Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Guided Hybrid Machine Learning for In-Season Potato Tuber Yield Forecasting Under Heat and Drought Stress with Conformal Uncertainty Quantification

Kandaswamy Prakash

prakashk@presidencyuniversity.in

Presidency University

Sudharsan Manoharan

Presidency University

Article

Keywords: Potato yield forecasting, Crop simulation model, Hybrid machine learning, Conformal prediction, Heat and drought stress, Late blight

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11734497/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Physics-Guided Hybrid Machine Learning for In-Season Potato Tuber Yield Forecasting Under Heat and Drought Stress with Conformal Uncertainty Quantification

Kandaswamy Prakash[1*] and Sudharsan Manoharan[1]

[1*]School of Information Science, Presidency University, Bengaluru, Karnataka, India.

*Corresponding author(s). E-mail(s): prakashk@presidencyuniversity.in;
Contributing authors: sudharsan.m@presidencyuniversity.in;

Abstract

In-season forecasts of potato (*Solanum tuberosum* L.) tuber yield support irrigation, contracting and storage planning, yet process-based crop models are biased by unmodelled processes such as late blight and soil water capacity, whereas data-driven learners ignore physiology and extrapolate poorly. We propose a physics-guided hybrid machine learning (PG-HML) framework: a LINTUL-type (light interception and utilization) potato model, driven by observed weather to the issue date and analog-year climatology thereafter, supplies a prior forecast and crop states; a gradient-boosted learner corrects its residual using stress indices; and conformalized quantile regression (CQR) yields distribution-free prediction intervals. The framework was evaluated on 3240 simulated seasons (three agro-climates, 40 warming years, three planting dates, maturity classes and irrigation regimes) generated by a reference system containing processes unknown to the forecaster. On 972 held-out seasons from the final 12 years, PG-HML attained a harvest-time root-mean-square error (RMSE) of 3.64 t ha^{-1} ($R^2 = 0.899$), 18.9% lower than a bias-recalibrated process model and 26.7% lower than the best purely data-driven learner (Wilcoxon $p < 10^{-4}$); at 60 days after planting it was 4.48 t ha^{-1}. CQR gave the lowest interval score at every issue time, with intervals 9.8% narrower than split conformal at 60 days, while coverage of 0.83–0.88 for a 0.90 target exposed the cost of climatic non-stationarity. Leave-one-site-out tests showed that hybrid gains do not transfer uniformly to unseen climates.

Keywords: Potato yield forecasting, Crop simulation model, Hybrid machine learning, Conformal prediction, Heat and drought stress, Late blight

1 Introduction

Potato is the most important non-cereal food crop and a key contributor to food security in temperate, subtropical and highland production systems (Devaux et al, 2020). Attainable yields vary strongly with temperature, radiation and water supply, and the crop is notably sensitive to both heat and drought: elevated night temperatures suppress tuber initiation and shift assimilate partitioning from tubers to haulm (Struik, 2007; Levy and Veilleux, 2007; Kim and Lee, 2019), while its shallow root system makes it vulnerable to transient soil water deficits (Obidiegwu et al, 2015). Under ongoing warming, the frequency of such stress episodes is increasing (Hijmans, 2003; Haverkort and Verhagen, 2008). Accurate *in-season* yield forecasts, issued weeks before harvest, therefore have direct operational value for irrigation, crop protection, contract fulfilment and storage logistics (Haverkort and Struik, 2015).

Two families of forecasting tools dominate. *Process-based crop models* such as LINTUL-POTATO (Kooman and Haverkort, 1995), SUBSTOR and related models encode phenology, canopy development, radiation use and water balance (Spitters and Schapendonk, 1990; Raymundo et al, 2014). They extrapolate physically, but intercomparisons reveal large inter-model spread and systematic biases when site-specific soil properties, pest and disease pressure or cultivar parameters are not well constrained (Fleisher et al, 2017; Asseng et al, 2013). *Data-driven learners*—random forests, gradient boosting and neural networks—have reached competitive accuracy in crop yield prediction (van Klompenburg et al, 2020; Jeong et al, 2016; Gómez et al, 2019), but they require large representative training sets, ignore known physiology, and degrade when the climate drifts beyond the training distribution (Lobell and Burke, 2010).

Physics-guided or theory-guided machine learning seeks the best of both worlds (Karpatne et al, 2017; Reichstein et al, 2019). In crop science, coupling crop-model outputs with machine learning has improved maize and wheat yield prediction (Shahhosseini et al, 2021; Feng et al, 2019), yet three gaps remain for potato. First, hybrid schemes have rarely been formulated for *in-season* forecasting, where the future weather is unknown and the crop model must be completed with a climatology. Second, the value of the hybrid decomposition has seldom been separated from the value of simply adding crop-model outputs as features. Third, forecasts are usually reported as point estimates; calibrated uncertainty, which risk-averse growers and processors need, is rarely quantified with finite-sample guarantees.

This paper addresses these gaps with the following contributions:

1. A *physics-guided hybrid machine learning* (PG-HML) framework that uses a LINTUL-type potato model, run with observed weather to the issue date and an analog-year ensemble thereafter, as a prior; a gradient-boosted learner corrects its residual using crop states and agro-meteorological stress indices (Sect. 3).
2. A distribution-free uncertainty layer based on conformalized quantile regression (CQR) (Romano et al, 2019), compared against uncalibrated quantile regression, split conformal prediction and physics-only ensemble spread.
3. A transparent, fully reproducible *in-silico benchmark* of 3240 potato seasons in which a reference system contains processes deliberately withheld from the forecaster (late blight, soil water capacity, stronger heat and drought sensitivity,

field-level radiation-use variation, measurement error), together with a warming trend that makes the test period non-stationary.
4. A systematic evaluation across six issue times, statistical significance tests, ablation, regime-wise analysis, permutation importance, leave-one-site-out transfer and a climate response surface, which together delineate where hybridization pays off and where it does not.

The remainder of the paper is organized as follows. Sect. 2 reviews related work, Sect. 3 describes the experimental system and the proposed method, Sect. 4 reports the results, Sect. 5 discusses implications and limitations, and Sect. 6 concludes.

2 Background and related work

2.1 Process-based potato models

Potato growth models typically derive daily dry-matter production from intercepted radiation and a radiation-use efficiency (RUE) (Monteith, 1977), modulated by temperature and water stress, and partition assimilates to tubers after tuber initiation (Kooman and Haverkort, 1995; Spitters and Schapendonk, 1990). Development is driven by thermal time computed with cardinal temperatures (Yin et al, 1995). Such models have been used to analyse yield gaps (Haverkort and Struik, 2015), breeding strategies (Spitters and Schapendonk, 1990) and climate impacts (Hijmans, 2003; Raymundo et al, 2014). The multi-model intercomparison of Fleisher et al (2017) showed that even calibrated potato models differ widely under contrasting climates, motivating data-driven bias correction.

2.2 Machine learning for yield prediction

Systematic reviews report random forests, gradient boosting and neural networks as the most frequently used yield learners (van Klompenburg et al, 2020). Random forests outperformed multiple regression for several crops at regional and global scale (Jeong et al, 2016), and satellite-based learners have been applied to potato (Gómez et al, 2019). Gradient-boosted decision trees (Friedman, 2001; Chen and Guestrin, 2016; Ke et al, 2017) are particularly effective on tabular agronomic data. However, statistical yield models are known to transfer poorly to novel climates (Lobell and Burke, 2010).

2.3 Hybrid and physics-guided models

Theory-guided data science uses domain models to constrain or inform learners (Karpatne et al, 2017), an approach advocated for Earth-system science (Reichstein et al, 2019). For maize, adding APSIM outputs as features to ML ensembles reduced prediction error (Shahhosseini et al, 2021); for wheat, combining APSIM with random forests improved the assessment of climate extremes (Feng et al, 2019). Residual learning—training a learner on the discrepancy between observations and a physical model—is a natural alternative that keeps the physical prior as the backbone of the forecast. We explicitly compare both strategies.

Table 1 Archetypal agro-climatic sites of the in-silico experiment

Property	S1	S2	S3
Archetype	Temperate maritime	Continental	Subtropical winter-season
Latitude (°N)	52	45	27
Annual mean T (°C)	10.0	11.5	24.0
Seasonal amplitude (°C)	7.0	12.0	8.5
Wet-day probability	0.50	0.30	0.08
Heat-wave rate (yr^{-1})	0.6	1.4	1.0
Planting dates (DOY)	100/115/130	95/110/125	285/300/315
True soil TAW (mm)†	120	85	75
Mean yield ± SD ($\mathrm{t\,ha}^{-1}$)	26.8±8.8	24.8±12.1	25.6±13.9

†Total available water of the reference system; the forecaster assumes 100 mm.

2.4 Uncertainty quantification

Conformal prediction provides finite-sample, distribution-free coverage guarantees under exchangeability (Vovk et al, 2005; Lei et al, 2018; Angelopoulos and Bates, 2023). CQR (Romano et al, 2019) combines conformal calibration with quantile regression, producing intervals that adapt to heteroscedasticity. Interval quality is assessed with proper scoring rules such as the interval score (Gneiting and Raftery, 2007). To our knowledge, conformal methods have not previously been evaluated for in-season potato yield forecasting.

3 Materials and methods

3.1 Experimental design

Field datasets rarely contain the factorial combinations of climate, management and stress needed to disentangle model behaviour, and their ground truth is confounded by unknown factors. We therefore designed a controlled in-silico experiment, a common practice for method development in crop modelling (Spitters and Schapendonk, 1990; Asseng et al, 2013), in which the data-generating process is known, but only partially revealed to the forecaster. Three archetypal agro-climatic sites (Table 1) were simulated for 40 years with a linear warming trend of 0.03 $^{\circ}\mathrm{C\,yr}^{-1}$. Each site-year was combined with three planting dates, three cultivar maturity classes (early, mid, late) and three irrigation regimes (rainfed; deficit irrigation triggered at 60% depletion; full irrigation triggered at 30% depletion), giving $3 \times 40 \times 3 \times 3 \times 3 = 3240$ seasons. Years 1–22 (1782 seasons) form the training set, years 23–28 (486) the conformal calibration set and years 29–40 (972) the test set. The chronological split mimics operational use and exposes all methods to climatic non-stationarity.

3.2 Stochastic weather generator

Daily weather was produced by a Richardson-type generator (Richardson, 1981). Rain occurrence follows a first-order Markov chain with wet-after-wet probability p_{ww} and

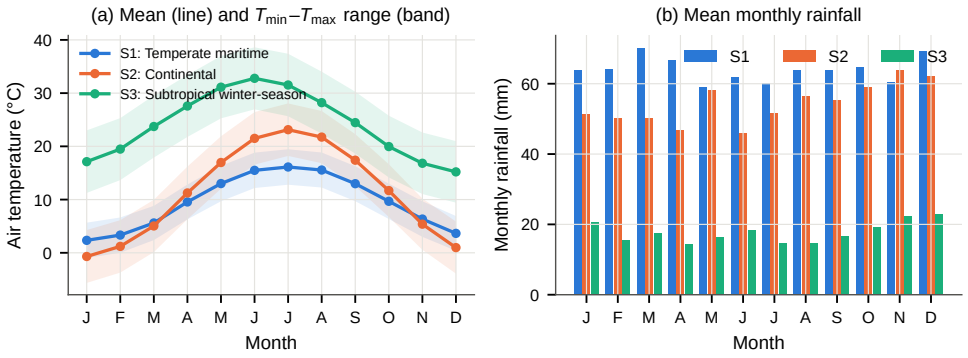

Fig. 1 Climatology of the three simulated sites (30 synthetic years). (a) Monthly mean air temperature with the $T_{\min}$–$T_{\max}$ envelope. (b) Mean monthly rainfall. S3 represents a dry-winter subtropical system where potato is grown from October to March.

wet-after-dry probability

$$p_{01} = \frac{\pi_w (1 - p_{ww})}{1 - \pi_w}, \tag{1}$$

where π_w is the stationary wet-day probability; wet-day amounts are gamma distributed with shape 0.75. Daily mean temperature on day-of-year j in year y is

$$T_j = \mu + 0.03(y - 1) + A \cos\!\left(\frac{2\pi(j - j_p)}{365}\right) + \varepsilon_j + \eta_j - 1.2\,\mathbb{I}_j^{\text{wet}}, \tag{2}$$

with AR(1) anomaly $\varepsilon_j = 0.72\,\varepsilon_{j-1} + \xi_j$, $\xi_j \sim \mathcal{N}(0, 1.9^2)$, and heat-wave anomalies η_j of 4–8 °C lasting 4–9 days, whose number follows a Poisson law with the site-specific rate. The diurnal range ΔT is reduced by 35% on wet days; $T_{\max}, T_{\min} = T \pm \Delta T/2$. Global radiation and reference evapotranspiration follow the Hargreaves–Samani formulation (Hargreaves and Samani, 1985; Allen et al, 1998):

$$R_s = 0.17\sqrt{\Delta T}\, R_a\, (1 - 0.25\,\mathbb{I}^{\text{wet}}), \tag{3}$$

$$ET_0 = 0.0023 \cdot 0.408\, R_a\, (T + 17.8)\sqrt{\Delta T}, \tag{4}$$

where R_a is extraterrestrial radiation (FAO-56). Fig. 1 summarizes the resulting climatologies.

3.3 LINTUL-type potato crop model

The process model $\mathcal{M}_\theta$ is a daily, vectorized LINTUL-type model (Kooman and Haverkort, 1995; Spitters and Schapendonk, 1990). Its nominal parameters θ are listed in Table 2.

Phenology: The daily thermal-time increment uses the beta function of Yin et al (1995),

$$\Delta TT = (T_o - T_b)\left[2x^a - x^{2a}\right], \quad x = \frac{T - T_b}{T_o - T_b}, \tag{5}$$

where $a = \ln 2 / \ln\!\big[(T_c - T_b)/(T_o - T_b)\big]$ and T is clipped to (T_b, T_c). Emergence occurs at $TT = TT_{em}$; TT_e denotes thermal time since emergence.

Canopy: During expansion ($TT_e < TT_{sen}$), leaf area index grows exponentially, limited by water stress K_s and a cultivar maximum,

$$LAI_{t+1} = \min\!\left\{ LAI_t \, e^{r_L \, \Delta TT \, K_s},\; LAI_{\max} \right\}, \tag{6}$$

and during senescence it declines at a rate accelerated by drought,

$$\Delta LAI = -\frac{\Delta TT}{TT_{mat} - TT_{sen}} \, LAI_{\max} \left[1 + \gamma (1 - K_s) \right]. \tag{7}$$

Growth: Daily dry-matter production follows the RUE concept (Monteith, 1977),

$$\Delta W = \epsilon \cdot 0.5 R_s \cdot \left(1 - e^{-k \, LAI} \right) \cdot f_T(T) \cdot K_s, \tag{8}$$

where f_T is a trapezoidal temperature response equal to one between 16 and 24 °C and zero below 5 and above 33 °C.

Partitioning and heat stress: After tuber initiation, the fraction of new assimilates allocated to tubers is

$$p_{tub} = p_{\max} \, \min\!\left\{ 1, \tfrac{(TT_e - TT_{ti})^+}{250} \right\} h(T_{\min}), \tag{9}$$

$$h(T_{\min}) = \operatorname{clip}\!\left(1 - (1 - h_0) \frac{T_{\min} - T_{lo}}{T_{hi} - T_{lo}},\; h_0,\; 1 \right), \tag{10}$$

which reproduces the well-documented inhibition of tuberization by warm nights (Struik, 2007; Levy and Veilleux, 2007; Rykaczewska, 2015). Fresh tuber yield is $Y = W_{tub} / (100 \, \rho_{DM})$ in $\mathrm{t\,ha^{-1}}$, where W_{tub} is in $\mathrm{g\,m^{-2}}$ and ρ_{DM} is the tuber dry-matter fraction.

Soil water: A single-layer bucket tracks root-zone depletion D_r (Allen et al, 1998):

$$D_{r,t} = \operatorname{clip}\!\big(D_{r,t-1} - P_t + K_c \, ET_{0,t} \, \phi(K_s) - I_t,\; 0,\; \mathrm{TAW} \big), \tag{11}$$

$$K_c = 0.15 + 0.95 \big(1 - e^{-k \, LAI} \big), \tag{12}$$

$$K_s = \operatorname{clip}\!\left(\frac{\mathrm{TAW} - D_r}{(1 - p) \, \mathrm{TAW}}, 0, 1 \right), \tag{13}$$

with $\phi(K_s) = 0.5 + 0.5 K_s$ under stress. Irrigation I_t refills the profile to 10% depletion whenever D_r exceeds the regime threshold.

3.4 Reference system and observation model

The "observed" yields are produced by a reference system that differs from the forecaster's model in ways that mirror real-world model error (Table 2): site-specific soil water capacity, a lower heat-stress threshold for night temperature, faster drought-induced senescence, stochastic RUE variation, and a late-blight module.

Table 2 Crop-model parameters of the forecaster (nominal) and of the reference system

Symbol	Description	Nominal	Reference
T_b, T_o, T_c	Cardinal temps. (°C)	2/22/32	2/22/32
TT_{em}	Planting–emergence (°Cd)	250	250
TT_{ti}	Emergence–tuber init. (°Cd)	200	200
ϵ	RUE (g MJ^{-1} PAR)	3.0	$3.0\,e^{u\,a}$
k	Extinction coefficient	0.75	0.75
r_L	LAI growth rate (°Cd^{-1})	0.018	0.018
$p_{\max}$	Max. tuber partitioning	0.85	0.85
T_{lo}, T_{hi}	Heat-stress $T_{\min}$ limits (°C)	15/22	13/22
h_0	Partitioning floor	0.25	0.25
TAW	Total available water (mm)	100	75–120
p	Depletion fraction	0.35	0.35
γ	Drought senescence factor	1.5	2.3
ρ_{DM}	Tuber dry-matter fraction	0.20	0.20
TT_{sen}/TT_{mat}	Early (°Cd)	700/1050	700/1050
	Mid (°Cd)	900/1350	900/1350
	Late (°Cd)	1100/1650	1100/1650
$LAI_{\max}$	Early/Mid/Late	3.4/4.0/4.6	3.4/4.0/4.6
–	Late-blight module	absent	present[b]

[a] $u = u_{\text{site-year}} + u_{\text{field}}$, $u_{\text{site-year}} \sim \mathcal{N}(0, 0.05^2)$, $u_{\text{field}} \sim \mathcal{N}(0, 0.04^2)$.

[b] See (14); susceptibility 1.25/1.0/0.7 for early/mid/late cultivars.

Blight-favourable days are defined, in the spirit of rain- and temperature-based blight rules (Hijmans et al, 2000), as days with rainfall > 0.5 mm and $10 < T < 24$ °C. Each favourable day after canopy closure ($TT_e > 400$ °Cd) reduces green LAI by 3% and adds to a cumulative loss ℓ; the observed yield is

$$y = Y_{\text{ref}} \cdot \text{clip}\big(1 - s\,\ell,\ 0.25,\ 1\big) \cdot (1 + \nu), \tag{14}$$

where s is a season-specific severity (cultivar susceptibility multiplied by a log-normal factor with $\sigma = 0.25$), reflecting the dominant role of *Phytophthora infestans* in potato yield loss (Fry, 2008), and $\nu \sim \mathcal{N}(0, 0.05^2)$ is multiplicative measurement error. The resulting yields range from 2.8 to 63.6 t ha^{-1} (mean 25.7, SD 11.8 t ha^{-1}; Fig. 4).

3.5 In-season forecasting problem

A forecast is issued at day $d \in \{30, 45, 60, 75, 90\}$ after planting (DAP) or at harvest ("H", all 150 days observed). Only weather up to day d is known. The process-model forecast completes the season with $K = 8$ analog years drawn from the training years for the same site and planting date:

$$\hat{y}_d^{\text{PM}} = \frac{1}{K} \sum_{k=1}^{K} \mathcal{M}_\theta\big(\mathbf{w}_{1:d} \oplus \mathbf{w}_{d+1:T}^{(k)}\big), \quad \sigma_d^{\text{PM}} = \text{sd}_k\{\cdot\}, \tag{15}$$

where $\oplus$ denotes concatenation of daily weather vectors.

3.6 Proposed physics-guided hybrid learner (PG-HML)

The feature vector $\mathbf{x} = [\mathbf{x}_w, \mathbf{x}_m, \mathbf{x}_p]$ comprises (i) agro-meteorological indices $\mathbf{x}_w$ accumulated to day d: growing degree days, rainfall, ET_0, climatic water deficit $\sum(ET_0 - P)$, radiation, mean $T_{\max}$ and $T_{\min}$, counts of hot nights ($T_{\min} > 15$ °C), hot days ($T_{\max} > 30$ °C), dry days and blight-favourable days; (ii) management descriptors $\mathbf{x}_m$ (site, planting date, maturity class, irrigation regime); and (iii) physics features $\mathbf{x}_p$: $\hat{y}_d^{\mathrm{PM}}$, σ_d^{PM}, and the model states on day d (LAI, thermal time, total and tuber biomass, seasonal mean K_s, mean heat-stress index $1-h$ and mean light interception). The PG-HML forecast is

$$\hat{y}_d(\mathbf{x}) = \hat{y}_d^{\mathrm{PM}} + g_\phi(\mathbf{x}), \tag{16}$$

where the residual learner g_ϕ is a histogram-based gradient-boosted tree ensemble (Friedman, 2001; Ke et al, 2017),

$$g_\phi(\mathbf{x}) = \sum_{m=1}^{M} \eta\, f_m(\mathbf{x}), \tag{17}$$

fitted by minimizing

$$\mathcal{L}(\phi) = \sum_{i \in \mathcal{D}_{tr}} \left(y_i - \hat{y}_{d,i}^{\mathrm{PM}} - g_\phi(\mathbf{x}_i)\right)^2 + \lambda \sum_m \|\mathbf{w}_m\|_2^2. \tag{18}$$

Because the learner models only the discrepancy, the forecast reverts towards the physical prior where the residual is poorly constrained, and the trees need not re-learn the first-order response of yield to radiation, temperature and water. Hyperparameters (learning rate $\eta \in \{0.02, 0.04, 0.08\}$, maximum leaf nodes $\in \{15, 31, 63\}$) were selected by five-fold year-grouped cross-validation on the training set at $d = 60$; the best setting was $\eta = 0.02$ with 15 leaves (CV RMSE 4.86 t ha^{-1}), used with 600 boosting iterations, minimum leaf size 15 and $\lambda = 1$.

3.7 Conformalized uncertainty quantification

Two quantile heads $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ (pinball loss) are trained on the same residual target and features, and shifted by $\hat{y}_d^{\mathrm{PM}}$. On the calibration set $\mathcal{D}_{cal}$ ($n = 486$) the conformity scores are

$$E_i = \max\left\{\hat{q}_{\alpha/2}(\mathbf{x}_i) - y_i,\ y_i - \hat{q}_{1-\alpha/2}(\mathbf{x}_i)\right\}, \tag{19}$$

and the CQR interval for a new season is (Romano et al, 2019)

$$\hat{C}(\mathbf{x}) = \left[\hat{q}_{\alpha/2}(\mathbf{x}) - \hat{Q},\ \hat{q}_{1-\alpha/2}(\mathbf{x}) + \hat{Q}\right], \tag{20}$$

where $\hat{Q}$ is the $\lceil (n+1)(1-\alpha) \rceil / n$ empirical quantile of $\{E_i\}$. Under exchangeability, $\Pr\{y \in \hat{C}(\mathbf{x})\} \geq 1-\alpha$ (Lei et al, 2018; Angelopoulos and Bates, 2023). We use $\alpha = 0.10$

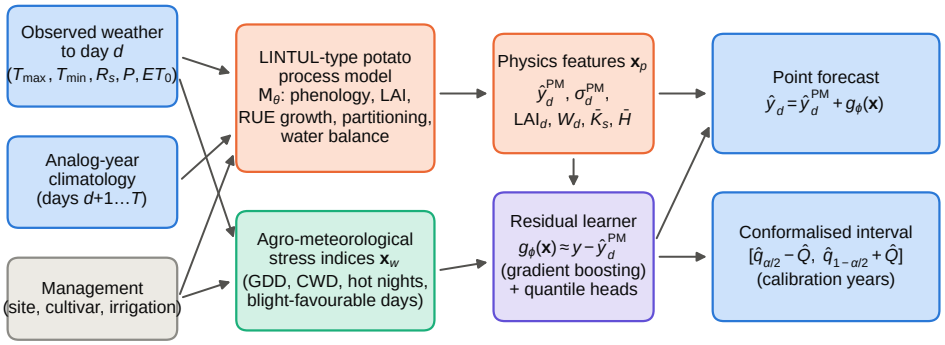

Fig. 2 Proposed PG-HML framework. The process model, run with observed weather to the issue day d and analog-year climatology thereafter, provides a prior forecast and crop states; a gradient-boosted residual learner corrects the prior using stress indices; conformalized quantile heads deliver calibrated 90% prediction intervals.

Algorithm 1 PG-HML training and in-season forecasting

Input: Training/calibration seasons $\mathcal{D}_{tr}, \mathcal{D}_{cal}$; issue day d; analog pool; K; α
Output: Point forecast $\hat{y}_d$ and interval $\hat{C}$

for each season i in $\mathcal{D}_{tr} \cup \mathcal{D}_{cal} \cup \{\text{new}\}$ **do**
  Run $\mathcal{M}_\theta$ on $\mathbf{w}_{1:d}$; store states on day d
  Draw K analog years; compute $\hat{y}_{d,i}^{\mathrm{PM}}, \sigma_{d,i}^{\mathrm{PM}}$
  Assemble $\mathbf{x}_i = [\mathbf{x}_w, \mathbf{x}_m, \mathbf{x}_p]$
end for
Fit g_ϕ on $\{(\mathbf{x}_i,\ y_i - \hat{y}_{d,i}^{\mathrm{PM}})\}_{\mathcal{D}_{tr}}$
Fit quantile heads $\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}$ on the same targets
Compute E_i on $\mathcal{D}_{cal}$; set $\hat{Q} \leftarrow \mathrm{quantile}_{\lceil (n+1)(1-\alpha) \rceil / n}(E)$
return $\hat{y}_d = \hat{y}_d^{\mathrm{PM}} + g_\phi(\mathbf{x})$, $\hat{C} = [\hat{q}_{\alpha/2} - \hat{Q}, \hat{q}_{1-\alpha/2} + \hat{Q}]$

and also evaluate a Mondrian variant with site-specific $\hat{Q}$, split conformal prediction with absolute residuals of (16), uncalibrated quantile regression, and the physics-only interval $\hat{y}_d^{\mathrm{PM}} \pm 1.645\, \sigma_d^{\mathrm{PM}}$. The full procedure is summarized in Algorithm 1 and Fig. 2.

3.8 Baselines

Seven point forecasters were compared: (i) **PM**, the raw process-model forecast; (ii) **PM-cal**, PM with linear recalibration $a + b\,\hat{y}_d^{\mathrm{PM}}$ fitted on the training set, a strong and commonly used physics-based baseline; (iii) **Ridge** regression; (iv) **RF**, random forest (400 trees) (Breiman, 2001); (v) **GBM**, gradient boosting with the same hyperparameters as PG-HML; (vi) **MLP**, a two-layer (64–32) neural network with early stopping; and (vii) **HF-GBM**, a "hybrid-feature" GBM that receives the same physics

features as PG-HML but predicts yield directly. Learners (iii)–(vi) use only $[\mathbf{x}_w, \mathbf{x}_m]$. All models were implemented in Python with scikit-learn (Pedregosa et al, 2011).

3.9 Evaluation metrics and statistical tests

For test observations y_i and forecasts $\hat{y}_i$, $e_i = \hat{y}_i - y_i$:

$$\mathrm{RMSE} = \sqrt{\tfrac{1}{n}\textstyle\sum_i e_i^2}, \quad \mathrm{MAE} = \tfrac{1}{n}\textstyle\sum_i |e_i|, \quad \mathrm{Bias} = \bar{e}, \tag{21}$$

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2}, \quad \mathrm{nRMSE} = 100\,\frac{\mathrm{RMSE}}{\bar{y}}, \tag{22}$$

$$d = 1 - \frac{\sum_i e_i^2}{\sum_i \left(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|\right)^2}, \tag{23}$$

where d is Willmott's index of agreement (Willmott, 1981). Intervals $[l_i, u_i]$ are evaluated by empirical coverage, mean width and the interval score (Gneiting and Raftery, 2007)

$$\mathrm{IS}_\alpha = \tfrac{1}{n}\textstyle\sum_i \Big[(u_i - l_i) + \tfrac{2}{\alpha}(l_i - y_i)^+ + \tfrac{2}{\alpha}(y_i - u_i)^+\Big]. \tag{24}$$

Differences in RMSE were assessed with a paired bootstrap (2000 resamples, 95% CI) and differences in absolute error with the Wilcoxon signed-rank test. Permutation importance (Breiman, 2001; Fisher et al, 2019) (10 repeats) was computed on the test set at $d = 60$.

4 Results

4.1 Simulated crop behaviour

Fig. 3 illustrates reference-system trajectories for contrasting scenarios. Under full irrigation at S1 the canopy closes about 47 DAP and tubers bulk steadily to 41 t ha^{-1}. Rainfed S2 shows severe water stress ($K_s < 0.2$) from 48 DAP, premature senescence and only 4 t ha^{-1}, whereas the irrigated counterpart reaches 32 t ha^{-1} but suffers repeated heat-stress episodes. Across all seasons (Fig. 4), irrigation regime explained most variation at the drier sites S2 and S3 (rainfed means 12.5 and 10.5 t ha^{-1} versus 32.7 and 35.6 t ha^{-1} under full irrigation), while at the humid site S1 the rainfed penalty was smaller (23.1 vs. 29.9 t ha^{-1}) and blight losses mattered more. These magnitudes are consistent with reported potato yield levels and stress responses (Haverkort and Struik, 2015; Obidiegwu et al, 2015).

4.2 Point-forecast accuracy

Table 3 reports test-set accuracy at 60 DAP and at harvest. The raw process model was strongly biased (+7.9 t ha^{-1} at harvest) because it ignores blight and overestimates soil water at S2 and S3; linear recalibration (PM-cal) removed the bias and reduced RMSE from 9.58 to 4.49 t ha^{-1}. Purely data-driven learners reached 4.76–5.04 t ha^{-1} (Ridge: 6.83). PG-HML achieved the lowest harvest-time error (RMSE 3.64 t ha^{-1},

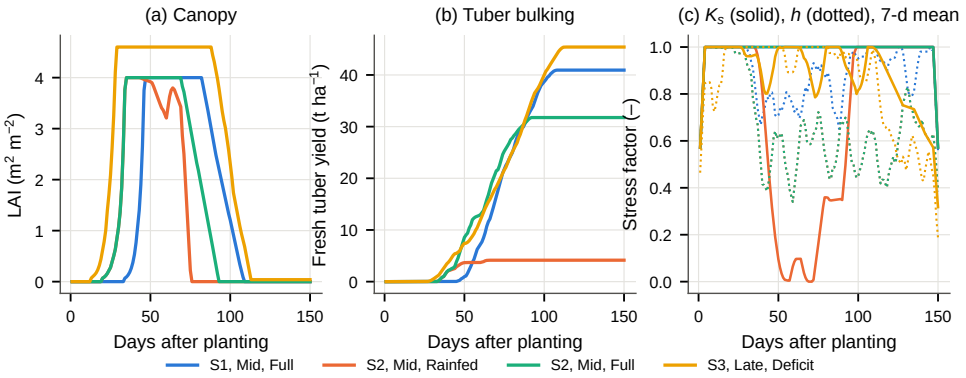

Fig. 3 Reference-system crop dynamics for four representative seasons (year 35; shown without the blight and RUE-noise terms). (a) Leaf area index. (b) Fresh tuber yield accumulation. (c) Water-stress factor K_s (solid) and heat partitioning factor h (dotted), 7-day running means.

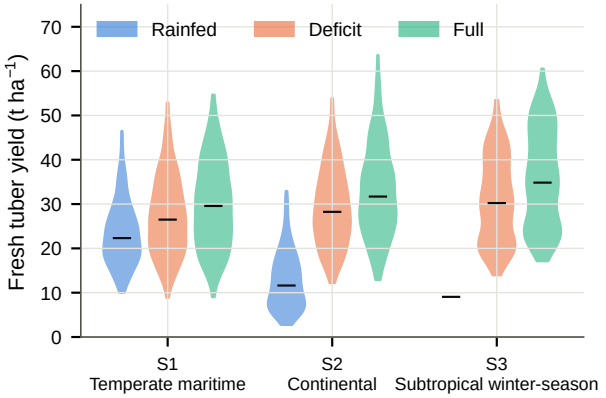

Fig. 4 Distribution of observed fresh tuber yield by site and irrigation regime (3240 seasons). Horizontal bars mark medians.

MAE 2.72 t ha^{-1}, $R^2 = 0.899$, $d = 0.974$), i.e., 18.9% below PM-cal, 26.7% below GBM and 23.5% below MLP. Fig. 5 shows that PG-HML removes the regime-dependent scatter that dominates the data-driven learners, particularly for low-yielding rainfed seasons.

4.3 Skill versus issue time

Table 4 and Fig. 6 show RMSE at six issue times. All learners improve as the season unfolds. Early in the season (30 DAP) the forecasting skill is limited by unknown future weather and the models are statistically close (5.72–6.20 t ha^{-1}, except Ridge). From 45 DAP onwards, the two physics-informed learners separate from the rest, and PG-HML attains its minimum at 90 DAP (3.45 t ha^{-1}). A notable finding is that

Table 3 Test-set accuracy (972 seasons, years 29–40) at 60 days after planting (DAP) and at harvest. Best value per column in bold

Model	60 DAP					Harvest				
	RMSE (t ha^{-1})	MAE (t ha^{-1})	Bias (t ha^{-1})	R^2	d	RMSE (t ha^{-1})	MAE (t ha^{-1})	Bias (t ha^{-1})	R^2	d
PM (raw)	11.00	9.21	9.06	0.077	0.833	9.58	7.95	7.87	0.301	0.866
PM-cal	5.11	3.90	**0.94**	0.801	0.946	4.49	3.38	**0.01**	0.846	0.957
Ridge	6.88	5.41	1.24	0.639	0.887	6.83	5.43	0.31	0.644	0.886
RF	4.76	3.74	1.31	0.827	0.954	5.04	3.99	1.11	0.806	0.946
GBM	4.84	3.79	1.40	0.821	0.953	4.97	3.90	0.58	0.812	0.949
MLP	4.98	3.93	1.41	0.811	0.950	4.76	3.70	0.21	0.827	0.952
HF-GBM	**4.44**	**3.31**	1.07	**0.849**	**0.962**	3.86	2.85	0.29	0.886	0.971
PG-HML	4.48	**3.31**	1.04	0.847	**0.962**	**3.64**	**2.72**	0.24	**0.899**	**0.974**

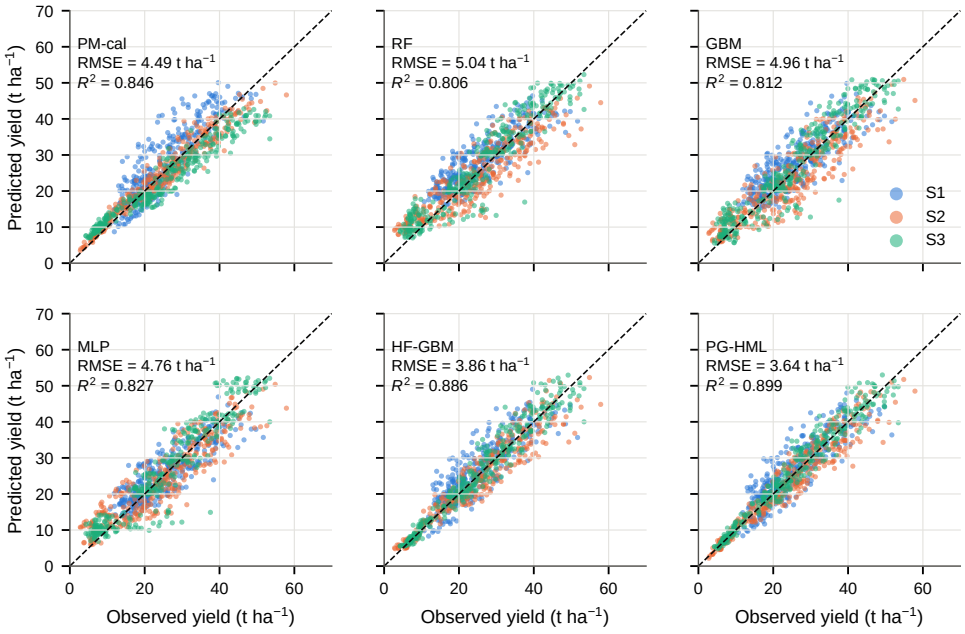

Fig. 5 Observed versus predicted harvest-time yield on the 972 test seasons, coloured by site. Dashed line: 1:1.

the purely data-driven learners *deteriorate* between 90 DAP and harvest (e.g., GBM 4.42→4.96 t ha^{-1}): their window-aggregated weather features accumulate weather after crop maturity, which is irrelevant to yield. Models that receive process-model states, which stop evolving at maturity, are far less affected.

4.4 Statistical significance

Table 5 confirms that PG-HML is significantly more accurate than PM-cal and all purely data-driven learners at both 60 DAP and harvest (all bootstrap CIs exclude

Table 4 Test root-mean-square error (RMSE, t ha^{-1}) versus forecast issue time (DAP)

Model	30	45	60	75	90	Harvest
PM (raw)	11.89	11.61	11.00	10.36	9.81	9.58
PM-cal	6.20	5.78	5.11	4.70	4.45	4.49
Ridge	7.10	7.03	6.88	6.67	6.63	6.83
RF	**5.72**	5.48	4.76	4.42	4.50	5.04
GBM	5.82	5.74	4.84	4.37	4.42	4.96
MLP	5.96	5.73	4.98	4.54	4.47	4.76
HF-GBM	5.73	5.58	**4.44**	**3.75**	3.55	3.86
PG-HML	5.85	**5.47**	4.48	3.82	**3.45**	**3.64**

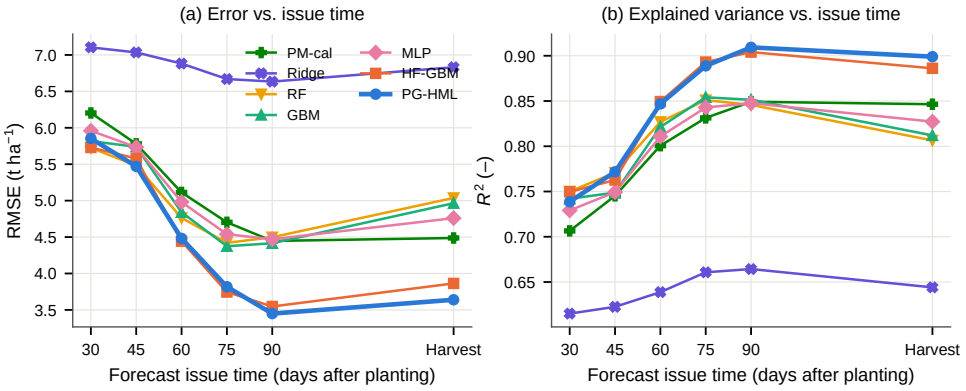

Fig. 6 Forecast skill as a function of issue time (days after planting; H = harvest). (a) RMSE. (b) Coefficient of determination.

zero, Wilcoxon $p < 10^{-4}$). Against HF-GBM, PG-HML is significantly better at harvest (ΔRMSE = 0.22 t ha^{-1}, $p = 2 \times 10^{-4}$) but statistically indistinguishable at 60 DAP ($p = 0.71$). Hence the dominant gain stems from injecting physics into the learner; the residual formulation adds a further, smaller benefit that grows late in the season.

4.5 Regime-wise performance

Pooled metrics can hide regime-specific behaviour, because a large part of the pooled variance lies between management regimes. Table 6 breaks down harvest-time accuracy by site and irrigation. PG-HML was the most accurate model in five of nine regimes, with the largest gains where the forecaster's physics is most wrong: blight-prone, irrigated S1 seasons (RMSE 4.23 vs. 6.03 t ha^{-1} for PM-cal) and the low-TAW subtropical site S3 (e.g., rainfed 1.60 vs. 2.40 t ha^{-1}). Conversely, at the continental site S2, where the soil-water error is smaller and blight pressure is low, recalibrated physics alone was best in all three regimes, and PG-HML was within 0.1–0.9 t ha^{-1}

Table 5 Paired comparison of PG-HML with competing models on the test set. ΔRMSE > 0 favours PG-HML

Issue time	Comparator	ΔRMSE (t ha^{-1})	95% bootstrap CI (t ha^{-1})	Wilcoxon p
60 DAP	PM-cal	0.63	[0.40, 0.86]	$< 10^{-4}$
	RF	0.28	[0.10, 0.47]	$< 10^{-4}$
	GBM	0.36	[0.16, 0.55]	$< 10^{-4}$
	MLP	0.50	[0.27, 0.73]	$< 10^{-4}$
	HF-GBM	-0.04	[-0.11, 0.04]	0.71
Harvest	PM-cal	0.85	[0.63, 1.06]	$< 10^{-4}$
	RF	1.40	[1.17, 1.62]	$< 10^{-4}$
	GBM	1.33	[1.10, 1.54]	$< 10^{-4}$
	MLP	1.12	[0.88, 1.37]	$< 10^{-4}$
	HF-GBM	0.22	[0.14, 0.30]	2×10^{-4}

Table 6 Harvest-time test RMSE (t ha^{-1}) by site and irrigation regime

Site	Regime	Mean y	PM-cal	GBM	PG-HML
S1	Rainfed	20.9	**3.75**	5.36	3.87
	Deficit	26.2	6.51	5.24	**5.11**
	Full	29.3	6.03	4.95	**4.23**
S2	Rainfed	11.9	**1.66**	5.14	1.77
	Deficit	29.1	**3.10**	4.78	3.66
	Full	33.1	**3.53**	5.71	4.39
S3	Rainfed	11.2	2.40	5.42	**1.60**
	Deficit	29.8	5.23	3.63	**3.22**
	Full	34.0	5.52	4.10	**3.40**

of it, while GBM was markedly worse. Within-regime R^2 values (0.56–0.94) are lower than the pooled value, a caveat that should accompany any pooled yield-forecasting metric.

4.6 Ablation study

Table 7 isolates the components of PG-HML. Removing the process-model state features (keeping only the residual target) increased the harvest RMSE by 0.25 t ha^{-1}; removing the stress indices (hot nights, hot days, dry days, blight-favourable days) increased it by 0.14 t ha^{-1}. Learning yield directly instead of the residual (HF-GBM) cost 0.22 t ha^{-1}. The analog-ensemble spread σ_d^{PM} did not help point accuracy (removing it slightly lowered the 60-DAP RMSE to 4.43 t ha^{-1}; at harvest it is identically zero), so it is retained only as an interpretable diagnostic.

Table 7 Ablation of PG-HML components (test RMSE, t ha^{-1})

Variant	60 DAP	Harvest
PG-HML (full)	4.48	**3.64**
w/o process-model states	4.52	3.88
w/o stress indices	4.52	3.78
w/o ensemble spread σ_d^{PM}	**4.43**	**3.64**
direct target instead of residual (HF-GBM)	4.44	3.86
no physics at all (GBM)	4.84	4.96

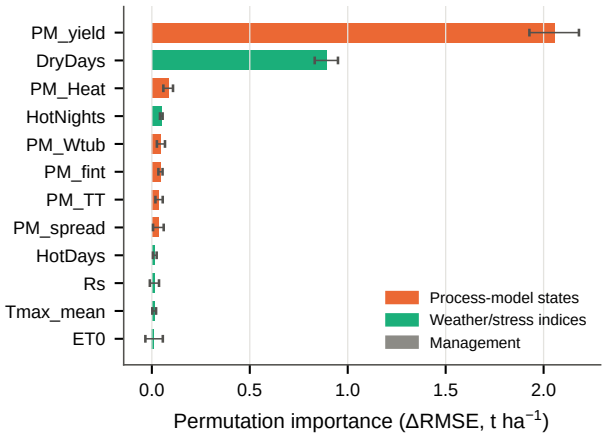

Fig. 7 Permutation importance of the 12 most influential features of the PG-HML residual learner at 60 DAP (mean ± SD over 10 repeats).

4.7 Interpretability

Permutation importance at 60 DAP (Fig. 7) is dominated by the process-model forecast $\hat{y}_d^{\mathrm{PM}}$ ($\Delta\mathrm{RMSE} = 2.05 \pm 0.13$ t ha^{-1}), followed by the number of dry days (0.89 ± 0.06 t ha^{-1}). The latter is plausible: in this system dry days are the complement of the wet days that drive both blight risk and soil-water recharge—the two processes missing from the forecaster's model. Heat-related features (mean heat-stress index, hot nights) rank next, consistent with the lower heat threshold of the reference system. The learner thus attends to precisely the processes that the physical prior misrepresents, which supports the interpretability of the hybrid decomposition.

4.8 Uncertainty quantification

Table 8 and Fig. 8 compare interval methods. Physics-only ensemble intervals were severely overconfident (coverage 0.15 at 60 DAP and collapsing to zero at harvest) because the analog spread represents only weather uncertainty, not model error. Uncalibrated quantile regression under-covered (0.74–0.82). Split conformal prediction achieved coverage closest to the nominal 0.90 (0.87–0.91) at the price of wide,

Table 8 Prediction intervals ($1 - \alpha = 0.90$) on the test set: coverage, mean width (t ha^{-1}) and interval score (IS)

	60 DAP			Harvest		
Method	Cov.	Width	IS	Cov.	Width	IS
PM ensemble	0.153	8.48	115.8	0.000	0.00	159.0
QR (uncal.)	0.782	10.73	19.43	0.796	8.63	15.54
Split-CP	**0.901**	14.29	20.27	**0.877**	10.82	16.47
CQR (proposed)	0.881	12.89	**18.03**	0.874	10.19	14.58
Mondrian CQR	0.876	12.87	18.10	0.874	10.56	**14.23**

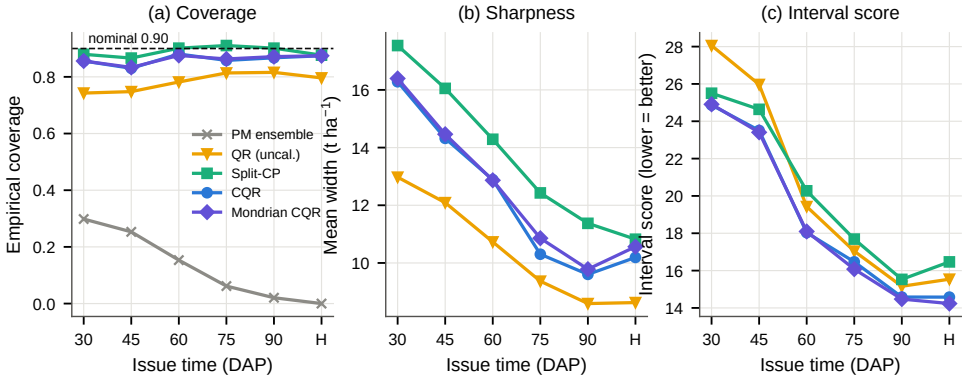

Fig. 8 Interval quality versus issue time. (a) Empirical coverage (dashed: nominal 0.90). (b) Mean interval width. (c) Interval score (lower is better). The PM ensemble is omitted from (b)–(c) for scale.

constant-width intervals. CQR produced the sharpest calibrated intervals—9.8% narrower than split conformal at 60 DAP (12.89 vs. 14.29 t ha^{-1})—and the lowest interval score at every issue time, with coverage of 0.83–0.88. The Mondrian (site-wise) variant behaved similarly and attained the best harvest interval score (14.23). The shortfall from 0.90 arises because the calibration years precede the warmer test years, violating exchangeability; the observed mean yield declined from 26.4 t ha^{-1} in training to 25.0 t ha^{-1} in the test years. Fig. 9 illustrates how the CQR interval narrows as the season progresses and contains the observed yield in three representative test seasons.

4.9 Transfer to unseen climates

In leave-one-site-out (LOSO) experiments (Fig. 10, Table 9), site indicators were removed and the models were trained on two sites and tested on all years of the third. No method dominated. At harvest, PG-HML was best for held-out S1 (5.95 t ha^{-1}) and, second to PM-cal for S2 (3.90 vs. 3.25 t ha^{-1}), whereas for the climatically distinct S3 the pure GBM was best (4.79 t ha^{-1}) and PG-HML showed a negative bias (-3.0 t ha^{-1}). At 60 DAP the raw process model was in fact the best forecaster for

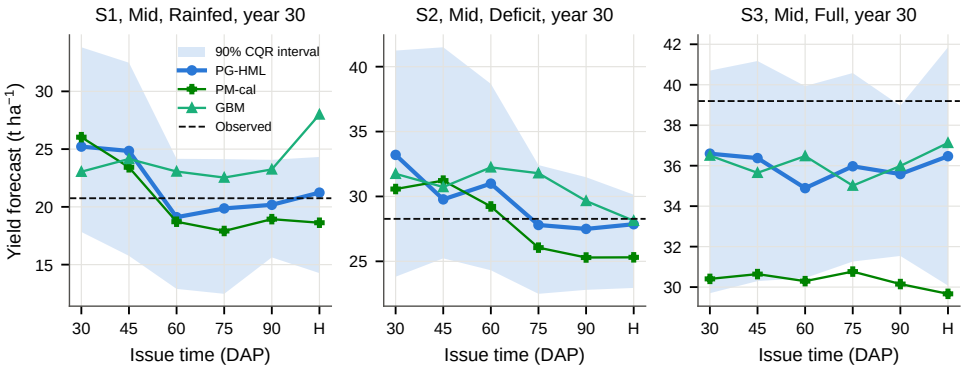

Fig. 9 In-season forecast trajectories for three test seasons. Shaded band: 90% CQR interval around the PG-HML forecast; dashed line: observed yield.

Table 9 Leave-one-site-out RMSE (t ha^{-1}); models trained on two sites and tested on the third

	60 DAP			Harvest		
Model	S1	S2	S3	S1	S2	S3
PM (raw)	14.41	10.31	**5.64**	13.47	9.15	5.30
PM-cal	7.68	**4.29**	6.84	7.39	**3.25**	6.41
GBM	**6.84**	6.74	7.22	6.98	7.58	**4.79**
PG-HML	7.21	5.07	8.49	**5.95**	3.90	6.04

S3. The residual learner therefore transfers the corrections it learned at the training sites—for instance blight-related corrections at humid S1—to a dry subtropical site where they do not apply. Hybrid corrections are site-specific, and new environments require at least a small amount of local calibration data.

4.10 Climate sensitivity of the reference system

To place forecast errors in agronomic context, Fig. 11 shows the simulated response of rainfed, mid-maturity potato at the continental site S2 to uniform warming and seasonal rainfall scaling (reference-system physics, 40 years). Yield declined by about 10% per °C of warming at baseline rainfall (−10.4% at +1 °C, −20.1% at +2 °C, −39.5% at +4 °C), driven jointly by night-temperature effects on partitioning and higher evaporative demand. A 30% rainfall reduction lowered yield by 25.6%, and a 60% reduction by 51.1%, while a 60% increase could offset roughly 3 °C of warming. These sensitivities exceed the PG-HML harvest RMSE (3.64 t ha^{-1}, 14.5% of the mean yield), indicating that the forecaster resolves agronomically meaningful differences.

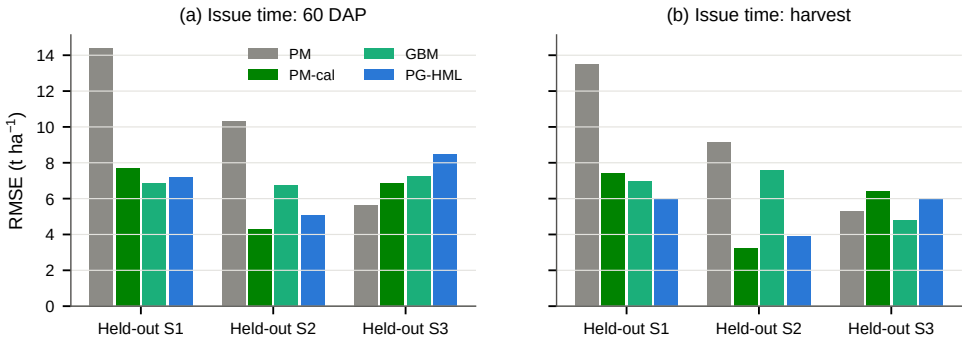

Fig. 10 Leave-one-site-out RMSE at (a) 60 DAP and (b) harvest.

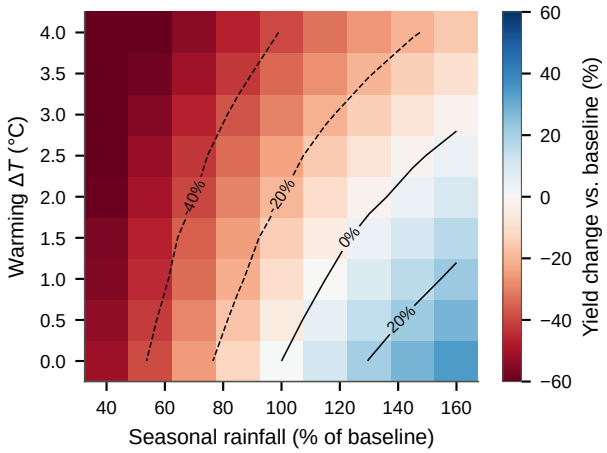

Fig. 11 Simulated yield change of rainfed mid-maturity potato at S2 versus warming and seasonal rainfall (reference-system physics, 40-year means).

5 Discussion

5.1 Why the hybrid works

The physical prior encodes first-order yield responses—radiation capture, phenology, water balance—that data-driven learners must otherwise infer from a limited number of seasons. Errors of the prior, however, are structured: they correlate with the weather patterns that drive unmodelled processes. A residual learner exploits exactly this structure, as revealed by the importance analysis. Our results suggest a practical hierarchy: (i) recalibrate the process model, which already halves the raw error; (ii) feed its outputs and states to a learner, which yields most of the remaining gain; and (iii) formulate the learner on the residual, which adds a significant late-season

improvement and keeps the forecast anchored to physics. The robustness of physics-informed models to post-maturity weather noise (Fig. 6) is an additional, practically relevant benefit.

5.2 Operational implications

At 60 DAP—typically shortly after tuber initiation and early bulking—PG-HML forecasts were within about 4.5 t ha^{-1} RMSE with honest 90% intervals of about 13 t ha^{-1}, sufficient to inform late irrigation and fertilization decisions, contracting volumes and storage allocation. CQR intervals adapt to the season: they are wide for rainfed, drought-prone crops and narrow for irrigated crops in benign years. For risk-averse users, split conformal provides closer adherence to the nominal level at the cost of sharpness, and recency-weighted or adaptive conformal schemes could be used to counter the coverage loss caused by warming.

5.3 Limitations and future work

This study is deliberately conducted in silico; its strength is a known ground truth and a factorial design, its weakness that real systems contain errors in weather inputs, management records and yield measurements, as well as processes beyond those simulated (nitrogen supply, other pests and diseases, tuber quality). The reference system shares the structure of the forecaster's model, which probably favours physics-informed methods compared with real-world structural error. The absolute RMSE values should therefore be read as relative benchmarks rather than as expected field performance. The LOSO results further show that hybrid corrections do not transfer freely across climates. Future work will (i) validate PG-HML on multi-location potato variety and irrigation trials, (ii) assimilate remotely sensed canopy cover into the process model before residual learning (Gómez et al, 2019), (iii) investigate adaptive conformal methods under non-stationarity, and (iv) integrate operational late-blight decision-support outputs as explicit features.

6 Conclusion

A physics-guided hybrid machine learning framework for in-season potato yield forecasting was developed and benchmarked. Coupling a LINTUL-type process model with a gradient-boosted residual learner reduced harvest-time RMSE to 3.64 t ha^{-1} ($R^2 = 0.899$), significantly outperforming a recalibrated process model (18.9% lower error) and the best purely data-driven learner (26.7% lower), and delivered 4.48 t ha^{-1} at 60 DAP. Conformalized quantile regression gave the sharpest calibrated intervals with the best interval score, while exposing the coverage cost of a warming climate. The hybrid's advantages were largest where the process model misrepresents disease and soil water, and limited when transferring to unseen climates. The open simulation benchmark and code provide a basis for rigorous comparison of potato yield-forecasting methods before field deployment.

Supplementary information. The Python source code (weather generator, crop model, experiment and figure scripts) and all result tables are provided as supplementary material.

Statements and Declarations

Funding. The authors declare that no funds, grants or other support were received during the preparation of this manuscript.

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval. Not applicable. The study used only simulated data and did not involve human participants or animals.

Consent to participate and for publication. Not applicable.

Data availability. All data analysed in this study were generated synthetically by the simulation code described in Sect. 3. The complete dataset (3240 simulated seasons) and all derived metrics are included in the supplementary material and can be regenerated exactly with the fixed random seed 2026.

Code availability. The Python code (NumPy, SciPy, pandas, scikit-learn, Matplotlib) that reproduces every number, table and figure is provided as supplementary material.

Author contributions. All authors contributed to the study conception and design. Simulation, model development and analysis were performed by Kandaswamy Prakash and Sudharsan Manoharan. The first draft of the manuscript was written by Kandaswamy Prakash and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

ORCID. Kandaswamy Prakash: 0000-0003-4073-5207; Sudharsan Manoharan: 0000-0002-0528-3076.

References

Allen RG, Pereira LS, Raes D, et al (1998) Crop evapotranspiration: guidelines for computing crop water requirements. Irrigation and Drainage Paper 56, FAO, Rome

Angelopoulos AN, Bates S (2023) Conformal prediction: a gentle introduction. Foundations and Trends in Machine Learning 16(4):494–591

Asseng S, et al (2013) Uncertainty in simulating wheat yields under climate change. Nature Climate Change 3(9):827–832

Breiman L (2001) Random forests. Machine Learning 45(1):5–32

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, pp 785–794

Devaux A, Goffart JP, Petsakos A, et al (2020) Global food security, contributions from sustainable potato agri-food systems. In: Campos H, Ortiz O (eds) The Potato Crop. Springer, Cham, p 3–35

Feng P, et al (2019) Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agricultural and Forest Meteorology 275:100–113

Fisher A, Rudin C, Dominici F (2019) All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research 20(177):1–81

Fleisher DH, et al (2017) A potato model intercomparison across varying climates and productivity levels. Global Change Biology 23(3):1258–1281

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of Statistics 29(5):1189–1232

Fry W (2008) *Phytophthora infestans*: the plant (and *R* gene) destroyer. Molecular Plant Pathology 9(3):385–402

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102(477):359–378

Gómez D, Salvador P, Sanz J, et al (2019) Potato yield prediction using machine learning techniques and Sentinel 2 data. Remote Sensing 11(15):1745

Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1(2):96–99

Haverkort AJ, Struik PC (2015) Yield levels of potato crops: recent achievements and future prospects. Field Crops Research 182:76–85

Haverkort AJ, Verhagen A (2008) Climate change and its repercussions for the potato supply chain. Potato Research 51(3–4):223–237

Hijmans RJ (2003) The effect of climate change on global potato production. American Journal of Potato Research 80(4):271–279

Hijmans RJ, Forbes GA, Walker TS (2000) Estimating the global severity of potato late blight with GIS-linked disease forecast models. Plant Pathology 49(6):697–705

Jeong JH, et al (2016) Random forests for global and regional crop yield predictions. PLoS ONE 11(6):e0156571

Karpatne A, et al (2017) Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Transactions on Knowledge and Data Engineering 29(10):2318–2331

Ke G, et al (2017) LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems

Kim YU, Lee BW (2019) Differential mechanisms of potato yield loss induced by high day and night temperatures during tuber initiation and bulking: photosynthesis and tuber growth. Frontiers in Plant Science 10:300

van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: a systematic literature review. Computers and Electronics in Agriculture 177:105709

Kooman PL, Haverkort AJ (1995) Modelling development and growth of the potato crop influenced by temperature and daylength: LINTUL-POTATO. In: Haverkort AJ, MacKerron DKL (eds) Potato Ecology and Modelling of Crops under Conditions Limiting Growth. Kluwer Academic Publishers, Dordrecht, p 41–59

Lei J, G'Sell M, Rinaldo A, et al (2018) Distribution-free predictive inference for regression. Journal of the American Statistical Association 113(523):1094–1111

Levy D, Veilleux RE (2007) Adaptation of potato to high temperatures and salinity—a review. American Journal of Potato Research 84(6):487–506

Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology 150(11):1443–1452

Monteith JL (1977) Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London B 281(980):277–294

Obidiegwu JE, Bryan GJ, Jones HG, et al (2015) Coping with drought: stress and adaptive responses in potato and perspectives for improvement. Frontiers in Plant Science 6:542

Pedregosa F, et al (2011) Scikit-learn: machine learning in Python. Journal of Machine Learning Research 12:2825–2830

Raymundo R, Asseng S, Cammarano D, et al (2014) Potato, sweet potato, and yam models for climate change: a review. Field Crops Research 166:173–185

Reichstein M, et al (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566(7743):195–204

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research 17(1):182–190

Romano Y, Patterson E, Candès EJ (2019) Conformalized quantile regression. In: Advances in Neural Information Processing Systems

Rykaczewska K (2015) The effect of high temperature occurring in subsequent stages of plant development on potato yield and tuber physiological defects. American Journal of Potato Research 92(3):339–349

Shahhosseini M, Hu G, Huber I, et al (2021) Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports 11:1606

Spitters CJT, Schapendonk AHCM (1990) Evaluation of breeding strategies for drought tolerance in potato by means of crop growth simulation. Plant and Soil 123(2):193–203

Struik PC (2007) Responses of the potato plant to temperature. In: Vreugdenhil D (ed) Potato Biology and Biotechnology: Advances and Perspectives. Elsevier, Amsterdam, p 367–393

Vovk V, Gammerman A, Shafer G (2005) Algorithmic Learning in a Random World. Springer, New York

Willmott CJ (1981) On the validation of models. Physical Geography 2(2):184–194

Yin X, Kropff MJ, McLaren G, et al (1995) A nonlinear model for crop development as a function of temperature. Agricultural and Forest Meteorology 77(1–2):1–16