Received: 6 April 2025 | Accepted: 17 March 2026

DOI: 10.1002/ppj2.70079

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Drone-based phenotyping of maize for multiple disease resistance and yield in breeding field trials

Danilo E. Moreta[1] | **Nina Blahut**[2] | **Nicholas Kaczmar**[1] | **Judith M. Kolkman**[1] | **Gary Bergstrom**[1] | **Margaret Smith**[1] | **Michael A. Gore**[1] | **Rebecca Nelson**[1]

[1]School of Integrative Plant Science, Cornell University, Ithaca, New York, USA

[2]Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York, USA

Correspondence
Rebecca Nelson and Danilo E. Moreta, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA.
Email: rjn7@cornell.edu and dem324@cornell.edu

Assigned to Associate Editor Jennifer Lachowiec.

Funding information
U.S. Department of Agriculture, National Institute of Food and Agriculture, Hatch Accession, Grant/Award Number: #1020798; Genomes to Fields (G2F) Initiative

Abstract

Improving selection for multiple disease resistance (MDR) and yield in maize (*Zea mays* L.) requires high-throughput, objective phenotyping tools, particularly under field conditions where several foliar diseases co-occur. We evaluated drone-based multispectral vegetation indices (VIs) for predicting resistance to northern leaf blight (NLB; inoculated), northern leaf spot (NLS; natural), anthracnose top dieback (ATD; natural), and for predicting grain yield across 2 years in near-isogenic inbreds, near-isogenic hybrids, and a diverse hybrid panel. VIs showed lower coefficients of variation but broad-sense heritability ranging from 0.09 to 0.95, compared with 0.30 to 0.99 for visual disease scores and 0.21 to 0.97 for yield. Correlations between VIs and ground traits were strongest in near-isogenic hybrids, particularly for early-season yield prediction ($r = 0.98$–0.99 in 2018; $r = 0.87$–0.90 in 2019), and moderate for total disease severity (e.g., $r = -0.61$ to -0.68 in 2018). Associations were weaker and less consistent in the diverse hybrid panel (yield $r = 0.11$–0.28). Disease-specific signals were temporally structured: NLS correlated most strongly with early-season VIs ($r = -0.59$ to -0.75), whereas ATD was best detected mid-season ($r = -0.63$ to -0.66) along with NLB ($r = -0.66$ to -0.75). Overall, multispectral VIs captured meaningful canopy variation related to MDR and yield, with predictive performance depending on germplasm structure and flight timing. These findings highlight the potential of drone-based temporal phenotyping to complement visual assessments and improve selection efficiency in maize breeding programs.

Abbreviations: ATD, anthracnose top dieback; BLUP, best linear unbiased predictor; CV, coefficient of variation; DAP, days after planting; DLA, diseased leaf area; G2F, Genomes to Fields; GY_{raw}, raw grain yield; MCARI, modified chlorophyll absorption in reflectance indices; MDR, multiple disease resistance; NDVI, normalized difference vegetation index; NIL, near-isogenic line; NLB, northern leaf blight; NLS, northern leaf spot; OSAVI, optimized soil-adjusted vegetation index; QTL, quantitative trait locus; rAUDPC, relative area under the disease progress curve; RDVI, renormalized difference vegetation index; tDS, total disease severity; VI, vegetation index.

Plain Language Summary

Maize (*Zea mays* L.) plants often face several leaf diseases at the same time, making it hard for breeders to choose the best varieties. In this study, we used drones to take repeated images of maize fields and measured simple color- and light-based indicators from the plants. We compared these drone measurements with traditional field ratings of three major diseases and with grain yield. The drone data were highly repeatable and were able to predict disease levels and yield well, in some cases even before symptoms were easy to see. These results show that drone technology can help breeders evaluate plants faster, more accurately, and with less labor, especially when multiple diseases occur together.

1 | INTRODUCTION

Sometimes, "we see things not as they are, but as we are. Because it is the 'I' behind the 'eye' that does the seeing." This aphorism by Anaïs Nin provides a metaphor for the difficulty of measuring a trait objectively only by visual observation. In plant breeding, screening for disease resistance traditionally relies on visual assessment of field-grown lines. While conventional and accepted, this method is time-consuming and prone to inaccuracies from observer subjectivity, limiting its precision and scalability (Bock et al., 2008, 2009, 2020; Poland & Nelson, 2011). Such imprecision can substantially affect genetic analyses, causing large variations in quantitative trait locus (QTL) effect estimates (Poland & Nelson, 2011). Furthermore, these methods typically target single diseases, despite the well-recognized importance of multiple disease resistance (MDR) for robust varietal performance (Wiesner-Hanks & Nelson, 2016).

In field trials, plots artificially inoculated with a target pathogen are frequently colonized by additional, naturally occurring diseases. Under such conditions, visual severity estimates can become less accurate or precise, as co-occurring pathogens create overlapping symptoms and complex spatiotemporal disease dynamics that challenge rater consistency. These complications are further amplified by the fact that resistance (or susceptibility) to one disease may also influence responses to others (Kolkman et al., 2023). As a result, objective selection for MDR (Wiesner-Hanks & Nelson, 2016) becomes particularly difficult. Likewise, selecting for yield under disease pressure without robust disease quantification can lead to misleading conclusions about tolerance and overall performance in breeding programs.

Remote sensing detects and classifies objects based on their spectral signatures and has become a key geospatial technology for overcoming the phenotyping bottleneck in digital agriculture. When combined with machine learning, it offers breeders powerful tools to accelerate crop improvement, including for disease resistance (DeSalvio et al., 2022; Furbank & Tester, 2011; Ghosal et al., 2018; Herr et al., 2023; Shakoor et al., 2017). Spectral imaging captures reflected electromagnetic energy across visible (400–700 nm), near-infrared (700–1300 nm), and, depending on the sensor, shortwave-infrared (1400–3000 nm) wavelengths. These measurements, recorded as pixels with distinct spectral signatures, provide information on physiological and biochemical traits, including leaf water status, pigment composition, and structural biomass properties (Curran, 1989; Peñuelas & Filella, 1998; Yendrek et al., 2017). Spectral remote sensing data are often summarized using vegetation indices (VIs), which are transformations of two or more spectral bands. VIs can detect subtle changes in plant health that are not readily visible to the human eye, making them valuable tools for monitoring stress responses. However, because biotic and abiotic stresses can trigger similar spectral signatures, remote sensing-based assessments must carefully account for potential confounding among stressors (Ghosal et al., 2018; Simko et al., 2017).

Unmanned aerial vehicles (UAVs, or drones) equipped with sensors enable rapid data collection over large field trials and across phenological stages (Ashapure, Jung, Chang, et al., 2019; Ashapure, Jung, Yeom, et al., 2019; Chang et al., 2017; Dash et al., 2017; Enciso et al., 2019; Jung et al., 2018; Mahlein, 2016; Yeom et al., 2018; Washburn et al., 2024). Drone-based spectral remote sensing has been successfully deployed for high-throughput phenotyping of key agronomic traits such as yield, plant height, and biomass across several crop species (Gano et al., 2024).

Spectral analysis is increasingly being explored for disease phenotyping across crop species (Mahlein, 2016; Shakoor et al., 2017; Simko et al., 2017; Song et al., 2021). Hyperspectral approaches have achieved high accuracy in distinguishing healthy from diseased tissue and in quantifying severity at both leaf and field scales in several pathosystems, including *Cercospora* leaf spot, rust, and powdery mildew in sugar beet (*Beta vulgaris* L.) (Mahlein et al., 2012; Rumpf et al., 2010), root rot (*Phymatotrichopsis omnivora*) in cotton (*Gossypium*

hirsutum L.) (Yang et al., 2010), and stripe rust (*Puccinia striiformis*) in wheat (*Triticum aestivum* L.) (Devadas et al., 2015). More affordable multispectral imaging technologies have also been used to monitor crop health, with successful applications in detecting powdery mildew and leaf rust in wheat (Franke & Menz, 2007), Huanglongbing (*Candidatus Liberibacter* spp.) in citrus (Garcia-Ruiz et al., 2013; Sankaran et al., 2013), and cotton root rot (Yang et al., 2010).

Few drone-based disease phenotyping studies have been conducted in maize (*Zea mays* L.), and they cover only a limited set of pathosystems, including the tar spot complex (TSC) (Loladze et al., 2019; Oh et al., 2021), maize streak virus (Chivasa et al., 2021), northern leaf blight (NLB) (DeChant et al., 2017; Stewart et al., 2019; Wiesner-Hanks et al., 2018), and rust (DeSalvio et al., 2022). For TSC and maize streak virus, high-altitude drone flights combined with plot-level spectral VIs effectively distinguished diseased from healthy canopies and showed strong agreement with ground disease scores and yield. However, these studies were conducted under highly controlled conditions (large four-row plots, a single manually inoculated pathogen, drone flights aligned with ground scoring, and fungicide treatment of non-inoculated plots), which are uncommon in routine breeding trials. By contrast, NLB studies have relied on machine learning to detect individual lesions from low-altitude imagery, an approach that is currently impractical for plot-level high-throughput phenotyping.

The use of drone-derived VIs has enabled accurate prediction of both maize diseases and yield. For example, multi-temporal VIs have been used to predict maize streak virus (Chivasa et al., 2021) and southern rust (DeSalvio et al., 2022) before major symptoms were visible. The TSC has also been predicted effectively using multi-temporal VIs (Oh et al., 2021) and integrated variables derived from these time series (Loladze et al., 2019). Notably, the same VIs used as disease predictors also exhibit strong predictive power for grain yield. In non-disease maize trials, multispectral VIs explain a substantial proportion of yield variation across environments (Guo et al., 2024; Ramos et al., 2020). Recent evidence further shows that high-temporal-resolution UAV phenotyping enhances predictive performance, whether leveraging full-season VI trajectories or a limited number of flights at key phenological stages such as pre-flowering (Washburn et al., 2024).

Despite progress in remote sensing-based phenotyping of individual plant diseases, applications targeting mixed foliar infections and naturally occurring disease complexes remain largely unexplored. In this study, we evaluated the potential of drone-based multispectral remote sensing to phenotype resistance to three co-occurring maize foliar diseases: artificially inoculated northern leaf blight (NLB_{art}), naturally occurring northern leaf spot (NLS_{nat}), and anthracnose top dieback (ATD_{nat}). We assessed the ability of drone-derived VIs to predict single and MDR as well as grain yield, using both single-time-point and temporally-averaged VIs from two flights. We further examined performance across three germplasm groups: near-isogenic inbred lines, near-isogenic hybrid lines, and a diverse hybrid panel field-tested at one location over two years. All evaluations were conducted under realistic breeding conditions, using two-row field nurseries without fungicide treatments and with asynchronous timing between drone flights and ground-based visual assessments.

Core Ideas

- Drone-based imaging captured diseases and yield reliably across germplasm and environments.
- Temporal drone data predicted disease severity and yield with strong correlations to field scores.
- Near-isogenic lines showed clearer disease signal and higher trait heritability than diverse hybrids.
- Multiple vegetation indices provided consistent yield prediction and time-specific disease detection.
- Drone phenotyping enables efficient selection for multiple disease resistance in breeding programs.

2 | MATERIALS AND METHODS

2.1 | Germplasm

Near-isogenic inbred and hybrid lines of maize (*Z. mays* L.) and a panel of diverse hybrids were field-tested for NLB resistance and grain yield in Aurora, NY, in 2018 and 2019. The near-isogenic line (NIL) inbreds were selected from a characterized introgression library consisting of a collection of 412 BC5F4 lines, created from backcrosses between 18 diverse nested association mapping founder lines (donors) and maize inbred line B73 (recurrent parent) (Kolkman et al., 2020). For the 2018 field trial, we selected 10 NILs based on both phenotypic criteria (specifically with statistically significantly more resistance or susceptibility to NLB than B73) and genotypic criteria (selecting lines with introgressions carrying certain loci that had been identified as NLB QTL; Kolkman et al., 2020; Poland et al., 2011). The 10 selected NIL inbreds carried an average introgression size of 1.14% of the donor maize genome (Kolkman et al., 2020). The single-cross NIL hybrids were developed by crossing each of the 10 NIL inbreds to a single tester (inbred line LH82). This tester was chosen because it had shown good heterosis with B73 and is important as a founder inbred line used in commercial maize breeding programs (Z. Li et al., 2018). For the 2019 field sea-

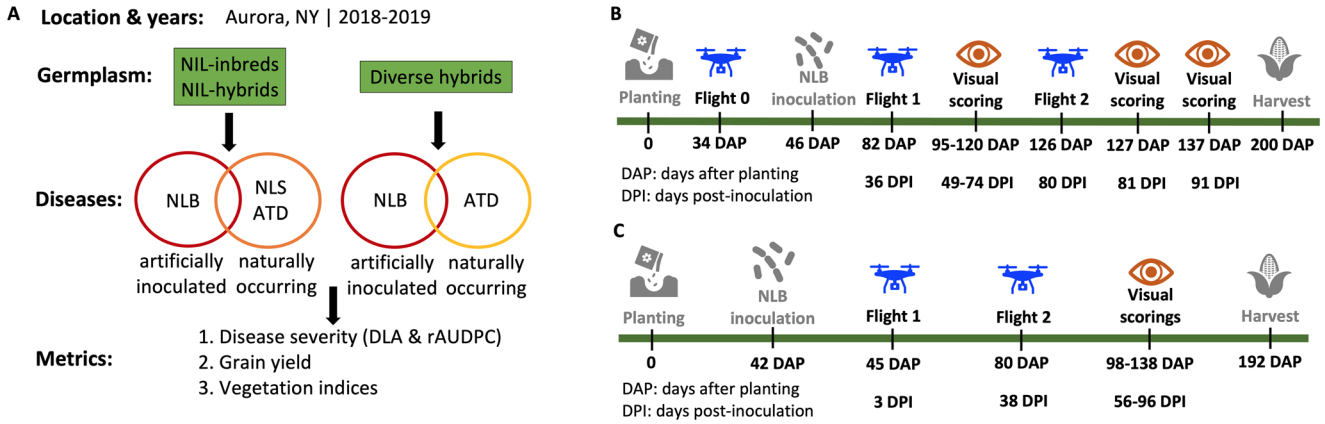

FIGURE 1 Experimental overview and research flow of ground-based visual assessments and high-throughput screening of maize foliar diseases throughout the 2018 and 2019 growing seasons. (A) Types of germplasm (mostly different lines between years within each type), diseases, and measurements. Relative timing of the asynchronous ground scoring of diseases and drone flights (remote sensing) (B) in 2018 and (C) in 2019. The Genomes to Fields (G2F) hybrids were inoculated with NLB 5 days earlier than the near-isogenic lines (NILs) in 2019 (not shown). ATD, anthracnose top dieback; DAP, days after planting; DLA, diseased leaf area; DPI, days post-inoculation; NLB, northern leaf blight; NLS, northern leaf spot; rAUDPC, relative area under the disease progress curve.

son, we selected NILs with introgressions that harbored four key NLB QTLs based on the field evaluation results of the previous year. We tested two distinct alleles per locus in 2019 for a total of eight NIL inbreds plus a subset of five NIL hybrids, mostly with a single donor-specific NLB QTL.

The diverse hybrid panel comprised a set of 250 hybrids under the Genomes to Fields (G2F) initiative and were categorized as the NY1 G2F population (AlKhalifah et al., 2018). The complete G2F hybrid panel was also field-tested for NLB resistance and yield both in 2018 and 2019, with some overlapping of hybrids among years. The G2F modern hybrids had more genetic diversity compared to the NILs (inbreds and hybrids). The G2F hybrid germplasm was generally composed of uniform taller plants with denser canopies than the NIL germplasm.

2.2 | Experimental design and implementation

The study was originally designed for simultaneous disease and yield evaluations using ground-based visual assessments of foliar diseases. We took advantage of this established experiment to explore the feasibility of drone-based high-throughput phenotyping to measure foliar disease severities (artificially inoculated and naturally occurring) and yield under disease pressure (Figure 1A).

The trials were planted in late May 2018 and early June 2019 at Cornell University's Robert B. Musgrave Research Farm in Aurora, NY (43 43′ 44.53″ N, 76 39′ 08.28″ W) at 261 m above sea level. Farm soils are silt loam with a pH of 7.3–8.0. The NIL inbreds and the NIL hybrids were planted in two side-by-side blocks following an alpha lattice experimental design with three randomized replicates in 2018 and four in 2019. The G2F diverse hybrid population of 250 hybrid lines was planted in a modified randomized complete block design with two replicates in both 2018 and 2019. The NILs and the G2F diverse hybrids were planted in contiguous areas in the same field.

The experimental unit for both germplasm groups was a two-row plot with two 6.40-m-long rows spaced 0.76 m apart. Spacer rows were used to separate the different field sections and germplasm groups (NILs v. G2F diverse hybrids), the NIL sets (NIL inbreds v. NIL hybrids), and the NIL conditions (inoculated v. uninoculated). No fungicides were applied to the uninoculated experimental plots. The fertilization regime was the same for all plots. The NILs were irrigated as required, while the G2F hybrids were rainfed. Grain yield was measured from each two-row plot using a plot combine harvester, adjusted to a standard grain moisture content of 15.5%, and then expressed in metric tons (t) per hectare per year.

2.3 | Inoculation, natural infections, and disease assessment

Setosphaeria turcica inoculum was prepared as described by Chung et al. (2010). Except for the uninoculated block of the NILs, all germplasm sets were inoculated with the NLB pathogen *S. turcica* race 1, isolate NY001. Individual plants were inoculated at the V7–V8 leaf stage in 2018 (46 days after planting [DAP]) and V6–V7 leaf stage in 2019 (42 DAP). The NILs were inoculated only with liquid inoculum in 2018 (0.5 mL of spore suspension, 4×10^3 conidia per mL, 0.02% Tween 20) and with both liquid and solid (1/4 teaspoon, ~1.3 mL of colonized sorghum grains) inoculum in 2019. The

G2F hybrids were inoculated only with solid inoculum in both years. When both were applied, the liquid and solid inoculum were placed in the whorl of each plant a day apart. Water was applied using a sprinkler only over the NILs 1 day after inoculation to promote NLB infection in the inoculated plants.

Two naturally occurring fungal foliar diseases were the predominant diseases in the inoculated NIL plots and in the uninoculated plots in 2018 and 2019: northern leaf spot (NLS, caused by *Cochliobolus carbonum* race 3) and anthracnose top dieback (ATD, caused by *Colletotrichum graminicola*). ATD also prevailed in the NLB-inoculated G2F plots in both years, while NLS symptoms were not observed in the G2F plots. ATD in the experimental plots was associated with an ATD epidemic that affected much of New York in September 2018 (Cummings, 2018) and that was observed in the surrounding farmers' fields.

Disease severity was visually assessed on a plot basis and scored independently for NLB, NLS, and ATD. Diseased leaf area (DLA) was recorded at up to three temporally spaced assessments per growing season to capture disease progression, except in the 2019 G2F hybrid trial, where ATD was scored twice due to late-season phenotypic constraints. Ratings were conducted at 10- to 16-day intervals post-anthesis, when symptoms of artificially inoculated NLB and naturally occurring NLS and ATD were prevalent. Disease severity was quantified as percent DLA (0%–100%) at each assessment. Repeated DLA measurements were summarized in two ways: (i) a disease severity index calculated as the mean of the available DLA scores, and (ii) the relative area under the disease progress curve (rAUDPC), which integrates disease intensity over time. Higher DLA and rAUDPC values indicate greater disease severity. Spatially corrected best linear unbiased predictors (BLUPs) were estimated from DLA-derived traits and rAUDPC values and used in subsequent analyses. rAUDPC was calculated following Fry (1978), with minor modifications, as follows:

$$\text{rAUDPC}_{\text{DLA}} = \frac{1}{100\ (t_n - t_1)} \sum_{i=1}^{n-1} \frac{y_i + y_{i+1}}{2}\ (t_{i+1} - t_i) \tag{1}$$

where y_i = DLA rating at time i, with i = 1–3 ratings; t_i = DLA rating day; $t_{i+1} - t_i$ = day interval between two DLA ratings; and n = number of ratings. rAUDPC values were then multiplied by 100 to express them as percentages.

We also calculated total disease severity (tDS) on a plot basis by summing either average DLA ratings or rAUDPCs of NLB, NLS, and ATD based on the following equation:

$$\text{tDS}\ (\%) = \text{NLB}_{\text{art}} + \text{NLS}_{\text{nat}} + \text{ATD}_{\text{nat}} \tag{2}$$

where tDS = total disease severity expressed in either percent DLA or percent rAUDPC, NLB_{art} = artificially inoculated northern leaf blight, NLS_{nat} = naturally occurring northern leaf spot, and ATD_{nat} = naturally occurring anthracnose top dieback. Therefore, tDS represents the cumulative quantification of different foliar diseases that overlapped in space and time (Supporting Information S1).

2.4 | Aerial image acquisition and data processing

As the established field trials were not initially intended for drone-based disease phenotyping, drone flights were conducted at arbitrary dates over the contiguous NIL and G2F field experiments in 2018 (Figure 1B) and 2019 (Figure 1C). The timing of these flights did not coincide with the timing of the ground-based visual estimates of each disease due to logistical constraints.

The flights were carried out using a DJI Matrice 600 unoccupied aerial systems (DJI) weighing ~10 kg with a camera and batteries. Flight plans were designed using Mission Planner (ArduPilot, https://ardupilot.org/) and executed using Litchi (https://flylitchi.com/). The drone flew a total of 11 times every 6–18 days in 2018 and 13 times every 4–14 days in 2019. Each flight lasted a maximum of 20 min at a cruise speed of 2 m/s. The flights were conducted under sunny conditions at 20 and 25 m above the ground, covering and extending beyond an experimental area of 0.87 ha in 2018 and 0.98 ha in 2019. Eight ground control points, measured using a Trimble RTK-GPS, were utilized for georeferencing and accuracy.

The experimental area contained a total of 644 experimental two-row plots in 2018 and 668 plots in 2019 (Supporting Information S2). The aerial images were acquired using the RedEdge-M multispectral camera (MicaSense) equipped with five individual sensors that captured blue (475 nm, 20 nm full-width at half maximum [FWHM]), green (560 nm, 20 nm FWHM), red (668 nm, 10 nm FWHM), red-edge (717 nm, 10 nm FWHM), and near infrared (840, 40 nm FWHM). Images were collected automatically based on GPS positions to give 75% forward overlap and 75% sidelap to ensure high coverage of images. We used the spectral information of a small subset of the flights (three in 2018 and two in 2019) based on flight and image quality. Thus, multispectral UAV imagery was acquired from at least two growth stages during each growing season to capture temporal variation in canopy condition.

The raw multispectral images were processed using the Pix4Dmapper software (v3.3.24; Pix4D) (https://www.pix4d.com/) to obtain composite geometrically corrected images (orthomosaics). The QGIS software (v2.2) (https://www.qgis.org/en/site/) was used to create rectangular boundaries to define experimental plots that were captured in the orthomosaics. R scripts from the FIELDimageR package (https://www.opendronemap.org/fieldimager/) were used to extract

spectral information from the orthomosaics. Spectral information was then aggregated at the plot-level based on the rectangular plot boundaries drawn in QGIS. In this study, we selected and extracted five plot-level multispectral VIs—normalized difference vegetation index (NDVI) (Rouse et al., 1973), renormalized difference vegetation index (RDVI) (Roujean & Breon, 1995), optimized soil-adjusted vegetation index (OSAVI) (Rondeaux et al., 1996), and modified chlorophyll absorption in reflectance indices (MCARI1 and MCARI2) (Haboudane et al., 2004) (Supporting Information S3)—to comprehensively capture the structural and physiological impacts of NLB, NLS, and ATD on the maize canopy. These indices quantify canopy greenness, leaf area development, and pigment degradation, key indicators of photosynthetic integrity and overall plant health under biotic stress. NDVI and RDVI capture variations in canopy density and photosynthetic activity, while OSAVI minimizes soil background influence, improving sensitivity in heterogeneous canopies. The MCARI1 and MCARI2 indices were included for their direct sensitivity to chlorophyll content, allowing for the detection of chlorosis, a primary symptom of the diseases under investigation. For temporal integration, VI values from individual flights were averaged to represent overall canopy status during the disease assessment period. To account for spatial field heterogeneity, spatially corrected BLUPs were estimated from raw VI data and used in all subsequent analyses.

Our choice of these specific VIs is strongly supported by the work of Loladze et al. (2019), who successfully deployed the same set of indices to phenotype TSC resistance in maize. In their study, these indices demonstrated a strong relationship with the area under the disease progress curve (AUDPC) of TSC, proving their efficacy in quantifying disease severity and resistance in a maize breeding context. This precedent validates our approach of using a targeted, multi-index framework to dissect the complex canopy changes caused by NLB, NLS, and ATD.

2.5 | Statistical analysis

All analyses were conducted in R (R Core Team, 2020) using RStudio (RStudio Team, 2022). Spatial correction of field phenotypes was performed with the SpATS package (Rodríguez-Álvarez et al., 2024), which implements the PS-ANOVA P-spline methodology of Rodríguez-Álvarez et al. (2018). For the NIL trial, inbreds and hybrids were analyzed together within each year with replication and germplasm type as fixed effects, whereas G2F hybrids were analyzed separately by year. Spatially corrected genotype BLUPs from SpATS were used for heritability estimation and correlations between ground-truth traits and VIs. For each plot i, the spatial mixed model fitted was as follows:

$$y_i = \mu + g_{k(i)} + f(x_i, z_i) + R_{r(i)} + C_{c(i)} + e_i \quad (3)$$

where y_i denotes the phenotype, μ is the overall mean, $g_{k(i)}$ is the random effect of genotype k, $R_{r(i)}$ and $C_{c(i)}$ are random row and column effects, $f(x_i, z_i)$ is a two-dimensional smooth surface modeled using tensor-product P-splines to account for spatial heterogeneity across the field, and e_i is the residual effect.

Broad-sense heritability was estimated using the generalized definition of Cullis et al. (2006):

$$H^2 = 1 - \frac{\bar{v}_\Delta}{2\,\sigma_g^2} \quad (4)$$

where σ_g^2 is the genetic variance and $\bar{v}_\Delta$ is the average prediction error variance of pairwise BLUP differences, as returned by the SpATS heritability() function.

The spatial mixed model was fitted separately for each flight (and its corresponding VIs) and for each individual disease assessment. We also applied the model to aggregated trait values, including mean DLA (from 2 to 3 ratings), AUDPC (from three ratings), mean VIs (from two flights), and grain yield. Broad-sense heritability was then estimated for both the individual time-point data and the averaged traits.

The coefficient of variation (CV) was calculated from the replicated raw values of each trait to assess the consistency of ground-based disease ratings, grain yield, and VIs. Pearson correlation coefficients (r) were computed using spatially corrected BLUPs from both individual time points and averaged traits to quantify the relationships between ground-based disease severity, grain yield, and drone-derived VIs. Ground-based severity inputs included BLUPs for DLA and rAUDPC for each disease (NLB_{art}, NLS_{nat}, or ATD_{nat}) as well as the total foliar disease metric (Equation 2). Statistical significance of the correlations was declared based on a p-value $\leq \alpha$ ($\alpha = 0.05$).

3 | RESULTS

3.1 | Consistency and heritability of ground-based disease scorings, yield, and vegetation indices

Across both years, VIs exhibited substantially lower CVs than the conventional phenotypic metrics (DLA and raw grain yield [GY_{raw}]) in the NIL germplasm. In 2018, the mean CV of VIs across inbreds and hybrids was 7%, increasing slightly to 9% in 2019. In contrast, DLA displayed much higher variability, with mean CVs of 43% in 2018 and 62% in 2019, while GY_{raw} showed intermediate CVs of 35% and 26%, respectively (Figure 2). A similar pattern was observed in the G2F diverse hybrid panel. In 2018, VIs showed a lower overall

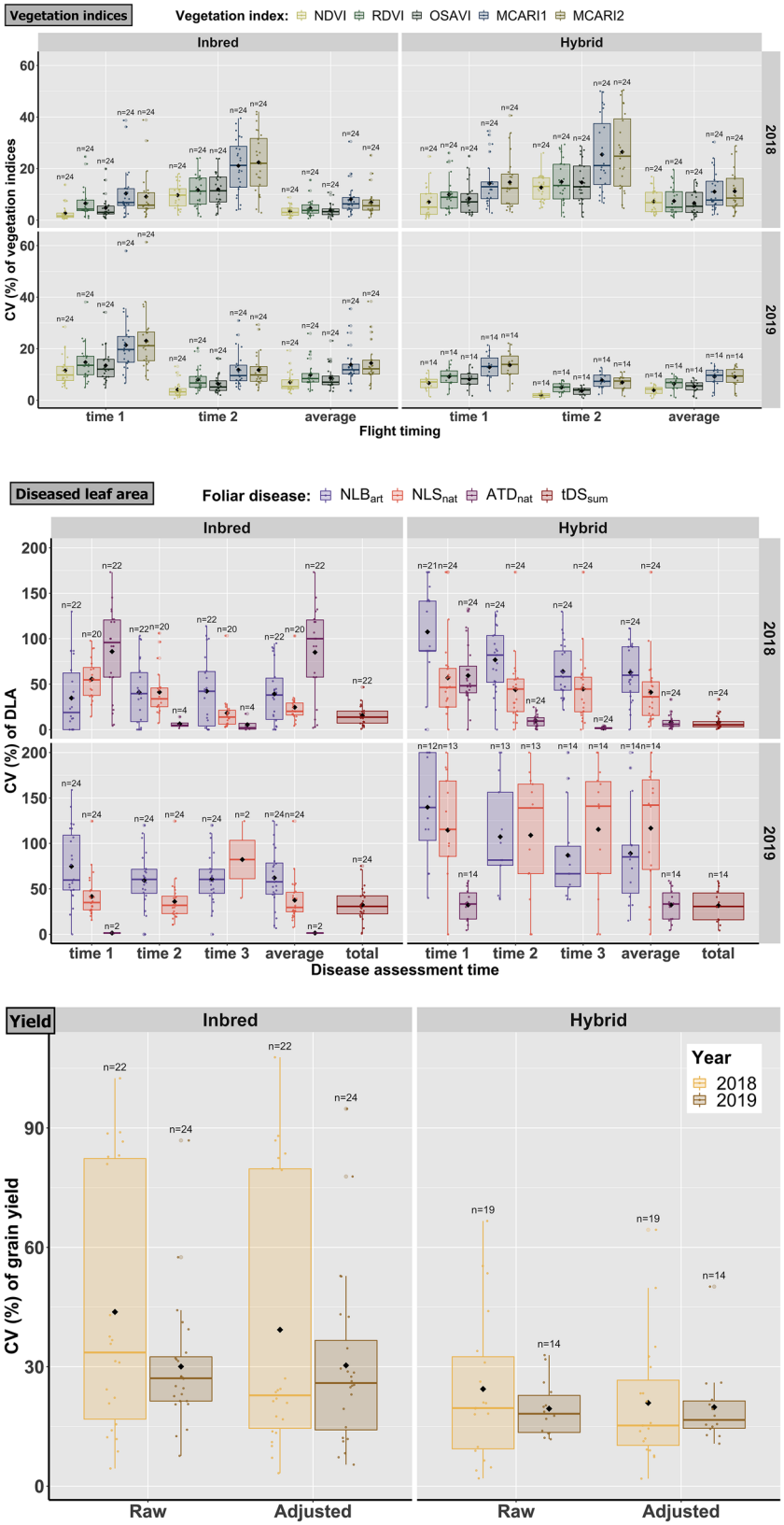

FIGURE 2 Distribution of coefficients of variation (CVs) for phenotypic traits in the near-isogenic line (NIL) germplasm. Boxplots represent the CV for vegetation indices (VIs), diseased leaf area (DLA), and grain yield across inbreds and hybrids evaluated in 2018 and 2019. VIs and DLA distributions include measurements from multiple time points and their respective seasonal averages. Evaluated diseases include artificially inoculated northern leaf blight (NLB_{art}), naturally occurring northern leaf spot (NLS_{nat}), and naturally occurring anthracnose top dieback (ATD_{nat}).

(Continues)

FIGURE 2 (Continued)
Total disease severity (tDS_{sum}) represents the sum of NLB_{art}, NLS_{nat}, and ATD_{nat}. Calculated VIs include NDVI (normalized difference vegetation index), RDVI (renormalized difference vegetation index), OSAVI (optimized soil-adjusted vegetation index), and MCARI1 and MCARI2 (modified chlorophyll absorption in reflectance indices). Adjusted grain yield was included as a control. Black diamonds indicate mean CV values; total observation counts (n) for each trait are provided above the respective plots.

mean CV (21%) than DLA (46%) but a higher CV than GY_{raw} (15%). In 2019, the VIs again had the lowest variability, with a mean CV of 7%, compared with 37% for DLA and 14% for GY_{raw} (Figure 3).

At the individual time points assessed in both germplasm panels and across both years, the mean CVs of the VIs consistently remained lower than those of the corresponding DLA measurements, further underscoring the greater stability of the VI-based assessments over time. Together, these results demonstrate that multispectral VIs captured by UAV platforms provide markedly more stable phenotypic signals than both DLA scores and raw yield measurements, underscoring their potential as reliable, low-noise indicators for high-throughput disease and yield phenotyping across germplasm and years.

Broad-sense heritability (H^2) estimates from spatially corrected BLUPs showed clear differences between ground-measured traits and drone-derived VIs across years and germplasm groups. In both the NIL and G2F panels, averaged and aggregated disease phenotypes (NLB, NLS, ATD, tDS) and grain yield generally exhibited higher or comparable heritability than the average VIs (Tables 1 and 2), indicating stronger genetic signal in the visually scored traits. Exceptions included NLB in the 2018 NIL trial and GY_{raw} in the 2019 G2F trial, where several VIs matched or slightly exceeded ground-trait heritability, demonstrating that multispectral indices can sometimes capture genotype-specific canopy stress with similar reliability. Across both years, NIL germplasm typically showed higher heritability than the G2F panel.

Patterns at individual time points mirrored those of the averaged and aggregated traits. In the NIL germplasm, VIs from flight 1 in 2018 and flights 1–2 in 2019 showed heritability similar to the individual DLA scores, whereas VIs from flight 2 in 2018 declined. In the G2F germplasm, individual DLA scores consistently outperformed VIs across flights in both years. Early-season VIs (flight 0, 2018) showed heritability comparable to DLA scores and were higher in the NIL than the G2F panel. VIs derived from the late-season flight at 82 DAP in 2018 exhibited the lowest heritability estimates (H^2 = 0.09–0.14) within the G2F diverse hybrid panel. This pattern suggests a reduced genetic signal at later growth stages, likely due to increased environmental noise and canopy saturation, particularly in genetically heterogeneous germplasm. In 2019, VIs from the comparable late-season flight at 80 DAP showed moderately higher heritability (H^2 = 0.26–0.34), but values remained lower than those observed at earlier time points. Heritability estimates of 1.0 for NLS_{DLA3} and NLS_{rAUDPC} in the 2019 NIL germplasm were considered model artifacts and excluded from biological interpretation.

Overall, these patterns indicate that while ground-based disease scores and yield traits generally retain stronger genetic signal, multispectral VIs (particularly at early developmental stages) can at times provide heritability comparable to conventional phenotyping, underscoring their targeted value for capturing genotype-specific canopy responses.

3.2 | Relationship of yield and disease severity with vegetation indices

Across both years, correlations between VIs and ground-measured traits were generally stronger in the NIL germplasm (Table 3) than in the G2F panel (Table 4), likely due to the more homogeneous genetic background of the NIL inbreds/hybrids, in contrast to the highly diverse G2F hybrids. Correlation magnitudes varied across years, flight times (crop phenological stages), and diseases (NLB_{art}, NLS_{nat}, ATD_{nat}, and tDS), indicating that both temporal and disease-specific factors shaped the strength of VI–ground trait associations.

VI–yield correlations varied markedly across years, germplasm groups, and flight times (Tables 3 and 4). In 2018, the strongest associations were observed during the early flight 0 (34 DAP), with moderate correlations in NIL inbreds (r = 0.66–0.73) and very strong correlations in NIL hybrids (r = 0.98–0.99); correlations in the G2F panel were weaker (r = 0.24–0.28). Later-season flights (82 and 126 DAP) showed only slightly reduced correlations in NIL hybrids and were largely nonsignificant in the G2F panel. In 2019, flight 1 (45 DAP) produced the highest correlations in NIL hybrids (r = 0.87–0.90), whereas the G2F hybrids showed only weak associations (r = 0.14–0.17). When VIs were averaged across flight 1 and flight 2, strong correlations with yield were observed only in NIL hybrids, both in 2018 (r = 0.70–0.96) and 2019 (r = 0.83–0.86). These results suggest that early-season multispectral imaging can effectively capture genotype-driven yield variation, particularly in germplasm with well-defined and uniform genetic backgrounds such as NIL hybrids, and may therefore be most informative for high-throughput yield prediction.

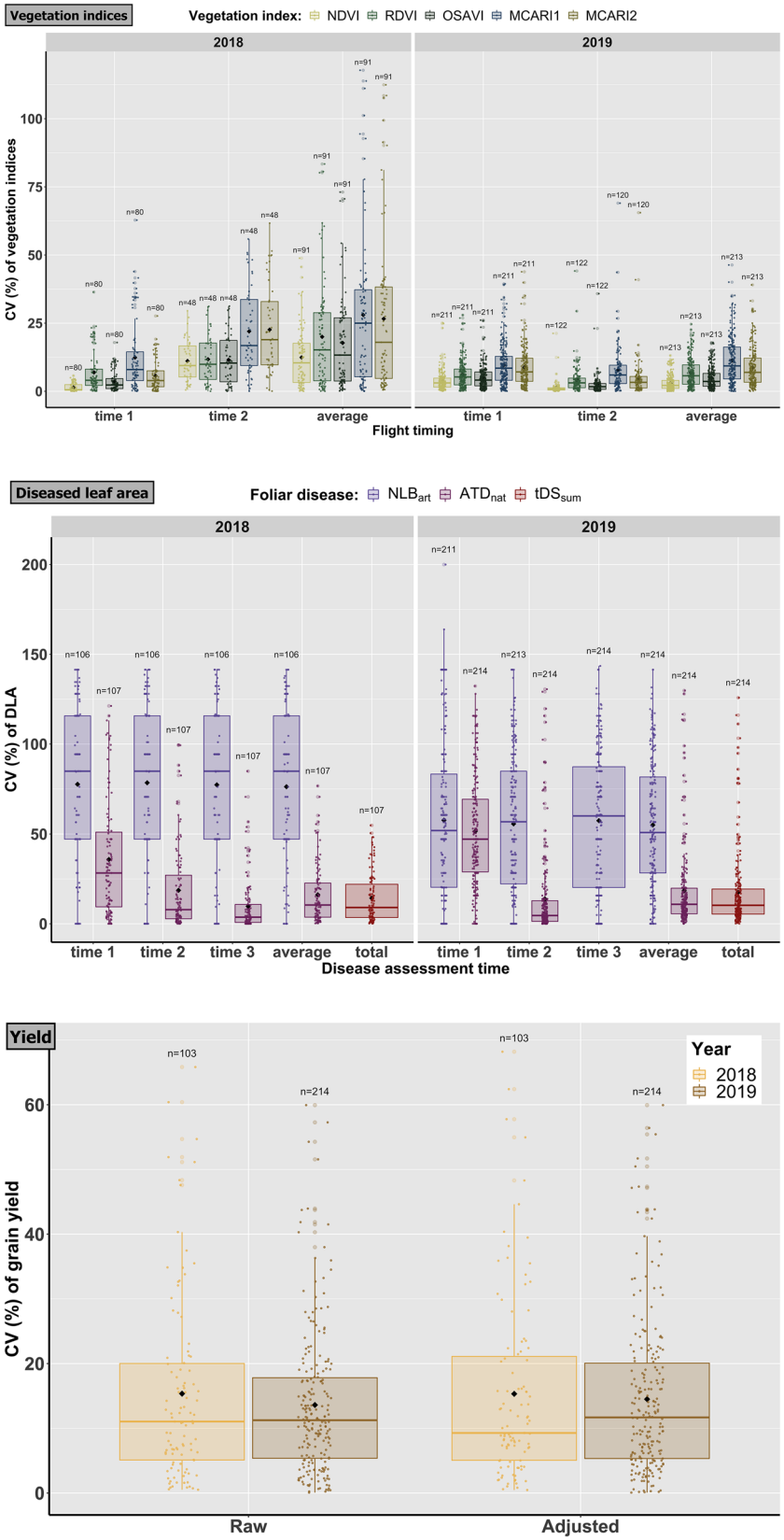

FIGURE 3 Distribution of coefficients of variation (CVs) for phenotypic traits in the Genomes To Fields (G2F) hybrid germplasm. Boxplots represent the CV for vegetation indices (VIs), diseased leaf area (DLA), and grain yield evaluated in 2018 and 2019. VIs and DLA distributions include measurements from multiple time points and their respective seasonal averages. Evaluated diseases include artificially inoculated northern leaf blight (NLB_{art}) and naturally occurring anthracnose top dieback (ATD_{nat}). Total disease severity (tDS_{sum}) represents the sum of NLB_{art} and

(Continues)

FIGURE 3 (Continued)
ATD_{nat}. Calculated VIs include NDVI (normalized difference vegetation index), RDVI (renormalized difference vegetation index), OSAVI (optimized soil-adjusted vegetation index), and MCARI1 and MCARI2 (modified chlorophyll absorption in reflectance indices). Adjusted grain yield was included as a control. Black diamonds indicate mean CV values; total observation counts (*n*) for each trait are provided above the respective plots.

TABLE 1 Broad-sense heritability (H^2) of ground-measured phenotypes (disease severity and yield) in a near-isogenic line (NIL) population of maize inbreds and hybrids and in a Genomes to Fields (G2F) population of diverse maize hybrids field tested in Aurora, NY, in 2018 and 2019.

	NY2018				NY2019			
Ground-based assessments	**DAP**	**NILs**	**DAP**	**G2F**	**DAP**	**NILs**	**DAP**	**G2F**
NLB_{DLA1}	95	0.74	113, 115	0.79	98	0.86	92, 93	0.30
NLB_{DLA2}	106	0.72	120, 122	0.70	114	0.86	105, 106	0.79
NLB_{DLA3}	116	0.72	127, 131	0.73	124	0.86	120	0.81
NLB_{DLAavg}	NA	0.72	NA	0.78	NA	0.86	NA	0.76
NLB_{rAUDPC}	NA	0.72	NA	0.76	NA	0.86	NA	0.78
NLS_{DLA1}	100	0.89	NA	NA	114	0.88	NA	NA
NLS_{DLA2}	110	0.90	NA	NA	124	0.92	NA	NA
NLS_{DLA3}	120	0.96	NA	NA	138	1.0	NA	NA
NLS_{DLAavg}	NA	0.95	NA	NA	NA	0.92	NA	NA
NLS_{rAUDPC}	NA	0.95	NA	NA	NA	1.0	NA	NA
ATD_{DLA1}	116	0.95	113, 115	0.73	138	0.86	120	0.78
ATD_{DLA2}	127	0.89	120, 122	0.76	NA	NA	134	0.90
ATD_{DLA3}	137	0.99	127, 131	0.67	NA	NA	NA	NA
ATD_{DLAavg}	NA	0.99	NA	0.81	NA	NA	NA	0.89
ATD_{rAUDPC}	NA	0.94	NA	0.81	NA	NA	NA	NA
tDS_{DLA}	NA	0.96	NA	0.80	NA	0.95	NA	0.90
tDS_{rAUDPC}	NA	0.97	NA	0.80	NA	0.74	NA	0.77
GY_{raw}	200	0.97	200	0.42	192	0.83	192	0.21
GY_{adj}	200	0.49	200	0.33	192	0.56	192	0.45

Note: NA indicates that the measurement was not recorded due to ambiguity or subjectivity in the visual assessment (e.g., coalesced lesions from two to three diseases) or due to the absence of symptoms in the field (e.g., no NLS observed in the G2F trials). Disease severity was measured using a percent estimate of diseased leaf area (DLA). The severity index per year for each disease was obtained by averaging three different temporally spaced DLA ratings.
Abbreviations: ATD, naturally occurring anthracnose top dieback; DAP, days after planting; GY_{adj}, adjusted grain yield for stand counts; GY_{raw}, raw grain yield (without adjustment for stand counts); NLB, artificially inoculated northern leaf blight; NLS, naturally occurring northern leaf spot; rAUDPC, relative area under the disease progress curve; tDS, total disease severity (sum of NLB, NLS, and ATD).

Correlations between VIs and both total and individual foliar disease severities varied across years, germplasm groups, and flight times (Tables 3 and 4). In the NIL hybrids, total disease severity (tDS)—the sum of NLB, NLS, and ATD—showed moderate negative correlations with flight 1 VIs in 2018 (82 DAP; $r = -0.64$ to -0.66) and flight 2 VIs (126 DAP; $r = -0.61$ to -0.68). In 2019, tDS was significantly associated only with NDVI from flight 2 (80 DAP; $r = -0.60$). In the G2F hybrids, tDS (NLB + ATD) displayed moderate–weak correlations with 2018 flight 2 VIs (126 DAP; $r = -0.46$ to -0.54) and weak correlations with 2019 flight 1 VIs (45 DAP; $r = -0.15$ to -0.16).

Individual diseases also showed heterogeneous association patterns with VIs. NLB severity correlated with two flight 2 VIs in the 2018 NIL inbreds (126 DAP; $r = -0.58$ to -0.59), with both flight 1 (45 DAP; $r = -0.62$ to -0.63) and flight 2 VIs (80 DAP; $r = -0.66$ to -0.75) in the 2019 NIL inbreds, and with flight 1 in the 2019 G2F hybrids where correlations were notably weaker (45 DAP; $r = -0.14$ to -0.16). NLS was associated with most early-season flight 0 VIs in the 2018 NIL inbreds (34 DAP; $r = -0.59$ to -0.61) and NIL hybrids ($r = -0.72$ to -0.75), and with most flight 1 VIs in the 2018 NIL hybrids (82 DAP; $r = -0.58$ to -0.64). ATD severity was linked to flight 0 NDVI in 2018 NILs (34 DAP; $r = -0.58$) and to flight 1 VIs (82 DAP; $r = -0.63$ to -0.66). In the 2018 G2F hybrids, ATD showed moderate correlations with flight 2 VIs (126 DAP; $r = -0.45$ to -0.52).

TABLE 2 Broad-sense heritability (H^2) of drone phenotypes (vegetation indices [VIs]) in a near-isogenic line (NIL) population of maize inbreds and hybrids and in a Genomes to Fields (G2F) population of diverse maize hybrids field tested in Aurora, NY, in 2018 and 2019.

		NY2018			NY2019		
Flight	**Drone-derived VIs**	**DAP**	**NILs**	**G2F**	**DAP**	**NILs**	**G2F**
flight 0	NDVI	34	0.93	0.58	NA	NA	NA
	RDVI		0.95	0.68		NA	NA
	OSAVI		0.94	0.65		NA	NA
	MCARI1		0.95	0.70		NA	NA
	MCARI2		0.95	0.68		NA	NA
flight 1	NDVI	82	0.79	0.14	45	0.78	0.24
	RDVI		0.92	0.13		0.76	0.36
	OSAVI		0.90	0.09		0.77	0.34
	MCARI1		0.94	0.12		0.74	0.38
	MCARI2		0.91	0.10		0.75	0.36
flight 2	NDVI	126	0.50	0.38	80	0.82	0.26
	RDVI		0.57	0.50		0.83	0.32
	OSAVI		0.55	0.49		0.83	0.34
	MCARI1		0.53	0.56		0.82	0.30
	MCARI2		0.45	0.56		0.82	0.34
Avg[a]	NDVI	NA	0.37	0.27	NA	0.80	0.21
	RDVI		0.93	0.15		0.81	0.18
	OSAVI		0.91	0.21		0.81	0.23
	MCARI1		0.94	0.08		0.81	0.13
	MCARI2		0.93	0.16		0.80	0.19

Note: NA indicates that the corresponding unmanned aerial vehicle (UAV) flight was not conducted.

Abbreviations: DAP, days after planting; MCARI1 and MCARI2, modified chlorophyll absorption in reflectance indices; NDVI, normalized difference vegetation index; OSAVI, optimized soil-adjusted vegetation index; RDVI, renormalized difference vegetation index.

[a]Averaged from flight 1 and flight 2, both conducted after artificial inoculation with *S. turcica* (NLB causal agent).

When averaged VIs from flight 1 and flight 2 were used, tDS, NLS, and ATD displayed moderate correlations in the 2018 NIL hybrids, while NLB showed moderate correlations with averaged VIs in the 2019 NIL inbreds. In the G2F hybrids (2018), only tDS and ATD exhibited weak associations with a small subset of averaged VIs.

These results suggest that while multispectral VIs are effective for capturing foliar disease dynamics, their predictive accuracy is highly dependent on the synchronization of flight/phenological timing with specific disease progression stages and the genetic background of the germplasm.

4 | DISCUSSION

Our findings demonstrate that drone-based multispectral remote sensing offers strong potential as a cost- and time-effective approach for high-throughput phenotyping of both single and MDR (including NLB, NLS, and ATD) as well as yield under disease pressure in realistic breeding field trials. However, the performance of this technology (as disease and yield indicators) was influenced by several factors intrinsic to breeding programs: differences in germplasm type and genetic background, the timing of flights relative to crop phenology, and the degree of synchrony between UAV acquisitions and ground-based disease assessments.

4.1 | Consistency and heritability of visually scored phenotypes and drone phenotypes

Across germplasm groups and years, drone-derived VIs showed generally greater consistency than ground-based visual ratings of disease severity and yield, as reflected by their lower CVs (Figures 2 and 3). This pattern aligns with evidence that visual assessments are prone to observer bias and reduced precision (Bock et al., 2008, 2009; Bock et al., 2020; Poland & Nelson, 2011). In our study, the co-occurrence of NLB, NLS, and ATD (each with distinct canopy positions and symptom progressions) likely further diminished rating consistency, consistent with known challenges such as

TABLE 3 Significant Pearson correlations (*r*) between ground-based traits (grain yield and diseases) and vegetation indices (VIs) in the near-isogenic line (NIL) germplasm (2018–2019).

		NY2018 NIL inbreds					NY2018 NIL hybrids				
				DLA_{avg}					DLA_{avg}		
Flight	**VI**	**GY_{raw}**	**tDS**	**NLB**	**NLS**	**ATD**	**GY_{raw}**	**tDS**	**NLB**	**NLS**	**ATD**
flight 0	NDVI	0.66			−0.59		0.98			−0.72	−0.58
	RDVI	0.71			−0.59		0.99			−0.74	
	OSAVI	0.69			−0.59		0.98			−0.73	
	MCARI1	0.73			−0.61		0.99			−0.75	
	MCARI2	0.71					0.99			−0.75	
flight 1	NDVI						0.81	−0.66			−0.63
	RDVI						0.94	−0.65		−0.61	−0.65
	OSAVI						0.90	−0.66		−0.58	−0.66
	MCARI1						0.96	−0.64		−0.64	−0.64
	MCARI2						0.93	−0.64		−0.61	−0.66
flight 2	NDVI							−0.61			
	RDVI			−0.59			0.84	−0.68			
	OSAVI			−0.58			0.82	−0.67			
	MCARI1						0.80	−0.68			
	MCARI2						0.74	−0.68			
Avg[a]	NDVI						0.70				
	RDVI						0.96	−0.64		−0.64	−0.58
	OSAVI						0.94	−0.62		−0.62	
	MCARI1						0.97	−0.65		−0.65	−0.59
	MCARI2						0.96	−0.65		−0.63	−0.59
		NY2019 NIL inbreds					**NY2019 NIL hybrids**				
				DLA_{avg}					DLA_{avg}		
Flight	**VI**	**GY_{raw}**	**tDS**	**NLB**	**NLS**	**ATD**	**GY_{raw}**	**tDS**	**NLB**	**NLS**	**ATD**
flight 1	**NDVI**			**−0.63**			**0.87**				
	RDVI			−0.62			0.89				
	OSAVI			−0.62			0.89				
	MCARI1			−0.63			0.90				
	MCARI2			−0.62			0.90				
flight 2	NDVI		−0.60	−0.75							
	RDVI			−0.68			0.82				
	OSAVI			−0.71			0.79				
	MCARI1			−0.66			0.82				
	MCARI2			−0.70			0.79				
Avg[a]	NDVI			−0.68			0.83				
	RDVI			−0.65			0.86				
	OSAVI			−0.67			0.85				
	MCARI1			−0.65			0.86				
	MCARI2			−0.67			0.85				

Note: Correlations were computed using spatially corrected best linear unbiased predictor (BLUP) values. Only statistically significant correlations at a *p*-value $\leq \alpha$ ($\alpha = 0.05$) are reported. Note that flight 0, flight 1, and flight 2 were conducted at 34, 82, and 126 days after planting (DAP) in 2018, while flight 1 and flight 2 were conducted at 45 and 80 DAP in 2019.

Abbreviations: DLA, diseased leaf area; MCARI1 and MCARI2, modified chlorophyll absorption in reflectance indices; NDVI, normalized difference vegetation index; OSAVI, optimized soil-adjusted vegetation index; GY_{raw}, raw grain yield (without adjustment for stand counts); RDVI, renormalized difference vegetation index.

[a]Averaged from flight 1 and flight 2, both conducted after artificial inoculation with *S. turcica* (NLB causal agent).

TABLE 4 Significant Pearson correlations (r) between ground-based traits (grain yield and diseases) and vegetation indices (VIs) in the Genomes to Fields (G2F) germplasm (2018–2019).

		NY2018 G2F hybrids				NY2019 G2F hybrids			
				DLA_{avg}				DLA_{avg}	
Flight	**VI**	**GY_{raw}**	**tDS**	**NLB**	**ATD**	**GY_{raw}**	**tDS**	**NLB**	**ATD**
flight 0	NDVI	0.28				NA	NA	NA	NA
	RDVI	0.26				NA	NA	NA	NA
	OSAVI	0.27				NA	NA	NA	NA
	MCARI1	0.24				NA	NA	NA	NA
	MCARI2	0.25				NA	NA	NA	NA
flight 1	NDVI							−0.14	
	RDVI					0.16	−0.15	−0.16	
	OSAVI					0.14	−0.15	−0.15	
	MCARI1	0.11				0.17	−0.16	−0.16	
	MCARI2	0.12				0.16	−0.16	−0.16	
flight 2	NDVI		−0.46		−0.45				
	RDVI		−0.53		−0.52				
	OSAVI		−0.52		−0.51				
	MCARI1		−0.54		−0.52				
	MCARI2		−0.53		−0.52				
Avg[a]	NDVI		−0.14		−0.15				
	RDVI								
	OSAVI		−0.12		−0.13				
	MCARI1								
	MCARI2				−0.10				

Note: Correlations were computed using spatially corrected best linear unbiased predictor (BLUP) values. Only statistically significant correlations at a p-value $\leq \alpha$ ($\alpha = 0.05$) are reported. Note that flight 0, flight 1, and flight 2 were conducted at 34, 82, and 126 days after planting (DAP) in 2018, while flight 1 and flight 2 were conducted at 45 and 80 DAP in 2019. NA indicates that the corresponding UAV flight was not conducted.

Abbreviations: DLA, diseased leaf area; MCARI1 and MCARI2, modified chlorophyll absorption in reflectance indices; NDVI, normalized difference vegetation index; OSAVI, optimized soil-adjusted vegetation index; GY_{raw}, raw grain yield (without adjustment for stand counts); RDVI, renormalized difference vegetation index.

[a]Averaged from flight 1 and flight 2, both conducted after artificial inoculation with *S. turcica* (NLB causal agent).

overlapping lesions, coalescence, and senescence that complicate visual scoring and can substantially alter genetic effect estimates (Poland & Nelson, 2011).

In contrast, drone-based VIs provided more stable indicators of canopy greenness and vigor, even though they primarily sample upper-canopy signals. Their improved consistency aligns with previous work demonstrating the robustness of multi-temporal spectral traits for disease detection and yield prediction (DeSalvio et al., 2022; Guo et al., 2024; Oh et al., 2021; Wiesner-Hanks et al., 2018). These results highlight the value of drone-based spectral sensing as an objective and reliable complement to conventional scoring in the heterogeneous conditions typical of resistance-breeding trials.

Although drone-derived VIs exhibited lower CVs than ground-measured disease severity and yield, broad-sense heritability (H^2) was generally higher for the ground traits, indicating that spatially corrected visual scores, despite their subjectivity, remain highly effective at capturing genetic variation in foliar disease resistance under field conditions. Similar patterns of strong genetic signal in ground phenotypes relative to spectral traits have been noted in prior maize remote-sensing studies (Chivasa et al., 2021; DeSalvio et al., 2022).

VIs nonetheless showed moderate to high heritability, consistent with earlier observations of their temporal stability and physiological relevance (Chivasa et al., 2021; Washburn et al., 2024). Their objectivity and low variability make them valuable complements to visual scores: ground traits provide strong genetic resolution, whereas VIs offer scalable, high-throughput indicators of canopy health. Combining both can thus enhance selection for MDR and yield.

Higher heritability in the NIL germplasm than in the diverse G2F hybrids reflects the more uniform genetic background of near-isogenic materials. This distinction underscores the importance of germplasm composition when

interpreting the performance of both ground-based and drone-derived phenotypes.

4.2 | Relationship between visually scored disease severity, yield, and drone-derived phenotypes

We found promising correlations between VIs and disease resistance or yield in certain germplasm groups and years, though these associations were not consistent across all conditions.

Across both years, VI–trait correlations were generally stronger in the NIL germplasm than in the diverse G2F hybrids. This is consistent with previous findings suggesting that HTP-derived reflectance patterns more effectively capture genetic variance in populations with structured genetic backgrounds (Rutkoski et al., 2016) and under standardized field-based breeding scenarios (Zaman-Allah et al., 2015). Early-season flights produced the strongest VI–yield associations (particularly in NIL hybrids), supporting the premise that early canopy vigor and biomass establishment are critical determinants of final yield (Bänziger et al., 2000). Our findings align with recent evidence that multispectral imaging effectively captures these early-season physiological drivers to predict performance in maize breeding pipelines (Prasanna et al., 2021). The exceptionally high early-season VI–yield correlations observed in NIL hybrids (r = 0.98–0.99) likely reflect the strong phenotypic contrast between near-isogenic resistant and susceptible classes and, therefore, may not be directly reproducible in more genetically diverse breeding populations.

Lower VI–yield correlations in the diverse G2F panel (r = 0.11–0.28) likely reflect greater phenological and architectural heterogeneity relative to the NIL germplasm. Differences in flowering time (55–72 vs. 61–78 days to anthesis) combined with fixed flight schedules (34, 82, and 126 DAP in 2018; 45 and 80 DAP in 2019) resulted in temporal misalignment between spectral measurements and crop developmental stage, reducing predictive power. This is consistent with prior work showing that canopy reflectance is highly sensitive to growth stage and structural variation (Asner, 1998; Hatfield et al., 2008; L. Li et al., 2014), underscoring the need to align drone flight acquisitions with phenology in diverse breeding populations.

The five selected VIs performed similarly for yield prediction, in line with reports of functional redundancy among common multispectral indices (Hunt et al., 2013), whereas a few VIs were more informative for specific diseases (Mahlein, 2016) and flight times, especially in the NIL inbreds. VI–disease associations were highly dependent on flight timing and disease progression stage, echoing prior findings that each flight date captures physiologically unique information (Mahlein, 2016; Washburn et al., 2024). Multitemporal averaging modestly improved VI–trait correlations in the NILs. Overall, while VIs captured meaningful variation in disease severity and yield, their predictive reliability depended strongly on germplasm composition and on the temporal alignment of imaging with crop and disease development.

5 | CONCLUSION

This study demonstrates that drone-based multispectral phenotyping can provide scalable, objective, and temporally responsive measurements that capture meaningful variation in maize disease severity and yield under realistic breeding conditions. Although the strength of VI–trait associations varied across germplasm groups, years, and flight times, early- to mid-season multispectral data proved particularly informative in genetically uniform materials, enabling reliable prediction of both single and MDR as well as yield. Importantly, temporal VI data captured disease- and phenology-specific canopy dynamics that are difficult to quantify consistently through visual scoring, offering a path toward more objective selection in environments where multiple foliar diseases co-occur. Together, these findings highlight the value of integrating drone-derived temporal phenotypes with traditional assessments to enhance the accuracy, throughput, and robustness of selection for MDR and yield in maize breeding programs.

AUTHOR CONTRIBUTIONS

Danilo E. Moreta: Conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Nina Blahut**: Data curation. **Nicholas Kaczmar**: Data curation; investigation. **Judith M. Kolkman**: Investigation. **Gary Bergstrom**: Resources; supervision. **Margaret Smith**: Resources; supervision. **Michael A. Gore**: Resources. **Rebecca Nelson**: Funding acquisition; project administration; supervision; writing—original draft.

ACKNOWLEDGMENTS

We thank Cornell staff for their valuable help with the field trials, especially Paul Stachowski (former Musgrave Research Farm Manager), Keith Payne, Sherrie Norman, Daniel Fisher, and Ace Repka (former Nelson Lab manager). The authors also thank the Gore Lab manager, Liam Wickes-Do, for providing access to the image analysis software. We gratefully acknowledge the support of the Hatch program and the Genomes to Fields (G2F) Initiative. This research was supported in part by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Hatch Accession #1020798.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Danilo E. Moreta https://orcid.org/0000-0001-6866-9200
Nina Blahut https://orcid.org/0000-0002-9255-2333
Judith M. Kolkman https://orcid.org/0000-0001-7388-7245
Gary Bergstrom https://orcid.org/0000-0001-9613-270X
Margaret Smith https://orcid.org/0000-0002-0001-7860
Michael A. Gore https://orcid.org/0000-0001-6896-8024
Rebecca Nelson https://orcid.org/0000-0002-9026-7803

REFERENCES

Alkhalifah, N., Campbell, D. A., Falcon, C. M., Gardiner, J. M., Miller, N. D., Romay, M. C., Walls, R., Walton, R., Yeh, C.-T., Bohn, M., Bubert, J., Buckler, E. S., Ciampitti, I., Flint-Garcia, S., Gore, M. A., Graham, C., Hirsch, C., Holland, J. B., Hooker, D., … Lawrence-Dill, C. J. (2018). Maize Genomes to Fields: 2014 and 2015 field season genotype, phenotype, environment, and inbred ear image datasets. *BMC Research Notes*, *11*, Article 452. https://doi.org/10.1186/s13104-018-3508-1

Ashapure, A., Jung, J., Chang, A., Oh, S., Maeda, M., & Landivar, J. (2019). A comparative study of RGB and multispectral sensor-based cotton canopy cover modelling using multi-temporal UAS data. *Remote Sensing*, *11*(23), Article 2757. https://doi.org/10.3390/rs11232757

Ashapure, A., Jung, J., Yeom, J., Chang, A., Maeda, M., Maeda, A., & Landivar, J. (2019). A novel framework to detect conventional tillage and no-tillage cropping system effect on cotton growth and development using multi-temporal UAS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*, 49–64. https://doi.org/10.1016/j.isprsjprs.2019.04.003

Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, *64*(3), 234–253. https://doi.org/10.1016/S0034-4257(98)00014-5

Bänziger, M., Edmeades, G. O., Beck, D., & Bellon, M. R. (2000). *Breeding for drought and nitrogen stress tolerance in maize: From theory to practice*. CIMMYT.

Bock, C. H., Barbedo, J. G. A., Del Ponte, E. M., Bohnenkamp, D., & Mahlein, A.-K. (2020). From visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. *Phytopathology Research*, *2*, Article 9. https://doi.org/10.1186/s42483-020-00049-8

Bock, C. H., Parker, P. E., Cook, A. Z., & Gottwald, T. R. (2008). Characteristics of the perception of different severity measures of citrus canker and the relationships between the various symptom types. *Plant Disease*, *92*(6), 927–939. https://doi.org/10.1094/PDIS-92-6-0927

Bock, C. H., Parker, P. E., Cook, A. Z., Riley, T., & Gottwald, T. R. (2009). Comparison of assessment of citrus canker foliar symptoms by experienced and inexperienced raters. *Plant Disease*, *93*(4), 412–424. https://doi.org/10.1094/PDIS-93-4-0412

Chang, A., Jung, J., Maeda, M. M., & Landivar, J. (2017). Crop height monitoring with digital imagery from Unmanned Aerial System (UAS). *Computers and Electronics in Agriculture*, *141*, 232–237. https://doi.org/10.1016/j.compag.2017.07.008

Chivasa, W., Mutanga, O., & Burgueño, J. (2021). UAV-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial MSV inoculation. *Computers and Electronics in Agriculture*, *184*, Article 106128. https://doi.org/10.1016/j.compag.2021.106128

Chung, C. L., Jamann, T., Longfellow, J., & Nelson, R. (2010). Characterization and fine-mapping of a resistance locus for northern leaf blight in maize bin 8.06. *Theoretical and Applied Genetics*, *121*(1), 205–227. https://doi.org/10.1007/s00122-010-1303-z

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, *11*(4), Article 38. https://doi.org/10.1198/108571106X154443

Cummings, J. (2018). *Anthracnose top dieback prevalent across NY, September 2018*. NYS IPM Program. https://blogs.cornell.edu/ccefieldcropnews/2018/09/20/anthracnose-top-dieback-prevalent-across-ny-september-2018/

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, *30*(3), 271–278. https://doi.org/10.1016/0034-4257(89)90069-2

Dash, J. P., Watt, M. S., Pearse, G. D., Heaphy, M., & Dungey, H. S. (2017). Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak. *ISPRS Journal of Photogrammetry and Remote Sensing*, *131*, 1–14. https://doi.org/10.1016/j.isprsjprs.2017.07.007

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., Nelson, R. J., & Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*, *107*(11), 1426–1432. https://doi.org/10.1094/PHYTO-11-16-0417-R

DeSalvio, A. J., Adak, A., Murray, S. C., Wilde, S. C., & Isakeit, T. (2022). Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. *Scientific Reports*, *12*(1), Article 7571. https://doi.org/10.1038/s41598-022-11591-0

Devadas, R., Lamb, D. W., Backhouse, D., & Simpfendorfer, S. (2015). Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat. *Precision Agriculture*, *16*(4), 477–491. https://doi.org/10.1007/s11119-015-9390-0

Enciso, J., Avila, C. A., Jung, J., Elsayed-Farag, S., Chang, A., Yeom, J., Landivar, J., Maeda, M., & Chavez, J. C. (2019). Validation of agronomic UAV and field measurements for tomato varieties. *Computers and Electronics in Agriculture*, *158*, 278–283. https://doi.org/10.1016/j.compag.2019.02.011

Franke, J., & Menz, G. (2007). Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture*, *8*(3), 161–172. https://doi.org/10.1007/s11119-007-9036-y

Fry, W. E. (1978). Quantification of general resistance of potato cultivars and fungicide effects for integrated control of potato late blight. *Phytopathology*, *68*(11), 1650–1655. https://doi.org/10.1094/Phyto-68-1650

Furbank, R. T., & Tester, M. (2011). Phenomics—Technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, *16*(12), 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Gano, B., Bhadra, S., Vilbig, J. M., Ahmed, N., Sagan, V., & Shakoor, N. (2024). Drone-based imaging sensors, techniques, and applications in plant phenotyping for crop breeding: A comprehensive review. *Plant Phenome Journal*, *7*, Article e20100. https://doi.org/10.1002/ppj2.20100

Garcia-Ruiz, F., Sankaran, S., Maja, J. M., Lee, W. S., Rasmussen, J., & Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, *91*, 106–115. https://doi.org/10.1016/j.compag.2012.12.002

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the USA*, *115*(18), 4613–4618. https://doi.org/10.1073/pnas.1716999115

Guo, Y., Fu, Y. H., Chen, S., Hao, F., Zhang, X., de Beurs, K., & He, Y. (2024). Predicting grain yield of maize using a new multispectral-based canopy volumetric vegetation index. *Ecological Indicators*, *166*, Article 112295. https://doi.org/10.1016/j.ecolind.2024.112295

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, *90*(3), 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Hatfield, J. L., Gitelson, A. A., Schepers, J. S., & Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, *100*, S-117–S-131. https://doi.org/10.2134/agronj2006.0370c

Herr, A. W., Adak, A., Carroll, M. E., Elango, D., Kar, S., Li, C., Jones, S. E., Carter, A. H., Murray, S. C., Paterson, A., Sankaran, S., Singh, A., & Singh, A. K. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, *63*(4), 1722–1749. https://doi.org/10.1002/csc2.21028

Hunt, E. R., Jr., Doraiswamy, P. C., McMurtrey, J. E., Daughtry, C. S., Perry, E. M., & Akhmedov, B. (2013). A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 103–112. https://doi.org/10.1016/j.jag.2012.07.020

Jung, J., Maeda, M., Chang, A., Landivar, J., Yeom, J., & McGinty, J. (2018). Unmanned aerial system assisted framework for the selection of high yielding cotton genotypes. *Computers and Electronics in Agriculture*, *152*, 74–81. https://doi.org/10.1016/j.compag.2018.06.051

Kolkman, J. M., Moreta, D. E., Repka, A., Bradbury, P., & Nelson, R. J. (2023). Brown midrib mutant and genome-wide association analysis uncover lignin genes for disease resistance in maize. *The Plant Genome*, *16*(1), Article e20278. https://doi.org/10.1002/tpg2.20278

Kolkman, J. M., Strable, J., Harline, K., Kroon, D. E., Wiesner-Hanks, T., Bradbury, P. J., & Nelson, R. J. (2020). Maize introgression library provides evidence for the involvement of *liguleless1* in resistance to northern leaf blight. *G3: Genes, Genomes, Genetics*, *10*(10), 3611–3622. https://doi.org/10.1534/g3.120.401500

Li, L., Zhang, Q., & Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors*, *14*(11), 20078–20111. https://doi.org/10.3390/s141120078

Li, Z., Coffey, L., Garfin, J., Miller, N. D., White, M. R., Spalding, E. P., De Leon, N., Kaeppler, S. M., Schnable, P. S., Springer, N. M., & Hirsch, C. N. (2018). Genotype-by-environment interactions affecting heterosis in maize. *PLoS ONE*, *13*(1), Article e0191321. https://doi.org/10.1371/journal.pone.0191321

Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., & Boddupalli, M. P. (2019). Application of remote sensing for phenotyping tar spot complex resistance in maize. *Frontiers in Plant Science*, *10*, Article 552. https://doi.org/10.3389/fpls.2019.00552

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors—Parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, *100*(2), 241–251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Mahlein, A.-K., Steiner, U., Hillnhütter, C., Dehne, H.-W., & Oerke, E.-C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods*, *8*(1), Article 3. https://doi.org/10.1186/1746-4811-8-3

Marques Ramos, A. P., Prado Osco, L., Elis Garcia Furuya, D., Nunes Gonçalves, W., Cordeiro Santana, D., Pereira Ribeiro Teodoro, L., Antonio Da Silva Junior, C., Fernando Capristo-Silva, G., Li, J., Henrique Rojo Baio, F., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*, Article 105791. https://doi.org/10.1016/j.compag.2020.105791

Oh, S., Lee, D.-Y., Gongora-Canul, C., Ashapure, A., Carpenter, J., Cruz, A. P., Fernandez-Campos, M., Lane, B. Z., Telenko, D. E. P., Jung, J., & Cruz, C. D. (2021). Tar spot disease quantification using Unmanned Aircraft Systems (UAS) data. *Remote Sensing*, *13*(13), Article 2567. https://doi.org/10.3390/rs13132567

Peñuelas, J., & Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, *3*(4), 151–156. https://doi.org/10.1016/S1360-1385(98)01213-8

Poland, J. A., Bradbury, P. J., Buckler, E. S., & Nelson, R. J. (2011). Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the USA*, *108*(17), 6893–6898. https://doi.org/10.1073/pnas.1010894108

Poland, J. A., & Nelson, R. J. (2011). In the eye of the beholder: The effect of rater variability and different rating scales on QTL mapping. *Phytopathology*, *101*(2), 290–298. https://doi.org/10.1094/PHYTO-03-10-0087

Prasanna, B. M., Cairns, J. E., Zaidi, P. H., Beyene, Y., Makumbi, D., Gowda, M., Magorokosho, C., Zaman-Allah, M., Olsen, M., Das, A., Worku, M., Gethi, J., Vivek, B. S., Nair, S. K., Rashid, Z., Vinayan, M. T., Issa, A. B., San Vicente, F., Dhliwayo, T., & Zhang, X. (2021). Beat the stress: Breeding for climate resilience in maize for the tropical rainfed environments. *Theoretical and Applied Genetics*, *134*(6), 1729–1752. https://doi.org/10.1007/s00122-021-03773-7

R Core Team. (2020). R: A language and environment for statistical computing *(Version 4.0.0)* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rodríguez-Álvarez, M. X., Boer, M. P., Eilers, P., & van Eeuwijk, F. (2024). SpATS: Spatial Analysis of Field Trials with Splines (R Package Version 1.0-19) [Computer software]. https://CRAN.R-project.org/package=SpATS

Rodríguez-Álvarez, M. X., Boer, M. P., Van Eeuwijk, F. A., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, *23*, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Roujean, J. L., & Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, *51*(3), 375–384. https://doi.org/10.1016/0034-4257(94)00114-3

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation (NASA/GSFC Type III final report)*. NASA.

RStudio Team. (2022). RStudio: Integrated development environment for R *(Version 2022.02.0)* [Computer software]. RStudio, PBC. http://www.rstudio.com/

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, *74*(1), 91–99. https://doi.org/10.1016/j.compag.2010.06.009

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics*, *6*(9), 2799–2808. https://doi.org/10.1534/g3.116.032888

Sankaran, S., Maja, J. M., Buchanon, S., & Ehsani, R. (2013). Huanglongbing (citrus greening) detection using visible, near infrared and thermal imaging techniques. *Sensors*, *13*(2), 2117–2130. https://doi.org/10.3390/s130202117

Shakoor, N., Lee, S., & Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Current Opinion in Plant Biology*, *38*, 184–192. https://doi.org/10.1016/j.pbi.2017.05.006

Simko, I., Jimenez-Berni, J. A., & Sirault, X. R. R. (2017). Phenomic approaches and tools for phytopathologists. *Phytopathology*, *107*(1), 6–17. https://doi.org/10.1094/PHYTO-02-16-0082-RVW

Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *The Crop Journal*, *9*(3), 633–645. https://doi.org/10.1016/j.cj.2021.03.015

Stewart, E. L., Wiesner-Hanks, T., Kaczmar, N., Dechant, C., Wu, H., Lipson, H., Nelson, R. J., & Gore, M. A. (2019). Quantitative phenotyping of northern leaf blight in UAV images using deep learning. *Remote Sensing*, *11*(19), Article 2209. https://doi.org/10.3390/rs11192209

Washburn, J. D., Adak, A., DeSalvio, A. J., Arik, M. A., & Murray, S. C. (2024). High temporal resolution unoccupied aerial systems phenotyping provides unique information between flight dates. *The Plant Phenome Journal*, *7*(1), 20113. https://doi.org/10.1002/ppj2.20113

Wiesner-Hanks, T., & Nelson, R. (2016). Multiple disease resistance in plants. *Annual Review of Phytopathology*, *54*, 229–252. https://doi.org/10.1146/annurev-phyto-080615-100037

Wiesner-Hanks, T., Stewart, E. L., Kaczmar, N., Dechant, C., Wu, H., Nelson, R. J., Lipson, H., & Gore, M. A. (2018). Image set for deep learning: Field images of maize annotated with disease symptoms. *BMC Research Notes*, *11*(1), Article 440. https://doi.org/10.1186/s13104-018-3548-6

Yang, C., Everitt, J. H., & Fernandez, C. J. (2010). Comparison of airborne multispectral and hyperspectral imagery for mapping cotton root rot. *Biosystems Engineering*, *107*(2), 131–139. https://doi.org/10.1016/j.biosystemseng.2010.07.011

Yendrek, C. R., Tomaz, T., Montes, C. M., Cao, Y., Morse, A. M., Brown, P. J., McIntyre, L. M., Leakey, A. D. B., & Ainsworth, E. A. (2017). High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiology*, *173*(1), 614–626. https://doi.org/10.1104/pp.16.01447

Yeom, J., Jung, J., Chang, A., Maeda, M., & Landivar, J. (2018). Automated open cotton boll detection for yield estimation using unmanned aircraft vehicle (UAV) data. *Remote Sensing*, *10*(12), Article 1895. https://doi.org/10.3390/rs10121895

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., Hornero, A., Albà, A. H., Das, B., Craufurd, P., Olsen, M., Prasanna, B. M., & Cairns, J. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods*, *11*(1), Article 35. https://doi.org/10.1186/s13007-015-0078-2

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Moreta, D. E., Blahut, N., Kaczmar, N., Kolkman, J. M., Bergstrom, G., Smith, M., Gore, M. A., & Nelson, R. (2026). Drone-based phenotyping of maize for multiple disease resistance and yield in breeding field trials. *The Plant Phenome Journal*, *9*, e70079. https://doi.org/10.1002/ppj2.70079