A Systematic Evaluation of Spectral-Peak-Relative Temporal Alignment for Satellite-Based Field-Level Wheat Grain Protein Prediction

Vlasios Mangidis

vmangidi@purdue.edu

Purdue University West Lafayette

German Mandrini

Purdue University West Lafayette

Ignacio Antonio Ciampitti

Purdue University West Lafayette

Research Article

Keywords: grain protein concentration, winter wheat, Sentinel-2, spectral-peak-relative alignment, spatial cross-validation, machine learning, grain filling

Posted Date: August 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9999512/v2

Additional Declarations: Competing interest reported. Ignacio A. Ciampitti reports administrative support and equipment, materials, or supplies were provided by General Mills and by the United Sorghum Checkoff Program. The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A Systematic Evaluation of Spectral-Peak-Relative Temporal Alignment for Satellite-Based Field-Level Wheat Grain Protein Prediction

Vlasios Mangidis[1*], German Mandrini[1] and Ignacio Antonio Ciampitti[1*]

[1*]Department of Agronomy, College of Agriculture, Purdue University, 915 W. State Street, West Lafayette, 47907, IN, United States.

*Corresponding author(s). E-mail(s): vmangidi@purdue.edu; iciampit@purdue.edu;

Abstract

Satellite-based prediction of grain protein concentration (GPC) in wheat typically composites spectral observations over fixed calendar windows, implicitly assuming phenological synchrony across fields. We present a systematic evaluation of whether aligning multi-source remote sensing time series to field-specific, spectral-peak-relative windows improves field-level GPC prediction, for a quality trait whose physiology, senescence-linked nitrogen remobilization, contrasts with the season-integrating behavior of yield. Integrating Sentinel-2 imagery (31 vegetation indices, 10 spectral bands), ERA5-Land reanalysis, gSSURGO soil properties, and USGS 3DEP topography across 228 commercial winter wheat fields in western Kansas (2024–2025), we compared six temporal strategies (peak-relative vs. calendar × monthly, biweekly, growth-stage) using three ensemble tree models under nested cross-validation with Boruta feature selection. A single 30-day post-peak window (peak + [16,45] days) was the top-performing and most consistently selected window, chosen in 4 of 5 outer folds, reproducing prior accuracy under random cross-validation ($R^2 \approx 0.28$); though its advantage over the best calendar window was not statistically significant (paired bootstrap $p = 0.08$). Under leave-county spatial cross-validation, however, this skill did not transfer across counties (Sentinel-2–only $R^2 \approx 0.01$; per-county median $R^2 = -0.23$), indicating the satellite signal supports within-region interpolation but not spatial

extrapolation to unseen counties; ablation shows that neither the spectral nor the static features transfer across counties on their own, and the residual cross-county skill emerges only from their combination. A near-real-time application at ∼3 weeks before harvest retains most within-region skill at a modest accuracy cost. The results delineate where spectral-peak-relative alignment helps, concentrating a senescence-linked signal within region, and where it does not, providing an honest operational baseline for satellite-based grain-quality monitoring.

Keywords: grain protein concentration, winter wheat, Sentinel-2, spectral-peak-relative alignment, spatial cross-validation, machine learning, grain filling

Nomenclature

AWC	Available Water Capacity (mm)
BOA	Bottom-of-Atmosphere
CCC	Concordance Correlation Coefficient
CV	Cross-Validation
DOY	Day of Year
GDD	Growing Degree Days (°C day)
GEE	Google Earth Engine
GCVI	Green Chlorophyll Vegetation Index
GPC	Grain Protein Concentration (%)
gSSURGO	Gridded Soil Survey Geographic database
HRW	Hard Red Winter wheat
K_{sat}	Saturated Hydraulic Conductivity (μm s^{-1})
KGE	Kling–Gupta Efficiency
MBE	Mean Bias Error (% protein)
MSI	Multi-Spectral Instrument (Sentinel-2 sensor)
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
PET	Potential Evapotranspiration (mm day^{-1})
RF	Random Forest
RMSE	Root Mean Squared Error (% protein)
RRMSE	Relative Root Mean Squared Error (%)
SHAP	SHapley Additive exPlanations
SWIR	Shortwave Infrared
3DEP	3D Elevation Program (USGS)
VI	Vegetation Index
VPD	Vapor Pressure Deficit (kPa)

1 Introduction

Wheat (*Triticum aestivum* L.) is the most widely cultivated cereal crop globally, providing approximately 20% of dietary calories and protein for the world's population (Asseng et al, 2019). In the United States, hard red winter wheat (HRW) is the dominant market class, and grain protein concentration (GPC) is the primary quality determinant that governs end-use suitability and market price premiums (Guttieri

et al, 2015). Fields producing grain above 12% protein typically receive price premiums, while low-protein lots (below 12% protein) are discounted or blended, creating strong economic incentives for early and accurate GPC prediction (Lollato and Knapp, 2021). Despite its commercial importance, GPC remains difficult to predict before harvest because it depends not on total biomass accumulation, as yield does, but on the ratio of nitrogen to carbohydrate partitioning into the grain during a relatively narrow developmental window (Martre et al, 2003; Bogard et al, 2010; De Oliveira Silva et al, 2021).

Satellite remote sensing offers a scalable, non-destructive means of monitoring crop biophysical status over large areas. Multispectral vegetation indices (VIs) derived from platforms such as Sentinel-2 are sensitive to canopy chlorophyll content, leaf area, and water status, proxies for the nitrogen dynamics that ultimately determine GPC (Clevers and Gitelson, 2013; Zhao et al, 2019). Several studies have demonstrated that satellite-derived spectral features can predict wheat GPC at experimental-plot to regional scales, with reported accuracies ranging from $R^2 = 0.28$ to $R^2 = 0.86$ depending on spatial scale, sensor type, sample size, and the availability of multi-year calibration data (Zhao et al, 2019; Tan et al, 2020; Wolters et al, 2022; Longmire et al, 2023; Xu et al, 2024). A recent synthesis-analysis of 84 studies found that prediction accuracy varies widely with sensor proximity, wavelength range, and crop species, yet satellite-based approaches at field-level remain underexplored (Bastos et al, 2021). Beyond wheat, satellite-based prediction of seed protein concentration has also been demonstrated for soybean using growth-stage-aligned imagery and ensemble machine learning (Hernandez et al, 2023). To date, most wheat GPC studies have relied on spectral observations acquired at fixed calendar dates or within calendar-defined compositing windows.

The first open question is whether aligning observations with crop phenology improves GPC prediction over fixed calendar dates. In practice, wheat fields within even a small production region can differ by several weeks in their crop phenological development due to differences in planting date, cultivar maturity class, soil moisture, and microclimate (McMaster and Wilhelm, 1997; Bolton and Friedl, 2013). When spectral observations are composited over calendar-fixed windows, the same temporal bin may correspond to different growth stages across fields, thereby introducing phenological noise that obscures the biophysical signals relevant to GPC. This problem is well recognized in yield prediction, where phenology-normalized time series have been shown to outperform calendar-based alternatives by 10-40% for maize and soybean (Bolton and Friedl, 2013; Sakamoto et al, 2014; Zeng et al, 2020; Meng et al, 2022), and integrated satellite–weather–field approaches have demonstrated the value of phenological timing for crop monitoring (Schwalbert et al, 2018; Nieto et al, 2021). More recently, (Zhang et al, 2025) demonstrated that phenology-aligned deep learning improves corn yield estimates over the US Midwest, increasing the explained variance from 61% to 71% relative to non-aligned approaches. However, to the extent of our knowledge, no study has systematically compared static calendar windows with dynamic, field-specific spectral-peak-relative temporal alignment for GPC prediction, nor quantified the magnitude of this improvement.

A second complementary question concerns the temporal resolution of the observation window, that is, whether spectral features should be aggregated at monthly, biweekly, or growth-stage intervals (whereas the first question asks *how* windows are aligned, this one asks *how long* each window should be). Unlike yield, which integrates photosynthetic production over the entire growing season, GPC is governed primarily by nitrogen remobilization from senescing vegetative tissues during grain filling (Gooding et al, 2003; Masclaux-Daubresse et al, 2008; Kichey et al, 2007). This physiological concentration of the GPC signal in a narrow developmental phase suggests that a single short window around grain filling may be more informative than features aggregated over longer periods, but this hypothesis has not been tested across multiple temporal resolutions or normalization strategies within a single consistent framework.

The individual components of this pipeline, peak detection, temporal normalization, and ensemble learning, are established. The contribution of this study is not a new method but a systematic, honestly-validated evaluation of an existing alignment strategy on a quality trait whose physiology, unlike yield, concentrates the relevant signal in a narrow developmental window rather than integrating it over the season. The same evaluation logic applies to other crop traits expressed within limited developmental phases.

The present study addresses these gaps by systematically comparing six temporal strategies for GPC prediction in winter wheat, defined by the factorial combination of two normalization approaches (spectral-peak-relative windows vs. calendar) and three temporal resolutions (monthly, biweekly, and growth stages). We evaluated all six strategies under a unified pipeline based on ensemble tree models, chosen for their robustness to collinearity and strong performance on tabular geospatial data (Random Forest, XGBoost, LightGBM), nested cross-validation, and automated feature selection via Boruta, applied to 228 commercial winter wheat fields across western Kansas during the 2024–2025 growing season.

The specific objectives of this study are to: (1) quantify the predictive advantage of spectral-peak-relative versus calendar-based temporal compositing for field-level GPC prediction; (2) identify the optimal temporal resolution and observation window within the spectral-peak-relative framework; (3) characterize the spectral, edaphic, meteorological, and topographic features most informative for GPC through SHAP-based feature importance analysis; and (4) evaluate the practical utility of the predictions for three-class grain quality classification. By providing the first systematic temporal strategy comparison for satellite-based GPC prediction, this work aims to establish a methodological baseline for future operational GPC mapping efforts.

2 Methodology

2.1 Study Area and Ground-Truth Data

The initial dataset encompasses 240 commercial winter wheat fields distributed across 19 counties in western Kansas, spanning the 2024–2025 growing season (Figure 1). The sampled region covers a roughly 220 km north–south by 250 km east–west extent of the semi-arid High Plains (37.37°–39.33° N, 98.91°–102.04° W), at elevations ranging from 551 to 1189 m (mean 758 m). Western Kansas is among the most productive

HRW zones in the United States; Kansas consistently ranks as the leading wheat-producing state, accounting for approximately 18% of total U.S. wheat output, with HRW representing 98% of the state's wheat crop (USDA-NASS, 2024). Production in the region is predominantly rainfed, reflecting the semi-arid climate of the western Great Plains (Schlegel et al, 2017). Of the 240 fields, 213 (88.8%) relied on dryland management while 27 (11.2%) were irrigated (208 and 20, respectively, among the 228 fields retained for analysis), as classified by the Landscape Irrigation Dataset (LANID; Xie et al, 2019).

GPC was determined at harvest from composite grain samples. Observed GPC ranged from 8.86% to 16.11%, with a mean of $12.15 \pm 1.36\%$ and a median of 12.10% (mean $12.14 \pm 1.34\%$ across the 228 fields retained after quality control; Figure S1), which falls within the expected range for commercial HRW wheat in the central Great Plains (Guttieri et al, 2015).

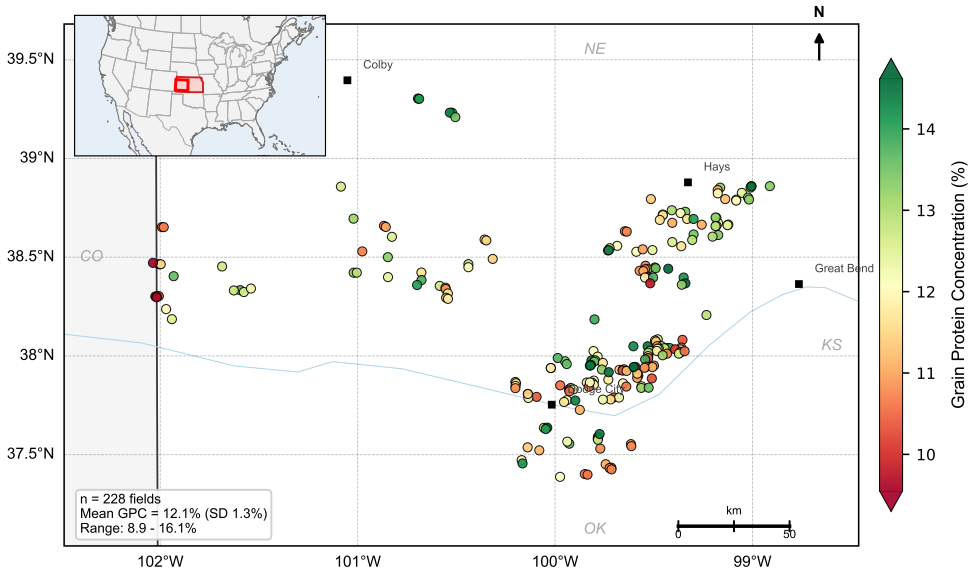

Fig. 1 Spatial distribution of the 228 winter wheat fields retained for analysis across western Kansas (2024–2025 growing season). Points are color-coded by measured grain protein concentration (%). The inset shows the location of the study area within the conterminous United States. Descriptive statistics of the observed GPC distribution are shown in the lower-left corner.

2.2 Remote Sensing and Ancillary Data

The predictor feature space was constructed by integrating four data sources: satellite multispectral imagery, gridded weather data, a national soil survey database, and a high-resolution digital elevation model. An overview of the complete pipeline, from raw data acquisition through model evaluation, is provided in Figure 2.

2.2.1 Sentinel-2 Multispectral Imagery

Multispectral surface reflectance data were obtained from the Sentinel-2A and Sentinel-2B satellites of the Copernicus programme (Drusch et al, 2012). These twin polar-orbiting sensors carry the Multi-Spectral Instrument, which acquires data in 13 spectral bands spanning the visible, red-edge, near-infrared (NIR), and shortwave infrared (SWIR) regions at 10–20 m spatial resolution, with a combined revisit interval of approximately five days at mid-latitudes.

We used the Level-2A (L2A) bottom-of-atmosphere (BOA) surface reflectance product from the harmonized Sentinel-2 collection (COPERNICUS/S2_SR_HARMONIZED) available through Google Earth Engine (GEE; Gorelick et al, 2017). L2A products are generated by the Sen2Cor processor (Main-Knorn et al, 2017), which performs atmospheric correction, terrain correction, and cirrus removal from the Level-1C top-of-atmosphere data. All acquisitions between September 1, 2024 and June 30, 2025 that intersected the study area were ingested. Ten reflectance bands were retained: B2 (blue, 490 nm), B3 (green, 560 nm), B4 (red, 665 nm), B5 (red-edge 1, 705 nm), B6 (red-edge 2, 740 nm), B7 (red-edge 3, 783 nm), B8 (NIR, 842 nm), B8A (narrow NIR, 865 nm), B11 (SWIR-1, 1610 nm), and B12 (SWIR-2, 2190 nm). Bands natively acquired at 20 m (B5–B7, B8A, B11, B12) were implicitly resampled to the 10 m analysis grid via nearest-neighbor interpolation, the default method in GEE, during zonal statistic computation. Digital number values were converted to surface reflectance by applying the collection scale factor of 10^{-4}, and resulting values were constrained to the physically valid $[0, 1]$ range.

Cloud and cloud-shadow masking followed a two-stage approach. First, scenes with a metadata cloud cover percentage exceeding 30% were excluded. For the remaining scenes, the Scene Classification Layer (SCL) produced by Sen2Cor was used to flag and mask pixels classified as no data (class 0), saturated or defective (1), dark area (2), cloud shadow (3), medium-probability cloud (8), high-probability cloud (9), thin cirrus (10), or snow/ice (11). A morphological dilation with a 100 m radius was then applied around all flagged pixels to suppress residual artifacts along cloud and shadow edges (Frantz et al, 2018).

A total of 31 distinct spectral vegetation indices (VIs) were computed from the cloud-free reflectance mosaics (Supplementary Table S1); NDWI and LSWI, which reduce to the same NIR/SWIR1 normalized difference for the Sentinel-2 band set, are counted once. The index suite was designed to capture complementary biophysical dimensions: canopy greenness and photosynthetic capacity (e.g., NDVI, EVI2, GCVI), chlorophyll and nitrogen status via the red-edge domain (e.g., NDRE, CIre, MTCI, S2REP, CCCI), canopy water content (e.g., NDWI, MSI), pigment composition and senescence dynamics (e.g., PSRI, mARI, CVI), and structural properties (e.g., SAVI, OSAVI, PVI). We deliberately emphasized the red-edge spectral region, as Sentinel-2's three dedicated red-edge bands (B5–B7) have shown strong sensitivity to foliar nitrogen and chlorophyll concentration in wheat canopies (Clevers and Gitelson, 2013; Prey and Schmidhalter, 2019; Zhao et al, 2019). The complete list of indices, their formulas, and original references is provided in Supplementary Table S1.

For each Sentinel-2 acquisition, the spatial mean of every band and VI was calculated across all valid (unmasked) pixels within each field polygon, yielding one spectral

observation per field per overpass. Over the 12-month study period, this procedure generated a multi-temporal series with approximately 10–12 observations per month per field, depending on cloud conditions and orbital overlap.

2.2.2 Meteorological Data

Daily meteorological variables were sourced from the ERA5-Land reanalysis (Muñoz-Sabater et al, 2021), a global land-surface dataset produced at 0.1° (∼9 km) resolution by the European Centre for Medium-Range Weather Forecasts (ECMWF). Data were accessed via GEE (ECMWF/ERA5_LAND/DAILY_AGGR) and spatially matched to each field centroid. Seven primary variables were extracted: mean, minimum, and maximum 2 m air temperature (T_{mean}, T_{min}, T_{max}; °C), total precipitation (P; mm day^{-1}), potential evapotranspiration (PET; mm day^{-1}), 2 m dewpoint temperature (T_{dew}; °C), and surface downwelling shortwave radiation (SSRD; MJ m^{-2} day^{-1}).

Three derived agroclimatic variables were additionally computed. Vapor pressure deficit (VPD; kPa) was calculated from T_{mean} and T_{dew} using the Tetens formula: $e_s = 0.6108 \times \exp(17.27\,T/(T + 237.3))$. VPD is a well-established driver of wheat transpiration and has been linked to grain protein accumulation, particularly under terminal heat stress (Asseng et al, 2019). Growing degree days (GDD; °C day) were computed as $\max(0,\ (T_{\text{max}} + T_{\text{min}})/2 - T_{\text{base}})$ with $T_{\text{base}} = 0$ °C, following standard thermal time conventions for winter wheat (McMaster and Wilhelm, 1997). Finally, a daily water deficit (mm) was obtained as the difference between PET and precipitation, since post-anthesis moisture stress is known to accelerate nitrogen remobilization to the grain, thereby increasing GPC (Gooding et al, 2003).

2.2.3 Soil Properties

Soil physical, chemical, and hydraulic properties were extracted from the Gridded Soil Survey Geographic (gSSURGO) database (Soil Survey Staff, USDA-NRCS), a widely used gridded soil product for the conterminous United States. A total of 43 variables were obtained at two depth intervals, topsoil (0–30 cm) and subsoil (30–100 cm) along with profile-integrated values. The extracted properties included particle size fractions (clay, sand, silt), organic matter content, bulk density, cation exchange capacity, available water capacity (AWC), and saturated hydraulic conductivity (K_{sat}). Water retention parameters, saturation (SAT), drained upper limit (DUL), and lower limit of plant-available water (LL15) were derived using the Saxton and Rawls (2006) pedotransfer functions (Saxton and Rawls, 2006). Soil organic matter, clay content, and water-holding capacity are all known to regulate nitrogen mineralization and plant uptake (Raun and Johnson, 1999), making these properties plausible auxiliary predictors for GPC estimation. An irrigation flag (0 = rainfed, 1 = irrigated) was also assigned to each field from the LANID dataset (Xie et al, 2019).

2.2.4 Topographic Data

Terrain variables were derived from the USGS 3D Elevation Program (3DEP) digital elevation model at 1/3 arc-second (∼10 m) resolution, accessed through GEE. Zonal statistics were computed for each field polygon: mean, minimum, maximum, standard

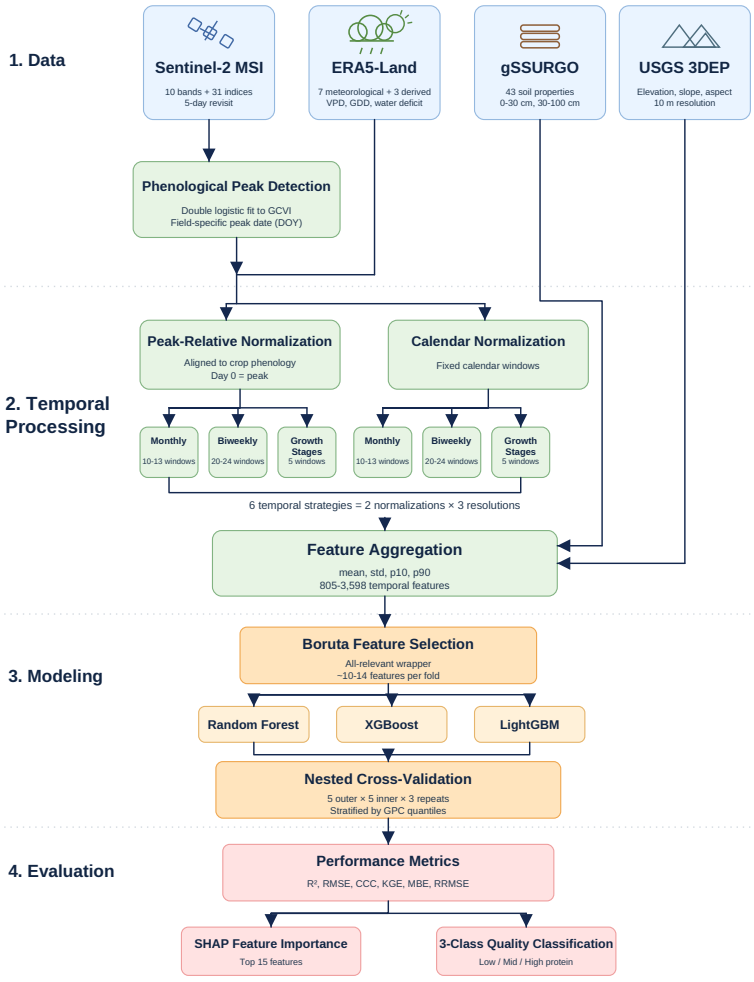

Fig. 2 Overview of the methodological framework for field-level wheat GPC prediction. The pipeline comprises four stages. **Stage 1:** Multi-source data acquisition from Sentinel-2 multispectral imagery (10 bands, 31 vegetation indices), ERA5-Land meteorological reanalysis (7 primary + 3 derived variables), gSSURGO soil properties (43 variables at two depth intervals), and USGS 3DEP topographic data. **Stage 2:** Phenological peak detection via double logistic fitting of GCVI time series, followed by temporal normalization (spectral-peak-relative or calendar) and aggregation at three resolutions (monthly, biweekly, growth stages), yielding six temporal strategies. Features are summarized as mean, standard deviation, and 10th/90th percentiles per window (805–3,598 temporal features + 70 static features). **Stage 3:** Boruta all-relevant feature selection (∼10–14 features per fold), followed by training of three ensemble tree models (Random Forest, XGBoost, LightGBM) within a nested cross-validation framework (5 outer × 5 inner × 3 repeats). **Stage 4:** Model evaluation using six complementary metrics (R^2, RMSE, CCC, KGE, MBE, RRMSE), SHAP-based feature importance analysis, and three-class grain quality classification (Low/Mid/High protein).

deviation, and range of elevation (m), as well as mean slope (°) and mean aspect (° from north). Field centroid coordinates were used only to extract these zonal statistics and to define the spatial cross-validation groups (Section 2.6.2); they were not included as predictors in any model. For the public dataset release these coordinates are aggregated to the county level to protect grower confidentiality (see the *Data and code availability* statement). Topographic position influences soil moisture redistribution and local microclimate, both of which can indirectly affect nitrogen availability and grain quality outcomes (Kravchenko and Bullock, 2000).

2.3 Phenological Peak Detection

A central methodological element of this work is the estimation of field-specific phenological timing from satellite data, which subsequently allows temporal normalization of the feature set. For each of the 240 fields, we estimated the date of maximum canopy greenness and used it as a temporal anchor (day 0) in the peak-relative normalization described in Section 2.4.

2.3.1 Reference Vegetation Index

The Green Chlorophyll Vegetation Index (GCVI = NIR/Green −1; Gitelson et al, 2003) was chosen as the phenological reference variable. GCVI exhibits an approximately linear relationship with canopy chlorophyll content and does not saturate at high biomass, overcoming a well-known limitation of NDVI. This characteristic makes GCVI particularly well suited for reliably detecting peak greenness timing in winter wheat.

2.3.2 Time-Series Preprocessing

The GCVI time series for each field was first restricted to the February 1 – June 30, 2025 window, which captures greenup, peak, and senescence while excluding the largely dormant winter period. Gaps in the time series were filled by linear interpolation to create a daily grid. The interpolated series was then smoothed using a Savitzky–Golay convolution filter (window = 7 days, 2nd-order polynomial; Savitzky and Golay, 1964; Chen et al, 2004) and resampled to 3-day composites. This preprocessing suppresses residual noise from cloud contamination while preserving the timing and shape of key phenological transitions.

2.3.3 Double Logistic Curve Fitting

A seven-parameter double logistic function (Beck et al, 2006) was fitted to each field's smoothed GCVI profile to obtain a continuous representation of seasonal canopy dynamics:

$$\mathrm{GCVI}(t) = m_1 + (m_2 - m_7 \cdot t)\,[\sigma_1(t) - \sigma_2(t)] \tag{1}$$

where:

$$\sigma_1(t) = \frac{1}{1 + \exp\left(\dfrac{m_3 - t}{m_4}\right)}, \qquad \sigma_2(t) = \frac{1}{1 + \exp\left(\dfrac{m_5 - t}{m_6}\right)} \tag{2}$$

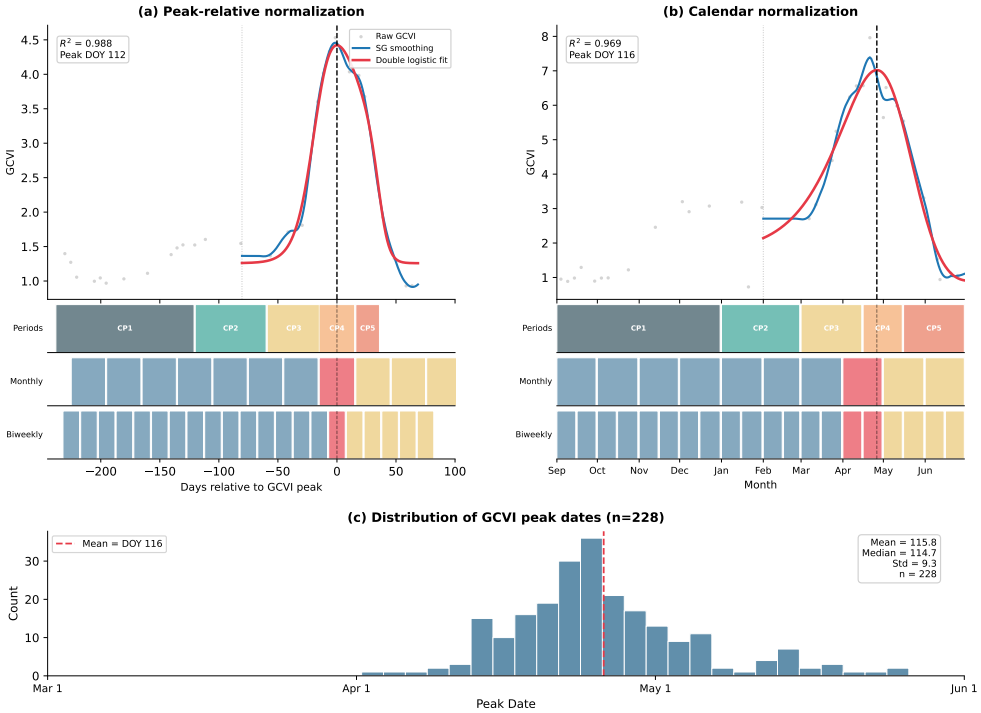

Fig. 3 Phenological peak detection and temporal aggregation framework. (**a**,**b**) GCVI time series for two representative fields showing raw observations (grey), Savitzky–Golay smoothing (blue), and double logistic fit (red), with the three aggregation resolutions displayed below under spectral-peak-relative and calendar normalization, respectively. (**c**) Distribution of detected GCVI peak dates across all 228 fields.

The parameters encode baseline GCVI (m_1), seasonal amplitude (m_2), greenup inflection point and rate (m_3, m_4), senescence inflection point and rate (m_5, m_6), and a linear drift correction (m_7). Estimation was carried out using the Levenberg–Marquardt algorithm as implemented in the *lmfit* library (Newville et al, 2014), with up to 2,000 function evaluations permitted. Initial parameter values were set heuristically from observed GCVI percentiles and timing of extrema to facilitate convergence toward biologically plausible solutions.

The peak date for each field was defined as the day of year (DOY) at which the fitted GCVI curve attained its maximum. This date serves as the phenological anchor point (day 0) for all subsequent peak-relative analyses.

2.3.4 Quality Control

Fields were excluded if their estimated peak fell outside the April–May window (DOY 91–152, consistent with the expected heading period for winter wheat in the central Great Plains (Lollato and Knapp, 2021)), which would indicate anomalous phenological timing for winter wheat in this region, or if the double logistic fit explained less than half the observed variance ($R^2 < 0.50$). These criteria removed 12 fields (5.0%),

leaving a final dataset of 228 fields. Among retained fields, peak dates averaged DOY 116 (late April) with a standard deviation of 9.3 days (range: DOY 92–146), reflecting relatively synchronous but spatially variable crop development across the study area (Figure 3). The mean curve-fit R^2 was 0.954 (range: 0.650–0.996), indicating that the double logistic model captured the observed GCVI trajectory well in the vast majority of fields.

2.4 Temporal Normalization and Aggregation

As established in Section 1, no study has systematically compared spectral-peak-relative and calendar-based temporal windows for GPC prediction. To address this gap, we first identified each field's phenological peak (Section 2.3) and then evaluated six temporal strategies in a factorial design crossing two normalization approaches with three aggregation resolutions, producing six distinct feature generation pipelines (Figure 3).

2.4.1 Temporal Normalization

Two normalization strategies were compared:

1. **Peak-relative (spectral-peak-relative) normalization.** Observation dates were re-expressed as days relative to each field's individually estimated GCVI peak date (Section 2.3), so that day 0 always corresponds to maximum canopy greenness. This effectively removes inter-field phenological variability arising from differences in planting date, cultivar maturity, soil thermal properties, and local microclimate, projecting all fields into a common developmental reference frame. Phenology-driven temporal alignment has shown promise in crop yield forecasting (Sakamoto et al, 2014) and vegetation monitoring (Zeng et al, 2020), although its potential for grain quality prediction has not been systematically explored.
2. **Calendar (fixed-date) normalization.** Dates were retained as absolute calendar dates without any field-specific adjustment. This approach is operationally simpler, since it does not require prior phenological estimation, and is more directly applicable in near-real-time scenarios. However, it implicitly assumes that all fields are at comparable developmental stages on any given date, which is rarely the case in practice.

Within each normalization framework, three levels of temporal resolution were evaluated:

1. **Growth stages (5 windows).** Five broad windows covering major developmental phases. Under peak-relative normalization, periods were defined in days from peak: CP1 [−240, −121], CP2 [−120, −60], CP3 [−59, −16], CP4 [−15, +15], and CP5 [+16, +36], corresponding approximately to fall establishment, overwintering, spring greenup, peak canopy, and early grain filling; these coarse windows are phenologically motivated and vary in length by design. Under calendar normalization, approximate fixed-date windows were used: CP1 [Sep–Dec 2024], CP2 [Jan–Feb 2025], CP3 [Mar 1–Apr 15], CP4 [Apr 16–May 15], and CP5 [May 16–Jun 30].

Table 1 Aggregation functions applied to each predictor variable type during temporal feature generation.

Variable type	Aggregation functions
Vegetation indices (31)	mean, standard deviation, 10th and 90th percentiles
Spectral bands (10)	mean
Temperature variables (mean, min, max)	mean
Precipitation	sum (period accumulation)
PET, GDD	sum (period accumulation)
VPD, SSRD	mean

2. **Monthly (10–13 windows).** For peak-relative normalization, 13 monthly windows were created: nine 30-day pre-peak windows, a central peak window ([−15, +15] days), and three 30-day post-peak windows. The calendar counterpart used the 10 standard months from September 2024 to June 2025.
3. **Biweekly (20–24 windows).** Fifteen-day intervals: 18 pre-peak biweeks, one peak biweek ([−7, +7] days), and 5 post-peak biweeks in the peak-relative version (24 in total); 20 fixed half-month intervals in the calendar version.

The spectral-peak-relative schemes consistently yield slightly more windows than their calendar counterparts because they insert a dedicated window centered on each field's estimated peak date, a period that, by construction, does not coincide with any fixed calendar boundary, and because partitioning the season into fixed-length blocks relative to a variable anchor date does not produce the same number of intervals as a fixed calendar partition.

2.4.2 Feature Aggregation

Daily time-series observations within each temporal window were aggregated into summary statistics using variable-appropriate functions, as summarized in Table 1.

This step converted the unevenly spaced daily time series into a fixed-length feature vector per field. The resulting number of raw features ranged from approximately 805 for the 5-period strategies (fewest windows) to 3,598 for the biweekly strategies (most windows), in addition to 70 static features (soil and topography) shared across all strategies.

2.4.3 Temporal Window Combination Optimization

Rather than selecting a single observation period a priori, we performed an exhaustive search over all possible contiguous window combinations within each of the six strategies. For a strategy defined by k temporal windows, the number of contiguous subsets is $k(k+1)/2$, yielding between 15 (5-period strategies) and 300 (biweekly strategies) candidate combinations. The full modeling pipeline (Sections 2.5–2.7) was executed independently for each combination, and the best-performing one was identified based on the highest cross-validated R^2 achieved by any of the three models. This data-driven approach identifies which phases of the growing season carry the strongest predictive signal for GPC, without relying on a priori assumptions about optimal sensing windows.

2.5 Feature Selection

The high dimensionality of the feature space relative to the sample size ($n = 228$) necessitated rigorous feature selection to mitigate overfitting. We adopted a two-stage procedure combining unsupervised filtering with a supervised wrapper method.

2.5.1 Unsupervised Pre-screening

An initial, target-independent screening was applied to the full dataset prior to cross-validation (CV). Two sequential filters were used: (i) a variance threshold of 10^{-6}, removing near-constant features; and (ii) a collinearity filter that, for each pair of features with Pearson $|r| > 0.90$, retained the member with higher variance. Because this step operates solely on feature distributions without reference to the target variable, it does not introduce information leakage into the cross-validation framework (Hastie et al, 2009). Depending on the temporal strategy, this pre-screening typically eliminated 60–70% of the input features.

2.5.2 Boruta Feature Selection

The pre-screened feature set was then subjected to the Boruta all-relevant feature selection algorithm (Kursa and Rudnicki, 2010), executed independently within each outer cross-validation fold using only training-fold data. Boruta wraps a Random Forest regressor and identifies features whose importance scores consistently exceed those of randomly permuted "shadow" features. A feature was confirmed as relevant if it passed the statistical test at $\alpha = 0.05$ within 100 iterations; tentative features, those neither clearly confirmed nor rejected, were conservatively retained to avoid discarding potentially informative variables in this data-limited setting. The underlying Random Forest estimator used 200 trees with a maximum depth of 7.

Across strategies and folds, Boruta typically selected 10–14 features, achieving a roughly two-order-of-magnitude reduction from the initial feature space. This compression was critical for stable model training given the modest sample size.

2.6 Machine Learning Models

The choice of ensemble tree–based machine learning over classical parametric approaches such as multiple linear regression or partial least squares regression (PLSR) was dictated by three characteristics of this prediction problem. First, the relationship between canopy spectral response and grain protein content is non-linear and is further modulated by cultivar, soil, and management effects that interact in ways that cannot be specified a priori. Second, stacking multiple spectral indices across several temporal windows produces a high-dimensional, strongly collinear feature space in which linear estimators become unstable and prone to over-fitting; tree-based ensembles are largely insensitive to feature collinearity and scale, and perform implicit feature selection at every split (Breiman, 2001; Hastie et al, 2009). Third, with 228 fields the sample size is sufficient to train regularized ensembles but insufficient to reliably fit deep neural architectures, which typically require one to two orders of magnitude more samples to outperform gradient-boosted trees on tabular remote-sensing data

(Grinsztajn et al, 2022; Shwartz-Ziv and Armon, 2022). Random Forest, XGBoost, and LightGBM were therefore selected because they represent the current state of the art on tabular geospatial regression tasks, have been applied successfully to closely related satellite-based seed protein prediction problems (Hernandez et al, 2023), and expose feature-importance diagnostics (e.g., gain, SHAP) that allow the contribution of individual phenological windows to be interpreted, an analytical requirement of this study. Each algorithm and its hyperparameter search space is detailed below:

1. **Random Forest (RF; Breiman, 2001).** An ensemble of decorrelated decision trees trained on bootstrap samples. Predictions are averaged across trees, which reduces variance and provides intrinsic robustness to overfitting. Hyperparameters tuned included number of trees (100–500), maximum depth (5–20, or unconstrained), minimum samples per leaf (1–4), and maximum features per split ($\sqrt{p}$, $\log_2 p$, or $0.5p$).
2. **XGBoost (Chen and Guestrin, 2016).** A gradient-boosted tree algorithm that sequentially fits weak learners to residual errors, with L1 and L2 regularization to control complexity. The tuned hyperparameter space covered boosting rounds (100–600), learning rate (0.01–0.1), tree depth (3–10), row and column subsample ratios (0.7–1.0), and minimum child weight (1–5).
3. **LightGBM (Ke et al, 2017).** A histogram-based gradient boosting framework that uses gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) for improved computational efficiency on high-dimensional inputs. Tuned hyperparameters included boosting rounds (100–600), learning rate (0.01–0.1), tree depth (5–20, or unlimited), number of leaves (31–127), and row and column subsample ratios (0.7–1.0).

Remaining missing values, if any, were imputed with the training-fold median.

2.6.1 Nested Cross-Validation

Training and evaluation followed a nested cross-validation protocol, which is considered essential for obtaining unbiased performance estimates when the dataset is small (Varma and Simon, 2006). Standard (non-nested) CV is known to yield optimistically biased accuracy figures because the same data inform both model selection and evaluation (Vabalas et al, 2019); nesting the loops prevents this information leakage.

The outer loop employed a 5-fold stratified split, where the protein distribution was discretized into five quantile bins for stratification. Each fold thus contained approximately 46 test samples and 182 training samples.

In the inner loop, a 5-fold repeated CV with 3 repeats (15 inner evaluations per configuration) was used to select optimal hyperparameters. Because the tuning grid for each model (Supplementary Table S3) exceeded 15 combinations, hyperparameters were selected by randomized search, sampling 15 configurations per model per outer fold (exhaustive grid search was reserved for grids of 15 or fewer combinations, which none of the three models reached). The scoring metric was negative RMSE. Importantly, Boruta feature selection (Section 2.5.2) was executed independently on each outer fold's training set before the inner hyperparameter search, maintaining strict

separation between feature selection, hyperparameter tuning, and final performance assessment.

Once the best hyperparameters were identified, the model was retrained on the full outer-fold training set and evaluated on the corresponding held-out test fold. Repeating this across all five outer folds yielded five independent, unbiased performance estimates per model per temporal strategy. Additionally, a simple averaging ensemble was computed by taking the arithmetic mean of the three base model predictions on each test fold. All random number generator seeds were fixed to 42 throughout to ensure reproducibility.

2.6.2 Spatial cross-validation and robustness analyses

Because the sampled fields are spatially clustered and several predictors carry environmental gradients, the random stratified outer split above estimates *within-region interpolation* but not spatial transfer. We therefore repeated the full nested protocol with the outer split grouped by *county*, using both spatial k-fold (counties partitioned into five groups) and leave-one-county-out cross-validation; field centroids were used only to assign these groups and were never predictors. To keep model selection honest, all strategy contrasts in this subsection use a single pre-specified estimator (Random Forest), and results are summarized both as pooled out-of-fold R^2 and, for leave-one-county-out, as the distribution of per-county R^2.

Four additional analyses support the interpretation. (i) *Feature-group ablations*, run on identical folds, quantify the contribution of each data source (Sentinel-2 only; meteorology only; soil and topography; with and without absolute elevation) and include a dimension-matched control in which a random subset of features equal in size to the M+1 set is drawn from the full stack. (ii) The advantage of the best spectral-peak-relative window over the best calendar window was assessed with a paired test on per-field errors and a paired bootstrap on the R^2 difference (20,000 resamples), and the window itself was re-selected *inside* each outer fold (training data only) to obtain a selection-unbiased estimate. (iii) A *near-real-time* evaluation re-estimated the phenological peak and rebuilt the M+1 features using only observations available at successive forecast cut-off dates, quantifying the online-versus-retrospective peak error and the associated change in prediction accuracy. (iv) The stability of the greenness peak as an alignment anchor was examined by computing the peak-to-senescence offset (using the senescence inflection m_5 of the double logistic) and by re-running the alignment anchored on m_5 at a matched window. Full code for all analyses is available in the repository.

2.7 Evaluation Metrics

Model performance was quantified using six metrics computed on the out-of-fold test predictions:

1. **Coefficient of determination** (R^2): the proportion of variance in observed GPC that is explained by the predictions.
2. **Lin's Concordance Correlation Coefficient** (CCC; Lin, 1989): jointly evaluates the precision and accuracy of agreement between predicted and observed values. A

CCC of 1.0 indicates perfect concordance, while values below the Pearson r signal the presence of systematic bias.

3. **Root Mean Squared Error** (RMSE; % protein): the magnitude of a typical prediction error, in the same units as the target.
4. **Relative RMSE** (RRMSE; %): RMSE normalized by the mean observed GPC, allowing comparison across studies with different protein ranges.
5. **Mean Bias Error** (MBE; % protein): the average signed prediction error, indicating systematic tendencies toward over- or underestimation.
6. **Kling–Gupta Efficiency** (KGE; Gupta et al, 2009): a composite metric that balances correlation (r), variability ratio ($\alpha = \sigma_{\text{pred}}/\sigma_{\text{obs}}$), and bias ratio ($\beta = \bar{y}_{\text{pred}}/\bar{y}_{\text{obs}}$) into a single score. KGE = 1 represents perfect agreement, and values are penalized when the model compresses or inflates the predicted range relative to observations.

All metrics are reported as mean ± standard deviation over the five outer CV folds to characterize both overall performance and inter-fold variability.

2.8 Model Interpretation

To identify which features drive the GPC predictions, SHAP (SHapley Additive exPlanations; Lundberg and Lee, 2017) values were computed for the best-performing model and temporal combination using the TreeExplainer algorithm. Rooted in cooperative game theory (Shapley, 1953), SHAP assigns each feature a contribution to each individual prediction, allowing us to quantify not only feature importance but also the direction and magnitude of each feature's effect. We generated beeswarm plots to visualize the distribution of SHAP values across all 228 fields, which facilitates the identification of the most influential spectral, meteorological, and edaphic predictors and their functional relationship with grain protein concentration.

3 Results

3.1 Phenological Peak Detection

Of the 240 fields, 228 (95.0%) passed the phenological quality control filters (minimum fit $R^2 \geq 0.50$; peak within April–May) and were retained for subsequent analysis. The double logistic fits achieved a mean fit R^2 of 0.95 (range: 0.65–0.996) (Figure S2).

Estimated GCVI peak dates ranged from DOY 92 to DOY 146, with a mean of 116 ± 9.3 days (Figure 3c). The standard deviation of only 9.3 days indicates that phenology was relatively synchronous across the region; the wider 54-day extreme range is set by a small number of early- and late-peaking fields. This modest dispersion is central to interpreting the temporal-alignment results (Section 4): where most fields peak within a narrow window, calendar and peak-relative compositing largely coincide, leaving alignment little room to outperform a fixed calendar.

3.2 Effect of Temporal Normalization and Resolution

Table 2 summarizes the best-performing temporal window for each combination of normalization, resolution, and model. The Random Forest peak-relative monthly strategy achieved the highest R^2 across all 18 strategy–model combinations ($R^2 = 0.30 \pm 0.05$), followed by the calendar biweekly strategy (LightGBM, $R^2 = 0.25 \pm 0.09$) and calendar growth-stage aggregation (Random Forest, $R^2 = 0.24 \pm 0.15$). Within the peak-relative normalization, the coarse growth-stage aggregation ranked lowest (Random Forest, $R^2 = 0.19$).

The advantage of spectral-peak-relative alignment was most pronounced at the monthly resolution, where peak-relative ($R^2 = 0.30$) exceeded calendar ($R^2 = 0.23$) by an absolute difference of 0.077 in per-fold mean R^2 (the paired-bootstrap estimate of this difference, $\Delta R^2 = 0.088$, and its statistical significance are reported in Section 3.6). At finer (biweekly) and coarser (period) resolutions, performance differences between the two normalization strategies were smaller and did not consistently favor either approach, suggesting that the 30-day monthly window represents a resolution where spectral-peak-relative alignment yields the greatest benefit. All values in this section are obtained under random (within-region) cross-validation; the statistical significance of the peak-relative advantage, and its behavior under spatial cross-validation, are examined in Section 3.6.

The best peak-relative monthly combination was the single post-peak window M+1 (16–45 days after peak GCVI, $R^2 = 0.30$). The best calendar monthly window was June ($R^2 = 0.23$), which corresponds approximately to the grain-filling period in most fields. The best calendar biweekly combination spanned late April through mid-May (Apr2 to May2, $R^2 = 0.22$), covering a similar phenological phase.

The complete cross-validated R^2 landscape across all tested contiguous combinations, separately for each of the six strategies, is provided in Figure S4 (Supplementary Material). This broader screening shows that the strongest predictive signal was concentrated in relatively narrow temporal windows rather than in broad seasonal aggregations.

3.3 Temporal Concentration of the GPC Signal

Single-period analysis under the peak-relative monthly strategy (Figure 4a) showed that M+1 alone ($R^2 = 0.30$, Random Forest) outperformed the peak month ($R^2 = 0.16$) and all pre-peak periods. The second-best single period was M−1 ($R^2 = 0.17$), ahead of the peak month ($R^2 = 0.16$) and M−5 ($R^2 = 0.151$). Under calendar normalization, the best single month was June ($R^2 = 0.23$).

Broader temporal aggregations did not exceed M+1 in this dataset: the full 13-period peak-relative stack (M−9 to M+3; 1,981 features) yielded $R^2 = 0.23 \pm 0.065$ (Random Forest) versus 0.30 ± 0.05 for M+1 alone (217 features). We caution, however, that this specific comparison is asymmetric (M+1 is the maximum of a search over many contiguous window combinations evaluated on the same folds, whereas the full stack is a single unselected configuration), and that window width is confounded with feature dimensionality. We therefore do not claim that M+1 outperforms any broader aggregation. Two controls bound this concern. First, a dimension-matched control

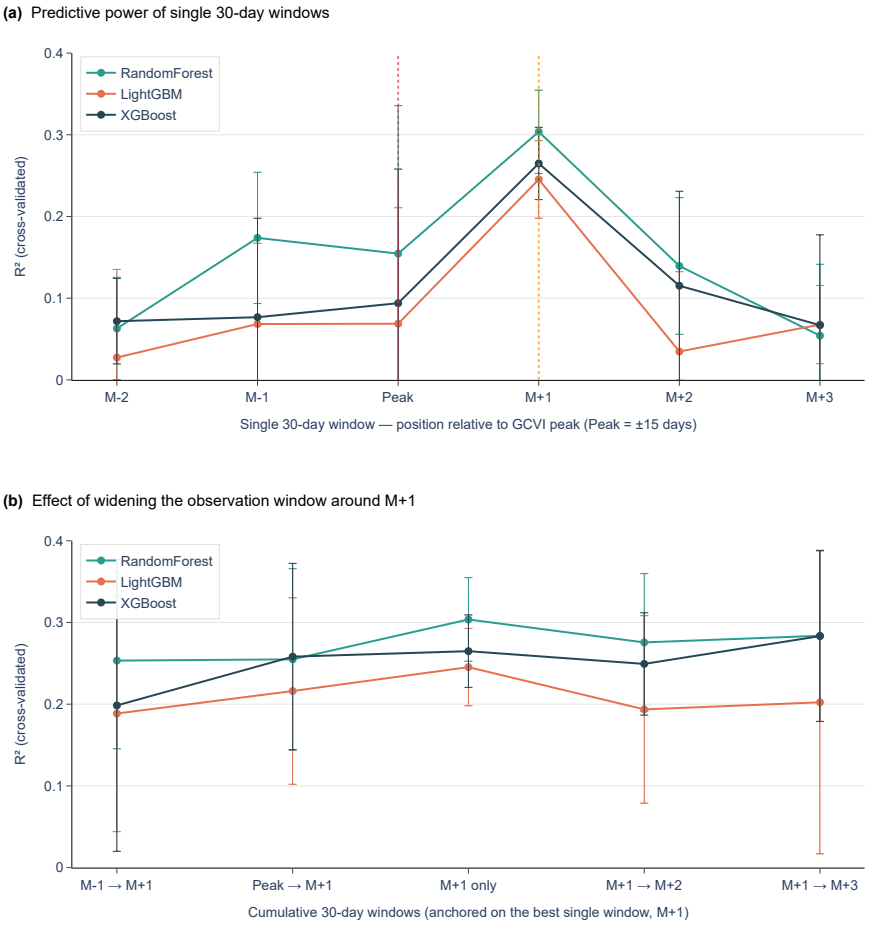

Fig. 4 Temporal sensitivity analysis for the peak-relative monthly strategy. All values represent the mean R^2 across the five outer folds of the nested cross-validation; error bars indicate ± 1 standard deviation across folds. (**a**) Single-period predictive power: each point shows the R^2 obtained when using a single 30-day monthly window in isolation as input to the model. The post-peak window M_{+1} (16–45 days after maximum canopy greenness) achieves the highest R^2 ($= 0.30$, Random Forest), substantially outperforming both the peak month itself ($R^2 = 0.16$) and all earlier or later individual windows. (**b**) Expanding window around the best single window (M+1): each point shows the R^2 obtained when M+1 is combined with one or more adjacent windows. Moving left from the center corresponds to adding preceding windows (Peak, then M−1); moving right corresponds to adding subsequent windows (M+2, then M+3). In this dataset, the single 30-day post-peak window M+1 alone (center, $R^2 = 0.30$, Random Forest) was not exceeded by the broader combinations tested. Because this panel is a search maximum evaluated on the same folds and window width is confounded with feature dimensionality, we do not interpret it as evidence that M+1 outperforms any broader aggregation (Section 3.6); panel (a), an equal-dimensionality single-window comparison, is the basis for the temporal-concentration finding.

Table 2 Complete cross-validated GPC prediction performance for all 18 temporal strategy–model combinations (6 strategies × 3 models). Values are mean ± standard deviation across five outer folds. The best result per metric is shown in **bold**. Results are sorted by temporal normalization (peak-relative, then calendar) and resolution (monthly, biweekly, growth stages).

Norm.	Resol.	Model	Best Window	R^2	RMSE (%)	RRMSE (%)	CCC	KGE
Peak-relative	Monthly	RF	Post-peak (M+1)	**0.30 ± 0.05**	**1.11 ± 0.08**	**9.16 ± 0.62**	0.46 ± 0.04	0.36 ± 0.05
		XGBoost	Post-peak (M+1)	0.27 ± 0.04	1.15 ± 0.11	9.43 ± 0.87	**0.47 ± 0.04**	**0.39 ± 0.05**
		LightGBM	Post-peak (M+1)	0.25 ± 0.05	1.16 ± 0.09	9.54 ± 0.74	0.42 ± 0.03	0.33 ± 0.04
	Biweekly	RF	0–75 days post-peak	0.22 ± 0.11	1.18 ± 0.05	9.67 ± 0.40	0.40 ± 0.07	0.31 ± 0.06
		XGBoost	0–75 days post-peak	0.19 ± 0.12	1.20 ± 0.07	9.85 ± 0.54	0.40 ± 0.11	0.32 ± 0.12
		LightGBM	0–75 days post-peak	0.17 ± 0.12	1.21 ± 0.06	9.98 ± 0.53	0.36 ± 0.10	0.27 ± 0.10
	Growth stages	RF	Peak −120 to +15 days	0.19 ± 0.09	1.20 ± 0.07	9.86 ± 0.63	0.39 ± 0.06	0.30 ± 0.06
		XGBoost	Peak −120 to +15 days	0.10 ± 0.10	1.26 ± 0.06	10.39 ± 0.54	0.35 ± 0.03	0.29 ± 0.06
		LightGBM	Peak −120 to +15 days	0.14 ± 0.14	1.23 ± 0.07	10.14 ± 0.59	0.36 ± 0.07	0.28 ± 0.05
Calendar	Monthly	RF	Jun	0.23 ± 0.15	1.16 ± 0.09	9.59 ± 0.75	0.42 ± 0.09	0.33 ± 0.07
		XGBoost	Jun	0.20 ± 0.17	1.18 ± 0.08	9.73 ± 0.69	0.43 ± 0.09	0.35 ± 0.06
		LightGBM	Jun	0.16 ± 0.16	1.21 ± 0.06	9.99 ± 0.47	0.36 ± 0.09	0.26 ± 0.07
	Biweekly	RF	Apr2–May2	0.22 ± 0.11	1.17 ± 0.02	9.64 ± 0.15	0.39 ± 0.08	0.29 ± 0.07
		XGBoost	Apr2–May2	0.20 ± 0.13	1.19 ± 0.02	9.78 ± 0.19	0.41 ± 0.07	0.33 ± 0.05
		LightGBM	Apr2–May2	0.25 ± 0.09	1.15 ± 0.05	9.46 ± 0.45	0.45 ± 0.07	0.36 ± 0.07
	Growth stages	RF	Jan–Jun	0.24 ± 0.15	1.16 ± 0.07	9.53 ± 0.59	0.43 ± 0.10	0.34 ± 0.07
		XGBoost	Jan–Jun	0.19 ± 0.13	1.19 ± 0.05	9.81 ± 0.43	0.45 ± 0.08	0.38 ± 0.07
		LightGBM	Jan–Jun	0.17 ± 0.19	1.20 ± 0.09	9.91 ± 0.66	0.41 ± 0.10	0.33 ± 0.08

shows the effect is not merely dimensional: a random 217-feature subset of the full stack reached only $R^2 = 0.14$, far below M+1's 0.28 at the same dimensionality, so M+1's advantage reflects signal concentration rather than feature count. Second, when the window is selected inside each outer fold (training data only), M+1 is chosen in 4 of 5 folds and the honestly-nested performance is $R^2 = 0.24$ (versus 0.28 for the fixed M+1), indicating a modest optimistic bias but a stable, reproducible choice of the post-peak window.

3.4 Model Performance

Under the optimal configuration (peak-relative monthly, M+1), Random Forest achieved the highest R^2 (0.30 ± 0.05), followed by XGBoost ($R^2 = 0.27 \pm 0.04$) and LightGBM ($R^2 = 0.25 \pm 0.05$). XGBoost attained the highest CCC (0.47 ± 0.04) and KGE (0.39 ± 0.05). All three models produced near-zero mean bias error (MBE $\leq |0.033|\%$).

Concatenated out-of-fold predictions from the Random Forest model ($n = 228$; Figure 5a) yielded $R^2 = 0.31$,[1] CCC = 0.48, RMSE = 1.11%, and MBE = −0.03%. The RRMSE was 9.2%.

To assess operational relevance for executing early-season quality segregation and marketing decisions, predictions were discretized into three industry-aligned quality classes, Low (<11.4%), Mid (11.4–12.8%), and High (>12.8%), yielding classification

[1]Three R^2 conventions appear in this paper, all computed on the same data and differing only in aggregation and model selection, not in underlying performance: the per-fold mean (0.30 ± 0.05, Table 2); the pooled out-of-fold R^2 (0.31, this section), which pools all 228 predictions into a single regression and is marginally higher because it retains between-fold variance in the observed mean; and the pooled R^2 of the pre-specified pipeline used for the spatial analyses (0.28, Section 3.6), which fixes the M+1 window and a single Random Forest in advance rather than selecting the best of three models. The secondary analyses (near-real-time, m_5-anchor, and retrospective control) re-run this pre-specified pipeline on marginally different inputs and therefore report values within a few hundredths of the headline figures (spatial M+1 R^2 between 0.13 and 0.15; random between 0.28 and 0.30); these small differences reflect the specific configuration, not changes in performance.

accuracies of 47–49% versus a 33% random baseline (Figure 5b). Misclassifications were concentrated between adjacent classes, with few Low–High confusions.

3.5 Feature Importance

Boruta selected 10–14 features per fold from ∼200 candidates (after unsupervised pre-screening), achieving >95% dimensionality reduction. SHAP analysis of the best-performing model (Random Forest, M+1) identified predictors from four domains (Figure 6):

Soil properties. Topsoil organic matter content was the most influential predictor (mean |SHAP| = 0.16). Subsoil saturated hydraulic conductivity (K_{sat}) also appeared among the top features.

SWIR-based spectral indices. NBR2 temporal standard deviation (|SHAP| = 0.15) and mean MIRBI (|SHAP| = 0.15) during M+1 ranked second and third.

Red-edge indices. MTCI, CVI, and S2REP variability during M+1 contributed to predictions.

Meteorological and topographic features. Maximum air temperature and dewpoint during grain filling appeared among the top 15 predictors. Minimum field elevation was the most consistently selected feature across folds (5/5).

The four most stable features across cross-validation folds, NBR2_std, MIRBI_mean, minimum elevation, and GVI 10th percentile (each selected in 4–5 of 5 folds; Table S5), spanned three of the four predictor domains.

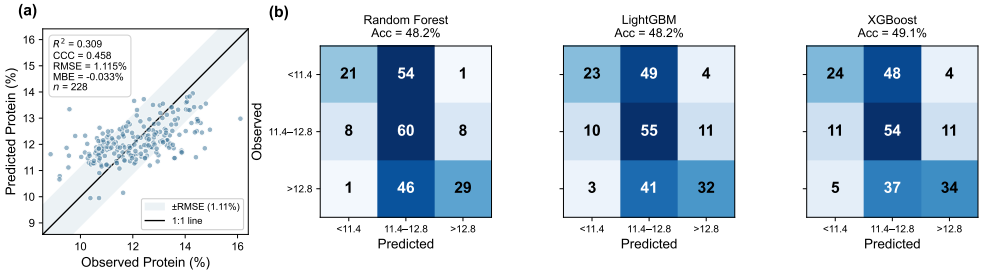

Fig. 5 Model evaluation under the optimal configuration (peak-relative monthly, M+1). (**a**) Predicted versus observed GPC for the Random Forest model (concatenated out-of-fold predictions, $n = 228$). The shaded band indicates ±RMSE around the 1:1 line. (**b**) Confusion matrices for three-class quality discretization (Low <11.4%, Mid 11.4–12.8%, High >12.8%) for all three models.

3.6 Spatial generalization, feature contributions, and robustness

The performance reported above reflects within-region interpolation under random cross-validation. When the outer split was grouped by county so that entire counties were held out, predictive skill dropped sharply (Figure 7a): the M+1 model fell from a pooled R^2 of 0.28 under random cross-validation to 0.13 under spatial cross-validation, and the best calendar strategy from 0.20 to 0.08. Leave-one-county-out

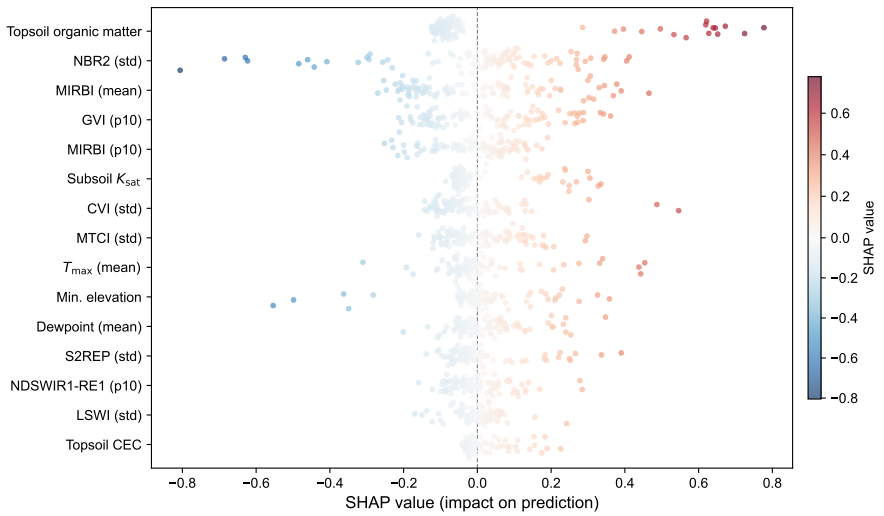

Fig. 6 SHAP beeswarm plot for the Random Forest model under the optimal temporal configuration (M+1). Each dot represents one field; horizontal position indicates the SHAP value (impact on prediction); color encodes the SHAP value magnitude and direction (red = positive, blue = negative). The 15 most important features are shown, ranked by mean absolute SHAP value.

validation agreed with the spatial k-fold estimate but revealed the underlying instability directly: the median per-county R^2 was -0.23, with 8 of 12 evaluable counties negative (Figure 7b; counties with at least three fields were evaluated, and the remaining 7 of the 19 counties, each with only one or two fields, were too small for a per-county estimate). For a previously unseen county, the model therefore has essentially no predictive skill.

Feature-group ablations on identical folds probed the source of the residual transferable signal (Figure 8). Under spatial cross-validation, Sentinel-2 spectral features alone had negligible skill ($R^2 = 0.01$), and meteorological features alone were slightly negative ($R^2 \approx -0.09$ under both random and spatial cross-validation), indicating that the coarse, field-level weather aggregates carry no independent GPC signal and act as noise out-of-sample; with all static context removed (spectral and meteorological features only), performance was $R^2 = 0.03$. The static features were not sufficient on their own either: soil and topography alone reached only $R^2 = 0.03$ (0.04 with absolute elevation excluded), essentially the same as removing them. The modest residual cross-county skill (full model $R^2 = 0.13$) therefore arose from the combination of the spectral signal with the static soil and topographic context, rather than from either block alone. Removing the topographic features (absolute elevation and slope) did not reduce spatial performance (pooled R^2 0.13 with all features versus 0.16 without; the per-field error difference was not significant, paired Wilcoxon two-sided $p = 0.057$), indicating that absolute elevation contributed no genuine transferable skill. A dimension-matched control confirmed that the M+1 advantage is not an artifact of feature count: a random 217-feature subset of the full stack reached only $R^2 = 0.14$, far below the 0.28 obtained by the targeted M+1 set at the same dimensionality. As

reference points, a mean predictor gives $R^2 \approx 0$ by construction, and a transparent linear baseline (ElasticNet) was substantially weaker than the tree ensembles on the M+1 set ($R^2 = 0.13$ random, 0.02 spatial), so the reported skill is not an artifact of model flexibility either.

The advantage of the best spectral-peak-relative window over the best calendar window, although consistent in direction, was not statistically significant. For the equal-dimensionality M+1 versus June comparison, the paired bootstrap on the R^2 difference gave $\Delta R^2 = +0.088$ (95% CI $[-0.035, +0.218]$, $p = 0.079$) under random cross-validation and $+0.054$ ($p = 0.198$) under spatial cross-validation. When the window was re-selected inside each outer fold, however, M+1 was chosen in 4 of 5 folds, and the selection-unbiased performance ($R^2 = 0.24$) was close to the fixed-M+1 value (0.28), indicating that the post-peak window is a stable, reproducible choice rather than a chance search maximum.

Under a near-real-time protocol, the field-specific peak became reliably estimable (mean absolute error ≤ 7 days) only around 1 June (Supplementary Figure S5a), which nonetheless precedes the close of the M+1 window (∼11 June); the binding constraint is thus the window rather than the detection delay. Rebuilding the M+1 features from the online (1 June) peak retained most of the within-region skill (random R^2 0.30 → 0.22; spatial 0.15 → 0.07; Supplementary Figure S5b), indicating that a near-real-time application a few weeks before harvest is feasible at a modest accuracy cost, whereas spatial transfer does not survive. Field-specific harvest dates were not available for these commercial fields, so lead time is expressed relative to the regional hard red winter wheat harvest (late June–July) rather than to each field's own harvest date.

Finally, the greenness peak was a defensible, if imperfect, alignment anchor. The peak-to-senescence offset (m_5 minus peak) averaged 26 days (SD 11), with 88% of fields senescing within the M+1 window (Supplementary Figure S6a); the offset covaried with grain-filling temperature ($r = -0.33$, Supplementary Figure S6b) but not with terminal water deficit ($r = -0.08$, $p = 0.25$) or with GPC ($r = 0.09$, $p = 0.16$), so the residual misalignment is not correlated with the target and does not induce a structural bias. Cultivar identity was not available for these commercial fields, so covariation of the offset with cultivar maturity class could not be tested directly. Re-anchoring the alignment on the senescence inflection m_5 at a matched window did not improve prediction (random R^2 0.29 for the peak versus 0.18 for m_5; spatial 0.12 versus -0.02), consistent with the greenness peak being estimated more stably than the inflection parameter. We also examined whether the peak marks a consistent thermal time. Expressed in common units, the field-to-field spread of peak timing was comparable whether measured as a calendar date (standard deviation 9.3 days) or as thermal time (growing degree-day accumulation to the peak, standard deviation 153 °C day, equivalent to ∼9 days at the local accumulation rate of ∼17 °C day^{-1}). Thermal-time anchoring would therefore not tighten peak timing relative to the calendar; absent independent phenological observations, we describe the anchor as spectral-peak-relative rather than claiming it coincides with a specific developmental stage.

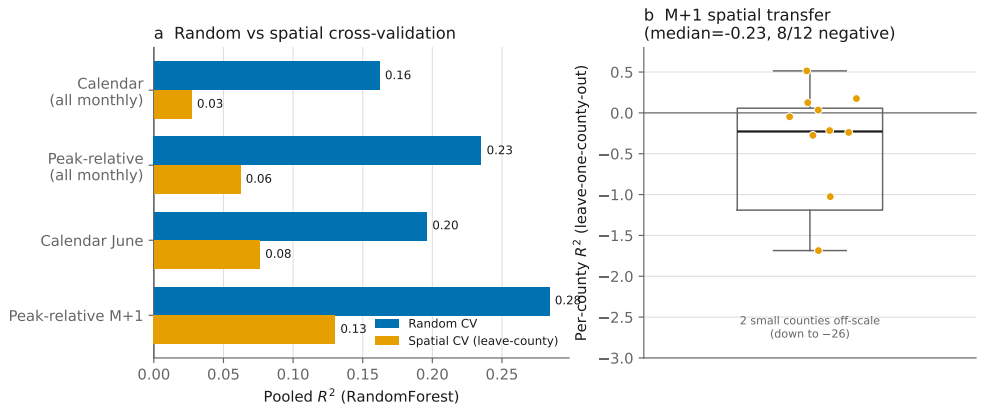

Fig. 7 Spatial generalization of GPC prediction. (**a**) Pooled R^2 (Random Forest) under random versus leave-county spatial cross-validation, for the peak-relative and calendar strategies. (**b**) Distribution of per-county R^2 for the M+1 model under leave-one-county-out validation; the model has near-zero or negative skill for a typical held-out county (two small counties fall off-scale).

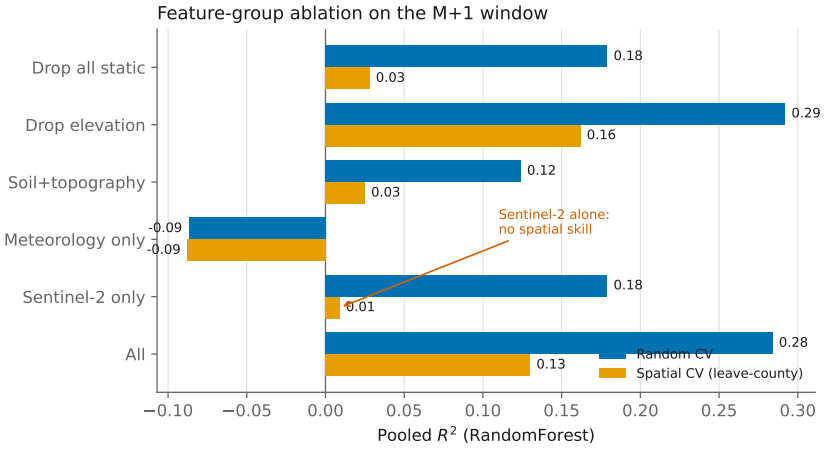

Fig. 8 Feature-group ablation on the M+1 window (Random Forest), under random and spatial cross-validation. Sentinel-2 spectral features alone have negligible spatial skill, and the static soil/topographic features alone are also insufficient; the modest residual cross-county skill arises only from their combination.

4 Discussion

This study provides a novel systematic comparison of crop phenology- and calendar-based temporal windows using remotely sensed satellite data to build predictive models of wheat grain protein concentration. Analogous benefits of a crop phenology approach have been documented for yield prediction in maize and soybean (Bolton and Friedl, 2013; Sakamoto et al, 2014; Zeng et al, 2020) but they have not yet been demonstrated for grain quality traits (except for soybean, Hernandez et al, 2023), which

are governed by fundamentally different physiological processes (Gooding et al, 2003; Martre et al, 2003). Our predictive accuracy is comparable to that of a methodologically similar study using multispectral data derived from UAV (Wolters et al, 2022), and considerably lower than multi-year regional models (Tan et al, 2020; Xu et al, 2024, 2020). Our single-season approach captures within-season field-level variability rather than the inter-annual signal exploited by multi-year models (Lobell and Ortiz-Monasterio, 2006), representing a fundamentally different prediction scope. The physics-based approach of (Longmire et al, 2023), which leveraged harvester-mounted protein measurements at an entirely different scale, represents a different framework.

Despite the moderate continuous accuracy, the models distinguished low- from high-protein fields within region (Figure 5), suggesting potential utility for early-season quality segregation in grain elevators and for marketing decisions (Diacono et al, 2013), subject to the important caveat that this skill does not transfer across counties (Section 3.6). Under a near-real-time protocol the field-specific peak becomes reliable only a few weeks before harvest, and prediction accuracy declines modestly when the online rather than the retrospective peak is used (Section 3.6); we therefore present the approach as having near-real-time potential rather than as a demonstrated operational forecast.

A key limitation is that field-level management variables (nitrogen rate and timing, cultivar, planting date, and irrigation) were not available for these commercial fields. The unexplained prediction variance therefore cannot be decomposed, and in particular cannot be attributed to management effects on the basis of the present satellite-only data. Obtaining such records is a priority for future work.

The temporal-resolution dependence of our models, adjusted by crop phenology, helps establish a practical guideline for designing future GPC monitoring frameworks. Importantly, spectral-peak-relative alignment may be especially valuable for remotely sensed crop traits that are not season-integrative but instead emerge within narrow developmental windows. Consistently, (Wolters et al, 2022) and (Graf et al, 2023) support the integration of crop phenology as priors for relevant crop trait characterization derived from satellite data. In such cases, calendar compositing can dilute the relevant signal by mixing observations from non-equivalent growth stages across fields (Wang et al, 2025). Within region and under random cross-validation, two observations support this view (Figure 4). First, a narrow post-peak window performed at least as well as broader seasonal aggregations; we interpret the window-width comparison cautiously, as it is confounded with feature dimensionality, but a dimension-matched control indicates the effect reflects signal concentration rather than feature count (Section 3.6). Second, aligning the window to each field's greenness peak rather than to fixed calendar dates gave a modest gain over the best calendar-monthly equivalent, although this advantage was not statistically significant. This non-significance is itself interpretable: field-to-field phenological dispersion in this single-region, single-season dataset was small (peak-date standard deviation of only 9.3 days), so the calendar and peak-relative windows target largely the same physiological period, with the post-peak M+1 window (approximately DOY 132–161 for the average field) overlapping the best calendar month, June. Where phenology is this synchronous, alignment has little room to outperform a fixed calendar. We therefore read the small, non-significant

ΔR^2 not as a failure of alignment but as a testable prediction: spectral-peak-relative alignment should confer a larger advantage where phenological spread is greater, such as across wider latitudinal or management gradients, multiple sowing dates, or multiple seasons, in which calendar compositing mixes more disparate growth stages. More broadly, the benefit of spectral-peak-relative alignment should extend to crop traits whose spectral signal is concentrated within a limited developmental phase, including quality traits, stress responses, and post-anthesis physiological processes.

A central qualification concerns spatial transfer. All of the accuracy figures above are obtained under random (within-region) cross-validation; when entire counties are held out, predictive skill does not transfer (median per-county $R^2 = -0.23$; Section 3.6). Ablation shows that neither the Sentinel-2 signal nor the static soil and topographic features transfer across counties on their own (each reaching $R^2 \approx 0.03$ under spatial cross-validation), and that the modest residual transferability of the full model ($R^2 = 0.13$) arises only from their combination. We therefore frame the results as evidence for *within-region interpolation* (predicting unsampled fields within a production region where the model has seen nearby, environmentally similar fields) rather than for spatial extrapolation to new regions. This scope is consistent with the single-season, single-region design and with the modest sample size; establishing transferable, operational GPC prediction will require multi-region, multi-year data and, in all likelihood, field-level management covariates that satellites cannot observe. Framed this way, the value of the study is a systematic, honestly-validated map of where spectral-peak-relative alignment helps for a senescence-limited quality trait and where it does not.

This spectral-timing specificity marks a distinction between GPC modeling and season-integrative yield prediction. Unlike yield prediction, where performance typically improves with longer observation windows (Sakamoto et al, 2014; Lobell et al, 2015), within region a single 30-day post-peak window captured as much or more GPC-relevant information than the broader aggregations tested here (equal-dimensionality comparison; the broader-window contrast in Figure 4b is a search maximum and is not over-interpreted). This divergence is physiologically coherent: yield integrates biomass accumulation, whereas GPC is governed by canopy nitrogen status and nitrogen remobilization from senescing vegetative tissues during grain filling (Gooding et al, 2003; Masclaux-Daubresse et al, 2008; Kichey et al, 2007). The spectral signatures most sensitive to this process, such as SWIR reflectance, which is linked to canopy water and dry matter fraction (Gao, 1996), and red-edge shifts, linked to chlorophyll loss (Clevers and Gitelson, 2013; Frampton et al, 2013), are concentrated precisely in this narrow post-peak window. Including earlier phenological stages likely diluted the signal with information unrelated to nitrogen partitioning (Zhu et al, 2026). The secondary predictive contribution from the late booting–heading phase likely reflects a snapshot of pre-anthesis nitrogen status that constrains the total nitrogen pool available for subsequent grain filling (Kichey et al, 2007).

The SWIR-based indices ranked above traditional red-edge indicators. Connected to these indices, NBR2, combining Sentinel-2's 1610 and 2190 nm bands, is sensitive to the cellulose-to-water ratio that increases during senescence (Gao, 1996; Serrano et al, 2002), while the temporal variability of this index during the post-peak window

captures the rate of canopy desiccation linked to nitrogen remobilization (Gooding et al, 2003). Red-edge indices (MTCI, CVI, S2REP) provided complementary information on chlorophyll dynamics. These results highlight the underexploited potential of Sentinel-2's SWIR bands for grain quality estimation (Clevers and Gitelson, 2013; Zhao et al, 2019; Perich et al, 2021; Bossung et al, 2022; Pancorbo et al, 2023; Jung et al, 2026). Regarding meteorological variables, the maximum air temperature during grain filling appeared among the top predictors. Although terminal heat stress is known to accelerate senescence and concentrate grain nitrogen (Stone and Nicolas, 1994; Asseng et al, 2019; Shi et al, 2024; Spiertz et al, 2006; Viswanathan and Khanna-Chopra, 2001), we note that in our data grain-filling maximum temperature was not positively associated with GPC at the field level (weak, marginally negative bivariate correlation), so this predictor should be interpreted as an environmental covariate rather than as evidence of a heat-driven protein increase in this dataset. Among the static predictors, topsoil organic matter was the most influential. This factor likely functions as an integrator of a broader suite of correlated field characteristics, such as greater water retention capacity, and more intensive management histories, rather than acting as a direct causal driver of GPC variability (Raun and Johnson, 1999). Finally, minimum field elevation was selected as an important feature across all cross-validation folds, as also reported by (Tilse et al, 2025), likely acting as a proxy for landscape position and local moisture redistribution (Kravchenko and Bullock, 2000; Kumhálová et al, 2011). At this inter-field scale, absolute elevation does not represent local micro-topography but rather acts as a spatial proxy for the macro-climatic gradient. In western Kansas, the east-to-west increase in elevation is tightly correlated with increasing regional aridity, allowing the model to capture broad geographic shifts in baseline productivity and environmental stress. Crucially, this proxy operates as a quasi-identifier of location: it aids prediction under random cross-validation, where environmentally similar neighboring fields appear in both the training and test folds, but it cannot transfer to entirely unseen counties. This is consistent with the ablation in Section 3.6, in which removing absolute elevation did not reduce spatial performance, and it clarifies that elevation supports within-region interpolation rather than spatial generalization.

A few limitations can be identified with potential for future research: (i) the single-region design limits generalizability; (ii) the modest sample size, rigorously handled through nested cross-validation (Varma and Simon, 2006; Vabalas et al, 2019), inherently limits the model complexity and may inflate performance metrics; and (iii) the absence of detailed field-level management data, i.e., the timing of N fertilization and specific cultivar information (e.g., growth cycle and protein target), all limit the predictive accuracy. Future consideration of other satellite sources such as the Sentinel-2 Next Generation (S2NG) mission, which has additional shortwave infrared (SWIR) bands and improved spatial resolution, could further exploit the grain-filling window identified here. Lastly, exploring other sensor sources (e.g., HLS) and fusion with Landsat thermal infrared bands could provide direct canopy temperature estimates linked to nitrogen dynamics during grain filling (Asseng et al, 2019).

5 Conclusions

This study systematically evaluated six temporal strategies, defined by the factorial combination of two normalization approaches (spectral-peak-relative vs. calendar) and three temporal resolutions (monthly, biweekly, growth stages), for Sentinel-2-based prediction of grain protein concentration in 228 commercial winter wheat fields across western Kansas. The main conclusions are: (i) within region, a single 30-day post-peak window was the top-performing and most consistently selected temporal window, reproducing prior accuracy ($R^2 \approx 0.28$ under random cross-validation), although its advantage over the best calendar window was suggestive rather than statistically significant, in part because the region's synchronous phenology (peak-date standard deviation of 9.3 days) leaves little separation between calendar and peak-relative windows, so the benefit of alignment is expected to grow with phenological dispersion; (ii) consistent with the senescence-driven physiology of GPC, and in contrast to the season-integrating behavior of yield, the predictive signal was temporally concentrated in the early grain-filling window rather than spread across the season; (iii) under leave-county spatial cross-validation, predictive skill did not transfer to unseen counties (the Sentinel-2 signal alone had negligible spatial skill), so the results support within-region interpolation rather than spatial extrapolation, with the modest residual transferability arising from the combination of spectral and static soil/topographic features rather than from either block alone; and (iv) a near-real-time application a few weeks before harvest retained most within-region skill at a modest accuracy cost, indicating potential rather than demonstrated utility for early grain segregation.

Rather than a new or generally transferable method, the contribution is a systematic, honestly-validated assessment of where phenological (spectral-peak-relative) alignment helps for a senescence-limited quality trait and where it does not. Future work should prioritize multi-region and multi-year data to test spatial and temporal transfer, larger samples to resolve the modest peak-relative advantage statistically, and integration of field-level management records (nitrogen, cultivar, planting date, irrigation) that the present satellite-only design cannot capture.

Acknowledgements. We appreciate the support from General Mills and the Sorghum Checkoff to complete this project. The authors also thank the European Space Agency (Copernicus), the European Centre for Medium-Range Weather Forecasts (ERA5-Land), USDA-NRCS (gSSURGO), and USGS (3DEP) for free and open access to the Earth observation and ancillary datasets that made this analysis possible. Computational resources were provided by the Gautschi community cluster at Purdue University.

Declarations

Funding. The authors received no financial funding (funds or grants) for conducting this study. Non-financial support is disclosed under Competing interests.

Competing interests. Ignacio A. Ciampitti reports administrative support and equipment, materials, or supplies were provided by General Mills and by the United Sorghum Checkoff Program. The remaining authors declare that they have no known

competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval. Not applicable.

Consent to participate. Not applicable.

Consent for publication. All authors have read and approved the submitted version of the manuscript and consent to its publication.

Data availability. Sentinel-2, ERA5-Land, gSSURGO, and USGS 3DEP datasets used in this study are publicly available through their respective providers and were accessed through Google Earth Engine. Field-level grain protein concentration measurements were obtained from commercial fields through an industry partnership under a data-sharing agreement (yield data were also collected but are not analyzed in the present study). To support reproducibility while preserving grower confidentiality, a de-identified version of the dataset is released alongside the code repository: exact field-centroid coordinates are removed, partner-issued field identifiers are replaced by anonymous integers, and each record is tagged with its containing U.S. county (Federal Information Processing Standards code) rather than an exact location. Per-field measurements and features are otherwise retained at their original granularity, so that all model-fitting, cross-validation, and feature-importance analyses reported in the paper can be reproduced from the released subset. Researchers requiring access at finer spatial granularity may request it from the corresponding authors subject to the relevant data-sharing terms.

Code availability. The source code that implements the data-acquisition pipeline, the feature-engineering steps, the cross-validation and modeling framework, and the figure-generation scripts is available at https://github.com/Ciampitti-Lab/WheatGPCPipeline.

Author contributions. Vlasios Mangidis: Conceptualization, Methodology, Investigation, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **German Mandrini:** Conceptualization, Methodology, Investigation, Validation, Writing – review & editing, Supervision. **Ignacio Antonio Ciampitti:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Project administration.

References

Asseng S, Martre P, Maiorano A, et al (2019) Climate change impact and adaptation for wheat protein. Global Change Biology 25:155–173. https://doi.org/10.1111/gcb.14481

Bastos L, Reis A, Sharda A, et al (2021) Current status and future opportunities for grain protein prediction using on- and off-combine sensors: A synthesis-analysis of the literature. Remote Sensing 13:5027. https://doi.org/10.3390/rs13245027

Beck P, Atzberger C, Høgda K, et al (2006) Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. Remote Sensing of Environment 100:321–334. https://doi.org/10.1016/j.rse.2005.10.021

Bogard M, Allard V, Brancourt-Hulmel M, et al (2010) Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake by winter wheat. Journal of Experimental Botany 61:4303–4312. https://doi.org/10.1093/jxb/erq238

Bolton D, Friedl M (2013) Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agricultural and Forest Meteorology 173:74–84. https://doi.org/10.1016/j.agrformet.2013.01.007

Bossung C, Schlerf M, Machwitz M (2022) Estimation of canopy nitrogen content in winter wheat from Sentinel-2 images for operational agricultural monitoring. Precision Agriculture 23:2229–2252. https://doi.org/10.1007/s11119-022-09918-y

Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324

Chen J, Jönsson P, Tamura M, et al (2004) A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. Remote Sensing of Environment 91:332–344. https://doi.org/10.1016/j.rse.2004.03.014

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp 785–794, https://doi.org/10.1145/2939672.2939785

Clevers J, Gitelson A (2013) Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. International Journal of Applied Earth Observation and Geoinformation 23:344–351. https://doi.org/10.1016/j.jag.2012.10.008

De Oliveira Silva A, Jaenisch BR, Ciampitti IA, et al (2021) Wheat nitrogen, phosphorus, potassium, and sulfur uptake dynamics under different management practices. Agronomy Journal 113:2752–2769. https://doi.org/10.1002/agj2.20658

Diacono M, Rubino P, Montemurro F (2013) Precision nitrogen management of wheat. A review. Agronomy for Sustainable Development 33:219–241. https://doi.org/10.1007/s13593-012-0111-z

Drusch M, Del Bello U, Carlier S, et al (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment 120:25–36. https://doi.org/10.1016/j.rse.2011.11.026

Frampton W, Dash J, Watmough G, et al (2013) Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS Journal of Photogrammetry and Remote Sensing 82:83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Frantz D, Haß E, Uhl A, et al (2018) Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. Remote Sensing of Environment 215:471–481. https://doi.org/10.1016/j.rse.2018.04.046

Gao BC (1996) NDWI — a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58:257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gitelson A, Gritz Y, Merzlyak M (2003) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology 160:271–282. https://doi.org/10.1078/0176-1617-00887

Gooding M, Ellis R, Shewry P, et al (2003) Effects of restricted water availability and increased temperature on the grain filling, drying and quality of winter wheat. Journal of Cereal Science 37:295–309. https://doi.org/10.1006/jcrs.2002.0501

Gorelick N, Hancher M, Dixon M, et al (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031

Graf LV, Merz QN, Walter A, et al (2023) Insights from field phenotyping improve satellite remote sensing based in-season estimation of winter wheat growth and phenology. Remote Sensing of Environment 299:113860. https://doi.org/10.1016/j.rse.2023.113860

Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on typical tabular data? In: Advances in Neural Information Processing Systems, vol 35. Curran Associates, Red Hook, NY, USA, pp 507–520

Gupta H, Kling H, Yilmaz K, et al (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377:80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Guttieri M, Baenziger P, Frels K, et al (2015) Prospects for selecting wheat with increased zinc and decreased cadmium concentration in grain. Crop Science 55:1712–1728. https://doi.org/10.2135/cropsci2014.08.0559

Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning, 2nd edn. Springer, New York, NY, USA, https://doi.org/10.1007/978-0-387-84858-7

Hernandez C, Correndo A, Prestholt A, et al (2023) On-farm soybean seed protein and oil prediction using satellite data. Computers and Electronics in Agriculture 212:108096. https://doi.org/10.1016/j.compag.2023.108096

Jung JG, Kim JH, Kim C, et al (2026) Assessment of crude protein in wheat kernels using SWIR hyperspectral imaging combined with deep learning-based segmentation. Journal of Agriculture and Food Research 26:102679. https://doi.org/10.1016/j.jafr.2026.102679

Ke G, Meng Q, Finley T, et al (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol 30. Curran Associates, Red Hook, NY, USA, pp 3149–3157

Kichey T, Hirel B, Heumez E, et al (2007) In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. Field Crops Research 102:22–32. https://doi.org/10.1016/j.fcr.2007.01.002

Kravchenko A, Bullock D (2000) Correlation of corn and soybean grain yield with topography and soil properties. Agronomy Journal 92:75–83. https://doi.org/10.2134/agronj2000.92175x

Kumhálová J, Kumhála F, Kroulík M, et al (2011) The impact of topography on soil properties and yield and the effects of weather conditions. Precision Agriculture 12:813–830. https://doi.org/10.1007/s11119-011-9221-x

Kursa M, Rudnicki W (2010) Feature selection with the Boruta package. Journal of Statistical Software 36:1–13. https://doi.org/10.18637/jss.v036.i11

Lin LK (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–268. https://doi.org/10.2307/2532051

Lobell D, Ortiz-Monasterio J (2006) Regional importance of crop yield constraints: Linking simulation models and geostatistics to interpret spatial patterns of yield across Mexican wheat. Field Crops Research 98:196–206. https://doi.org/10.1016/j.fcr.2006.01.007

Lobell D, Thau D, Seifert C, et al (2015) A scalable satellite-based crop yield mapper. Remote Sensing of Environment 164:324–333. https://doi.org/10.1016/j.rse.2015.04.021

Lollato R, Knapp M (2021) Predicting winter wheat heading date: A simple model and its validation in Kansas. Journal of Applied Meteorology and Climatology 60:1653–1667. https://doi.org/10.1175/JAMC-D-21-0063.1

Longmire A, Poblete T, Hornero A, et al (2023) Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. ISPRS Journal of Photogrammetry and Remote Sensing 205:49–62. https://doi.org/10.1016/j.isprsjprs.2023.09.018

Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, vol 30. Curran Associates, Red Hook, NY, USA, pp 4765–4774

Main-Knorn M, Pflug B, Louis J, et al (2017) Sen2Cor for Sentinel-2. In: Proceedings of SPIE, vol 10427. SPIE, Bellingham, WA, USA, p 1042704, https://doi.org/10.1117/12.2278218

Martre P, Porter J, Jamieson P, et al (2003) Modeling grain nitrogen accumulation and protein composition to understand the sink/source regulations of nitrogen remobilization for wheat. Plant Physiology 133:1959–1967. https://doi.org/10.1104/pp.103.030585

Masclaux-Daubresse C, Reisdorf-Cren M, Orsel M (2008) Leaf nitrogen remobilisation for plant development and grain filling. Plant Biology 10:23–36. https://doi.org/10.1111/j.1438-8677.2008.00097.x

McMaster G, Wilhelm W (1997) Growing degree-days: One equation, two interpretations. Agricultural and Forest Meteorology 87:291–300. https://doi.org/10.1016/S0168-1923(97)00027-0

Meng L, Liu H, Ustin S, et al (2022) A generalized model to predict large-scale crop yields integrating satellite-based vegetation index time series and phenology metrics. Ecological Indicators 137:108693. https://doi.org/10.1016/j.ecolind.2022.108693

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, et al (2021) ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data 13:4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Newville M, Stensitzki T, Allen D, et al (2014) LMFIT: Non-linear least-square minimization and curve-fitting for Python. https://doi.org/10.5281/zenodo.11813

Nieto L, Schwalbert R, Prasad P, et al (2021) An integrated approach of field, weather, and satellite data for monitoring maize phenology. Scientific Reports 11:15711. https://doi.org/10.1038/s41598-021-95253-7

Pancorbo JL, Alonso-Ayuso M, Camino C, et al (2023) Airborne hyperspectral and Sentinel imagery to quantify winter wheat traits through ensemble modeling approaches. Precision Agriculture 24:1288–1311. https://doi.org/10.1007/s11119-023-09990-y

Perich G, Aasen H, Verrelst J, et al (2021) Crop nitrogen retrieval methods for simulated Sentinel-2 data using in-field spectrometer data. Remote Sensing 13:2404. https://doi.org/10.3390/rs13122404

Prey L, Schmidhalter U (2019) Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for estimation of grain yield in wheat. International Journal of Applied Earth Observation and Geoinformation 82:101890. https://doi.org/10.1016/j.jag.2019.101890

Raun W, Johnson G (1999) Improving nitrogen use efficiency for cereal production. Agronomy Journal 91:357–363. https://doi.org/10.2134/agronj1999.00021962009100030001x

Sakamoto T, Gitelson A, Arkebauer T (2014) Near real-time prediction of U.S. corn yields based on time-series MODIS data. Remote Sensing of Environment 147:219–231. https://doi.org/10.1016/j.rse.2014.03.008

Savitzky A, Golay M (1964) Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36:1627–1639. https://doi.org/10.1021/ac60214a047

Saxton K, Rawls W (2006) Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Science Society of America Journal 70:1569–1578. https://doi.org/10.2136/sssaj2005.0117

Schlegel A, Assefa Y, Haag L, et al (2017) Yield and soil water in three dryland wheat and grain sorghum rotations. Agronomy Journal 109:227–238. https://doi.org/10.2134/agronj2016.07.0387

Schwalbert R, Amado T, Nieto L, et al (2018) Forecasting maize yield at field scale based on high-resolution satellite imagery. Biosystems Engineering 171:179–192. https://doi.org/10.1016/j.biosystemseng.2018.04.020

Serrano L, Penuelas J, Ustin S (2002) Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. Remote Sensing of Environment 81:355–364. https://doi.org/10.1016/S0034-4257(02)00011-1

Shapley L (1953) A value for n-person games. In: Kuhn H, Tucker A (eds) Contributions to the Theory of Games, vol II. Princeton University Press, Princeton, NJ, USA, p 307–317, https://doi.org/10.1515/9781400881970-018

Shi J, Ding Z, Ge X, et al (2024) Compound extreme heat and drought stress alter the spatial gradients of protein and starch in wheat grains. Agricultural Water Management 303:109049. https://doi.org/10.1016/j.agwat.2024.109049

Shwartz-Ziv R, Armon A (2022) Tabular data: Deep learning is not all you need. Information Fusion 81:84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Spiertz JHJ, Hamer RJ, Xu H, et al (2006) Heat stress in wheat (Triticum aestivum L.): Effects on grain growth and quality traits. European Journal of Agronomy 25:89–95. https://doi.org/10.1016/j.eja.2006.04.012

Stone P, Nicolas M (1994) Wheat cultivars vary widely in their responses of grain yield and quality to short periods of post-anthesis heat stress. Australian Journal of Plant Physiology 21:887–900. https://doi.org/10.1071/PP9940887

Tan C, Zhou X, Zhang P, et al (2020) Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm. PLoS ONE 15:e0228500. https://doi.org/10.1371/journal.pone.0228500

Tilse MJ, Bishop TFA, Filippi P (2025) Predicting within-field grain protein content at scale using agronomic and remote sensing variables, and machine learning. Precision Agriculture 26. https://doi.org/10.1007/s11119-025-10267-9

USDA-NASS (2024) Crop production 2024 summary. Tech. rep., United States Department of Agriculture, National Agricultural Statistics Service, Washington, DC, USA

Vabalas A, Gowen E, Poliakoff E, et al (2019) Machine learning algorithm validation with a limited sample size. PLoS ONE 14:e0224365. https://doi.org/10.1371/journal.pone.0224365

Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7:91. https://doi.org/10.1186/1471-2105-7-91

Viswanathan C, Khanna-Chopra R (2001) Effect of heat stress on grain growth, starch synthesis and protein synthesis in grains of wheat (Triticum aestivum L.) varieties differing in grain weight stability. Journal of Agronomy and Crop Science 186:1–7. https://doi.org/10.1046/j.1439-037x.2001.00432.x

Wang F, Zhang J, Li W, et al (2025) Characterization of N variations in different organs of winter wheat and mapping NUE using low altitude UAV-based remote sensing. Precision Agriculture 26. https://doi.org/10.1007/s11119-025-10234-4

Wolters S, Söderström M, Piikki K, et al (2022) Predicting grain protein concentration in winter wheat (*Triticum aestivum* L.) based on unpiloted aerial vehicle multispectral optical remote sensing. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science 72:788–802. https://doi.org/10.1080/09064710.2022.2085165

Xie Y, Lark T, Brown J, et al (2019) Mapping irrigated cropland extent across the conterminous United States at 30 m resolution using a semi-automatic training approach on Google Earth Engine. ISPRS Journal of Photogrammetry and Remote Sensing 155:136–149. https://doi.org/10.1016/j.isprsjprs.2019.07.005

Xu X, Teng C, Zhao Y, et al (2020) Prediction of wheat grain protein by coupling multisource remote sensing imagery and ECMWF data. Remote Sensing 12:1349. https://doi.org/10.3390/rs12081349

Xu X, Zhou L, Taylor J, et al (2024) The 500-meter long-term winter wheat grain protein content dataset for China from multi-source data. Scientific Data 11:1008. https://doi.org/10.1038/s41597-024-03845-z

Zeng L, Wardlow B, Xiang D, et al (2020) A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sensing of Environment 237:111511. https://doi.org/10.1016/j.rse.2019.111511

Zhang J, Guan K, Chen Z, et al (2025) Aligning satellite-based phenology in a deep learning model for improved crop yield estimates over large regions. Agricultural and Forest Meteorology 372:110675

Zhao H, Song X, Yang G, et al (2019) Monitoring of nitrogen and grain protein content in winter wheat based on Sentinel-2A data. Remote Sensing 11:1724. https://doi.org/10.3390/rs11141724

Zhu H, Zhao C, Li J, et al (2026) Remote sensing estimation for winter wheat plant nitrogen based on nitrogen distribution model building and leaf nitrogen quantitatively inversion using vegetation index clustering method. Computers and Electronics in Agriculture 247:111698. https://doi.org/10.1016/j.compag.2026.111698

Supplementary Material

This document provides supplementary tables and figures referenced in the main text.

Table S1 Complete list of the 31 distinct spectral vegetation indices (NDWI and LSWI share the NIR/SWIR1 formulation and are listed once) computed from Sentinel-2 surface reflectance bands. Here MSI denotes the Moisture Stress Index, distinct from the Sentinel-2 Multi-Spectral Instrument listed in the Nomenclature. Band abbreviations: B2 (blue, 490 nm), B3 (green, 560 nm), B4 (red, 665 nm), B5 (red-edge 1, 705 nm), B6 (red-edge 2, 740 nm), B7 (red-edge 3, 783 nm), B8 (NIR, 842 nm), B8A (narrow NIR, 865 nm), B11 (SWIR-1, 1610 nm), B12 (SWIR-2, 2190 nm).

Index	Formula	Reference
Canopy Greenness		
NDVI	$(B8 - B4)/(B8 + B4)$	Rouse et al., 1974
EVI2	$2.5 \times (B8 - B4)/(B8 + 2.4 \times B4 + 1)$	Jiang et al., 2008
GNDVI	$(B8 - B3)/(B8 + B3)$	Gitelson et al., 1996
GCVI	$B8/B3 - 1$	Gitelson et al., 2003
SAVI	$1.5 \times (B8 - B4)/(B8 + B4 + 0.5)$	Huete, 1988
OSAVI	$1.16 \times (B8 - B4)/(B8 + B4 + 0.16)$	Rondeaux et al., 1996
ExG	$(2B3 - B4 - B2)/(B2 + B3 + B4)$	Woebbecke et al., 1995
PVI	$(B8 - 0.355 \times B4 - 0.149)/1.061$	Richardson & Wiegand, 1977
Chlorophyll / Red-Edge		
NDRE	$(B8A - B5)/(B8A + B5)$	Gitelson & Merzlyak, 1994
NDRE2	$(B7 - B5)/(B7 + B5)$	Barnes et al., 2000
CIre	$B8A/B5 - 1$	Gitelson et al., 2003
IRECI	$(B7 - B4)/(B5/B6)$	Frampton et al., 2013
MTCI	$(B6 - B5)/(B5 - B4)$	Dash & Curran, 2004
S2REP	$705 + 35 \times \frac{(B4+B7)/2-B5}{B6-B5}$	Frampton et al., 2013
RE_ratio	$B7/B5$	—
CCCI	$\frac{(B8A-B5)/(B8A+B5)}{(B8A-B4)/(B8A+B4)}$	Barnes et al., 2000
CVI	$B8 \times B4/B3^2$	Vincini et al., 2008
TCARI	$3[(B5 - B4) - 0.2(B5 - B3)(B5/B4)]$	Haboudane et al., 2002
mARI	$(1/B3 - 1/B5) \times B8$	Gitelson et al., 2001
PSRI	$(B4 - B2)/B6$	Merzlyak et al., 1999
DGCI	$[1 - (B3 - B4)/(B3 + B4 + B2)]/3$	Karcher & Richardson, 2003
Water / SWIR		
NDWI / LSWI	$(B8 - B11)/(B8 + B11)$	Gao, 1996; Xiao et al., 2004
MSI	$B11/B8$	Hunt & Rock, 1989
MNDWI	$(B3 - B11)/(B3 + B11)$	Xu, 2006
NBR	$(B8 - B12)/(B8 + B12)$	Key & Benson, 2006
NBR2	$(B11 - B12)/(B11 + B12)$	Key & Benson, 2006
SMMI	$(B8A - B11)/(B8A + B11)$	Xiao et al., 2004
MIRBI	$10 \times B12 - 9.8 \times B11 + 2$	Trigg & Flasse, 2001
NDSWIR1RedEdge1	$(B5 - B11)/(B5 + B11)$	—
Other		
GVI	$-0.2848B3 - 0.2435B4 + 0.5436B8 + 0.7243B11 + 0.0840B12$	Kauth & Thomas, 1976
DBSI	$\frac{B11-B3}{B11+B3} - \text{NDVI}$	Rasul et al., 2018

Table S2: Soil properties extracted from the gSSURGO database. Properties were aggregated at two depth intervals: topsoil (0–30 cm) and subsoil (30–100 cm). Additional derived variables are indicated with an asterisk (*).

Variable	Depth	Unit	Description
Clay	Top / Sub	%	Clay content

Table S2 continued

Variable	Depth	Unit	Description
Sand	Top / Sub	%	Sand content
Silt	Top / Sub	%	Silt content
Clay ratio*	Profile	—	Subsoil clay / topsoil clay ratio
OM	Top / Sub	%	Organic matter content
OM depth decay*	Profile	—	Subsoil OM / topsoil OM ratio
BD	Top / Sub	g cm^{-3}	Bulk density
CEC	Top / Sub	$cmol_c$ kg^{-1}	Cation exchange capacity
K_{sat} (log_{10})	Top / Sub	μm s^{-1}	Saturated hydraulic conductivity
AWC (volumetric)	Top / Sub	cm^3 cm^{-3}	Available water capacity (volumetric)
AWC (mm)	Top / Sub	mm	Available water capacity per depth interval
AWC total*	Profile	mm	Total profile available water capacity
PAW total*	Profile	mm	Plant-available water (Saxton–Rawls)
SR SAT	Top / Sub	cm^3 cm^{-3}	Saturation water content (Saxton–Rawls)
SR DUL	Top / Sub	cm^3 cm^{-3}	Drained upper limit (field capacity)
SR LL15	Top / Sub	cm^3 cm^{-3}	Lower limit at 15 bar (wilting point)
SR K_{sat}	Top / Sub	mm hr^{-1}	Saturated hydraulic conductivity (Saxton–Rawls)
SR AWC	Top / Sub	mm	Available water capacity (Saxton–Rawls)
Irrigated	Field	binary	Irrigation status (LANID dataset)

Table S3 Hyperparameter search spaces for the inner cross-validation (5-fold × 3 repeats). Because every grid exceeded 15 combinations, randomized search sampling 15 configurations per model per outer fold was used for all three models; the values below are the full grid sizes (search space), of which 15 were evaluated per fold.

Model	Hyperparameter	Search space
Random Forest	`n_estimators`	{100, 200, 400, 500}
	`max_depth`	{5, 10, 15, 20, None}
	`min_samples_leaf`	{1, 2, 4}
	`max_features`	$\{\sqrt{p}, \log_2 p, 0.5p\}$
	Grid size: $4 \times 5 \times 3 \times 3 = 180$	
XGBoost	`n_estimators`	{100, 300, 500, 600}
	`learning_rate`	{0.01, 0.03, 0.05, 0.1}
	`max_depth`	{3, 4, 5, 6, 7, 8, 10}
	`subsample`	{0.7, 0.8, 0.9}
	`colsample_bytree`	{0.7, 0.8, 1.0}
	`min_child_weight`	{1, 3, 5}
	Grid size: $4 \times 4 \times 7 \times 3 \times 3 \times 3 = 9{,}072$	
LightGBM	`n_estimators`	{100, 300, 500, 600}
	`learning_rate`	{0.01, 0.03, 0.05, 0.1}
	`max_depth`	{5, 10, 15, 20, −1}
	`num_leaves`	{31, 63, 127}
	`subsample`	{0.7, 0.8, 0.9}
	`colsample_bytree`	{0.7, 0.8, 1.0}
	Grid size: $4 \times 4 \times 5 \times 3 \times 3 \times 3 = 6{,}480$	

Table S4 Definitions of evaluation metrics used in this study. In all formulas, y_i and $\hat{y}_i$ denote observed and predicted GPC for sample i, $\bar{y}$ and $\bar{\hat{y}}$ their respective means, σ_y and $\sigma_{\hat{y}}$ their standard deviations, and n the number of samples.

Metric	Formula	Optimal value
R^2	$1 - \dfrac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	1 (higher is better)
RMSE	$\sqrt{\dfrac{1}{n}\sum(y_i - \hat{y}_i)^2}$	0 (lower is better)
RRMSE	$\text{RMSE} \,/\, \bar{y} \times 100$	0% (lower is better)
MBE	$\dfrac{1}{n}\sum(\hat{y}_i - y_i)$	0 (unbiased)
CCC	$\dfrac{2\,r\,\sigma_y\sigma_{\hat{y}}}{\sigma_y^2 + \sigma_{\hat{y}}^2 + (\bar{y} - \bar{\hat{y}})^2}$	1 (higher is better)
KGE	$1 - \sqrt{(r-1)^2 + \left(\dfrac{\sigma_{\hat{y}}}{\sigma_y} - 1\right)^2 + \left(\dfrac{\bar{\hat{y}}}{\bar{y}} - 1\right)^2}$	1 (higher is better)

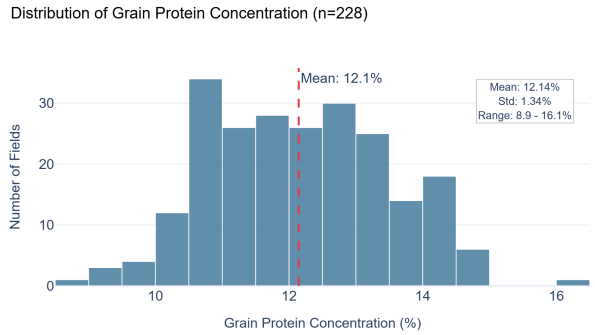

Fig. S1 Distribution of GPC across the 228 validated winter wheat fields. The dashed red line indicates the mean (12.14%). GPC ranged from 8.9% to 16.1% with a standard deviation of 1.34%.

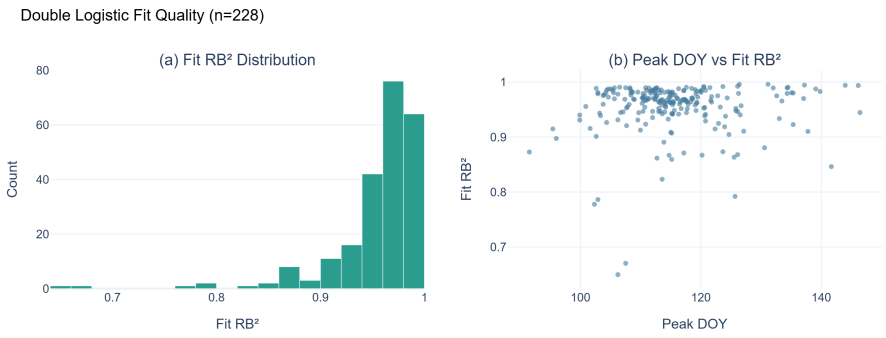

Fig. S2 Quality of the double logistic function fitted to the GCVI time series for phenological peak detection ($n = 228$ fields after QC filtering). (**a**) Distribution of fit R^2 values (median = 0.97). (**b**) Fit R^2 as a function of estimated peak day of year (DOY). Fields with $R^2 < 0.50$ were excluded during quality control (12 fields removed, 5% of initial 240).

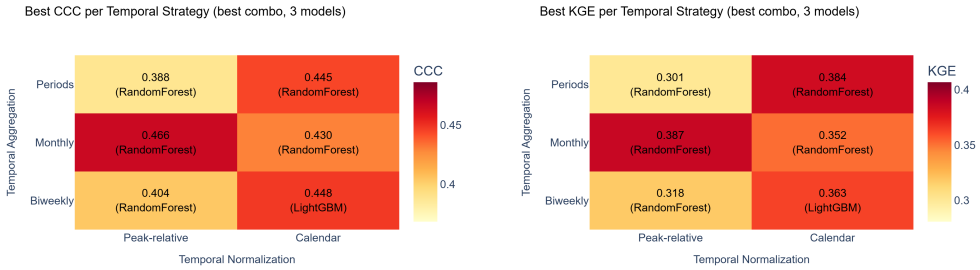

Fig. S3 Concordance correlation coefficient (CCC; left) and Kling–Gupta efficiency (KGE; right) heatmaps for the best-performing model within each temporal strategy. Each cell shows the highest metric value among Random Forest, XGBoost, and LightGBM for the corresponding normalization–resolution combination. These complement the R^2 results reported in the main text (Table 2).

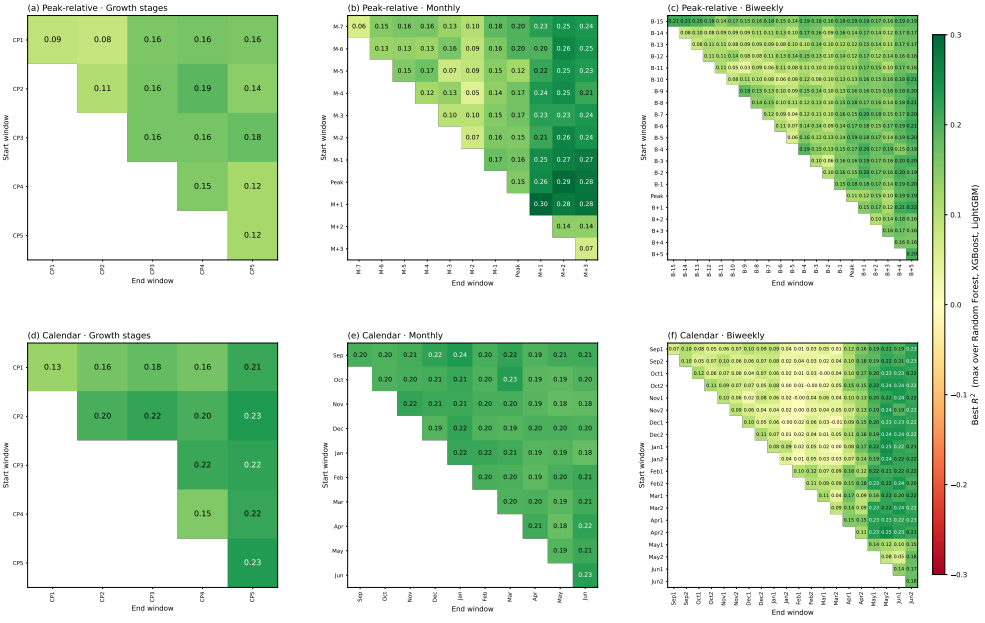

Fig. S4 Complete cross-validated R^2 landscape across all contiguous temporal window combinations tested for each of the six strategies defined in Section 2.4. Each cell shows the best R^2 achieved by any of the three models (Random Forest, XGBoost, LightGBM) when the feature set is restricted to windows from the row (*start*) through the column (*end*). Panels (a–c) correspond to peak-relative normalization (days from each field's estimated GCVI peak); panels (d–f) to calendar normalization. Columns correspond to increasing temporal resolution: growth stages, monthly, and biweekly; for legibility the monthly and biweekly panels display the central window range (M−7 to M+3 and B−15 to B+5) of the full search defined in Section 2.4. In the peak-relative monthly panel (b), the single post-peak window M+1 is among the strongest individual cells ($R^2 = 0.30$). This landscape is shown descriptively only: the equal-dimensionality single-window comparison, its paired significance test, and the nested within-fold window selection are reported in Section 3.6, and we do not interpret it as evidence that M+1 outperforms broader multi-window aggregation. The calendar panels (e,f) show a more diffuse pattern with the highest values clustered around the grain-filling period (May–June).

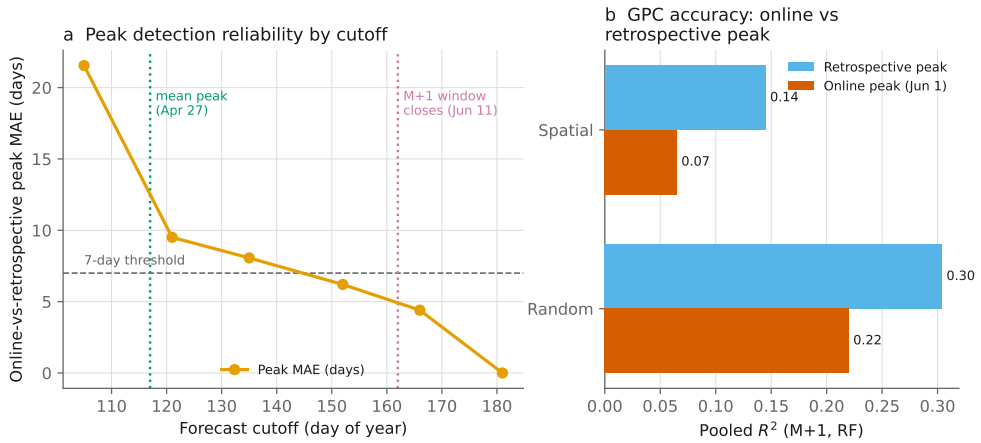

Fig. S5 Near-real-time evaluation. (**a**) Online-versus-retrospective peak error as a function of the forecast cut-off date; the peak becomes reliable ($\leq$ 7 days) before the M+1 window closes ($\sim$11 June). (**b**) GPC prediction accuracy (M+1, Random Forest) using the online (1 June) versus the retrospective peak, under random and spatial cross-validation.

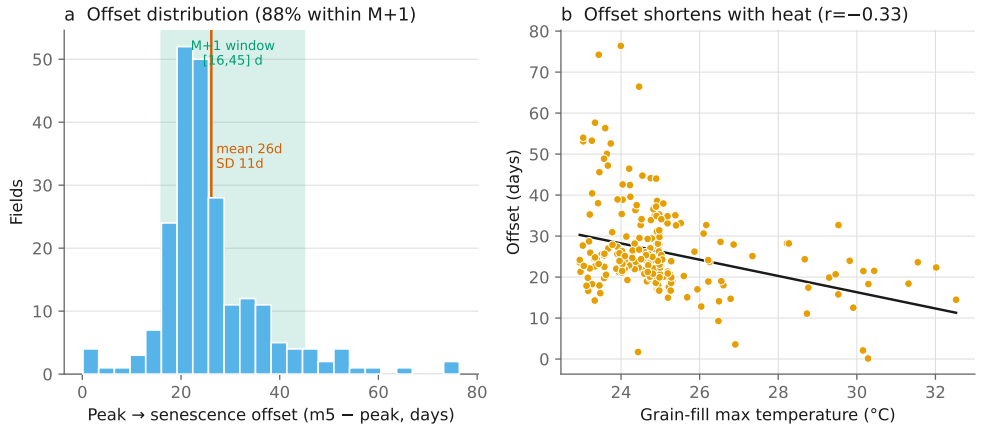

Fig. S6 Peak-to-senescence offset. (**a**) Distribution of the offset (m_5 minus peak, in days); 88% of fields senesce within the M+1 window (shaded). (**b**) The offset shortens with grain-filling maximum temperature ($r = -0.33$) but is uncorrelated with GPC.

Table S5 Top 20 features selected by Boruta across all five outer cross-validation folds under the optimal configuration (peak-relative monthly, M+1). "Appearances" indicates the number of fold–window combinations in which the feature was confirmed as relevant (maximum = 15, i.e., 5 folds × 3 top window combinations). "Combinations" indicates the number of the three top window combinations in which the feature appeared (maximum = 3). Features are ranked by total appearances.

Feature	Category	Appearances	Combinations	Description
elev_min	Topographic	15	3	Minimum field elevation
NBR2_std (M+1)	Spectral/SWIR	14	3	NBR2 temporal variability
MIRBI_mean (M+1)	Spectral/SWIR	14	3	Mean MIRBI during grain filling
CVI_std (M+1)	Spectral/RE	13	3	CVI temporal variability
soil_topsoil_om	Soil	12	3	Topsoil organic matter content
GVI_p10 (M+1)	Spectral/TC	10	3	GVI 10th percentile
NDSWIR1RE1_p90 (M+2)	Spectral/SWIR	9	2	NDSWIR1RedEdge1 90th pctl
MSI_p90 (M+2)	Spectral/Water	8	2	MSI 90th percentile
S2REP_std (M+1)	Spectral/RE	8	3	S2REP temporal variability
soil_sub_ksat	Soil	8	3	Subsoil saturated hyd. cond.
MIRBI_p10 (M+2)	Spectral/SWIR	7	2	MIRBI 10th percentile
GVI_p10 (Peak)	Spectral/TC	5	1	GVI 10th pctl at peak
GVI_p90 (Peak)	Spectral/TC	5	1	GVI 90th pctl at peak
dewpoint_mean (M+1)	Meteorological	5	2	Mean dewpoint during grain filling
MIRBI_p10 (M+1)	Spectral/SWIR	5	3	MIRBI 10th percentile
NBR2_p10 (M+2)	Spectral/SWIR	4	2	NBR2 10th percentile
t2m_max_mean (M+1)	Meteorological	4	1	Mean max temperature
LSWI_std (M+1)	Spectral/Water	4	3	LSWI temporal variability
GCVI_std (M+1)	Spectral/Green	4	2	GCVI temporal variability
MTCI_std (M+1)	Spectral/RE	4	1	MTCI temporal variability

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tables.zip