GNDVI, LCI) at 40 m AGL identified a subset of genotypes with consistently high performance across both structural and spectral dimensions (Figure 7). Of 498 genotypes with complete data, 38 (7.6%) exhibited $STI > 1.0$ simultaneously for all five traits, indicating above-average canopy development and photosynthetic capacity under both irrigation regimes. Among the top 30, all five STI values exceeded 1.0 with one exception (genotype 24-6, rank 27; height $STI = 0.74$); the mean height-based STI within this group was 1.61.

The top-ranked genotype (13-1) achieved the highest NDVI-based STI (1.28) and GNDVI-based STI (1.29) in the dataset while maintaining a height STI of 1.67, reflecting balanced canopy performance across structural and spectral traits under drought. Genotypes 21-20 (rank 3), 14-5 (rank 2), and 15-9 (rank 11) showed the highest height-based STI values within the top 30 (2.11, 1.95, and 2.05, respectively), indicating pronounced canopy

height stability under water stress. By contrast, genotype 7-2 (rank 28) recorded the highest height-based STI in the entire dataset (2.42) but ranked poorly across all spectral indices (NDVI rank 63, NDRE rank 51, GNDVI rank 82), demonstrating that structural resilience alone is insufficient to characterise overall drought tolerance when spectral indicators are considered jointly.

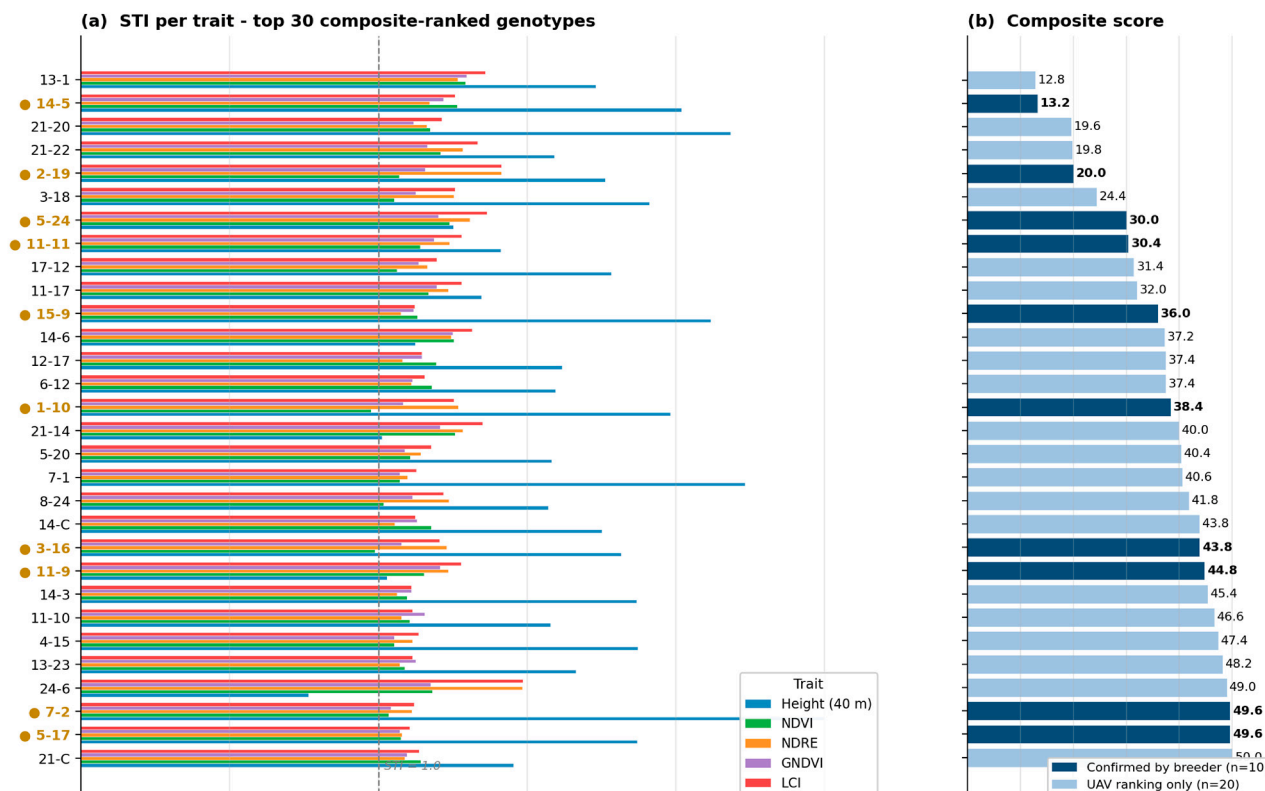


Figure 7. Composite drought ranking of the top 30 sesame M₂ genotypes (of 498 evaluated) from five UAV-derived traits at 40 m AGL across the two irrigation environments. (a) Stress Tolerance Index (STI) per trait—canopy height, NDVI, NDRE, GNDVI, LCI; dashed line, STI = 1.0 (population-mean performance). (b) Composite rank score (mean of the five trait ranks; lower = better). Gold circles and bold labels, and darker bars in (b), mark the 10 genotypes also present in the breeder's top 48.

A clear differentiation was evident between genotypes expressing drought tolerance primarily through structural vs. spectral dimensions. Genotypes 5-24 (rank 7), 21-22 (rank 4), and 11-9 (rank 22) showed moderate height STI values (1.21, 1.54, and 0.99, respectively) but ranked consistently high across all four SRIs, indicating stress tolerance expressed through maintenance of chlorophyll content and photosynthetic activity rather than biomass accumulation. These contrasting profiles illustrate the added value of the multi-trait composite, which captures a broader spectrum of drought adaptation strategies than any single-trait ranking.

3.5.2. Visual Field-Based Stress Ranking

Visual field evaluation of the 48 breeder-selected genotypes under ENV2 spanned three score levels on the 1–8 scale (Figure 8): 21 genotypes (43.8%) received stress tolerance grade 8, 22 (45.8%) received grade 7, and five (10.4%) received grade 6, indicating that the pre-selection process had already enriched the subset for stress-tolerant individuals.

Within the grade 8 group, eight genotypes received maximum productivity grade 8 simultaneously (1-10, 1-11, 1-14, 1-17, 1-9, 3-19, 9-15 and 9-18), six of which belonged to the medium-maturity class, suggesting that medium-maturity genotypes combine high stress

tolerance with agronomic productivity more consistently than late-maturing material. The remaining grade 8 genotypes showed productivity grades of 4–7; genotypes 2-4, 14-5, 15-16 and 15-18 combined grade 8 stress tolerance with the lowest productivity score (grade 4), confirming that stress tolerance and productivity are not always co-expressed and that selection on stress tolerance alone may overlook agronomically valuable genotypes.

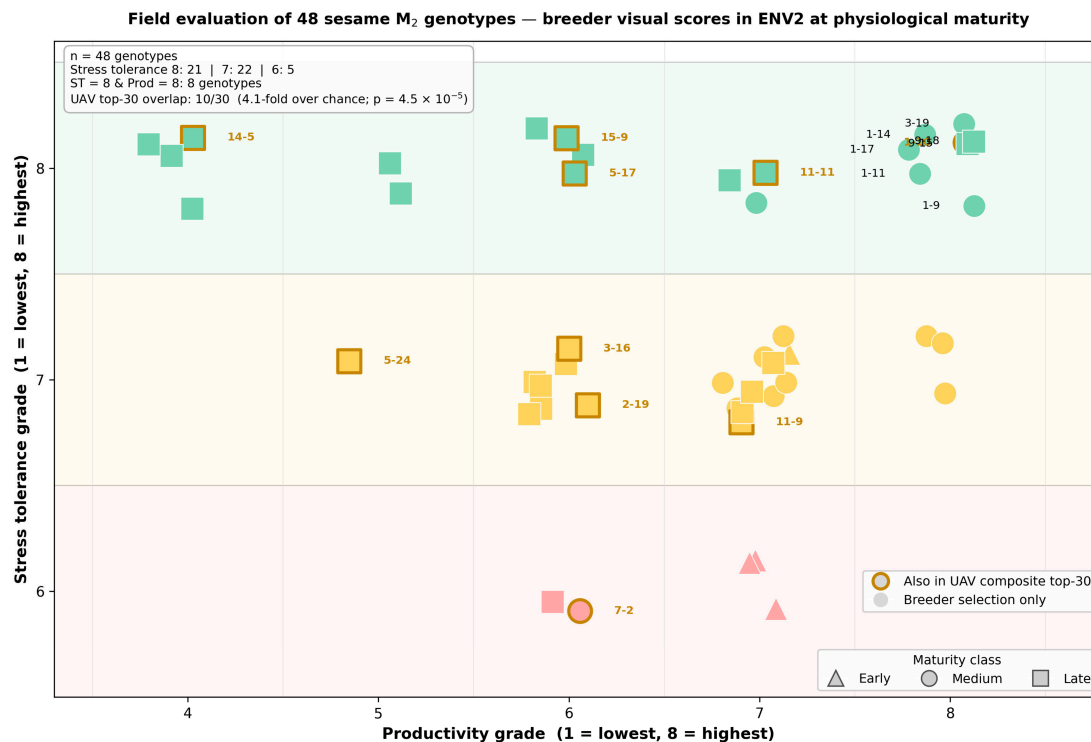


Figure 8. Field evaluation of 48 sesame M₂ genotypes selected by the breeder as candidates for drought tolerance at physiological maturity under ENV2 (water stress). Each marker represents one genotype, positioned by its productivity grade (x-axis) and stress-tolerance grade (y-axis); both traits were scored independently on a 1–8 scale (8 = highest). Background shading by stress-tolerance grade (green = 8, yellow = 7, pink = 6). Marker shape encodes maturity class relative to the standard reference (triangle = early, circle = medium, square = late). Gold-edged markers indicate the 10 genotypes also recovered in the UAV composite top 30 (4.1-fold enrichment over chance; hypergeometric $p = 4.5 \times 10^{-5}$). Labelled genotypes are those scoring 8 for both traits plus all UAV-confirmed genotypes.

Among the 22 grade 7 genotypes, three (19-4, 9-24, 9-25) achieved productivity grade 8, representing an intermediate stress-tolerance profile with high yield potential. The five grade 6 genotypes (13-18, 4-19, 8-3, 4-13, and 7-2) were predominantly of early or late maturity class, which may partly explain their reduced stress tolerance relative to the medium-maturity group. Jointly considering stress tolerance grade, productivity potential, and maturity cycle is therefore important when selecting M₂ genotypes for advancement to M₃.

3.5.3. Concordance Between UAV and Breeder Rankings

Of the 30 top-ranked genotypes in the UAV composite STI, 10 were independently confirmed within the breeder's 48-best selection from visual inspection of all 588 M₂ rows in ENV2 at physiological maturity. The two procedures are procedurally independent—sensor-based multitemporal phenotyping versus a single end-of-season expert walk-through—but observationally overlapping: the breeder evaluation was conducted in ENV2 at physiological maturity (1 September), the same canopy state captured by the final UAV

campaign of that day, so the two assessments share the late-season canopy signal. The hypergeometric null remains valid under exchangeability of the 588 rows, but some divergence is expected on traits not captured by either method alone. Under a random draw null, the expected overlap is 2.45 genotypes (hypergeometric $E[X]$, $SD = 1.46$); the observed 10 represents a 4.1-fold enrichment over chance (one-sided $p = 4.5 \times 10^{-5}$; Fisher odds ratio 6.84), stable across UAV top- k cut-offs of 10, 20, and 30 ($3.7 \times$ – $4.9 \times$; all $p \leq 0.006$; Figure 9). Of the 10 confirmed genotypes, five received breeder stress-tolerance grade 8 (14-5, 11-11, 15-9, 1-10, 5-17), four received grade 7, and one—genotype 7-2, the highest height STI entry of the entire dataset—received grade 6, giving 90% at grade ≥ 7 . Confirmation rates by UAV tier were 40% (ranks 1–10), 20% (ranks 11–20), and 40% (ranks 21–30). The 20 UAV top 30 genotypes not selected by the breeder were inspected and passed over, reflecting traits the UAV sensors do not capture—capsule load, branching architecture, stem strength—and the non-overlap is therefore informative rather than a failure of either method.

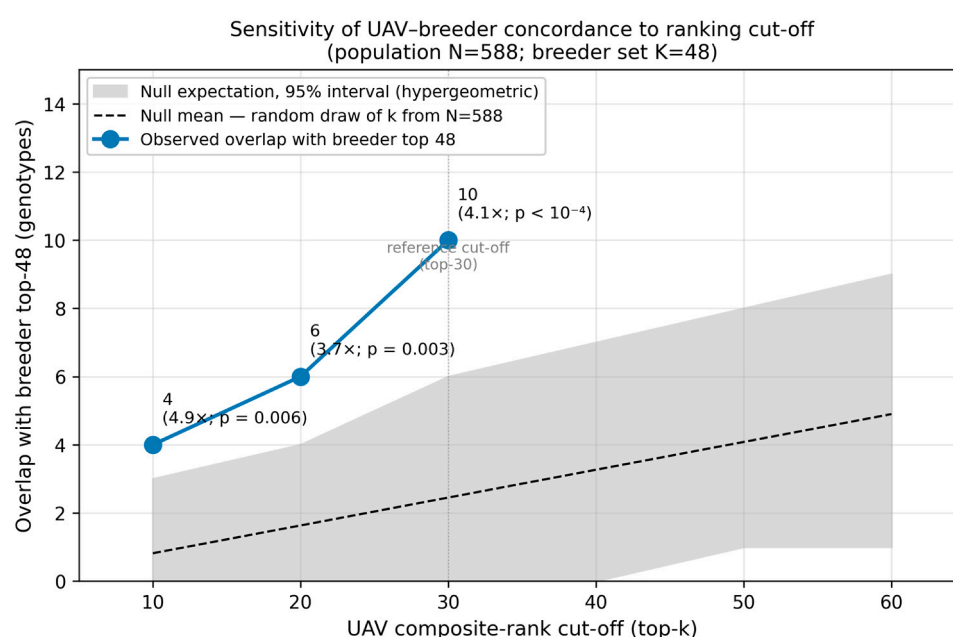


Figure 9. Concordance between the UAV composite Stress Tolerance Index ranking and the breeder’s independent visual selection across UAV top- k cut-offs. Blue points, observed overlap between the UAV top- k and the breeder’s 48-best genotypes (selected from all 588 M_2 rows); black dashed line, expected overlap under the null hypothesis of a random draw of k from the population (hypergeometric mean); grey band, 95% null interval. Annotations give the observed overlap, fold enrichment over null expectation, and one-sided hypergeometric p -value at each cut-off.

4. Discussion

4.1. Altitude-Dependent Accuracy of UAV Canopy Height Estimation

The strong agreement between UAV-derived canopy height at 40 m AGL and ground-measured plant height ($R^2 = 0.812$ in ENV1; 0.663 in ENV2) is consistent with the expectation that spatial resolution in structure-from-motion photogrammetry improves with decreasing flight altitude: the 40 m ground sampling distance resolves individual sesame plant apices within plot-level CHMs, whereas at 80 and 120 m, the increased footprint per pixel introduces sub-canopy mixing that attenuates height estimates. Although 80 m RMSE in ENV2 is nominally lower, the markedly reduced R^2 (0.209) indicates that the model captures substantially less genotypic variance.

The pronounced R^2 drop between 40 m and 80 m in ENV2 ($0.663 \rightarrow 0.209$) combines two effects: reduced canopy cover under drought decreases point-cloud density in sparse

canopies, and compression of the height range under water stress reduces the dynamic range available for regression. The magnitude is approximated by the canopy greenness proxy: mean NDVI in ENV2 was 15.9% lower than ENV1 on 6 August (0.626 vs. 0.744) and 36.1% lower on 1 September (0.387 vs. 0.606; Table S2), indicating substantially less green canopy cover precisely when the altitude gradient is steepest. Because reconstructable point density scales approximately with the inverse square of ground sampling distance (GSD 2.1, 4.2, 6.4 cm at 40, 80, 120 m, i.e., ~ 1.00 , 0.25, 0.11 of the 40 m density), already-sparse high-altitude point clouds lose canopy returns disproportionately over lower-cover stressed canopies, compounding the loss of explained variance at 80 and 120 m in ENV2. Similar altitude-dependent gradients have been reported for sorghum, wheat, and other crops [75,76], but the steepness observed here under stress underscores the importance of low-altitude protocols in populations with reduced canopy vigour. The RMSE of 0.112 m at 40 m in ENV1 falls within the range reported for cereals [25,27], validating this altitude as the operational standard for sesame phenotyping in this programme.

The systematic underestimation of tall plants at 120 m is a known artefact of reduced point-cloud density in SfM reconstruction [29]: with fewer image points reconstructing the uppermost canopy layer, the DSM truncates the tallest individuals. This is amplified in sesame because its indeterminate growth habit [54] produces a pointed canopy apex more susceptible to reconstruction error than the broad, flat tops of cereal canopies. The consistency across both environments and all four dates confirms this as a structural limitation of high-altitude flights rather than a site-specific artefact.

4.2. Temporal Dynamics of Spectral Reflectance Indices and Stress Response

The unimodal seasonal trajectory of all four SRIs in ENV1—peaking on 6 August and declining thereafter—is consistent with the phenological progression of sesame from vegetative growth through flowering and capsule filling to maturity-associated senescence. The temporal pattern of NDVI agrees with previous UAV-based monitoring of sesame under irrigation [9,31] and with ground-level spectroradiometer measurements in semi-arid environments. The convergence of ENV1 and ENV2 values on 22 July—with ENV2 marginally exceeding ENV1 across all four indices—indicates that the irrigation differential imposed from 7 July had not yet produced detectable canopy spectral divergence at this early stage. This temporal lag is agronomically interpretable: sesame has moderate root-system capacity for soil-water extraction [77], and the canopy spectral signal reflects primarily leaf-level photosynthetic status, which is buffered by short-term stomatal regulation. Divergence became pronounced from 6 August onward, coinciding with the onset of capsule filling and the largest accumulated irrigation differential (~ 135 mm by season end). This progression provides a physiological rationale for campaign timing: a single campaign before 22 July would have insufficient discriminatory power for stress screening, whereas campaigns from 6 August onward capture the maximal environment $\times$ genotype interaction [78].

Among the four indices, NDRE and LCI exhibited the steepest absolute decline under stress at the final date (NDRE ENV2: 0.051 vs. ENV1: 0.118; LCI ENV2: 0.070 vs. ENV1: 0.171). This is consistent with the physiological basis of red-edge indices [38,39]: chlorophyll degradation during drought-accelerated senescence causes a blue-shift in the red-edge reflectance feature, reducing both NIR and red-edge reflectance and disproportionately affecting NDRE and LCI relative to NDVI, which is less sensitive to chlorophyll degradation in the moderate-to-high concentration range. The greater sensitivity of NDRE and LCI to stress-induced canopy deterioration supports their inclusion in the multi-SRI height regression and the composite STI alongside NDVI and GNDVI.

4.3. Repeatability and Genotypic Discrimination Under Contrasting Water Regimes

The consistently higher ICC values under water stress—up to ~2-fold higher ICC(3,1) across all four SRIs—constitute one of the most operationally significant findings of this study. Under full irrigation, canopy closure in densely sown progeny rows tends to equalise apparent spectral response across genotypes, compressing inter-genotypic variance and attenuating the genotypic signal available for ranking. Under drought, differential stomatal regulation, leaf rolling, and accelerated senescence in sensitive genotypes amplify spectral differences, increasing the signal-to-noise ratio for genotype discrimination. The average-measure ICC(3,4) exceeding 0.84 for all indices in ENV2—indicating that the mean of four UAV observations produces highly reliable rankings—directly supports the operational validity of multitemporal UAV campaigns for early-generation drought tolerance screening. The same direction of effect has been reported across cereals including wheat, sorghum, and barley [39,41], indicating a general stress-amplified signal-to-noise mechanism rather than a sesame-specific artefact.

The moderate ICC values for NDRE and LCI in ENV1 (0.300 and 0.333) are consistent with the known variability of red-edge indices in dense, closed canopies where the signal saturates [36,38] and is subject to wind-induced canopy motion and mixed-pixel effects. Their improvement under ENV2 (0.573 and 0.592) supports the interpretation that partial defoliation and reduced canopy density improve signal fidelity. The single-season repeatability reported here quantifies within-season multitemporal ranking consistency and should not be interpreted as broad-sense heritability [22,23], (Section 4.7).

4.4. Predictive Modelling: Cross-Validation Performance and SRI Contribution

Under LOGO—the least stringent scheme—the full structural–spectral predictor set produced only a marginal improvement over canopy height at 40 m AGL alone (OLS R^2 : 0.828 $\rightarrow$ 0.833). RF and GB produced a further small but statistically significant gain over OLS ($R^2 = 0.840$; Wilcoxon $p \leq 0.01$ on absolute residuals); however, the effect size—a ~3 mm reduction in median residual at a mean canopy height of ~0.74 m—is agronomically negligible. Under LOEO, both ensembles underperformed OLS (full-set $R^2 = 0.711$ for RF and 0.722 for GB vs. 0.751 for OLS; $p < 10^{-6}$), reflecting the well-known tendency of tree ensembles to extrapolate poorly outside training-set support when a fixed-effect covariate is held out [46,79]. For within-season ranking the three algorithms are interchangeable; for cross-environment transfer, OLS is preferable.

LODO changes the picture substantively. With measurement date held out, R^2 of all three algorithms on the 40 m only predictor set drops to 0.50–0.56 (a 0.27–0.32 R^2 -unit drop relative to LOGO): the quantitative signature of date-leakage inflation in LOGO, which retains date in the design matrix and estimates its effect from ~1000 observations per date before the UAV predictors are consulted [80]. LODO severs this channel; honest generalisation to a new phenological stage is ≈ 0.55 —operationally useful but 30 percentage points below the LOGO-optimistic figure.

The full structural–spectral set introduces a failure mode specific to OLS: linear extrapolation beyond the SRI training range collapses to negative R^2 (pooled -0.203 ; date-3 hold-out -3.408). RF and GB do not exhibit this collapse, because recursive tree splitting bounds predictions to training-set support [81] and saturates at the observed maximum rather than extrapolating along a fitted slope. The collapse on 6 August occurred because NDRE (0.229 vs. training maximum 0.213) and LCI (0.339 vs. 0.313) exceeded their July training range while the date-3 categorical level was unrepresented. Such failures can be flagged in advance by an applicability-domain check on the predictor vector, and avoided by using tree ensembles or by constraining linear predictions to training support. The 40 m only structural model, in which plant height scales monotonically with growth stage

and training dates bound the hold-out distribution, remains comparable across all three algorithms under LODO and is therefore the robust default for across-stage deployment.

LOGO with date as a fixed-effect covariate should therefore be reported as an upper bound on generalisation, with LODO as the headline figure when the deployment target is a new phenological stage; for across-stage use, the 40 m only structural model is preferable regardless of estimator, and if the full set is required, a non-linear estimator should protect against linear extrapolation failure.

Permutation importance of the RF model under LOGO ranked UAV canopy height at 40 m AGL far above all other predictors (mean importance 0.130 ± 0.008), followed by 120 m height (0.030), date and environment dummies (0.007–0.009), 80 m height (0.007), and the four SRIs with NDRE highest (NDRE 0.004 > NDVI 0.002 > LCI 0.001 > GNDVI ≈ 0.000). The ordering is physiologically consistent with the greater sensitivity of the red-edge band to drought-induced chlorophyll degradation (Section 4.2); critically, the date and environment dummies—not the SRIs—carry the bulk of auxiliary variance under LOGO, which is precisely why LODO produces such a sharp drop when those dummies are inestimable. As the four SRIs together accounted for under 0.01 of total permutation importance, and their inclusion did not improve—and under LODO degraded—generalisation, a reduced-burden structural-only protocol (a single 40 m flight without multispectral acquisition or calibration) is justified for across-stage deployment with no loss of predictive accuracy.

4.5. Composite Multi-Trait STI and Drought Adaptation Strategies

The composite STI integrating season mean canopy height and four SRIs at 40 m AGL identified 38 M₂ genotypes (7.6% of 498) with above-average performance across all five traits under both regimes—consistent with the expected frequency of favourable mutation combinations in a γ -irradiation M₂ population pre-selected for agronomic vigour. The composite captures two biologically distinct drought adaptation modes that single-trait ranking would conflate: genotype 7-2—highest height STI (2.42) but low spectral indices—exhibits structural resilience via osmotic adjustment or stem-level water storage without maintaining photosynthetic activity; conversely, 5-24 and 21-22—consistently high spectral STIs but moderate height STI—maintain photosynthetic capacity at the cost of reduced elongation, consistent with isohydric regulation. The composite therefore identifies genotypes balancing structural and metabolic resilience, the combination most likely to translate into yield stability. The composite ranking is robust to trait weighting: across equal-weight, structural-emphasis, spectral-emphasis, and ICC-weighted schemes, the genotype ordering was almost unchanged (Spearman $\rho = 0.98$ –1.00, with 25–30 of the top 30 retained), degrading only when reduced to canopy height alone ($\rho = 0.75$; Table S5).

The STI framework of Fernandez [47], originally developed for yield under contrasting water regimes, extends naturally to UAV-derived canopy traits given plot-level measurements across environments and dates; using season mean values mitigates single-date noise, and a weighted composite emphasising the 6 August and 1 September dates—when environment $\times$ genotype interactions peak (Section 3.3)—may refine the ranking in future applications. Although the cumulative ENV2 deficit of 135 mm corresponds arithmetically to a ~40% reduction relative to ENV1, the imposed regime was severe rather than moderate: irrigation was withheld completely from DAS 46 (7 July, reproductive onset) through the remainder of the season—~85 days of zero water input under peak central-Mediterranean summer conditions, with only one rainfall event above 10 mm (19.6 mm on 31 July) during the stress phase. Sesame is acutely sensitive to water deficit at capsule filling [54,82], so the STI rankings reported here characterise genotypes under a genuinely severe terminal drought relevant to rain-fed and deficit-irrigated Mediterranean production.

4.6. Validation Against Independent Breeder Assessment

The recovery of 10 UAV top 30 genotypes within the breeder's 48-best selection is statistically robust (4.1-fold over null; hypergeometric $p = 4.5 \times 10^{-5}$; Figure 9) despite the two procedures measuring partly different phenotypic dimensions—the UAV composite integrates structural and spectral traits across four dates and two environments, while the breeder additionally weighs capsule load, branching, and stem architecture [83]. The observed enrichment indicates meaningful but not complete information sharing between sensor-based and expert-based selection.

Confirmed genotypes clustered at the extremes of the UAV top 30 (40% in ranks 1–10 and 40% in ranks 21–30 vs. 20% in the intermediate tier), suggesting high specificity at the top but moderate sensitivity in the middle band, where unbalanced trait profiles can score highly on one or two dimensions without consistent multi-trait performance. The high confirmation rate in ranks 1–10 is operationally encouraging: the UAV composite effectively pre-screens the upper tail of the distribution, the subset of immediate interest for M_3 advancement.

The 20 UAV top 30 genotypes not in the breeder's selection were seen and deliberately passed over. Phenotypically, these false positives were typically vigorous, tall, late-maturing plants with dense green canopies—high canopy height and red-edge STI—that the breeder down-weighted for low capsule load, excessive vegetative branching, lodging-prone stems, or late maturity exposing capsule filling to terminal drought. None of these reproductive and architectural traits is resolved by the four UAV-derived metrics; the UAV composite should therefore narrow the candidate pool for expert evaluation rather than replace it, and M_3 yield data will clarify which approach—or their combination—best predicts drought performance.

4.7. Limitations and Future Directions

The results derive from a single season (2025), and multi-year evaluation is required before top-ranked genotypes are advanced [22,84]. The M_2 generation is, however, a unique non-recurring opportunity: it is the first non-chimeric generation in which induced recessive mutations segregate at detectable frequency [17,85] across the full 588-row population, and stringent selection in subsequent generations narrows this diversity. The rankings are therefore interpreted as a calibration baseline rather than definitive recommendations, to be confirmed by replicated, multi-location M_3 yield trials in 2026 and M_4 multi-environment evaluation thereafter.

The unreplicated within-environment design precludes broad-sense heritability estimation, as the residual term confounds genotype $\times$ date interaction, measurement error, and unblocked micro-environmental variation. The ICC-based repeatability is therefore an operational measure of within-season multitemporal ranking consistency rather than a heritability ceiling across years.

The LODO analysis was restricted to the three dates with 40 m AGL imagery, with one fold requiring extrapolation to the seasonal peak. A fully informative test of temporal generalisation requires additional within-season campaigns or multi-year data spanning ascending, peak, and senescence phases—both planned for subsequent generations; including growing-degree-day accumulation as a continuous covariate may further improve cross-stage transferability.

The four-band multispectral payload could be extended with shortwave-infrared or thermal-infrared sensors, which capture canopy temperature and transpiration more directly than chlorophyll-based indices [30,86]. The absence of yield data is a further limitation: M_3 progeny-trial yield components will provide the most agronomically relevant validation of the composite STI, and integration with genomic marker data will identify

mutant loci underlying the superior drought tolerance profiles [87–89]. The composite STI also collapses heat tolerance, terminal drought escape via early maturity, and capsule shattering resistance onto one canopy-derived axis; resolving these dimensions will require thermal/SWIR sensing and phenological covariates rather than a univariate index.

5. Conclusions

This study establishes multi-altitude, multi-SRI UAV remote sensing as a robust framework for high-throughput phenotyping of an M₂ sesame mutagenesis population under contrasting irrigation regimes. The principal conclusions are as follows.

UAV canopy height at 40 m AGL was the most accurate proxy for ground-measured plant height across both environments and all four campaigns ($R^2 = 0.812$ in ENV1; 0.663 in ENV2). Higher altitudes (80 and 120 m) produced substantially degraded accuracy, particularly under water stress, and are not recommended as primary phenotyping altitudes for sesame.

Joint inclusion of four SRIs and two additional altitudes produced only a marginal LOGO improvement over 40 m alone (OLS R^2 : $0.828 \rightarrow 0.833$; RF and GB: 0.840). Under LODO, all three algorithms fell to $R^2 \approx 0.50$ – 0.56 on the 40 m only set—a 0.27 – 0.32 R^2 -unit drop that establishes the across-stage generalisation baseline. On the full structural–spectral set, OLS collapsed under LODO ($R^2 = -0.203$) due to linear extrapolation beyond the SRI training range, while RF and GB remained stable ($R^2 \approx 0.49$ – 0.53). Under LOEO, the stress-trained model transferred slightly better than the irrigation-trained model in the 2025 trial (ENV2 $\rightarrow$ ENV1 $R^2 = 0.812$ vs. 0.743), though with only two folds, this is descriptive rather than directional. LOGO $R^2 \approx 0.83$ should therefore be reported as an upper bound for within-season ranking; LODO $R^2 \approx 0.55$ is the operationally relevant figure for across-stage deployment, and the 40 m only structural model is the safer choice for that task.

Genotypic repeatability of all four SRIs was higher under water stress than under full irrigation (ICC(3,1) up to ~ 2 -fold greater; ICC(3,4) > 0.84 for all indices), supporting multitemporal UAV spectral indices as reliable selection criteria for early-generation drought screening.

The five-trait composite STI identified 38 M₂ genotypes (7.6% of 498) with above-average performance under both regimes. Of the UAV top 30, 10 were independently confirmed within the breeder's 48-best selection—a 4.1-fold enrichment over chance (hypergeometric $p = 4.5 \times 10^{-5}$), with 90% of confirmed genotypes at the two highest field grades. The 20 non-overlapping UAV candidates were inspected and deliberately passed over, reflecting reproductive and morphological traits that the UAV traits do not capture; the composite therefore narrows the candidate pool for expert evaluation without replacing it.

Combining low-altitude UAV structural measurements with multitemporal red-edge spectral indices offers a scalable phenotyping approach for drought-adaptive selection in sesame mutagenesis programmes. Results from this single unreplicated M₂ season require multi-year validation; the workflow will be carried forward to M₃ and later generations, where yield-based validation will determine whether the UAV composite, breeder assessment, or their combination best predicts drought performance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs18132181/s1>, Table S1: Regression performance of UAV canopy height estimates across flight altitudes under contrasting water regimes; Table S2: Spectral-index summary statistics across dates and irrigation treatments; Table S3: Cross-validated accuracy of plant height models under multiple predictor sets and validation schemes; Table S4: Per-fold performance of OLS, Random Forest, and Gradient Boosting under leave-one-date-out cross-validation, broken down by hold-out date; Table S5. Sensitivity of the composite Stress Tolerance Index ranking to trait weighting; Figure S1. Temporal trajectories of the four spectral reflectance indices (NDVI, NDRE,

GNDVI, LCI) across the four UAV campaigns under full irrigation (ENV1) and water stress (ENV2); Figure S2. Analytical workflow of the UAV-based phenotyping and drought-ranking pipeline.

Author Contributions: Conceptualisation, C.P., A.T. and D.N.V.; methodology, C.P., A.T., E.E. and D.N.V.; software, C.P. and A.T.; validation, C.P., A.T. and E.E.; formal analysis, C.P. and A.T.; investigation, C.P., A.T., E.E., A.K. and C.I.P.; resources, D.N.V.; data curation, C.P. and A.T.; writing—original draft preparation, C.P. and A.T.; writing—review and editing, C.P., A.T., E.E., C.I.P. and D.N.V.; visualisation, C.P. and A.T.; supervision, D.N.V.; project administration, D.N.V.; funding acquisition, D.N.V. All authors have read and agreed to the published version of the manuscript.

Funding: The research project is implemented in the framework of H.F.R.I call “3rd Call for H.F.R.I.’s Research Projects to Support Faculty Members and Researchers” (H.F.R.I. Project Number: 25758).

Data Availability Statement: The original data presented in the study are openly available in Zenodo at <https://zenodo.org/records/19923837> (accessed on 15 April 2026).

Acknowledgments: Special thanks to Eleni M. Abraham for the induction of genetic variability through gamma radiation in the genetic material that was evaluated.

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

Abbreviations

The following abbreviations are used in this manuscript:

AGL	Above ground level
ANOVA	Analysis of variance
CHM	Canopy height model
CV	Coefficient of variation
DAS	Days after sowing
DSM	Digital surface model
ENV1	Normal irrigation environment
ENV2	Water-stress environment
GB	Gradient Boosting
GDD	Growing degree days
GNDVI	Green Normalised Difference Vegetation Index
ICC	Intraclass correlation coefficient
LCI	Leaf Chlorophyll Index
LODO	Leave-one-date-out
LOEO	Leave-one-environment-out
LOGO	Leave-one-genotype-out
MAE	Mean absolute error
NDRE	Normalised Difference Red Edge
NDVI	Normalised Difference Vegetation Index
NTRIP	Networked Transport of RTCM via Internet Protocol
OLS	Ordinary least squares
P95	95th percentile
RF	Random Forest
RGB	Red, Green, Blue
RMSE	Root-mean-square error
RTCM	Radio Technical Commission for Maritime Services
RTK	Real-time kinematic
SfM	Structure-from-motion
SRI	Spectral reflectance index
STI	Stress Tolerance Index
UAV	Unmanned aerial vehicle

References

- Bedigian, D.; Harlan, J.R. Evidence for cultivation of sesame in the ancient world. *Econ. Bot.* **1986**, *40*, 137–154. [\[CrossRef\]](#)
- Teklu, D.H.; Shimelis, H.; Tesfaye, A.; Shayanowako, A.I.T. Analyses of genetic diversity and population structure of sesame (*Sesamum indicum* L.) germplasm collections through seed oil and fatty acid compositions and SSR markers. *J. Food Compos. Anal.* **2022**, *110*, 104545. [\[CrossRef\]](#)
- Li, M.; Luo, J.; Nawaz, M.A.; Stockmann, R.; Buckow, R.; Barrow, C.; Dunshea, F.; Suleria, H.A.R. Phytochemistry, Bioaccessibility, and Bioactivities of Sesame Seeds: An Overview. *Food Rev. Int.* **2023**, *40*, 309–335. [\[CrossRef\]](#)
- Mostashari, P.; Mousavi Khaneghah, A. Sesame Seeds: A Nutrient-Rich Superfood. *Foods* **2024**, *13*, 1153. [\[CrossRef\]](#) [\[PubMed\]](#)
- He, M.; Hu, J.; Deng, J.; Chen, X.; Olatunji, O.J.; Jayeoye, T.J. A comprehensive review on the anti-cancer properties of sesame lignans: Chemical composition, molecular mechanisms and prospects. *RSC Adv.* **2025**, *15*, 45636–45664. [\[CrossRef\]](#)
- Uzun, B.; Arslan, Ç.; Furat, Ş. Variation in Fatty Acid Compositions, Oil Content and Oil Yield in a Germplasm Collection of Sesame (*Sesamum indicum* L.). *J. Am. Oil Chem. Soc.* **2008**, *85*, 1135–1142. [\[CrossRef\]](#)
- FAOSTAT. *Crops and Livestock Products*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2024. Available online: <https://www.fao.org/faostat/> (accessed on 15 April 2026).
- CBI—Centre for the Promotion of Imports from Developing Countries. *The European Market Potential for Sesame Seeds*; Ministry of Foreign Affairs of the Netherlands: The Hague, The Netherlands, 2024. Available online: <https://www.cbi.eu/market-information/grains-pulses-oilseeds/sesame-seeds/market-potential> (accessed on 15 April 2026).
- Petsoulas, C.; Evangelou, E.; Tsitouras, A.; Aschonitis, V.; Kargiotidou, A.; Khah, E.M.; Pavli, O.I.; Vlachostergios, D.N. Spectral reflectance indices as a high-throughput selection tool in a sesame breeding scheme. *Remote Sens.* **2022**, *14*, 2629. [\[CrossRef\]](#)
- Stavridou, E.; Roussis, I.; Tsivelika, N.; Mavromatis, A.; Goulas, C.; Ninou, E.; Papadopoulos, I.; Sinapidou, E.; Tokatlidis, I.S. Characterization of the genetic diversity present in a diverse sesame landrace collection based on phenotypic traits and EST-SSR markers coupled with an HRM analysis. *Plants* **2021**, *10*, 656. [\[CrossRef\]](#) [\[PubMed\]](#)
- Teklu, D.H.; Shimelis, H.; Abady, S. Genetic Improvement in Sesame (*Sesamum indicum* L.): Progress and Outlook: A Review. *Agronomy* **2022**, *12*, 2144. [\[CrossRef\]](#)
- Qureshi, M.; Langham, D.R.; Lucas, S.J.; Uzun, B.; Yol, E. Breeding history for shattering trait in sesame: Classic to genomic approach. *Mol. Biol. Rep.* **2022**, *49*, 7185–7194. [\[CrossRef\]](#)
- Rauf, S.; Basharat, T.; Gebeyehu, A.; Elsayf, M.; Rahmatov, M.; Ortiz, R.; Kaya, Y. Sesame, an Underutilized Oil Seed Crop: Breeding Achievements and Future Challenges. *Plants* **2024**, *13*, 2662. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ashri, A. Sesame breeding. *Plant Breed. Rev.* **1998**, *16*, 179–228. [\[CrossRef\]](#)
- Qureshi, M.; Arslan, M.; Golukcu, M.; Bera, S.K.; Uzun, B.; Yol, E. Assessment of drought tolerance of sesame germplasm with agronomic and quality traits. *Crop Sci.* **2023**, *63*, 2763–2777. [\[CrossRef\]](#)
- Kouighat, M.; Hanine, H.; El Fechtali, M.; Nabloussi, A. First Report of Sesame Mutants Tolerant to Severe Drought Stress during Germination and Early Seedling Growth Stages. *Plants* **2021**, *10*, 1166. [\[CrossRef\]](#) [\[PubMed\]](#)
- IAEA. *Sesame Improvement by Induced Mutations, Proceedings of a Final Research Co-Ordination Meeting*; IAEA-TECDOC-1195; International Atomic Energy Agency: Vienna, Austria, 2001.
- Jhar, A.; Sabbagh, S.; Alkhatab, E.; Abrak, S.; Hatipoğlu, H.; Brennan, A.C. Optimising gamma irradiation seed treatment of sesame (*Sesamum indicum* L.) varieties for potential future application in mutation breeding. *Crop Breed. Genet. Genom.* **2024**, *6*, e240004. [\[CrossRef\]](#)
- Furbank, R.T.; Tester, M. Phenomics—Technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fischer, R.A.; Rebetzke, G.J. Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: A review. *Crop Pasture Sci.* **2018**, *69*, 439–459. [\[CrossRef\]](#)
- Wang, C.; Niu, J.; Miao, H.; Li, C.; Duan, Y.; Ju, M.; Cao, H.; Wei, L.; Wang, H.; Zhang, H. Genetic variation, correlation, and association mapping of seed yield and its component traits in sesame. *Genet. Resour. Crop Evol.* **2024**, *71*, 603–619. [\[CrossRef\]](#)
- Clarke, G.P.Y.; Stefanova, K.T. Optimal Design for Early-Generation Plant-Breeding Trials with Unreplicated or Partially Replicated Test Lines. *Aust. N. Z. J. Stat.* **2011**, *53*, 461–480. [\[CrossRef\]](#)
- Krause, M.D.; Piepho, H.-P.; Dias, K.O.G.; Singh, A.K.; Beavis, W.D. Models to Estimate Genetic Gain of Soybean Seed Yield from Annual Multi-Environment Field Trials. *bioRxiv* **2023**. [\[CrossRef\]](#)
- Bendig, J.; Bolten, A.; Bennertz, S.; Broscheit, J.; Eichfuss, S.; Bareth, G. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. *Remote Sens.* **2014**, *6*, 10395–10412. [\[CrossRef\]](#)
- Holman, F.H.; Riche, A.B.; Michalski, A.; Castle, M.; Wooster, M.J.; Hawkesford, M.J. High Throughput Field Phenotyping of Wheat Plant Height and Growth Rate in Field Plot Trials Using UAV Based Remote Sensing. *Remote Sens.* **2016**, *8*, 1031. [\[CrossRef\]](#)
- Volpato, L.; Pinto, F.; González-Pérez, L.; Thompson, I.G.; Borém, A.; Reynolds, M.; Gérard, B.; Molero, G.; Rodrigues, F.A. High Throughput Field Phenotyping for Plant Height Using UAV-Based RGB Imagery in Wheat Breeding Lines: Feasibility and Validation. *Front. Plant Sci.* **2021**, *12*, 591587. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Malambo, L.; Popescu, S.C.; Murray, S.C.; Putman, E.; Pugh, N.A.; Horne, D.W.; Richardson, G.; Sheridan, R.; Rooney, W.L.; Avant, R.; et al. Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 31–42. [\[CrossRef\]](#)
28. Watanabe, K.; Guo, W.; Arai, K.; Takanashi, H.; Kajiya-Kanegae, H.; Kobayashi, M.; Yano, K.; Tokunaga, T.; Fujiwara, T.; Tsutsumi, N.; et al. High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front. Plant Sci.* **2017**, *8*, 421. [\[CrossRef\]](#)
29. Madec, S.; Baret, F.; de Solan, B.; Thomas, S.; Dutartre, D.; Jezequel, S.; Hemmerlé, M.; Colombeau, G.; Comar, A. High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. *Front. Plant Sci.* **2017**, *8*, 02002. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Sahoo, M.M.; Tarshish, R.; Sabag, I.; Gadri, Y.; Morota, G.; Peleg, Z.; Alchanatis, V.; Herrmann, I. Sesame yield prediction using UAV-borne hyperspectral imagery, crop traits and genotype category. In *Precision Agriculture'25*; Wageningen Academic: Leiden, The Netherlands, 2025. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Sabag, I.; Bi, Y.; Sahoo, M.M.; Herrmann, I.; Morota, G.; Peleg, Z. Leveraging genomics and temporal high-throughput phenotyping to enhance association mapping and yield prediction in sesame. *Plant Genome* **2024**, *17*, e20481. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Maes, W.H.; Steppe, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Shi, Y.; Thomasson, J.A.; Murray, S.C.; Pugh, N.A.; Rooney, W.L.; Shafian, S.; Rajan, N.; Rouze, G.; Morgan, C.L.S.; Neely, H.L.; et al. Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* **2016**, *11*, e0159781. [\[CrossRef\]](#)
34. Torres-Sánchez, J.; López-Granados, F.; De Castro, A.I.; Peña-Barragán, J.M. Configuration and specifications of an unmanned aerial vehicle (UAV) for early site-specific weed management. *PLoS ONE* **2013**, *8*, e0058210. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [\[CrossRef\]](#)
36. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [\[CrossRef\]](#)
37. Barnes, E.M.; Clarke, T.; Richards, S.E.; Colaizzi, P.D.; Haberland, J.; Kostrzewski, M.; Waller, P.M.; Choi, C.Y.; Riley, E.; Thompson, T.L.; et al. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In *Proceedings of the Fifth International Conference on Precision Agriculture*, Bloomington, MN, USA, 16–19 July 2000.
38. Datt, B. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using *Eucalyptus* Leaves. *J. Plant Physiol.* **1999**, *154*, 30–36. [\[CrossRef\]](#)
39. Anderegg, J.; Yu, K.; Aasen, H.; Walter, A.; Liebisch, F.; Hund, A. Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm. *Front. Plant Sci.* **2020**, *10*, 01749. [\[CrossRef\]](#)
40. Tattaris, M.; Reynolds, M.P.; Chapman, S.C. A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. *Front. Plant Sci.* **2016**, *7*, 01131. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Condorelli, G.E.; Maccaferri, M.; Newcomb, M.; Andrade-Sanchez, P.; White, J.W.; French, A.N.; Sciara, G.; Ward, R.; Tuberosa, R. Comparative Aerial and Ground Based High Throughput Phenotyping for the Genetic Dissection of NDVI as a Proxy for Drought Adaptive Traits in Durum Wheat. *Front. Plant Sci.* **2018**, *9*, 00893. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Arais, J.L.; Cairns, J.E. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* **2014**, *19*, 52–61. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Elsayed, S.; Misteale, B.; Schmidhalter, U. Can changes in leaf water potential be assessed spectrally? *Funct. Plant Biol.* **2011**, *38*, 523–533. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Yue, J.; Yang, G.; Li, C.; Li, Z.; Wang, Y.; Feng, H.; Xu, B. Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models. *Remote Sens.* **2017**, *9*, 708. [\[CrossRef\]](#)
45. Bendig, J.; Yu, K.; Aasen, H.; Boltan, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87. [\[CrossRef\]](#)
46. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [\[CrossRef\]](#)
47. Fernandez, G.C.J. Effective selection criteria for assessing plant stress tolerance. In *Proceedings of the International Symposium on Adaptation of Vegetables and Other Food Crops in Temperature and Water Stress*, Taiwan, China, 13–18 August 1992; Kuo, C.G., Ed.; AVRDC: Shanhu, Taiwan, 1992; pp. 257–270.
48. Bouslama, M.; Schapaugh, W.T. Stress tolerance in soybeans. I. Evaluation of three screening techniques for heat and drought tolerance. *Crop Sci.* **1984**, *24*, 933–937. [\[CrossRef\]](#)
49. Fischer, R.A.; Maurer, R. Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust. J. Agric. Res.* **1978**, *29*, 897–912. [\[CrossRef\]](#)

50. Farshadfar, E.; Sabaghpour, S.H.; Zali, H. Comparison of parametric and non-parametric stability statistics for selecting stable and adapted chickpea genotypes under diverse environments. *Aust. J. Crop Sci.* **2012**, *6*, 514–524.
51. Mevlut, A.; Sinan, K.; Seymus, F.; Bekir, E. Evaluation of drought stress tolerance based on selection indices in Turkish oat landraces and cultivars. *Afr. J. Agric. Res.* **2009**, *4*, 299–306.
52. Roth, L.; Marzougui, A.; Walter, A. A review of the journey of field crop phenotyping: From trait stamp collections and fancy robots to phenomics-informed crop performance predictions. *J. Plant Physiol.* **2025**, *311*, 154542. [[CrossRef](#)] [[PubMed](#)]
53. Garcia-Oliveira, A.L.; Dwivedi, S.L.; Chander, S.; Nelimor, C.; Abd El Moneim, D.; Ortiz, R. Breeding smarter: Artificial intelligence and machine learning tools in modern breeding—A review. *Agronomy* **2026**, *16*, 137. [[CrossRef](#)]
54. Sayed, L.M.; Khaled, K.A.M.; Samaha, G.M.; Saber, A.A. Integrating phenotypic, molecular, and bioinformatics approaches for developing drought-tolerant sesame. *Sci. Rep.* **2025**, *15*, 21897. [[CrossRef](#)]
55. Sabag, I.; Pnini, S.; Morota, G.; Peleg, Z. Refining flowering date enhances sesame yield independently of day-length. *BMC Plant Biol.* **2024**, *24*, 711. [[CrossRef](#)] [[PubMed](#)]
56. Langham, D. Phenology of Sesame. In *Issues in New Crops and New Uses*; Janick, J., Whipkey, A., Eds.; ASHS Press: Alexandria, VA, USA, 2007.
57. SZ DJI Technology Co., Ltd. *DJI Terra*, Version 5.1.1; SZ DJI Technology Co., Ltd.: Shenzhen, China, 2025. Available online: <https://enterprise.dji.com/dji-terra> (accessed on 3 December 2025).
58. Lisein, J.; Pierrot-Deseilligny, M.; Bonnet, S.; Lejeune, P. A Photogrammetric Workflow for the Creation of a Forest Canopy Height Model from Small Unmanned Aerial System Imagery. *Forests* **2013**, *4*, 922–944. [[CrossRef](#)]
59. Xie, T.; Li, J.; Yang, C.; Jiang, Z.; Chen, Y.; Guo, L.; Zhang, J. Crop height estimation based on UAV images: Methods, errors, and strategies. *Comput. Electron. Agric.* **2021**, *185*, 106155. [[CrossRef](#)]
60. Hütt, C.; Bolten, A.; Hüging, H.; Bareth, G. UAV LiDAR Metrics for Monitoring Crop Height, Biomass and Nitrogen Uptake: A Case Study on a Winter Wheat Field Trial. *PFG-J. Photogramm. Remote Sens. Geoinf. Sci.* **2023**, *91*, 65–76. [[CrossRef](#)]
61. Liu, Y.; Nie, C.; Li, L.; Shi, L.; Liu, S.; Nan, F.; Cheng, M.; Yu, X.; Bai, Y.; Jia, X.; et al. Maize tasseling date forecast from canopy height time series estimated by UAV LiDAR data. *Crop J.* **2025**, *13*, 975–990. [[CrossRef](#)]
62. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
63. Donner, A.; Wells, G. A Comparison of Confidence Interval Methods for the Intraclass Correlation Coefficient. *Biometrics* **1986**, *42*, 401–412. [[CrossRef](#)]
64. Agresti, A. *Categorical Data Analysis*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013.
65. Boyle, E.I.; Weng, S.; Gollub, J.; Jin, H.; Botstein, D.; Cherry, J.M.; Sherlock, G. GO::TermFinder—Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **2004**, *20*, 3710–3715. [[CrossRef](#)] [[PubMed](#)]
66. Python Software Foundation. *Python Language Reference*, Version 3.14; Python Software Foundation: Wilmington, DE, USA, 2025. Available online: <https://www.python.org> (accessed on 7 October 2025).
67. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
68. The Pandas Development Team. *Pandas-Dev/Pandas: Pandas, (Latest Version)*; Zenodo: Geneva, Switzerland, 2020. [[CrossRef](#)]
69. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56.
70. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)] [[PubMed](#)]
71. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
72. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
73. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
74. Yemata, G.; Bekele, T. Evaluation of sesame (*Sesamum indicum* L.) varieties for drought tolerance using agromorphological traits and drought tolerance indices. *PeerJ* **2024**, *12*, e16840. [[CrossRef](#)] [[PubMed](#)]
75. ten Harkel, J.; Bartholomeus, H.; Kooistra, L. Biomass and Crop Height Estimation of Different Crops Using UAV-Based Lidar. *Remote Sens.* **2020**, *12*, 17. [[CrossRef](#)]
76. Ziliani, M.G.; Parkes, S.D.; Hoteit, I.; McCabe, M.F. Intra-Season Crop Height Variability at Commercial Farm Scales Using a Fixed-Wing UAV. *Remote Sens.* **2018**, *10*, 2007. [[CrossRef](#)]
77. Boureima, S.; Ouakroum, A.; Diouf, M.; Cisse, N.; Van Damme, P. Screening for drought tolerance in mutant germplasm of sesame (*Sesamum indicum*) probing by chlorophyll *a* fluorescence. *Environ. Exp. Bot.* **2012**, *81*, 37–43. [[CrossRef](#)]

78. Kyratzis, A.C.; Skarlatos, D.P.; Menexes, G.C.; Vamvakousis, V.F.; Katsiotis, A. Assessment of Vegetation Indices Derived by UAV Imagery for Durum Wheat Phenotyping under a Water Limited and Heat Stressed Mediterranean Environment. *Front. Plant Sci.* **2017**, *8*, 01114. [[CrossRef](#)] [[PubMed](#)]
79. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [[CrossRef](#)] [[PubMed](#)]
80. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillerá-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
81. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.M.; Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **2018**, *6*, e5518. [[CrossRef](#)] [[PubMed](#)]
82. Dossa, K.; Diouf, D.; Wang, L.; Wei, X.; Zhang, Y.; Niang, M.; Fonceka, D.; Yu, J.; Mmadi, M.A.; Yehouessi, L.W.; et al. The Emerging Oilseed Crop *Sesamum indicum* Enters the “Omics” Era. *Front. Plant Sci.* **2017**, *8*, 01154. [[CrossRef](#)] [[PubMed](#)]
83. Rebetzke, G.J.; Jimenez-Berni, J.A.; Bovill, W.D.; Deery, D.M.; James, R.A. High-throughput phenotyping technologies allow accurate selection of stay-green. *J. Exp. Bot.* **2016**, *67*, 4919–4924. [[CrossRef](#)] [[PubMed](#)]
84. Smith, A.B.; Cullis, B.R.; Thompson, R. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Sci.* **2005**, *143*, 449–462. [[CrossRef](#)]
85. Kar, B.; Sahoo, T.; Nath, B.; Swain, S.S.; Pradhan, B.K.; Pradhan, C.; Rout, G.R.; Panda, D. Induction of genetic variability through chemical and physical mutagenesis for sesame crop improvement. *Open Agric.* **2023**, *8*, 20220178. [[CrossRef](#)]
86. Deery, D.; Jimenez-Berni, J.; Jones, H.; Sirault, X.; Furbank, R. Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* **2014**, *4*, 349–379. [[CrossRef](#)]
87. Sabag, I.; Morota, G.; Peleg, Z. Genome-wide association analysis uncovers the genetic architecture of tradeoff between flowering date and yield components in sesame. *BMC Plant Biol.* **2021**, *21*, 549. [[CrossRef](#)] [[PubMed](#)]
88. Anderegg, J.; Zenkl, R.; Walter, A.; Hund, A.; McDonald, B.A. Combining High-Resolution Imaging, Deep Learning, and Dynamic Modeling to Separate Disease and Senescence in Wheat Canopies. *Plant Phenomics* **2023**, *5*, 0053. [[CrossRef](#)] [[PubMed](#)]
89. Cobb, J.N.; DeClerck, G.; Greenberg, A.; Clark, R.; McCouch, S. Next-generation phenomics: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* **2013**, *126*, 867–887. [[CrossRef](#)] [[PubMed](#)]

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.