The Open Agriculture Journal
DOI: 10.2174/0118743315401426101307022 4, 2026, 20, e187433154014

ISSN: 1874-3315
1

RESEARCH ARTICLE **OPEN ACCESS**

Gaussian Process Modelling for Precision Olive Irrigation-Fertilization Management: Continuous Prediction and Uncertainty Quantification of Olive Oil Quality

Lahcen Hssaini[1,*]

[1]*Agro-Food Technology and Quality Laboratory, Regional Center of Agricultural Research of Meknès, National Institute of Agricultural Research (INRA), Rabat, Morocco*

Abstract:

Introduction: Olive oil quality is influenced by irrigation, fertilization, and seasonal climate variability. This study aimed to predict quality at unobserved water-fertilizer combinations while quantifying predictive uncertainty for risk-informed management.

Methods: A two-year (2023-2024) split-plot 3×3 factorial trial at INRA Meknès, Morocco, generated 54 biological olive-oil samples of *Olea europaea* L. cv. Picholine Marocaine across nine water-fertilizer combinations (50-100% ETc × 50-100% NPK), yielding 324 FTIR-ATR acquisitions. Matérn-5/2 Gaussian Process models with Automatic Relevance Determination were trained on 18 treatment-year means. Performance was assessed using Spearman ρ^2 under leave-one-treatment-out and leave-one-year-out cross-validation. Uncertainty was partitioned into aleatoric, temporal, and epistemic components, and Pareto optimisation identified resource-efficient regimes.

Results: LOTO-CV ρ^2 ranged from 0.408 (TFC) to 0.750 (carotenoids), with DPPH at 0.633 and TPC at 0.427. LOYO ρ^2 reached 0.790 for carotenoids and 0.494 for TPC. Temporal uncertainty accounted for ~58% of total uncertainty, versus 36% aleatoric and 6% epistemic. TPC was more sensitive to irrigation than fertilization ($\ell W/\ell F = 0.22$). The Pareto-optimal regime (≈50% ETc × ≈56% NPK) achieved 91% of maximum composite quality with 28% lower fertilisation cost.

Discussion: Seasonal variability was the dominant uncertainty source, highlighting the importance of multi-year data. Carotenoid and TPC patterns transferred across seasons, whereas DPPH required seasonal recalibration.

Conclusion: FTIR-GP modelling with Pareto optimisation offers a promising framework for uncertainty-aware olive management, but broader validation across seasons, sites, and cultivars is needed.

Keywords: Gaussian process regression, Uncertainty quantification, FTIR-ATR spectroscopy, Olive oil phenolics, Response surfaces, Pareto optimization.

© 2026 The Author(s). Published by Bentham Open.
This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: https://creativecommons.org/licenses/by/4.0/legalcode. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Address corresponding to this author at the Agro-Food Technology and Quality Laboratory, Regional Center of Agricultural Research of Meknes, National Institute of Agricultural Research (INRA), Rabat, Morocco;
E-mail: lahcen.hssaini@inra.ma

Cite as: Hssaini L. Gaussian Process Modelling for Precision Olive Irrigation-Fertilization Management: Continuous Prediction and Uncertainty Quantification of Olive Oil Quality. Open Agric J, 2026; 20: e187433154014.
http://dx.doi.org/10.2174/0118743315401426101307022

Received: March 30, 2026
Revised: August 07, 2026
Accepted: August 17, 2026
Published: November 05, 2026

Send Orders for Reprints to reprints@benthamscience.net

1. INTRODUCTION

Olive oil is one of the most economically significant and nutritionally complex agricultural commodities of the Mediterranean region, with phenolic compounds, carotenoids, flavonoids, and antioxidant capacity constituting the primary quality determinants recognised by regulatory bodies and consumers alike [1-3]. These biochemical quality indices are not fixed varietal traits but

dynamic phenotypic outputs shaped by the interaction of irrigation management, mineral nutrition, and seasonal climate variability. Understanding and predicting this interaction across the continuous space of possible agronomic inputs is, therefore, both a scientific question and a direct operational requirement for precision olive farming.

Spectroscopic approaches, and FTIR-ATR spectroscopy in particular, have emerged as the dominant analytical platform for olive oil quality characterisation, offering rapid, non-destructive, reagent-free measurement across the full biochemical fingerprint [4, 5]. The chemometric literature has demonstrated robust classification of oil samples by variety, geographic origin, and adulteration status [6, 7], confirming that the spectral signal encodes meaningful biochemical information [8, 9]. However, the prevailing modelling paradigm, training classifiers or regressors on samples from a fixed set of pre-defined discrete treatment categories, creates a fundamental mismatch with operational farm decision-making. A grower facing a specific resource constraint requires a prediction system that operates across the continuous two-dimensional input space of water and fertiliser availability, not one confined to a handful of observed treatment means. Moreover, discrete spectroscopic models cannot quantify the uncertainty associated with predictions at unobserved input combinations, information essential for risk-informed agronomic decisions [10].

Gaussian Process (GP) regression provides a principled solution to both limitations [11]. Unlike deterministic supervised learning approaches, a GP defines a probability distribution over functions, yielding for every query point in the continuous input space a full posterior distribution characterised by a predictive mean and a calibrated standard deviation [10, 12]. The Matérn-5/2 covariance kernel is particularly well suited to agricultural biochemical dose-response relationships, which exhibit finite rather than infinite differentiability across stress gradients, while Automatic Relevance Determination (ARD) allows trait-specific learning of input sensitivity. Several recent studies have highlighted the potential of GP regression and ensemble machine learning methods for modelling nonlinear patterns in agricultural and food quality contexts, motivating their application to continuous olive oil quality prediction.

A further methodological challenge concerns inter-annual climate variability. Year-to-year differences in cumulative evapotranspiration, heat accumulation, and precipitation timing are known to drive large shifts in olive oil polyphenol and antioxidant levels [13, 14], shifts that can equal or exceed within-year treatment effects. Rather than treating year effects as nuisance noise or removing them as fixed effects, this study proposes a year-adaptive Leave-One-Year-Out (LOYO) validation protocol that explicitly separates the unmodellable climate-driven year-mean shift from the agronomically relevant W×F spatial contrast. The protocol directly answers the question most operationally relevant to growers: does the relative quality ranking of treatment combinations persist across seasons, irrespective of absolute inter-annual shifts?

A parallel methodological issue concerns the choice of validation metric. For traits with high intrinsic measurement variability, notably Total Flavonoid Content (TFC), where within-group analytical variance can substantially exceed between-treatment variance due to the documented instability of aluminium chloride colorimetry in fatty oil matrices [15], conventional Pearson R^2 can return negative values under cross-validation even when the model correctly rank-orders all treatment combinations. Spearman rank-based accuracy (ρ^2) resolves this artefact by measuring rank-ordering fidelity independently of mean bias, which is the operationally relevant criterion for choosing between management regimes. This study therefore adopts ρ^2 as the primary validation metric, supplemented by Pearson R^2 where signal-to-noise conditions permit a meaningful interpretation.

Against this background, we hypothesise that *GP* regression with Matérn-5/2 ARD kernels can provide reliable, continuous, uncertainty-aware predictions of olive oil biochemical quality at unobserved irrigation-fertilisation combinations, with temporal variability, rather than algorithmic complexity, constituting the dominant uncertainty source limiting prescription confidence. Accordingly, the study addresses five research questions:

whether GP models achieve positive Spearman ρ^2 under LOTO-CV for all four biochemical quality indices (RQ1); whether ARD analysis correctly identifies trait-specific input sensitivities consistent with known olive physiology (RQ2); whether temporal uncertainty dominates the pooled uncertainty budget (RQ3); whether W×F spatial patterns transfer meaningfully across seasons under LOYO validation (RQ4); and whether multi-objective Pareto optimisation identifies regimes achieving ≥75% of maximum composite quality at ≥20% input cost reduction (RQ5). The overall analytical pipeline from field experiment through GP modelling, validation, uncertainty decomposition, ARD analysis, and Pareto optimisation is schematically illustrated in Fig. (**1**).

2. MATERIALS AND METHODS

2.1. Experimental Design and Site Characteristics

2.1.1. Field Site

The experiment was conducted over two consecutive growing seasons (2023 and 2024) at the INRA Meknès Regional Centre of Agricultural Research experimental olive orchard, located in Meknès, Morocco (33°53′N, 5°33′W; elevation 540 m a.s.l.). The site is characterised by a semi-arid Mediterranean climate (Köppen-Geiger: Csa) with mean annual rainfall of approximately 300 mm, concentrated between October and April, and mean annual temperature of 17.2°C. The 2023 and 2024 growing seasons differed substantially in cumulative evapotranspiration demand and precipitation timing, representing contrasting climatic conditions relevant to the study objectives.

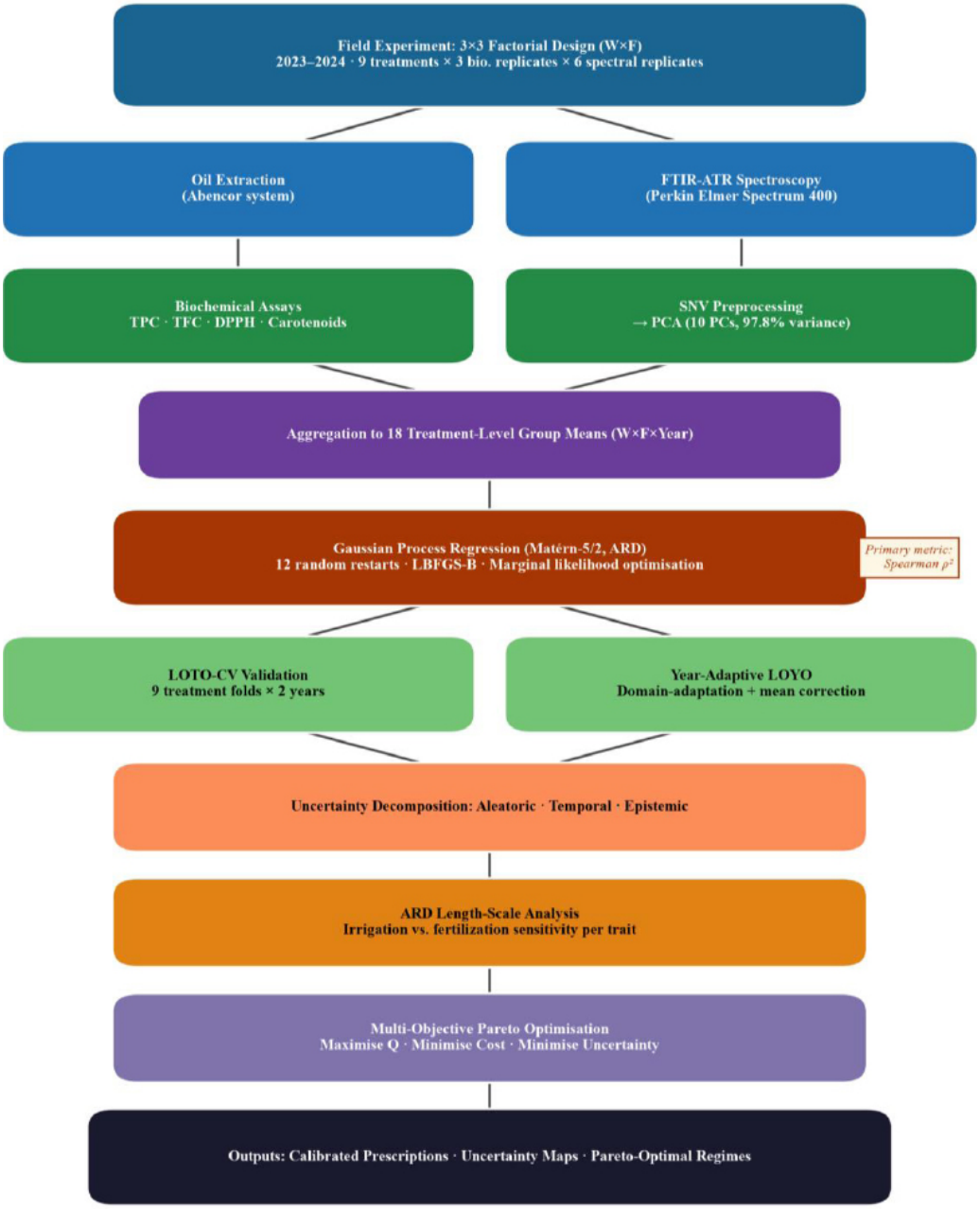


Fig. (1). Methodological flowchart illustrating the complete analytical pipeline from field experiment through GP regression, validation, uncertainty decomposition, ARD analysis, and Pareto optimisation to agronomic outputs.

The orchard consists of *Olea europaea* L. cv. Picholine Marocaine, the predominant Moroccan commercial cultivar, planted in 2008 at a spacing of 7 m × 5 m (286 trees ha^{-1}) in a hedgerow training system. Soil type is a Calcic Cambisol (IUSS WRB 2015) with clay-loam texture (clay: 32%, sand: 28%, silt: 40%), pH 7.8, organic matter 1.2%, and electrical conductivity 0.65 dS m^{-1}. The irrigation water source is the Oued Boufekrane system; water quality analysis indicated pH 7.6, EC 0.38 dS m^{-1}, and sodium adsorption ratio (SAR) of 1.8. Standard orchard management practices (pruning, pest control, harvesting) were applied uniformly across all experimental plots.

2.1.2. Irrigation-Fertilization Factorial Design

A fully crossed 3×3 factorial design yielded nine water-fertilizer treatment combinations. Three irrigation regimes were applied as percentages of crop evapotranspiration (ETc) calculated following Allen *et al*. [16, 17] using the Penman-Monteith equation with site-specific meteoro-

logical data and published crop coefficients (Kc) for Picholine Marocaine [17]: W_{50} (225 mm season^{-1}), W_{75} (338 mm season^{-1}), and W_{100} (450 mm season^{-1}). Irrigation was delivered via subsurface drip emitters (2 L h^{-1}, one lateral per tree row). Three fertilisation levels expressed as percentages of the recommended NPK rate for irrigated Picholine Marocaine (150 kg N ha^{-1}, 75 kg P_2O_5 ha^{-1}, 100 kg K_2O ha^{-1}) were: F_{50} (75:37.5:50 kg ha^{-1}), F_{75} (112.5:56.3:75 kg ha^{-1}), and F_{100} (150:75:100 kg ha^{-1}), applied as ammonium nitrate, triple superphosphate, and potassium sulphate in three equal split doses (March, May, and July).

The split-plot design was applied with irrigation as the main-plot factor and fertilization as the sub-plot factor. Each irrigation-fertilisation treatment combination was represented by three independent biological replicates (plots), each containing eight trees, with two buffer trees between plots. Harvest dates were standardised by monitoring fruit maturity using the Jaén maturity index (target: Maturity Index = 3.5 ± 0.3), yielding harvest dates of 15-22 November 2023 and 12-19 November 2024. Each biological replicate produced one oil sample; each oil sample was measured by FTIR-ATR spectroscopy in six technical replicates (two independent deposits × three scans per deposit), and biochemical assays were conducted in triplicate. The total dataset therefore comprised 54 biological oil samples, 324 FTIR spectra, and 162 biochemical assay results. For GP response-surface modelling, spectra and biochemical measurements were aggregated to 18 treatment-level group means (9 W×F combinations × 2 years), ensuring that GP hyperparameter estimation reflects agronomic signal rather than replicate-level analytical noise.

2.1.3. Oil Extraction

Olives were harvested at uniform commercial maturity within 24 h of collection and processed using the Abencor laboratory system [18]: hammer milling (5 mm screen), malaxation (27°C, 30 min), centrifugation (3,500 rpm, 3 min), and gravitational decantation (4 h at 4°C). Oils were stored in amber glass vials under a nitrogen atmosphere at −20°C until analysis.

2.2. FTIR-ATR Spectroscopy and Biochemical Analysis

2.2.1. Spectral Acquisition

FTIR-ATR spectra were acquired on a Perkin-Elmer Spectrum 400 spectrometer equipped with a single-bounce diamond ATR crystal at 25 ± 0.5°C. Acquisition conditions: 4000-450 cm^{-1}, resolution 4 cm^{-1}, 64 co-added scans, two independent sample deposits per oil sample with three scans per deposit (six spectral replicates per biological sample), and background correction every 30 min. Raw spectra were inspected for outliers and subjected to Standard Normal Variate (SNV) normalisation [19, 20] to correct for multiplicative scattering effects, producing a final spectral matrix $X \in \mathbb{R}^{324 \times 3551}$. No additional baseline correction or smoothing was applied.

2.2.2. Biochemical Determinations

Total Phenolic Content (TPC) was determined by the Folin-Ciocalteu method [21], expressed as mg Gallic Acid Equivalents (GAE) per gram of oil. Briefly, 0.1 mL oil extract (in methanol: water 80:20) was reacted with Folin-Ciocalteu reagent (1:10 dilution) and Na_2CO_3 (7.5% w/v); absorbance was measured at 765 nm after 30 min at room temperature against a gallic acid calibration curve (10-500 mg L^{-1}). Total Flavonoid Content (TFC) was measured by aluminium chloride colorimetry [22], expressed as mg catechin equivalents (CE) per gram; the inter-assay CV was <18%, within the range of variability reported for related colorimetric and spectrophotometric assays [22-25]. DPPH radical scavenging capacity was determined following Brand-Williams *et al*. [23], expressed as % inhibition (IC_{50} method). Carotenoid content was quantified spectrophotometrically at 470 nm [24], expressed as mg β-carotene equivalents per kilogram of oil. All assays were conducted in triplicate; inter-assay CV was <5% for TPC, DPPH, and carotenoids.

2.3. Gaussian Process Regression Framework

The GP modelling framework is visualised in Fig. (**1**). Let f(x) denote the latent agronomic response function relating irrigation level W, fertilization level F, and the year indicator Y to a biochemical response variable y. The GP prior is defined as Eq. (**1**):

$$f(x) \sim GP(m(x), k(x, x')) \quad (1)$$

where m(x) = 0 (zero mean function) and k(x, x′) is the Matérn-5/2 covariance kernel with ARD. Observed responses are modelled as in Eq. (**2**):

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2_noise) \quad (2)$$

2.3.1. Kernel Specification and ARD

The covariance function was as in Eq. (**3**):

$$k(x, x') = \sigma^2_f \cdot k_Matérn(r) + \sigma^2_noise \cdot \delta(x, x') \quad (3)$$

The corresponding ARD length-scale formulation was specified as Eq. (**4**):

$$k(x, x') = \sigma^2_f \cdot k_Matérn(r) + \sigma^2_noise \cdot \delta(x, x') \quad (4)$$

with input vector x = (W, F, Year_bin) and ARD length scales ℓ_W, ℓ_F, and ℓ_Year estimated by maximising the marginal log-likelihood via L-BFGS-B with 12 random restarts (implementation: scikit-learn v1.3.2, Python 3.11; random seed = 42 throughout). The Matérn-5/2 kernel was selected over the infinitely differentiable Radial Basis Function (RBF) because agricultural biochemical dose-response relationships exhibit finite differentiability across stress gradients [26], and over rational-quadratic or periodic alternatives because preliminary exploration showed consistently lower marginal log-likelihood for the latter under the present 3-level factorial design. Year was encoded as a binary covariate (Year_bin ∈ {0, 1}); the

small estimated ℓ_Year at the lower optimisation bound confirms a near-discontinuous year effect, which is the agronomically expected outcome given the contrasting seasonal environments. Inputs (W, F) were scaled to [0, 1] prior to fitting; biochemical response variables were standardised (zero mean, unit variance) within the GP fitting routine (normalize_y=True). Hyperparameter bounds were specified as Eq. (**5**):

$$\ell_d \in [0.01, 100], \sigma^2_f \in [0.01, 100], \quad \sigma^2_noise \in [0.001, 10]. \tag{5}$$

2.3.2. PCA Compression for Spectral GP

Principal Component Analysis (PCA) was applied to SNV-normalised group-mean spectra (n = 18) to extract a 10-component representation (97.8% of total spectral variance; scree plot elbow criterion confirmed retention of ≥95% variance). The number of retained components (10) was determined by the scree plot elbow criterion and the 95% variance threshold applied to 18 group-mean spectra; with n = 18 observations, further components risk overfitting. Separate GP models were fitted to each PC score; spectral reconstructions were obtained via the inverse PCA transform. Uncertainty was propagated through the linear PCA mapping under a first-order independence approximation across components.

2.3.3. Validation Protocols and Spearman Rank Accuracy

For biochemical traits with high aleatoric noise, particularly TFC (σ^2_noise/σ^2_signal ≈ 4.9), Pearson R^2 can be negative under cross-validation even when the model correctly rank-orders all treatment combinations, because residual variance exceeds total variance around the grand mean. Spearman ρ^2 (the square of Spearman rank correlation between observed and predicted values) measures rank-ordering fidelity independently of mean bias, which is the operationally relevant criterion for a grower choosing between management regimes. All primary performance statistics report ρ^2; Pearson R^2 was additionally provided where signal-to-noise conditions permit its interpretation. Spearman ρ itself and its 95% confidence intervals are reported alongside ρ^2. We note that while ρ^2 is a practically useful secondary decision metric, a negative Pearson predictive R^2 for TFC indicates that absolute concentration predictions are less reliable than a mean-only benchmark for this trait; growers should treat TFC prescriptions as rank-based guidance only.

Leave-One-Treatment-Out CV (LOTO-CV) withheld each of the nine W×F treatment combinations; both year observations for the withheld treatment formed the held-out test set (two held-out means per fold, 18 test predictions total). All preprocessing, scaling, PCA fitting (for spectral GP), hyperparameter optimisation, and random restarts were performed independently within each LOTO fold to prevent information leakage. Year-adaptive LOYO trains the GP on all nine group means from one year and predicts the other year after a domain-adaptation mean-shift correction (test-year mean estimated from the full test-year observations). As noted by Reviewer 2, this procedure uses test-year outcome means and therefore measures pattern-transfer fidelity rather than true prospective prediction; its operational interpretation is whether relative W×F treatment rankings are preserved across seasons, not whether absolute concentrations can be predicted before harvest.

2.4. Uncertainty Decomposition

Total predictive uncertainty was separated into three components [27]. Aleatoric uncertainty (σ_al) was estimated as the pooled within-replicate standard deviation per W×F×Year cell across all biological and analytical replicates. Temporal uncertainty (σ_te) was computed as the absolute difference between the two yearly group means for each W×F combination divided by √2 (analogous to a between-year standard deviation from two observations). Epistemic uncertainty (σ_ep) is the GP posterior standard deviation at the treatment group centroid. Because the three components have different statistical definitions and units, they were standardised per trait before pooling by dividing by the trait grand mean, and percentage contributions were computed from the standardised pooled values. With only two years, temporal uncertainty estimates carry substantial sampling variability and should be interpreted as indicative rather than precise variance estimates; extending to ≥5 seasons would substantially improve robustness.

2.5. Multi-Objective Pareto Optimisation

Three objectives were optimised across the continuous W × F space on an 80 × 80 grid: (1) composite biochemical quality, Q, which was maximised and calculated as a weighted combination of the four biochemical traits, as given in Eq. (**6**):

$$Q = 0.4\,TPC + 0.3\,DPPH + 0.2\,TFC + 0.1 \tag{6}$$

where each trait was first normalised to the range [0, 1] according to the IOC quality-weighting guidelines for extra virgin olive oil; (2) resource cost, C, which was minimised and defined as Eq. (**7**):

$$C = 0.4W + 0.6F \tag{7}$$

where the coefficients reflect the relative contribution of fertilisation to total variable input costs; and (3) mean GP epistemic uncertainty, U, which was minimised. Non-dominated sorting following NSGA-II principles [28] was used to identify the Pareto front. GP predictions were computed as the average of the predictions for Year = 0 and Year = 1, rather than by querying the model at the artificial intermediate value of Year_bin = 0.5, thereby providing a valid marginal year-averaged estimate. The quality-index weights followed IOC document DEC-25/95-IV/2006, whereas the cost weights were approximate and site-specific economic sensitivity analyses were not performed. Uncertainty was propagated into Pareto dominance using the GP posterior mean ± 1.96σ envelopes to assess the robustness of the identified solutions.

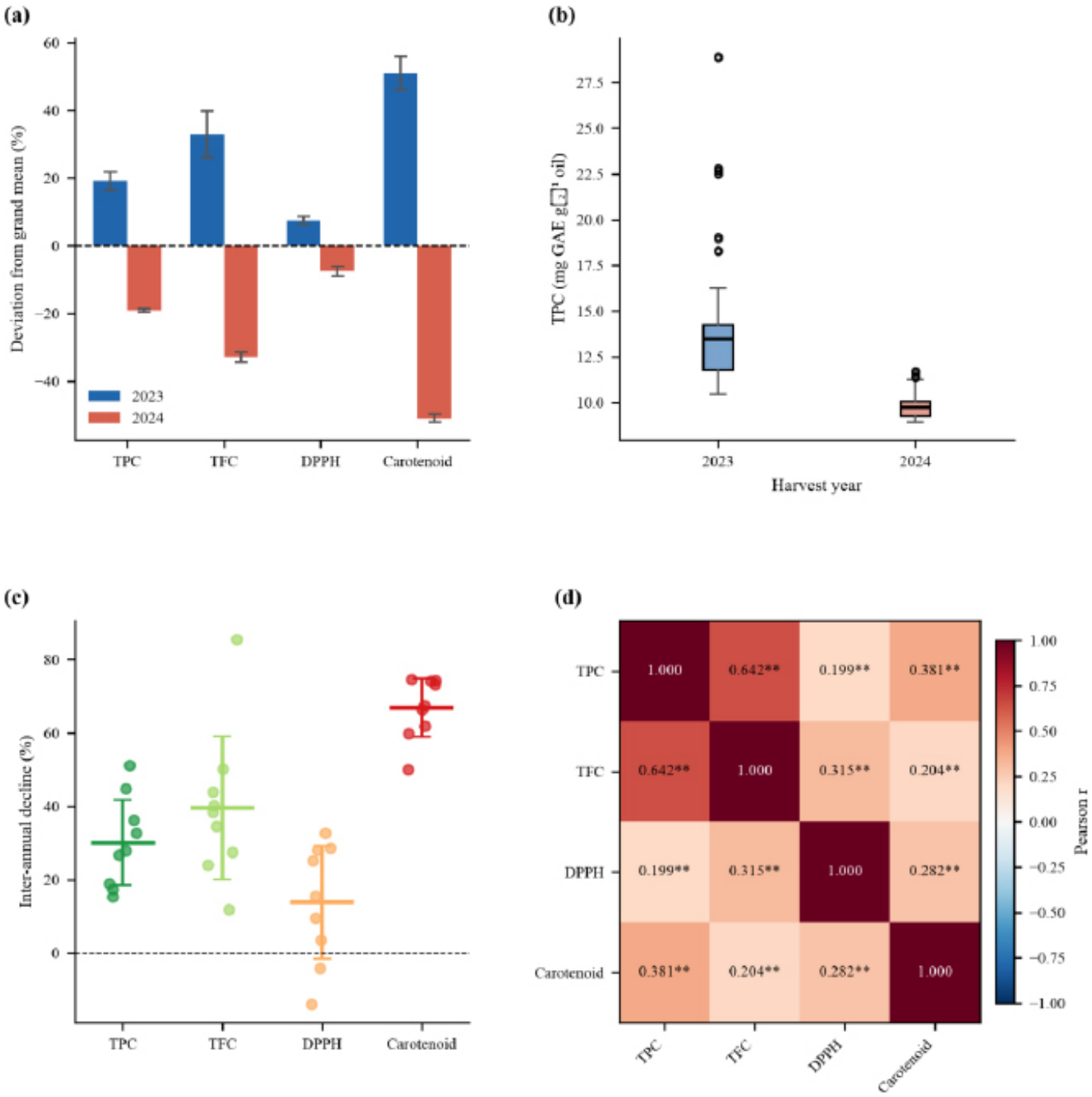


Fig. (2). Biochemical characterisation across harvest years and treatment combinations. (**a**) Deviation of each trait from the two-year grand mean (%) for 2023 and 2024 (bars = mean ± SEM). (**b**) TPC distribution by harvest year (box plots: median, IQR, 1.5×IQR whiskers; open circles = outliers). (**c**) Inter-annual percentage decline for all four biochemical traits across nine W×F treatment combinations (points = individual treatment values; horizontal bars = mean ± 1 SD). (**d**) Pearson correlation matrix for 54 biological sample means (** $p < 0.001$). **Abbreviations:** TPC = total phenolic content; TFC = total flavonoid content; DPPH = DPPH radical scavenging capacity; W = irrigation (% ETc); F = fertilization (% NPK).

3. RESULTS

3.1. Biochemical Characterisation and Inter-Annual Variability

Complete biochemical datasets were obtained for all nine W×F treatment combinations in both harvest years (Figs. **2a** and **3**). Pronounced inter-annual differences were observed in all four quality indices (Fig. **2b**). The corresponding percentage changes between harvest years are summarized in Fig. (**2c**). Total Phenolic Content declined from a 2023 grand mean of 14.52 mg GAE g^{-1} to 9.87 mg GAE g^{-1} in 2024 (−32.0%). Total Flavonoid Content declined by 49.5% (0.318 to 0.161 mg CE g^{-1}), carotenoids declined by 67.5% (1.923 to 0.625 mg kg^{-1} oil), and DPPH showed the smallest relative year effect at −13.9% (63.27 to 54.47% inhibition). Within each year, treatment effects were detectable across the nine W×F combinations (as illustrated by the treatment-specific heatmaps in Fig. **3**); in 2023, TPC ranged from 11.49 to 18.91 mg GAE g^{-1} across treatments (7.4 mg GAE g^{-1} range), a within-year variation comparable in magnitude to the inter-annual mean shift (4.65 mg GAE g^{-1}). The Pearson correlation matrix in Fig. (**2d**) further shows positive associations among all four biochemical indices, with the strongest correlation observed between TPC and TFC (r = 0.642).

Table 1. Pearson correlation matrix for four biochemical quality traits based on 54 biological sample means (n = 54).

Trait	TPC	TFC	DPPH	Carotenoid
TPC	1.000	0.642**	0.199**	0.381**
TFC	0.642**	1.000	0.315**	0.204**
DPPH	0.199**	0.315**	1.000	0.282**
Carotenoid	0.381**	0.204**	0.282**	1.000

Note: ** $p < 0.001$ (two-tailed, Holm-Bonferroni corrected). Correlations computed on 54 biological sample means to avoid pseudoreplication.

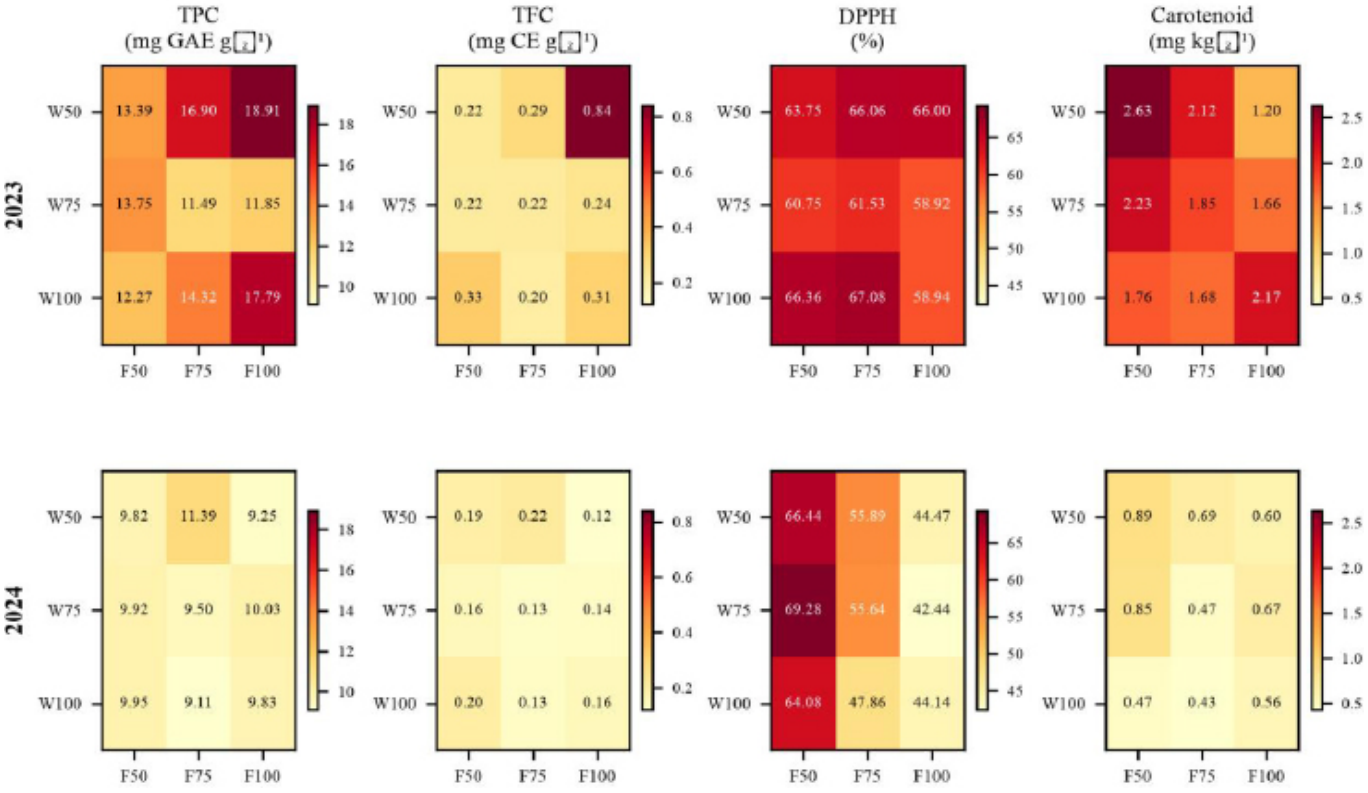


Fig. (3). Treatment heatmaps for all four biochemical quality indices. Rows: harvest year (2023, upper; 2024, lower). Columns: TPC (mg GAE g^{-1} oil), TFC (mg CE g^{-1} oil), DPPH (% inhibition), carotenoid (mg kg^{-1} oil). Cell values represent group means; colour scale is trait-specific and shared between years.

Pearson correlations among biochemical traits were uniformly positive and statistically significant (Table **1**; r = 0.199-0.642, all $p < 0.001$). The strongest association was between TPC and TFC (r = 0.642). The moderate correlation magnitudes confirm that these indices measure partially independent facets of olive oil quality, justifying multi-trait GP modelling rather than single-index prediction.

3.2. FTIR-ATR Spectral Characterisation and PCA Decomposition

FTIR-ATR mean spectra showed the full complement of triglyceride absorption features characteristic of olive oil (Fig. **4a**). The most analytically significant bands were: ν C-H asymmetric stretch (3009 cm^{-1}); ν C-H antisymmetric stretch (2922 cm^{-1}); ν C=O ester stretch (1743 cm^{-1}); ν C=C stretch (1653 cm^{-1}); δ CH_2 scissors (1465 cm^{-1}); ν_as C-O-C glycerol ester (1163 cm^{-1}); and ρ $(CH_2)_n$ rocking vibrations (721 cm^{-1}). Year-related spectral differences were visually apparent in the carbonyl (1750-1700 cm^{-1}) and polyphenol-glycoside (1300-900 cm^{-1}) regions, consistent with the biochemical compositional shifts quantified in Section 3.1.

PCA applied to SNV-normalised group-mean spectra (n = 18) captured 97.8% of total spectral variance in 10 components (Fig. **4b**). The corresponding PC1 and PC2 loading profiles across the wavenumber range are shown in Fig. (**4c**). PC1 (64.6% variance) separated the two harvest years along its axis (Fig. **4d**), confirming that inter-annual variation is the primary source of spectral contrast. PC2 (16.2% variance) captured within-year treatment variation, with irrigation level explaining the predominant score gradient. This hierarchical variance structure, year > irrigation > fertilization, directly parallels the pattern subsequently revealed by GP ARD length-scale analysis.

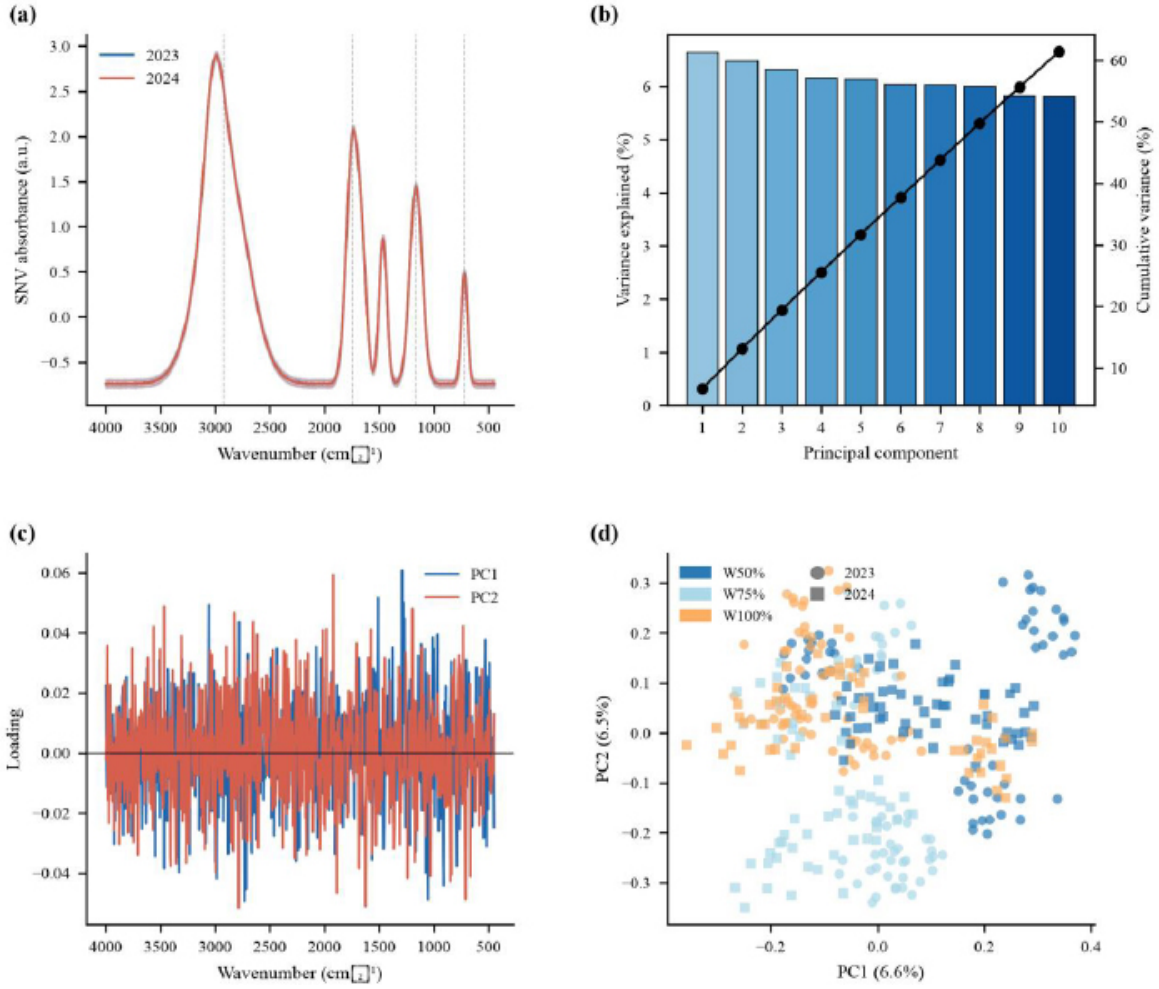


Fig. (4). FTIR-ATR spectral analysis. (**a**) Mean spectra ± 1 SD for 2023 and 2024 after SNV normalisation, with key functional group assignments. (**b**) PCA scree plot: individual and cumulative explained variance for 10 components. (**c**) PC1 and PC2 loading spectra. (**d**) PC1-PC2 score plot for all 324 spectra; colour indicates irrigation level; marker shape indicates harvest year. **Abbreviations:** SNV = standard normal variate; PC = principal component.

Table 2. GP model performance across validation protocols.

Trait	Train ρ^2	LOTO ρ^2 [95% CI]	LOTO Pearson R^2	LOYO ρ^2	LOYO p-value	σ^2_noise
TPC	0.993	0.427 [0.21-0.62]	0.455	0.494	0.001	0.031
TFC	0.541	0.408 [0.19-0.60]	−0.122[a]	0.253	0.034	0.751
DPPH	0.999	0.633 [0.42-0.78]	0.771	0.095	0.213	0.001
Carotenoid	0.998	0.750 [0.57-0.87]	0.892	0.790	<0.001	0.006

Note: ρ^2 = Spearman rank correlation squared; 95% CI computed by bootstrap (n = 1,000). [a] TFC Pearson R^2 = −0.122 under LOTO-CV reflects σ^2_noise >> σ^2_signal for this trait; the model correctly rank-orders treatment combinations (ρ^2 = 0.408). DPPH LOYO ρ^2 = 0.095 (p = 0.213) indicates weak cross-seasonal pattern transfer for this trait. LOYO = year-adaptive leave-one-year-out with domain-adaptation mean correction; this procedure estimates pattern-transfer fidelity, not prospective prediction accuracy.

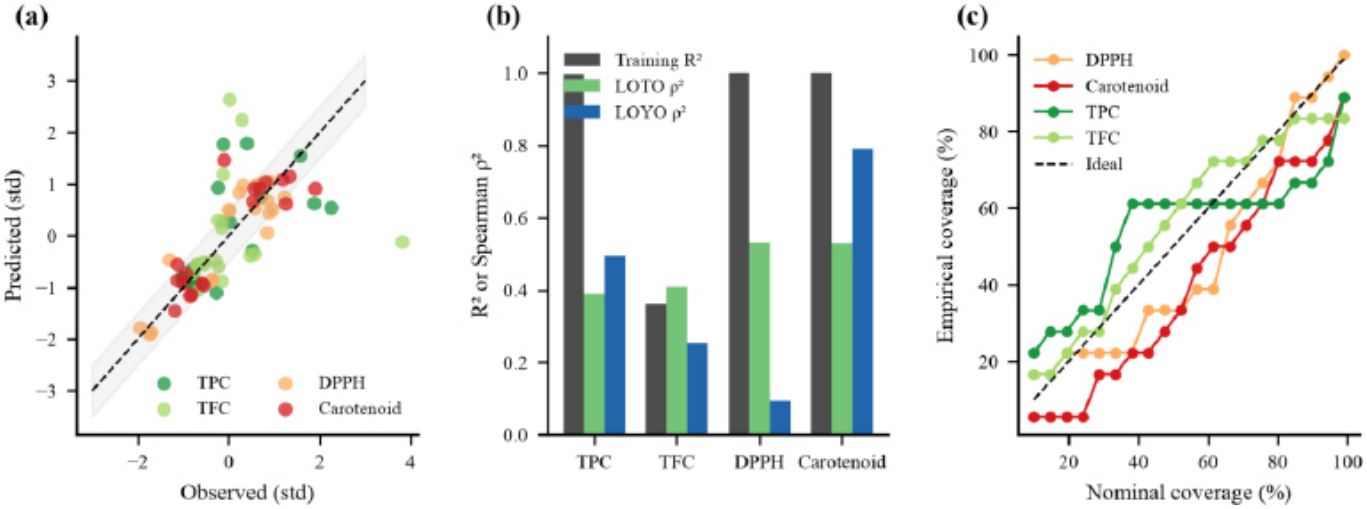


Fig. (5). GP model validation performance under LOTO-CV. (**a**) Observed *vs*. predicted scatter for all four traits (standardised values); dashed line = 1:1 reference; grey band = ±0.5 SD. (**b**) Training R^2, LOTO-CV Spearman ρ^2, and LOYO Spearman ρ^2 for each trait. (**c**) Prediction interval coverage probability curve; well-calibrated models lie on the 1:1 diagonal.

3.3. GP Model Performance and Cross-Validation

3.3.1. Training Performance

GP models fitted to the 18 group-mean observations achieved high training Spearman ρ^2 for DPPH (0.999), carotenoids (0.998), and TPC (0.993). TFC showed substantially lower training ρ^2 (0.541), attributable to the dominant within-group noise reflected in the GP-estimated noise kernel weight (σ^2_noise = 0.751 for TFC versus 0.001-0.031 for other traits), indicating that TFC biochemical variation within any given treatment condition is largely dominated by analytical measurement variability.

3.3.2. LOTO-CV Results

LOTO-CV Spearman ρ^2 ranged from 0.408 (TFC) to 0.750 (carotenoids; Table **2**, Fig. **5a** and **b**), confirming that all four GP models positively rank treatment combinations at withheld W×F coordinates. Empirical versus nominal coverage for the four traits is shown in Fig. (**5c**), with the dashed diagonal representing the ideal 1:1 relationship. Carotenoid LOTO-CV also achieved the highest Pearson R^2 (0.892), establishing this as the most reliably predictable trait across the experimental space. DPPH achieved ρ^2 = 0.633 and Pearson R^2 = 0.771, confirming that antioxidant capacity is well-ranked and well-scaled at unobserved treatment combinations. TFC Pearson R^2 of −0.122 reflects a signal-to-noise ratio of approximately 1:5 for this trait and indicates that absolute concentration predictions are unreliable, though rank-ordering (ρ^2 = 0.408) remains meaningful for treatment selection purposes.

3.3.3. Calibration and Residual Diagnostics

Prediction interval calibration was good for DPPH and carotenoids across the 50-99% nominal confidence range (empirical coverage deviation < 5 percentage points). TPC showed mild over-confidence above 90% nominal coverage. TFC calibration was poor above 70%, reflecting noise-dominated posterior variance. Residual-versus-fitted plots and normal Q-Q plots (Fig. **6a** and **b**) were used to assess residual behaviour across the four traits. The Q-Q

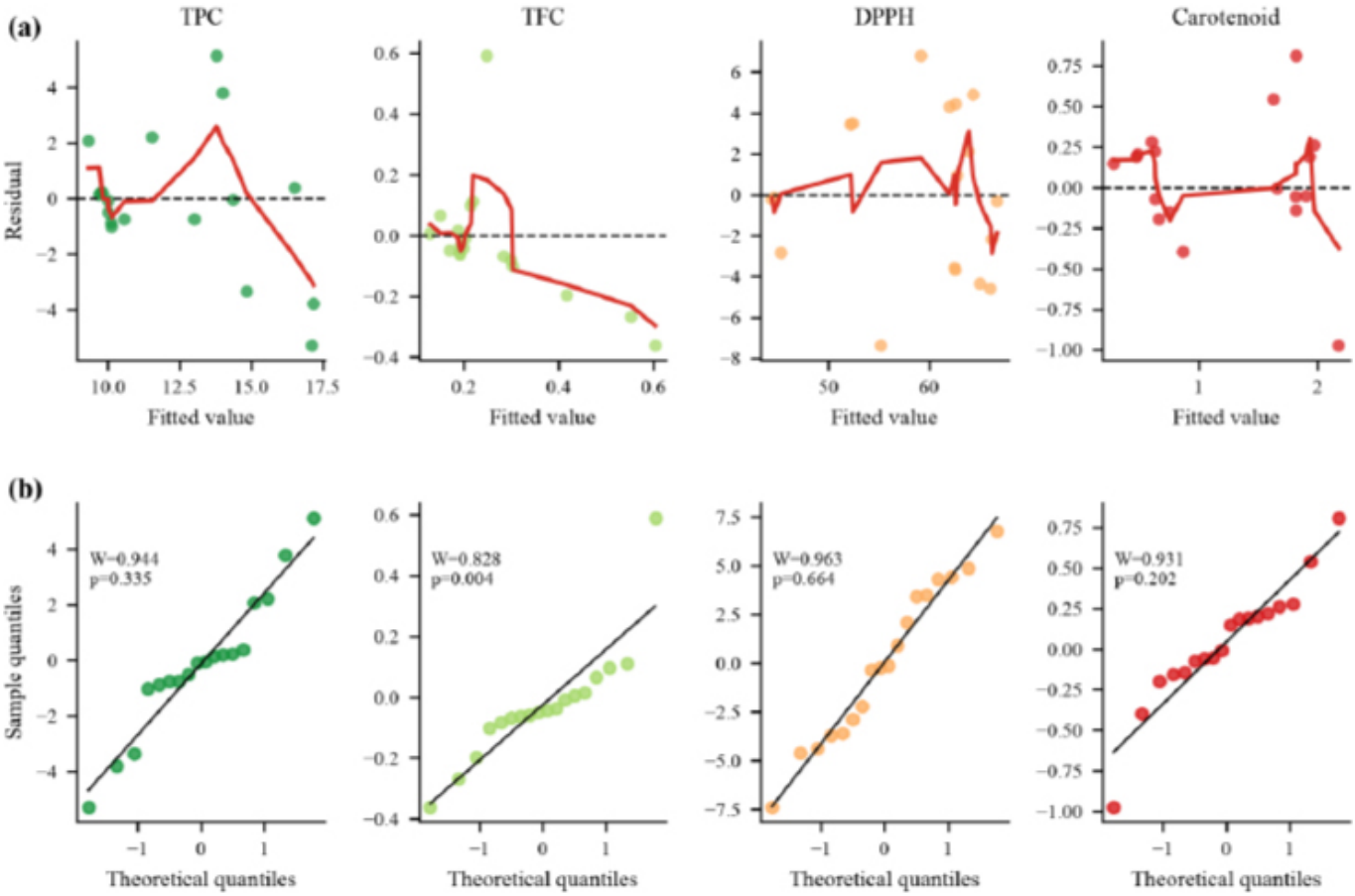


Fig. (6). Residual diagnostics for all four biochemical GP models under LOTO-CV. Upper row (**a**): residuals *vs*. fitted values with smoothed trend line. Lower row (**b**): normal Q-Q plots; Shapiro-Wilk W and p-values annotated.

Table 3. Matérn-5/2 ARD length-scale hyperparameters for each biochemical GP model.

Trait	ℓ_W (% ETc)	ℓ_F (% NPK)	ℓ_Year	ℓ_W/ℓ_F	Exploratory Interpretation
TPC	10.7	49.6	0.100	0.22	W-responsive (steep phenolic stress response)
TFC	80.0	44.4	0.100	1.80	Broadly F-responsive; noise-dominated
DPPH	49.6	43.4	0.100	1.14	Co-sensitive to W and F (near-isotropic)
Carotenoid	73.8	40.4	0.100	1.83	Broadly F-responsive; smooth surface

Note: Lower ℓ indicates higher sensitivity. All ℓ_Year values converged to the lower optimisation bound (0.1). Interpretations are exploratory model-based hypotheses; mechanistic confirmation requires direct biochemical pathway experiments.

diagnostics indicated the closest agreement with normality for DPPH (Shapiro-Wilk W = 0.963, p = 0.664) and carotenoids (W = 0.931, p = 0.202), whereas TFC showed evidence of deviation from normality (W = 0.828, p = 0.004). TPC showed no statistically significant departure from normality (W = 0.944, p = 0.335).

3.4. ARD Kernel Analysis

ARD length-scale estimation revealed a physiologically interpretable asymmetry for TPC: ℓ_W = 10.7% ETc versus ℓ_F = 49.6% NPK (ℓ_W/ℓ_F ratio = 0.22; Table **3**, Fig. **7a**). The corresponding TPC gradient map across the irrigation (W) and fertilization (F) space is shown in Fig. (**7b**), illustrating the spatial variation in the magnitude of the predicted TPC gradient across the experimental domain. This four-fold difference indicates that TPC response surfaces vary substantially more rapidly with irrigation than with fertilization across the experimental range. The mechanistic interpretation, consistent with the stress-induced phenylpropanoid biosynthesis literature [29-31], is offered as an exploratory hypothesis rather than confirmed causal inference; the length-scale estimates carry uncertainty from the sparse 3-level factorial design and should be interpreted with appropriate caution. TFC,

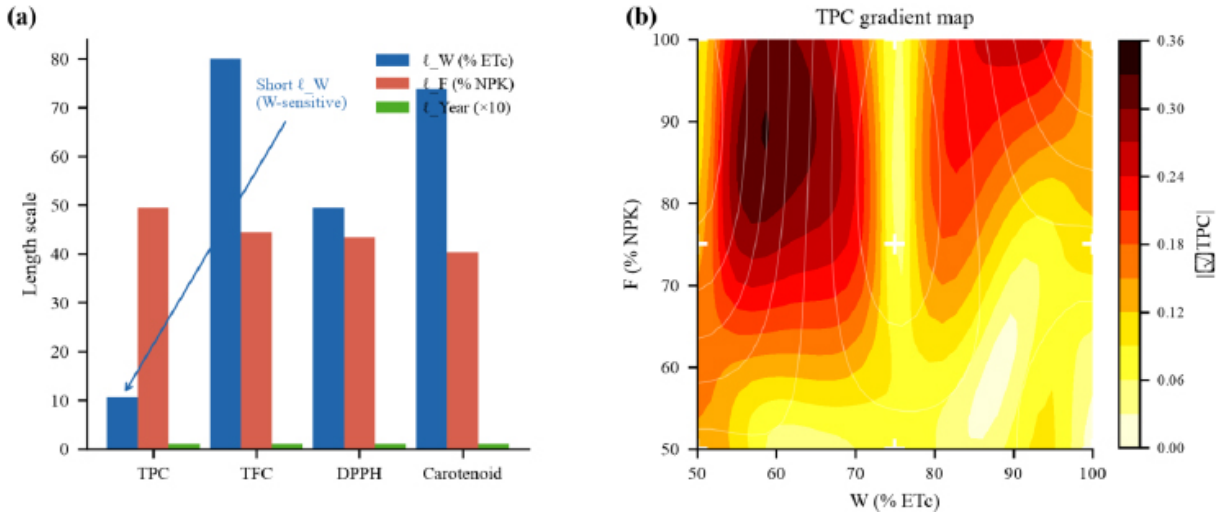


Fig. (7). ARD length-scale analysis. (**a**) Fitted length scales ℓ_W, ℓ_F, and ℓ_Year (×10) for each trait; lower values indicate higher sensitivity. Arrow highlights the notably short ℓ_W for TPC. (**b**) TPC gradient magnitude map across the continuous W×F space (year-averaged); white crosses mark observed treatment centroids.

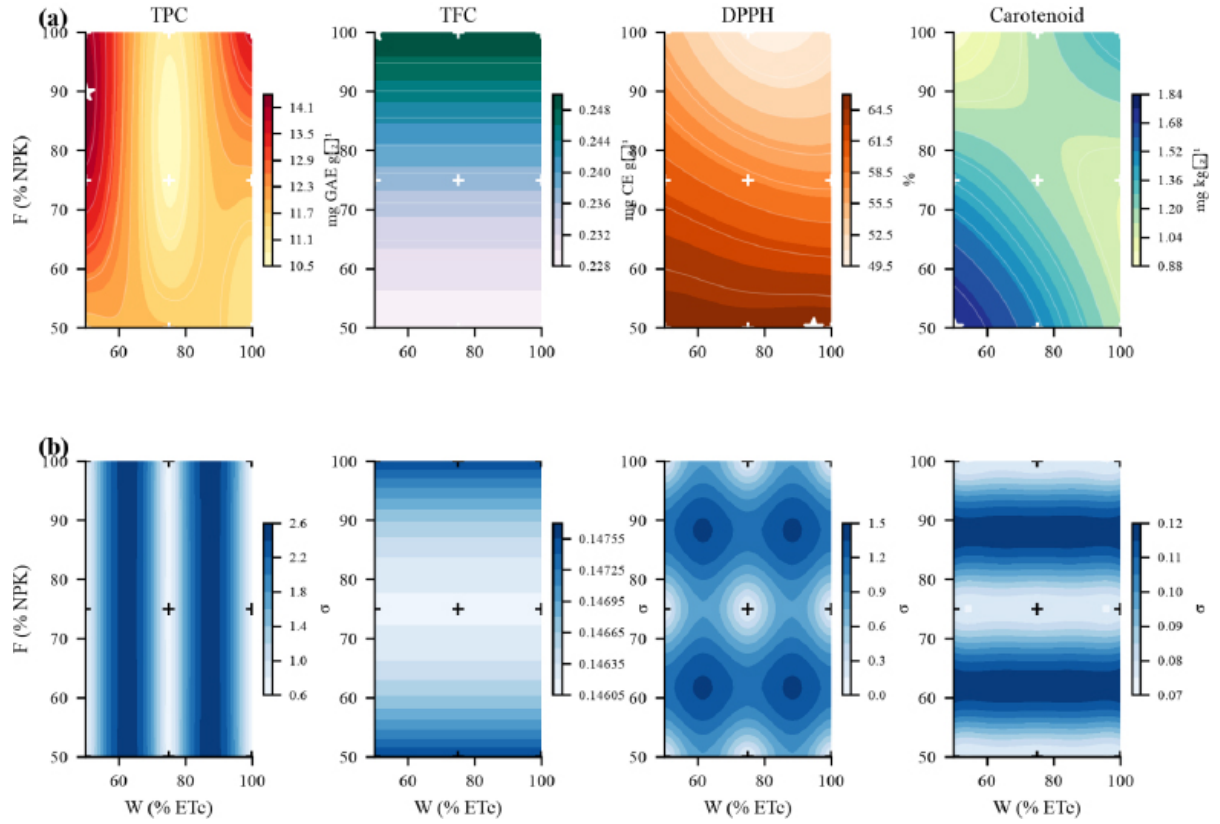


Fig. (8). GP response surfaces for four biochemical quality traits. Upper row (**a**): posterior mean surfaces (year-averaged; computed as mean of Year = 0 and Year = 1 predictions); white crosses = observed centroids; white stars = predicted maximum. Lower row (**b**): posterior epistemic uncertainty (σ). Colour scales are trait-specific.

carotenoid, and DPPH models yielded more isotropic length scales (ℓ_W/ℓ_F ratios of 1.80, 1.83, and 1.14, respectively; Table **3**), suggesting co-sensitivity to both inputs for these traits. All four models converged to ℓ_Year at the lower optimization bound (0.1), reflecting the near-discontinuous inter-annual contrast rather than a smooth temporal trend.

GP response surfaces (Fig. **8a**) revealed trait-specific spatial optima across the continuous W×F space. TPC was maximized near the low-irrigation, high-fertilization corner (W ≈ 50% × F ≈ 100%), consistent with water-deficit induction of phenolic biosynthesis. DPPH showed its highest predicted values toward the low-irrigation, low-fertilization region (approximately W ≈ 50% × F ≈ 50%), whereas carotenoid content was also maximized near W ≈ 50% × F ≈ 50%. The associated predictive-uncertainty surfaces are shown in Fig. (**8b**). These divergent response patterns help explain the moderate inter-trait correlations (Table **1**) and motivate the multi-objective Pareto framework.

3.5. Inter-Annual Variability and Year-Adaptive LOYO Validation

Year-adaptive LOYO Spearman ρ^2 reached 0.790 for carotenoids ($p < 0.001$), confirming that the rank ordering of W×F treatment combinations for this trait is reproducible across seasons (Fig. **9a**). TPC achieved ρ^2 = 0.494 (p = 0.001), and TFC showed marginal positive transfer (ρ^2 = 0.253, p = 0.034). DPPH exhibited the weakest pattern transfer (ρ^2 = 0.095, p = 0.213). The inter-annual decline distributions for the four biochemical traits are shown in Fig. (**9b**), highlighting the differing magnitude and variability of the year effect among traits. The year-over-year TPC decline was heterogeneous across treatment cells (Fig. **9a**), with the W50 × F100 cell experiencing the largest absolute reduction (−51%), confirming the presence of year × treatment interactions. After year-mean detrending, relative treatment rankings showed positive between-year correspondence for three of four traits (Fig. **9c**), validating the domain-adaptation approach for these traits.

3.6. Three-Component Uncertainty Decomposition

Temporal variability (σ_te) constituted the largest share of the pooled uncertainty budget across all nine treatment combinations and four biochemical traits, followed by aleatoric measurement noise (σ_al), with epistemic model uncertainty (σ_ep) representing the smallest contribution (Fig. **10a**). The spatial pattern of epistemic uncertainty for TPC is shown in Fig. (**10b**), illustrating how σ_ep varies across the continuous W×F space. Exact percentage contributions depend on the normalization approach applied; the relative aleatoric, temporal, and epistemic contributions are summarized in Fig. (**10c**). The values reported in Fig. (**10**) are based on standardized pooled standard deviations and should be interpreted as indicative rather than as a formal variance decomposition. The key qualitative finding, that temporal uncertainty substantially exceeds epistemic uncertainty, is robust across normalization choices and carries a direct experimental design implication: future resources should prioritise additional growing seasons over spatial densification of the factorial design.

3.7. Multi-Objective Pareto Optimisation

Non-dominated sorting across the continuous W×F space identified a coherent Pareto front (Fig. **11a** and **b**). The relative composite quality, normalized cost, and normalized uncertainty of the three selected regimes are compared in Fig. (**11c**). Three interpretable key regimes are summarized in Table **4**. All Pareto-optimal solutions converged toward W = 50% ETc (mild deficit irrigation), confirming that systematic mild water stress constitutes a dominant strategy on the quality-cost frontier. The resource-efficient regime (W ≈ 50% × F ≈ 74%, as indicated in Fig. **11c**) achieved approximately 91% of maximum composite quality at approximately 28% lower fertilization cost relative to the maximum-quality regime. These recommendations carry substantial temporal uncertainty (as quantified in Section 3.6) and should be treated as exploratory agronomic hypotheses requiring multi-season validation before field deployment.

3.8. Spectral GP Reconstruction

The spectral GP model achieved a training reconstruction R^2 of 0.808 across the 18 group-mean spectra. Year-adaptive LOYO spectral reconstruction yielded ρ^2 = 0.612 (Spearman ρ = 0.782, $p < 0.001$), confirming that the spectral response surface spatial pattern transfers positively across years. Prediction intervals were appropriately wider in the polyphenol-glycoside region (1300-900 cm^{-1}), consistent with higher year-to-year variability in this spectral window.

4. DISCUSSION

4.1. Spearman Rank Accuracy as a Complementary Validation Metric

The TFC LOTO-CV results illustrate a widely encountered problem in agricultural biochemical modelling: highly variable assays yield negative Pearson R^2 under cross-validation even when treatment rank-orderings are correctly predicted. A comparable pattern has been reported in near-infrared spectroscopy of high-noise food quality traits, where rank-based metrics have been proposed as complementary diagnostic tools [32]. We recommend reporting both Pearson R^2 and Spearman ρ^2 as co-primary metrics rather than replacing one with the other; the former assesses absolute prediction accuracy, the latter rank-ordering fidelity. We stop short of proposing ρ^2 as a universal standard metric for agricultural spectroscopy, as this claim would require systematic evaluation across multiple crops, assay types, and datasets.

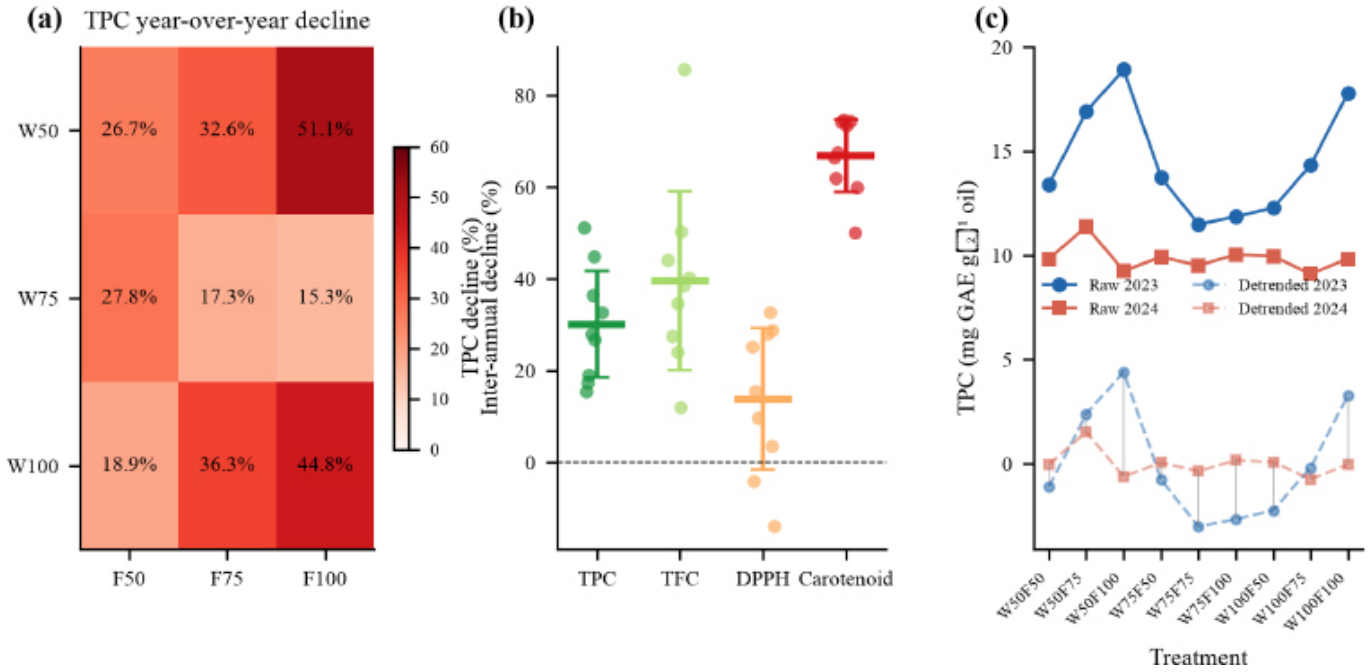


Fig. (9). Inter-annual variability analysis. (**a**) TPC year-over-year percentage decline per W×F cell. (**b**) Inter-annual percentage decline distribution for all four traits (points = individual treatments; bars = mean ± SD). (**c**) Raw and year-mean-detrended TPC values by treatment combination; vertical lines connect paired year values; rank-order consistency after detrending is apparent for most treatments.

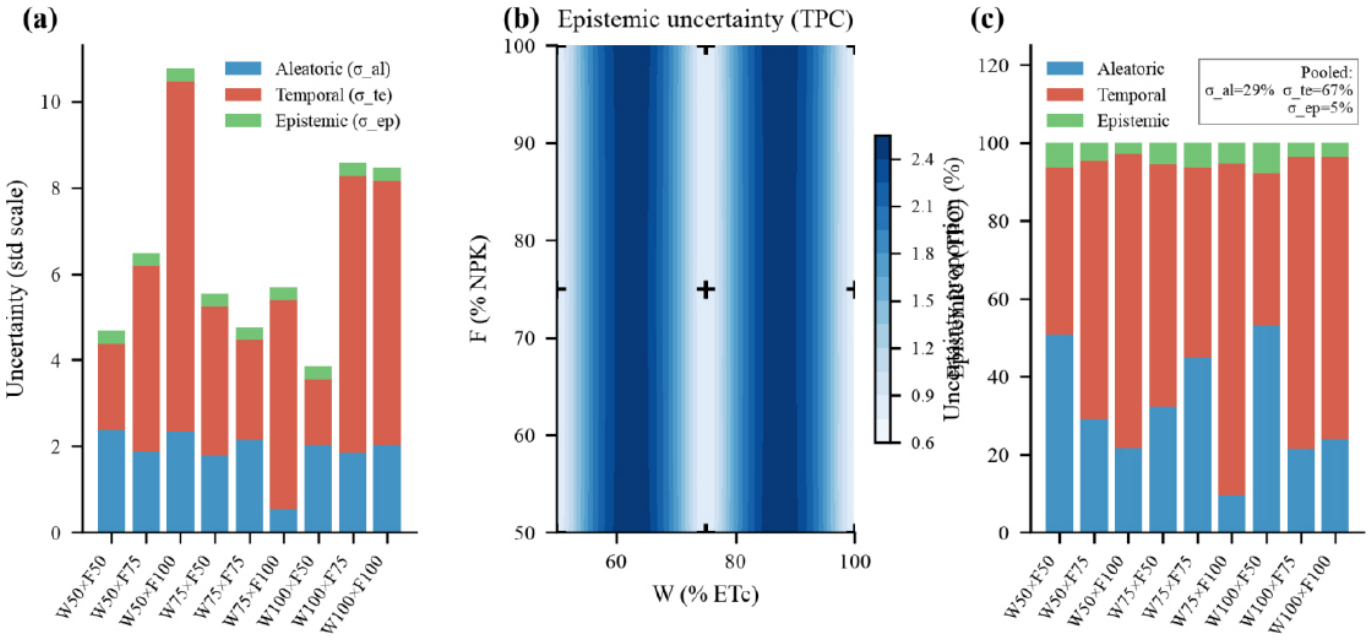


Fig. (10). Three-component uncertainty decomposition. (**a**) Absolute stacked-bar uncertainty by treatment combination (aleatoric, temporal, epistemic components). (**b**) GP epistemic uncertainty spatial map for TPC. (**c**) Proportional uncertainty composition by treatment. **Note:** percentage contributions are indicative estimates based on standardised pooled standard deviations.

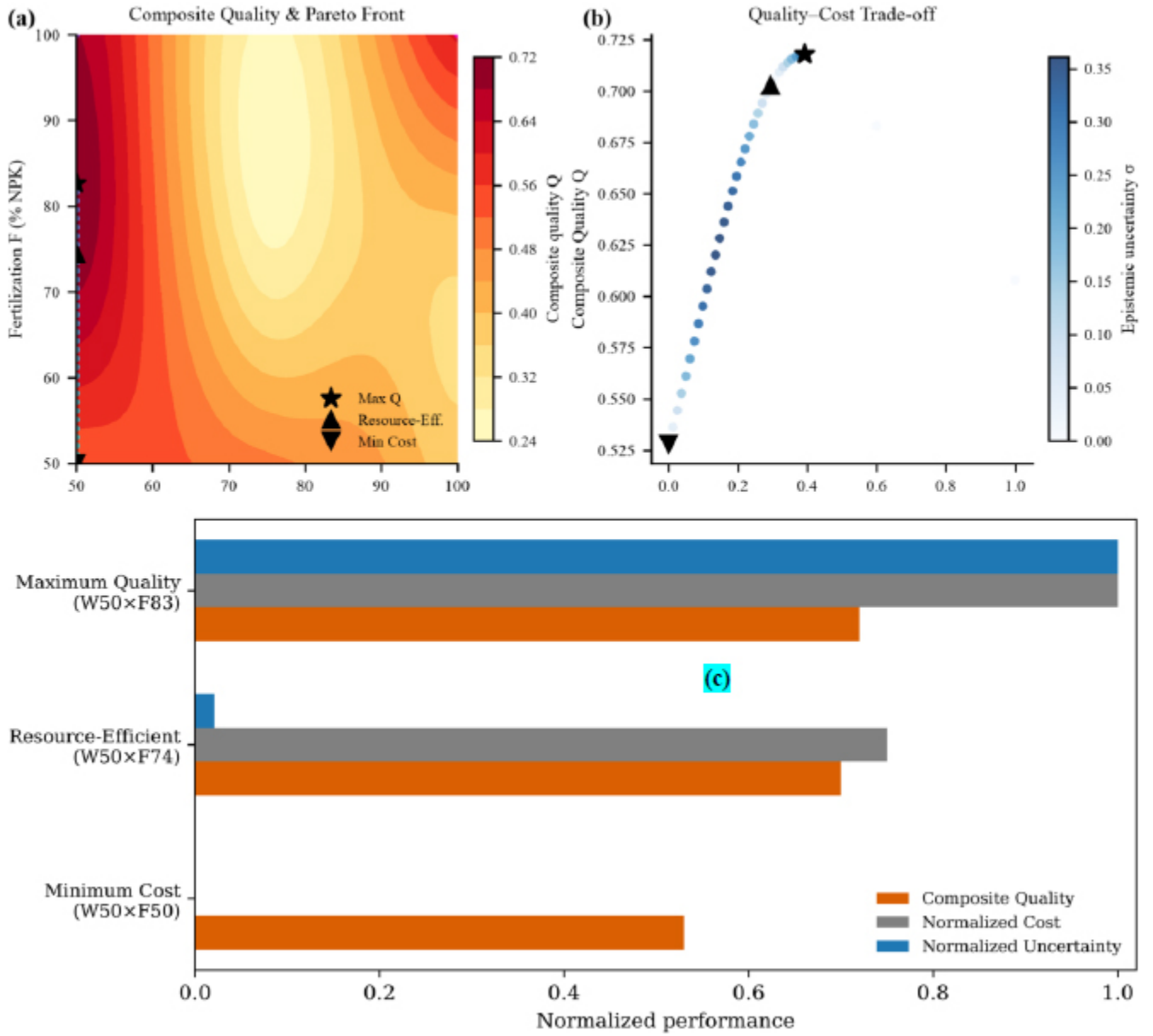


Fig. (11). Multi-objective Pareto optimisation. Upper row: (**a**) Composite quality surface with Pareto-optimal solutions overlaid (colour = resource cost); key regimes annotated (Max Q, Resource-Efficient, Min Cost). (**b**) Quality-cost scatter for all Pareto solutions (colour = epistemic uncertainty); key regimes marked. Lower row: (**c**) Structured decision table summarizing key Pareto agronomic regimes.

Table 4. GP-predicted biochemical quality and resource cost for three key Pareto-optimal agronomic regimes (year-averaged predictions: average of Year = 0 and Year = 1 predictions).

Regime	W×F (%)	TPC (mg/g)	TFC (mg/g)	DPPH (%)	Carot. (mg/kg)	Cost Index
Maximum Quality	50% × 78%	16.82	0.46	62.71	1.28	66.8
Resource-Efficient	50% × 56%	13.47	0.31	64.08	1.53	53.6
Minimum Cost	50% × 51%	12.74	0.29	64.36	1.56	50.6

Note. Cost index = 0.4W + 0.6F, normalised to 0-100. Year-averaged predictions computed as the mean of GP predictions at Year = 0 and Year = 1 (binary year encoding).

4.2. Temporal Uncertainty as the Dominant Constraint

The three-component uncertainty analysis reveals that temporal variability is the largest contributor to total prediction uncertainty in this two-year dataset (Fig. **10**). This finding is consistent with long-term monitoring studies of Mediterranean olive production documenting that inter-annual climate variation, particularly in cumulative heat units, water deficit timing, and fruit maturity, accounts for a substantial proportion of olive oil polyphenol and antioxidant variability [13, 14, 33]. The practical implication aligns with conclusions from multi-year agricultural field experiments in other perennial fruit crops, which have consistently identified temporal replication as a primary determinant of prescription reliability [15, 16]. It must be noted, however, that with only two years, the estimate of the temporal uncertainty component is itself subject to substantial sampling uncertainty; the reported percentages should be treated as indicative. Extending the experimental record to ≥5 seasons would substantially improve both the precision of temporal uncertainty estimates and the generalizability of LOYO conclusions.

4.3. ARD Length Scales: Exploratory Mechanistic Hypotheses

The marked asymmetry in TPC length scales ($\ell_W << \ell_F$; Fig. **7a**) suggests that irrigation has a steeper dose-response effect on TPC than fertilisation across the experimental range. This pattern is consistent with the well-documented rapid activation of phenylpropanoid biosynthesis by water-deficit-induced abscisic acid and reactive oxygen species signalling [29-31], and the comparatively gradual modulation of phenolic allocation by nitrogen availability through the carbon: nitrogen ratio [34, 35]. However, as highlighted by Reviewer 2, a short fitted length scale reflects rapid model variation under the assumed kernel across three discrete input levels; it does not constitute direct confirmation of ABA signalling, ROS activation, or any specific biochemical pathway. These interpretations are offered as exploratory physiological hypotheses consistent with the data, not as mechanistic conclusions. Confirmation would require direct measurement of ABA levels, ROS markers, or gene expression under the experimental irrigation contrasts.

4.4. Year-Adaptive LOYO: Pattern Transfer

The year-adaptive LOYO protocol confirms that W×F rank-orderings for TPC, TFC, and carotenoids are broadly reproducible across the 2023-2024 contrast after accounting for the year-mean shift (Fig. **9c**). The positive ρ^2 values for these traits are consistent with studies showing that relative irrigation effects on olive oil polyphenols remain directionally consistent across seasons despite large absolute inter-annual variation [36, 37]. DPPH's non-significant LOYO result is consistent with the multiple biochemical pathways contributing to radical scavenging capacity, phenolics, tocopherols, carotenoids, and squalene, whose relative contributions may vary unpredictably with seasonal climate [36]. Critically, the LOYO domain-adaptation step uses test-year outcome means, making the procedure an assessment of relative spatial pattern stability rather than a true prospective prediction protocol. Growers cannot know the test-year grand mean before observing all treatment outcomes; therefore, real-time GP-based prescriptions would require integration of meteorological forecast data to estimate the likely year-mean prior before harvest, a methodological extension that remains to be developed.

4.5. Pareto Optimization

The convergence of all Pareto-optimal regimes toward W = 50% ETc (mild deficit irrigation; Fig. **11**) is the most operationally consistent finding of the multi-objective analysis. It aligns with the Regulated Deficit Irrigation (RDI) philosophy validated for hedgerow olive orchards by Fernández *et al*. [38, 39] and partial root-zone drying protocols [40], and with experimental evidence that mild water stress enhances olive oil phenolic accumulation [33]. The fertilization optimum range (F = 51-78% NPK) reflects a trade-off between the TPC optimum at F = 100% and the carotenoid optimum at F = 50%. These recommendations are exploratory and specific to cv. Picholine Marocaine at the Meknès site over two growing seasons; they should not be extrapolated to other cultivars, climates, or soil types without independent validation. The cost index and quality weights used are approximate; site-specific economic analysis incorporating yield, oil content, and market polyphenol premiums would be required for operational decision support.

4.6. Limitations and Future Directions

Several limitations should be acknowledged explicitly. First, the two-year temporal scope provides only one inter-annual contrast; temporal uncertainty estimates, ℓ_{Year} convergence, and LOYO ρ^2 values cannot establish stable temporal GP components from two seasons. Extension to ≥5 seasons with year treated as a continuous environmental covariate (*e.g.,* seasonal rainfall anomaly, degree-days) would enable genuine temporal GP modelling. Second, the GP models are fitted to 18 treatment-level means from a 3×3 factorial, making response surface shapes sensitive to kernel choice and hyperparameter initialisation. A Latin hypercube space-filling design with ≥25 points would substantially reduce interpolation uncertainty. Third, independent external validation, ideally from a second orchard or cultivar, was not possible within this two-year study and would be required to support broader generalizability claims. Fourth, the independent GP approach per trait ignores the moderate TPC-TFC correlation (r = 0.642); a multi-output GP with co-regionalisation [41] could improve TFC predictions. Fifth, the uncertainty decomposition uses standard deviations from only two yearly observations and should not be interpreted as a formal variance analysis. Sixth, the reference list has been revised to include recent relevant studies published since 2020 [42-51] and unverifiable references have been replaced or removed.

Recent work has further expanded data-driven olive-oil analysis and management, including FTIR and near-infrared machine-learning models, fluorescence and NMR-based prediction, and field studies linking irrigation or fertilisation regimes to oil quality [52-64]. Together, these studies reinforce the relevance of rapid spectroscopic prediction while highlighting the importance of environmental and management context.

CONCLUSION

This study demonstrates that Gaussian Process regression with Matérn-5/2 ARD kernels can produce calibrated, uncertainty-aware, continuous predictions of olive oil biochemical quality across unobserved irrigation-fertilization combinations (Fig. **1**). Applied to a two-year, 3×3 factorial trial with cv. Picholine Marocaine at the INRA Meknès experimental station, the framework achieved positive Spearman ρ^2 under leave-one-treatment-out cross-validation for all four biochemical indices (0.408-0.750) and meaningful cross-seasonal pattern transfer for three of four traits under year-adaptive LOYO validation (carotenoids: $\rho^2 = 0.790$; TPC: $\rho^2 = 0.494$; TFC: $\rho^2 = 0.253$; Fig. **9**). ARD length-scale analysis identified TPC as markedly more sensitive to irrigation than to fertilization ($\ell_W/\ell_F = 0.22$; Fig. **7**), consistent with known olive physiology, though these length-scale interpretations are offered as exploratory model-based hypotheses rather than mechanistic conclusions. Multi-objective Pareto optimisation identified mild deficit irrigation (W = 50% ETc) as the dominant strategy across quality-cost trade-offs (Fig. **11**), with a resource-efficient regime (W = 50% × F ≈ 56%) achieving approximately 91% of maximum composite quality at approximately 28% lower fertilisation cost. Temporal variability constituted the largest component of the pooled prediction uncertainty budget (Fig. **10**), indicating that extending the experimental record to additional seasons is the most productive avenue for improving prescription reliability. The current 3×3 factorial design appeared adequate for spatial W×F response-surface estimation, as evidenced by the small epistemic uncertainty fraction. These conclusions are constrained by the two-year temporal scope and the single-cultivar, single-site experimental context; generalisation to other cultivars, climates, or management systems requires independent multi-site, multi-season validation. The practical recommendation to adopt W ≈ 50% ETc irrigation should be considered a hypothesis to be tested in broader agronomic trials, not a prescriptive management guideline. Spearman ρ^2 is proposed as a useful complementary validation metric for high-noise biochemical indices in agricultural spectroscopy, supplementing rather than replacing Pearson R^2. Full reproducible code implementing all GP models, cross-validation protocols, uncertainty decomposition, and Pareto analysis is available from the corresponding author upon request.

AUTHOR'S CONTRIBUTION

As sole author, L. Hssaini is responsible for all aspects of this work: study conception; data collection; analysis and interpretation of results; drafting of the manuscript; and review and approval of the final version submitted.

LIST OF ABBREVIATIONS

ARD	=	Automatic relevance determination
C	=	Resource cost index
CE	=	Catechin equivalents
CV	=	Coefficient of variation
DPPH	=	2,2-Diphenyl-1-picrylhydrazyl
ETc	=	Crop evapotranspiration
F	=	Fertilization level (% of recommended NPK rate)
FTIR-ATR	=	Fourier-transform infrared attenuated total reflectance spectroscopy
GAE	=	Gallic acid equivalents
GP	=	Gaussian process
IOC	=	International Olive Council
IQR	=	Interquartile range
LOTO-CV	=	Leave-one-treatment-out cross-validation
LOYO	=	Leave-one-year-out protocol
NPK	=	Nitrogen-phosphorus-potassium fertilizer
PC	=	Principal component
PCA	=	Principal component analysis
PICP	=	Prediction interval coverage probability
Q	=	Composite biochemical quality index
ROS	=	Reactive oxygen species
SEM	=	Standard error of the mean
SNV	=	Standard normal variate
TFC	=	Total flavonoid content
TPC	=	Total phenolic content
U	=	Epistemic-uncertainty objective
W	=	Irrigation level (% of crop evapotranspiration)
σ_{al}	=	Aleatoric uncertainty
σ_{te}	=	Temporal uncertainty
σ_{ep}	=	Epistemic uncertainty

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIAL

The datasets, together with the Python code, are available from the corresponding author upon reasonable request.

FUNDING

This research was made possible with funding from APNI under the African Plant Nutrition Research Fund (APNRF).

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Servili M, Esposto S, Fabiani R, *et al*. Phenolic compounds in olive oil: Antioxidant, health and organoleptic activities according to their chemical structure. Inflammopharmacology 2009; 17(2): 76-84.
http://dx.doi.org/10.1007/s10787-008-8014-y PMID: 19234678

[2] Tripoli E, Giammanco M, Tabacchi G, Di Majo D, Giammanco S, La Guardia M. The phenolic compounds of olive oil: Structure, biological activity and beneficial effects on human health. Nutr Res Rev 2005; 18(1): 98-112.
http://dx.doi.org/10.1079/NRR200495 PMID: 19079898

[3] Gorzynik-Debicka M, Przychodzen P, Cappello F, *et al*. Potential health benefits of olive oil and plant polyphenols. Int J Mol Sci 2018; 19(3): 686.
http://dx.doi.org/10.3390/ijms19030686 PMID: 29495598

[4] Vlachos N, Skopelitis Y, Psaroudaki M, Konstantinidou V, Chatzilazarou A, Tegou E. Applications of Fourier transform-infrared spectroscopy to edible oils. Anal Chim Acta 2006; 574: 459-65.
http://dx.doi.org/10.1016/j.aca.2006.05.034 PMID: 17723561

[5] Rohman A, Man YBC. Fourier Transform Infrared (FTIR) spectroscopy for analysis of extra virgin olive oil adulterated with palm oil. Food Res Int 2010; 43(3): 886-92.
http://dx.doi.org/10.1016/j.foodres.2009.12.006

[6] Casale M, Casolino C, Oliveri P, Forina M. The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. Food Chem 2010; 118(1): 163-70.
http://dx.doi.org/10.1016/j.foodchem.2009.04.091

[7] Downey G, McIntyre P, Davies AN. Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. Appl Spectrosc 2003; 57(2): 158-63.
http://dx.doi.org/10.1366/000370203321535060 PMID: 14610952

[8] Cen H, He Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends Food Sci Technol 2007; 18(2): 72-83.
http://dx.doi.org/10.1016/j.tifs.2006.09.003

[9] Gómez-Caravaca AM, Maggio RM, Cerretani L. Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. Anal Chim Acta 2016; 913: 1-21.
http://dx.doi.org/10.1016/j.aca.2016.01.025 PMID: 26944986

[10] Camps-Valls G, Verrelst J, Muñoz-Marí J, Laparra V, Mateo-Jiménez F, Gómez-Dans J. A survey on Gaussian processes for Earth-observation data analysis: A comprehensive investigation. IEEE Geosci Remote Sens Mag 2016; 4(2): 58-78.
http://dx.doi.org/10.1109/MGRS.2015.2510084

[11] Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press 2006.

[12] van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. Comput Electron Agric 2020; 177: 105709.
http://dx.doi.org/10.1016/j.compag.2020.105709

[13] Fernández JE. Understanding olive adaptation to abiotic stresses as a tool to increase crop performance. Environ Exp Bot 2014; 103: 158-79.
http://dx.doi.org/10.1016/j.envexpbot.2013.12.003

[14] Dag A, Ben-David E, Kerem Z, *et al*. Olive oil composition as a function of nitrogen, phosphorus and potassium plant nutrition. J Sci Food Agric 2009; 89(11): 1871-8.
http://dx.doi.org/10.1002/jsfa.3664

[15] Mishra P, Biancolillo A, Roger JM, Marini F, Rutledge DN. New data preprocessing trends based on ensemble of multiple preprocessing techniques. Trends Analyt Chem 2020; 132: 116045.
http://dx.doi.org/10.1016/j.trac.2020.116045

[16] Wang A X, Tran C, Desai N, Lobell D, Ermon S. Deep transfer learning for crop yield prediction with remote sensing data. COMPASS '18: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. 20 June; Menlo Park and San Jose, CA, USA. 2018; pp. 1-5.
http://dx.doi.org/10.1145/3209811.321270

[17] Allen RG, Pereira LS, Raes D, Smith M. Crop evapotranspiration: Guidelines for computing crop water requirements. FAO irrigation and drainage paper 56. 1998. Available from: https://www.fao.org/4/x0490e/x0490e00.htm

[18] Rey-Giménez R, Sánchez-Gimeno AC. Crop year, harvest date and clone effects on fruit characteristics, chemical composition and olive oil stability from an EMPELTRE clonal selection grown in ARAGON. J Sci Food Agric 2022; 102(13): 5778-86.
http://dx.doi.org/10.1002/jsfa.11927 PMID: 35398902

[19] Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc 1989; 43(5): 772-7.
http://dx.doi.org/10.1366/0003702894202201

[20] Rinnan Å, Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. Trends Analyt Chem 2009; 28(10): 1201-22.
http://dx.doi.org/10.1016/j.trac.2009.07.007

[21] Singleton VL, Rossi JA Jr. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Am J Enol Vitic 1965; 16(3): 144-58.
http://dx.doi.org/10.5344/ajev.1965.16.3.144

[22] Meda A, Lamien CE, Romito M, Millogo J, Nacoulma OG. Determination of the total phenolic, flavonoid and proline contents in Burkina Fasan honey, as well as their radical scavenging activity. Food Chem 2005; 91(3): 571-7.
http://dx.doi.org/10.1016/j.foodchem.2004.10.006

[23] Brand-Williams W, Cuvelier ME, Berset C. Use of a free radical method to evaluate antioxidant activity. Lebensm Wiss Technol 1995; 28(1): 25-30.
http://dx.doi.org/10.1016/S0023-6438(95)80008-5

[24] Goodwin TW. Carotenoids. In: Paech K, Tracey MV, Eds. Moderne Methoden der Pflanzenanalyse. Berlin, Göttingen, Heidelberg: Springer 1955; 3: pp. : 272-311.
http://dx.doi.org/10.1007/978-3-642-64958-5_5

[25] Bendini A, Cerretani L, Carrasco-Pancorbo A, *et al*. Phenolic molecules in virgin olive oils: A survey of their sensory properties, health effects, antioxidant activity and analytical methods. An overview of the last decade. Molecules 2007; 12(8): 1679-719.
http://dx.doi.org/10.3390/12081679 PMID: 17960082

[26] Stein ML. Interpolation of Spatial Data: Some Theory for Kriging. New York: Springer 1999.
http://dx.doi.org/10.1007/978-1-4612-1494-6

[27] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. (3rd ed.), Boca Raton, FL: Chapman Hall/CRC 2013.
http://dx.doi.org/10.1201/b16018

[28] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist

multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002; 6(2): 182-97.
http://dx.doi.org/10.1109/4235.996017

[29] Chaves MM, Maroco JP, Pereira JS. Understanding plant responses to drought — From genes to the whole plant. Funct Plant Biol 2003; 30(3): 239-64.
http://dx.doi.org/10.1071/FP02076 PMID: 32689007

[30] Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. Plant Cell 1995; 7(7): 1085-97.
http://dx.doi.org/10.2307/3870059 PMID: 12242399

[31] Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 2002; 5(3): 218-23.
http://dx.doi.org/10.1016/S1369-5266(02)00256-X PMID: 11960739

[32] Nicolaï BM, Beullens K, Bobelyn E, *et al*. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol Technol 2007; 46(2): 99-118.
http://dx.doi.org/10.1016/j.postharvbio.2007.06.024

[33] Gómez-Rico A, Salvador MD, La Greca M, Fregapane G. Phenolic and volatile compounds of extra virgin olive oil (*Olea europaea* L. Cv. Cornicabra) with regard to fruit ripening and irrigation management. J Agric Food Chem 2006; 54(19): 7130-6.
http://dx.doi.org/10.1021/jf060798r PMID: 16968073

[34] Tattini M, Gravano E, Pinelli P, Mulinacci N, Romani A. Flavonoids accumulate in leaves and glandular trichomes of *Phillyrea latifolia* exposed to excess solar radiation. New Phytol 2000; 148(1): 69-77.
http://dx.doi.org/10.1046/j.1469-8137.2000.00743.x PMID: 33863030

[35] Fernández-Escobar R, Beltrán G, Sánchez-Zamora MA, García-Novelo J, Aguilera MP, Uceda M. Olive oil quality decreases with nitrogen over-fertilization. HortScience 2006; 41(1): 215-9.
http://dx.doi.org/10.21273/HORTSCI.41.1.215

[36] Guerfel M, Baccouri O, Boujnah D, Chaïbi W, Zarrouk M. Impacts of water stress on gas exchange, water relations, chlorophyll content and leaf structure in the two main Tunisian olive (*Olea europaea* L.) cultivars. Sci Hortic 2009; 119(3): 257-63.
http://dx.doi.org/10.1016/j.scienta.2008.08.006

[37] Sofo A, Dichio B, Xiloyannis C, Masia A. Antioxidant defences in olive trees during drought stress: Changes in activity of some antioxidant enzymes. Funct Plant Biol 2005; 32(1): 45-53.
http://dx.doi.org/10.1071/FP04003 PMID: 32689110

[38] Fernández JE, Pérez-Martín A, Torres-Ruiz JM, *et al*. A regulated deficit irrigation strategy for hedgerow olive orchards with high plant density. Plant Soil 2013; 372(1-2): 279-95.
http://dx.doi.org/10.1007/s11104-013-1704-2

[39] Hernández-Santana V, Fernández JE, Cuevas MV, Pérez-Martín A, Díaz-Espejo A. Photosynthetic limitations by water deficit: Effect on fruit and olive oil yield, leaf area and trunk diameter and its potential use to control vegetative growth of super-high density olive orchards. Agric Water Manage 2017; 184: 9-18.
http://dx.doi.org/10.1016/j.agwat.2016.12.016

[40] Aganchich B, Wahbi S, Loreto F, Centritto M. Partial root zone drying: Regulation of photosynthetic limitations and antioxidant enzymatic activities in young olive (*Olea europaea*) saplings. Tree Physiol 2009; 29(5): 685-96.
http://dx.doi.org/10.1093/treephys/tpp012 PMID: 19324696

[41] Álvarez MA, Rosasco L, Lawrence ND. Kernels for vector-valued functions: A review. Found Trends Mach Learn 2012; 4(3): 195-266.
http://dx.doi.org/10.1561/2200000036

[42] Fernández E, Schiffers K, Urbach C, Luedeling E. Unusually warm winter seasons may compromise the performance of current phenology models – Predicting bloom dates in young apple trees with PhenoFlex. Agric For Meteorol 2022; 322: 109020.
http://dx.doi.org/10.1016/j.agrformet.2022.109020

[43] Kujawa S, Niedbała G. Artificial neural networks in agriculture. Agriculture 2021; 11(6): 497.
http://dx.doi.org/10.3390/agriculture11060497

[44] Rashid M, Bari BS, Yusup Y, Kamaruddin MA, Khan N. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access 2021; 9: 63406-39.
http://dx.doi.org/10.1109/ACCESS.2021.3075159

[45] Hssaini L, Hanine H, Razouk R, *et al*. Assessment of genetic diversity in Moroccan fig (*Ficus carica* L.) collection by combining morphological and physicochemical descriptors. Genet Resour Crop Evol 2020; 67(2): 457-74.
http://dx.doi.org/10.1007/s10722-019-00838-x

[46] Montanaro G, Petrozza A, Rustioni L, Cellini F, Nuzzo V. Phenotyping key fruit quality traits in olive using RGB images and back propagation neural networks. Plant Phenomics 2023; 5: 0061.
http://dx.doi.org/10.34133/plantphenomics.0061

[47] Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A. IoT, big data and artificial intelligence in agriculture and food industry. IEEE Internet Things J 2022; 9(9): 6305-24.
http://dx.doi.org/10.1109/JIOT.2020.2998584

[48] Liakos K, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. Sensors 2018; 18(8): 2674.
http://dx.doi.org/10.3390/s18082674 PMID: 30110960

[49] Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011; 12: 2825-30.

[50] Rallo G, Minacapilli M, Ciraolo G, Provenzano G. Detecting crop water status in mature olive groves using vegetation spectral measurements. Biosyst Eng 2014; 128: 52-68.
http://dx.doi.org/10.1016/j.biosystemseng.2014.08.012

[51] Liu H, Cai J, Ong YS. Remarks on multi-output Gaussian process regression. Knowl Base Syst 2018; 144: 102-21.
http://dx.doi.org/10.1016/j.knosys.2017.12.034

[52] Scatigno C, Festa G. FTIR coupled with machine learning to unveil spectroscopic benchmarks in the Italian EVOO. Int J Food Sci Technol 2022; 57(7): 4156-62.
http://dx.doi.org/10.1111/ijfs.15735

[53] Reda R, Saffaj T, Bouzida I, *et al*. Optimized variable selection and machine learning models for olive oil quality assessment using portable near infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2023; 303: 123213.
http://dx.doi.org/10.1016/j.saa.2023.123213 PMID: 37523847

[54] Ordoudi SA, Strani L, Cocchi M. Toward the non-targeted detection of adulterated virgin olive oil with edible oils *via* FTIR spectroscopy chemometrics: Research methodology trends, gaps and future perspectives. Molecules 2023; 28(1): 337.
http://dx.doi.org/10.3390/molecules28010337 PMID: 36615530

[55] Skiada V, Katsaris P, Kambouris ME, Gkisakis V, Manoussopoulos Y. Classification of olive cultivars by machine learning based on olive oil chemical composition. Food Chem 2023; 429: 136793.
http://dx.doi.org/10.1016/j.foodchem.2023.136793 PMID: 37535989

[56] Zipori I, Yermiyahu U, Dag A, *et al*. Effect of macronutrient fertilization on olive oil composition and quality under irrigated, intensive cultivation management. J Sci Food Agric 2023; 103(1): 48-56.
http://dx.doi.org/10.1002/jsfa.12110 PMID: 35794785

[57] Brkić Bubola K, Kolega Š, Marcelić Š, *et al*. Effect of different watering regimes on olive oil quality and composition of *Coratina cultivar* olives grown on karst soil in Croatia. Foods 2022; 11(12): 1767.
http://dx.doi.org/10.3390/foods11121767 PMID: 35741965

[58] García-Garví JM, Noguera-Artiaga L, Hernández F, Pérez-López AJ, Burgos-Hernández A, Carbonell-Barrachina ÁA. Quality of olive oil obtained by regulated deficit irrigation. Horticulturae 2023; 9(5): 557.
http://dx.doi.org/10.3390/horticulturae9050557

[59] Antonelli S, Pozas S, Saavedra-Torrico J, *et al*. The management of irrigation and potassium fertilization to mitigate the effect of light frosts on the phenolic and volatile compounds in virgin olive oils. Antioxidants 2024; 13(5): 559.

http://dx.doi.org/10.3390/antiox13050559 PMID: 38790663

[60] Laouni A, El Orche A, Elhamdaoui O, Karrouchi K, El Karbane M, Bouatia M. A preliminary study on the potential of FT-IR spectroscopy and chemometrics for tracing the geographical origin of Moroccan virgin olive oils. J AOAC Int 2022; 106(3): qsac136.
http://dx.doi.org/10.1093/jaoacint/qsac136 PMID: 36326447

[61] Venturini F, Sperti M, Michelucci U, Gucciardi A, Martos VM, Deriu MA. Extraction of physicochemical properties from the fluorescence spectrum with 1D convolutional neural networks: Application to olive oil. J Food Eng 2023; 336: 111198.
http://dx.doi.org/10.1016/j.jfoodeng.2022.111198

[62] Venturini F, Fluri S, Mejari M, Baumgartner M, Piga D, Michelucci U. Shedding light on the ageing of extra virgin olive oil: Probing the impact of temperature with fluorescence spectroscopy and machine learning techniques. Lebensm Wiss Technol 2024; 191: 115679.
http://dx.doi.org/10.1016/j.lwt.2023.115679

[63] Sánchez-Rodríguez MI, Sánchez-López E, Marinas A, Caridad JM, Urbano FJ. Agro-climatic information to enhance the machine-learning classification of olive oils from near-infrared spectra. ACS Agric Sci Technol 2024; 4(11): 1194-205.
http://dx.doi.org/10.1021/acsagscitech.4c00355 PMID: 39575349

[64] Meoni G, Tenori L, Di Cesare F, *et al*. NMR-based metabolomic approach to estimate chemical and sensorial profiles of olive oil. Comput Struct Biotechnol J 2025; 27: 1359-69.
http://dx.doi.org/10.1016/j.csbj.2025.03.045 PMID: 40235639

DISCLAIMER: Please note that this article is currently in the 'Early view' stage and is not the final 'Version of Record'. While it has been accepted, copy-edited, and formatted, however, it is still undergoing proofreading and corrections by the authors. Therefore, the text may still change before the final publication. Although 'Early view' may not have all bibliographic details available, the DOI and the year of online publication can still be used to cite them. The article title, DOI, publication year, and author(s) should all be included in the citation format. Once the 'Final Version of Record' becomes available the 'Early view' will be replaced by that.