



Research article

Time-ordered empirical framework for forecasting climate-driven oil palm yield in southern Thailand

Napat Jantaraprasit^a, Parichart Promchote^{a,*}, Shih-Yu Simon Wang^a, Chalermphol Phumichai^a, Piya Kittipadakul^a, Luthiene França^b

^a Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand

^b Department of Plants, Soils and Climate, S. J. Jessie E. Quinney College of Agriculture and Natural Resources, Utah State University, Logan UT 84322-4820, USA

Article Info

Article history:

Received 25 October 2025

Revised 15 April 2026

Accepted 18 April 2026

Available online 10 June 2026

Keywords:

Climate variability,
Crop prediction,
Regression models,
Tree crop,
Tropical agriculture

Abstract

Importance of the work: Oil palm yield in southern Thailand is strongly sensitive to climatic variations; however, robust intra-annual forecasting frameworks are largely absent.

Objectives: To quantify lagged climate-yield relationships and to build province-scale empirical models for reliable intra-annual yield forecasting.

Materials and Methods: Monthly yield data were compiled for 13 southern provinces (2005–2022) to derive province-averaged reanalysis climate variables with lag adjustments. Collinearity was reduced using variance inflation factor analysis and models were trained through exhaustive search with expanding-window cross-validation.

Results: Robust models with 2–8 predictors were produced for 11 provinces. Across test splits, the ranges were: coefficient of determination (R^2), 0.33–0.73; root mean square error, 0.14–0.30 t/ha; coefficient of variation of the root mean square error, 8.38–23.75 %; and mean bias error, within ± 0.13 t/ha. Dew-point temperature, soil moisture at 1.6 m and precipitation dominated the selected predictors, with common lags near 3 mth and 21 mth. Forecast lead time spanned 3–20 mth, most often 6–11 mth ahead of harvest. The lowest errors occurred in Nakhon Si Thammarat and Surat Thani provinces, while southernmost provinces had higher relative errors. Partial- R^2 attribution highlighted mid-season soil moisture (often a 7 mth lag) as the most influential predictor in 7 of the 11 provinces.

Main finding: A time-ordered empirical framework successfully forecast province-scale oil palm yields several months in advance using reanalysis climate and soil-moisture predictors. The results could support operational early warning for mills and agencies, while also clarifying the lagged climatic controls that shape yield variability in southern Thailand.

* Corresponding author.

E-mail address: parichart.pr@ku.th (P. Promchote)

online 2452-316X print 2468-1458/Copyright © 2026. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), production and hosting by Kasetsart University Research and Development Institute on behalf of Kasetsart University.

<https://doi.org/10.34044/j.anres.2026.60.3.08>

Introduction

Oil palm (*Elaeis guineensis* Jacq.) is the world's most land-efficient oil crop and serves as a cornerstone of Thailand's economy, with the country ranking as the third-largest global producer (Murphy, 2014; FAOSTAT, 2023). Optimal oil palm yields depend on specific climatic conditions: mean temperatures between 22–33°C; annual precipitation exceeding 2,000 mm; high relative humidity; and abundant solar radiation (Oettli et al., 2018). Consequently, over 80% of Thailand's output originates from southern provinces—regions that now face escalating climatic risks (OAE, 2023; Promchote et al., 2023; Kuinkel et al., 2024).

For example, global warming is intensifying the frequency and severity of droughts and floods, often exacerbated by El Niño-Southern Oscillation phases that disrupt monsoon rainfall and humidity patterns. In southern Thailand, El Niño-related warm phases can prolong dry-season conditions and increase the water deficit (Rahman et al., 2013; Promchote et al., 2023; Kuinkel et al., 2024). Oil palm is particularly sensitive to such climate variability because of its long and asynchronous reproductive phenology. Specifically, its inflorescence initiation begins approximately 2 yr before harvest, with sex determination occurring during this period, rendering reproductive development highly vulnerable to temperature and moisture stress (Adam et al., 2011). Consequently, climate anomalies within this 2–3 yr window can elevate inflorescence abortion, bias sex determination toward male inflorescences, reduce the number of fruit-bearing female inflorescences and ultimately decrease fresh fruit bunch number and weight while altering oil composition (Adam et al., 2011). These physiological disruptions propagate into measurable yield losses approximately 24–36 mth after the initial climate shock (Adam et al., 2011; Kamil and Omar, 2017; Chiarawipa et al., 2020), providing a clear mechanistic basis for incorporating long temporal lags in climate-yield modeling frameworks.

Despite this sensitivity, there is little use of intra-annual yield prediction using climate variables such as dew-point temperature and vapor-pressure deficit. Many evaluations rely on random splits or single holdout periods, which are inappropriate for time-series forecasting because temporal dependence can bias performance estimates, yielding either overly optimistic or overly pessimistic forecasting skills that do not reflect operational forecasting conditions. The current study aimed to better emulate real-world deployment based on validating model performance using time-ordered

expanding-window (walk-forward) cross-validation (Paudel et al., 2022), in which models are trained on all data available up to a given forecast origin and evaluated on subsequent observations. To address the data gaps, the feasibility of a province-scale forecasting framework was assessed to integrate observation-derived climate and soil-moisture predictors within a statistical modeling approach, with performance evaluated using expanding-window cross-validation. With the focus on 13 southern provinces, the aims were: 1) to quantify the temporal lags and magnitudes through which individual climate variables influence yield anomalies; and 2) to develop empirical models capable of accurate intra-annual prediction.

At the provincial scale, such intra-annual yield predictions could function as an early-warning indicator of climate-driven production variability several months ahead of harvest. This information could support regional authorities and industry stakeholders in anticipating supply shortfalls or surpluses, enabling proactive planning of milling capacity, harvest logistics and input distribution during climate anomalies. While the framework did not aim to prescribe farm-level management, it should a scalable decision-support tool for strategic planning under increasing climate variability.

Materials and Methods

Data collection

Monthly oil palm yield data were obtained for 13 southern provinces of Thailand (Office of Agricultural Economics, 2023). The locations of these provinces are shown in Fig. 5. The record spans from January 2005 to December 2022, covering 216 monthly observations that represent provincial averages. Oil palm cultivation areas in each province were extracted from the CROPGRIDS database (Tang et al., 2024). Non-arable lands, such as seas and reserved forests, were masked to ensure accurate representation of cultivation areas. The coordinates of these areas were matched to the nearest coordinate in the climate datasets and averaged provincially.

Monthly reanalysis climate datasets were sourced from Climate Prediction Center (CPC; van den Dool et al., 2003) and European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5; Hersbach et al., 2023). CPC provided soil moisture data at 1.6 m below ground surface (SM-1.6). Although typically, oil palm has a dense shallow root mat,

it can also develop deeper structural roots, with reported vertical rooting depths of approximately 1.5–5 m depending on soil conditions (Intara et al., 2018), providing a physiological rationale for considering deeper-layer soil moisture as an informative predictor (Intara et al., 2018). ERA5 provided dew-point temperature (DT), downward short-wave radiation (DSW), evaporation (E), mean sea level pressure (MSLP), potential evaporation (PE), precipitation (PR), relative humidity (RH), soil temperature level 2 (ST), surface latent heat flux (SLHF), surface sensible heat flux (SSHF) and temperature at 2 m (T), as shown in Table 1.

Land-surface elevation at each grid point was required to calculate the vapor pressure deficit (VPD), for which the Global 30 Arc-Second Elevation (GTOPO30) dataset was applied (Rodell et al., 2004), as shown in Table 1.

Elevation values were extracted using nearest-neighbor matching to the ERA5 grid and were used to compute saturation vapor pressure at 2 m, following Yuan et al. (2019), as shown in Equations 1–4:

$$VPD = SVP - AVP \quad (1)$$

$$SVP = 6.12 \times f_w \times e^{17.67T_a / T_a + 243.5} \quad (2)$$

$$f_w = 1 + 7 \times 10^{-4} + 3.46 \times 10^{-6} P_{mst} \quad (3)$$

$$P_{mst} = P_{msl} ((T_a + 273.16) / ((T_a + 273.16) + 0.0065 \times Z))^{5.625} \quad (4)$$

where the variables used followed Yuan et al. (2019): SVP is the saturated vapor pressure; AVP is the actual vapor pressure (measured in kilopascals); f_w is the pressure adjustment factor; T_a = temperature at 2 m (measured in degrees Celsius); Z = altitude (measured in meters); P_{mst} is the adjusted pressure (measured in hectopascals) and P_{msl} is the mean sea level pressure (1013.25 hPa).

Data preparation

Monthly ERA5 data, consisting of dew-point temperature (DT) and relative humidity (RH) at 0.25° resolution were used to calculate VPD (Eq. 1–4) following Yuan et al. (2019). Provincial climate time series were created by averaging grid cells corresponding to oil palm cultivation areas, as identified in the CROPGRIDS dataset (Tang et al., 2024). A 3 mth-centered moving average was applied to reduce short-term fluctuations in both the yield and climate series (Ang et al., 2022).

Because oil palm development responds to past environmental conditions with substantial delay, lagged versions were generated of each climate variable in the range 1–36 mth. Adam et al. (2011) noted that leaf development may require up to 36 mth to complete and inflorescence processes could span up to 24 mth. Accordingly, the lag range captures the full window of potential environmental influence on yield.

Table 1 Datasets used in this study

Data type	Data Source	Parameter	Spatial resolution
Cultivated area	CROPGRIDS	Oil palm	0.05 × 0.05 (5 km)
Oil palm	OAE	Yield	Provincial
Land surface	GTOPO30	Height above sea level	0.25 × 0.25 (27.5 km)
Reanalysis	CPC	Soil moisture at 1.6 m (SM-1.6)	0.5 × 0.5 (55 km)
Reanalysis	ERA5	Dew-point temperature (DT)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Downward short-wave radiation (DSW)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Evaporation (E)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Mean sea level pressure (MSLP)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Potential evaporation (PE)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Precipitation (PR)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Relative humidity (RH)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Soil temperature level 2 (ST)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Surface latent heat flux (SLHF)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Surface sensible heat flux (SSHF)	0.25 × 0.25 (27.5 km)
Reanalysis	ERA5	Temperature at 2 m (T)	0.25 × 0.25 (27.5 km)

References for data sources: CROPGRIDS = Tang et al. (2024); OAE = Office of Agricultural Economics (2023); GTOPO30 = Rodell et al., 2004; CPC = van den Dool et al., 2003; ERA5 = Hersbach et al., 2023.

Seasonal prediction regression model (Empirical model)

Initially, the candidate predictors were reduced to avoid excessive dimensionality. Then, a stepwise variance inflation procedure (VIF; Zuur et al., 2010) was applied to remove multicollinearity. For each iteration, the variable with $VIF > 5$ was dropped and the model was refitted, repeating until all remaining predictors had $VIF \leq 5$ (Fox and Weisberg, 2019). If more than 20 predictors remained after this step, only the 20 variables with p -values < 0.01 and the highest absolute correlations with yield were retained to reduce the computational burden associated with evaluating thousands of predictor combinations. This two-step screening followed the Principle of Parsimony by retaining a compact set of non-redundant predictors, improving interpretability for regional agricultural decision-makers and reducing the risk of overfitting during the exhaustive subset search.

Next, multiple linear regression (MLR) models were built for the 13 provinces using R, a language and environment for statistical computing and graphics (R Core Team, 2024) by splitting the data into subsets of 80% training, 10% validation and 10% testing (Jamshidi et al., 2024). The training and validation phases used expanding-window (or walk-forward) cross-validation (Paudel et al., 2022). The first 80% of observations were used to train the model and to forecast the next month in the validation set. Then the training window was extended by 1 mth, retrained and the subsequent month was predicted, continuing until all validation points had been covered (Fig. 1). All combinations of predictors were evaluated (limited to 10 variables per model), with each

subset's validation root mean square error (RMSE; Equation 5) being recorded:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

where y_i is the observed value, $\hat{y}$ is the predicted value, $\bar{y}$ is the mean of observed values and n is the number of observations.

Subsequently, the subset that minimized the RMSE was selected.

Once the optimal predictor set had been identified, the final MLR was refitted using the combined training and validation data and evaluated on the test split (Fig. 1). The test period outcome was quantified based on the coefficient of determination (R^2), RMSE, the coefficient of variation of the RMSE (CV-RMSE) and mean error bias (MBE) of the best model (Equations 5–8):

$$R^2 = 1 - (\sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2) \quad (6)$$

$$CV-RMSE = (RMSE / \bar{y}) \times 100 \quad (7)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (8)$$

Where y_i is the observed value, $\hat{y}$ is the predicted value, $\bar{y}$ is the mean of observed values and n is the number of observations.

The relative importance of each predictor was assessed using the Lindeman, Merenda and Gold (LMG) method (Groemping, 2006). Partial R^2 values were calculated to determine the contribution of each climatic variable to the overall model performance.



Fig. 1 Workflow of the expanding-window cross-validation, where blue bars show successive training windows (starting with first 80% of data), orange stars mark one-step-ahead forecasts on each of the 10% validation points, green star shows final prediction on the 10% test hold-out set, RMSE = root mean square error, R^2 = coefficient of determination, MAE = mean absolute error and MBE = mean bias error

Results

Climate variables associated with oil palm yield

Before fitting the regression models, the climate-yield correlations were assessed (Figs. 2 and 3). In each of the 13 provinces, Pearson correlations were computed between monthly oil palm yield and each climate variable at lags of 1–36 mth, with only those with $p < 0.01$ being retained. DT

featured most prominently among the top predictors, followed by SM-1.6 and PR (Fig. 2). Energy-balance variables (SSHF, E and RH) were commonly retained, whereas PE and VPD appeared only sporadically (Fig. 2). The lagged correlations spanned 1–36 mth (Fig. 2). A 21 mth lag was the most prevalent, followed by a 3 mth lag. Lags of 14 mth, 15 mth, 29 mth and 34 mth occurred with similar frequency, reflecting both seasonal carry-over and longer-term effects. In contrast, lags of 13 mth and 16–19 mth were seldom observed (Fig. 3).

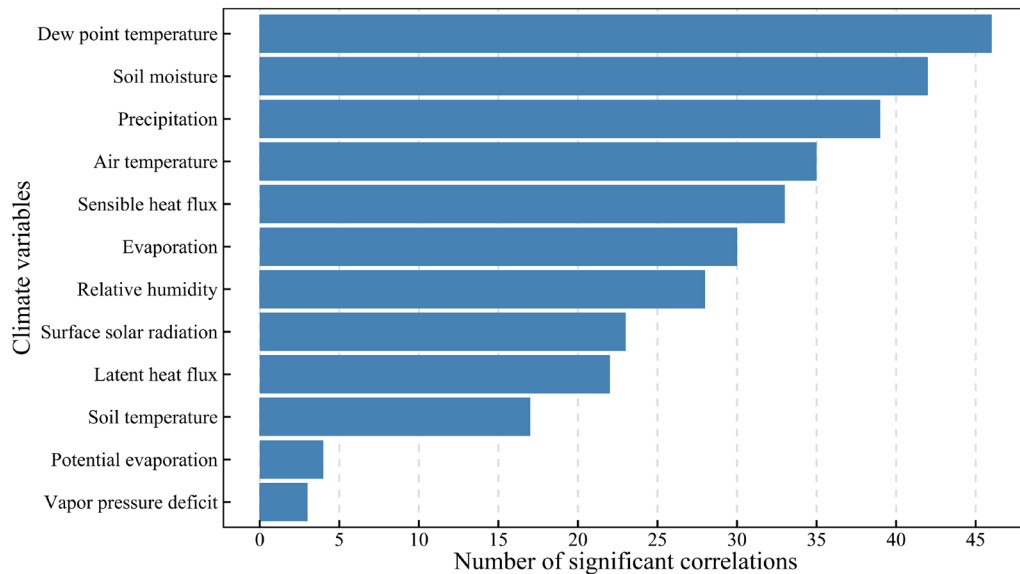


Fig. 2 Variable retention frequency across all southern provinces, showing how often each climate variable was retained across provinces after variable selection ($p < 0.01$, variance inflation factor ≤ 5). Dew-point temperature, soil moisture and precipitation were retained most frequently, whereas potential evaporation and vapor pressure deficit were retained infrequently.

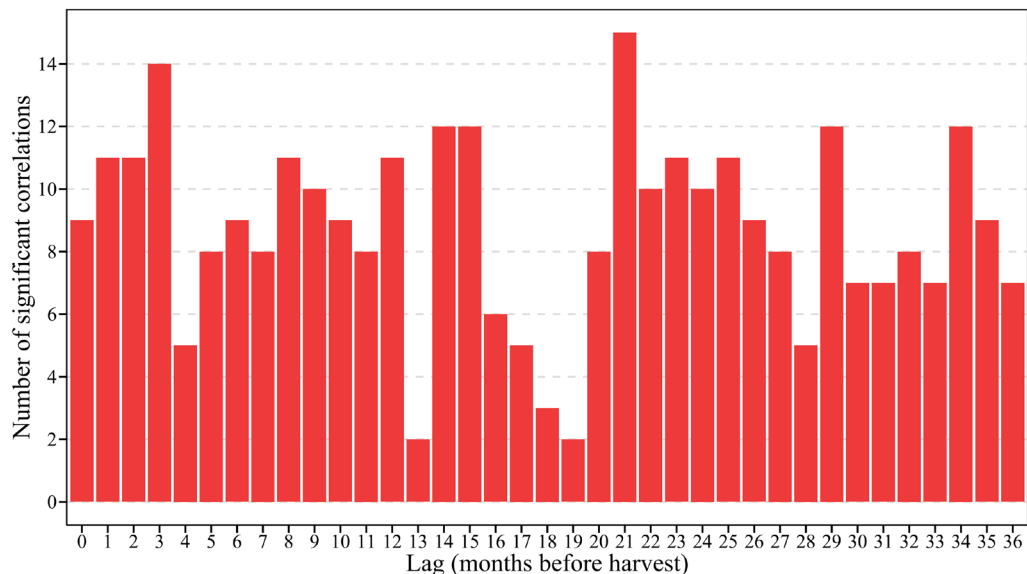


Fig. 3 Frequency of significant lags across southern provinces, with peaks at 21 mth, indicating long-term developmental effects, and 3 mth indicating short-term pre-harvest effects.

Development of models

The empirical models contained between two and eight predictors (Table 2), ranging from a minimum in Satun province (two) to a maximum in Phatthalung (eight). Models for Narathiwat and Phuket were excluded because of unreliable results, leaving 11 empirical models for evaluation. Most provinces contained 5–8 predictors, except for Satun, Ranong and Songkhla, which had only 2–3. Some provinces had long-term climatic influences on inflorescence development,

reflected in their extended lag values (notably Songkhla and Yala). Commonly selected variables included SM-1.6, DT and PR, though each province featured a distinct lag of highest relative importance according to the LMG test. For example, Chumphon's six-predictor model identified SM-1.6 with a 7 mth lag as the most influential, whereas in Nakhon Si Thammarat, DT at an 11 mth lag ranked the highest. Considering the top-ranked LMG contribution regardless of lag across the 11 robust models, SM-1.6 was the dominant predictor in 7 provinces, DT in 3 provinces and PR in 1 province, (Table 2, Fig. 4).

Table 2 Empirical models by province including number of predictors, formula and predictor with highest Lindeman-Merenda-Gold (LMG) relative importance

Province	Number of predictors	Formula	Highest LMG	Forecast lead time (mth)
Chumphon	6	$SM_{-7} + DSW_{-10} + VPD_{-10} + PR_{-11} + DT_{-12} + DT_{-23}$	SM_{-7}	7
Krabi	7	$SM_{-7} + SM_{-8} + SM_{-9} + RH_{-9} + T_{-9} + PR_{-11} + SM_{-20}$	SM_{-7}	7
Nakhon Si Thammarat	6	$SM_{-7} + ST_{-7} + SSHF_{-8} + DT_{-11} + DT_{-12} + DT_{-24}$	DT_{-11}	7
Phangnga	5	$DSW_{-8} + SM_{-9} + SM_{-21} + SSHF_{-21} + DT_{-23}$	SM_{-9}	8
Phatthalung	8	$VPD_{-6} + ST_{-6} + SM_{-7} + RH_{-7} + DT_{-12} + VPD_{-13} + SM_{-14} + DT_{-24}$	SM_{-7}	6
Ranong	3	$DT_{-11} + SM_{-21} + E_{-23}$	DT_{-11}	11
Satun	2	$SM_{-7} + DT_{-23}$	DT_{-23}	7
Songkhla	5	$PR_{-7} + DT_{-12} + T_{-13} + ST_{-26} + T_{-36}$	PR_{-7}	7
Surat Thani	5	$SM_{-6} + T_{-8} + DSW_{-8} + PR_{-10} + SSHF_{-14}$	SM_{-6}	6
Trang	6	$PR_{-3} + DT_{-4} + PR_{-9} + DT_{-15} + SM_{-20} + DT_{-22}$	SM_{-20}	3
Yala	3	$ST_{-20} + SSHF_{-21} + SM_{-36}$	SM_{-36}	20

DSW = downward shortwave radiation; DT = dew-point temperature; E = evaporation; PR = precipitation; RH = relative humidity; SM = soil moisture at 1.6 m depth; SSHF = surface sensible heat flux; ST = soil temperature level 2; T = temperature at 2 m; VPD = vapor pressure deficit.

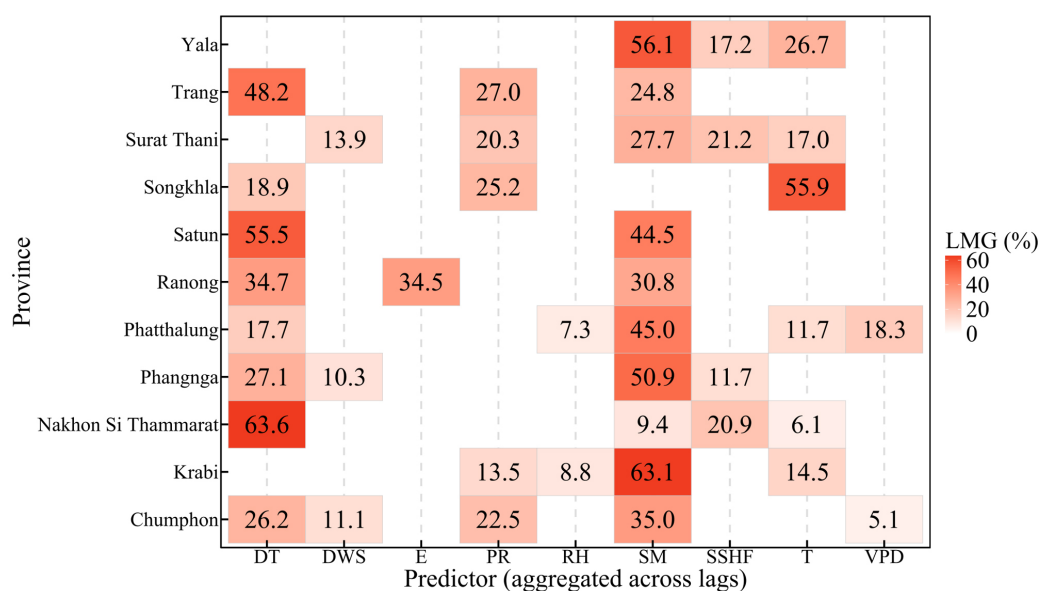


Fig. 4 Lindeman–Merenda–Gold (LMG) heatmap showing the relative importance of climate predictors in the final multiple linear regression models for each province. Values represent the aggregated LMG contribution of each predictor family across all selected lagged terms within a province (contributions from multiple lags of same variable are summed). DSW = downward shortwave radiation; DT = dew-point temperature; E = evaporation; PR = precipitation; RH = relative humidity; SM = soil moisture at 1.6 m depth; SSHF = surface sensible heat flux; T = temperature at 2 m; VPD = vapor pressure deficit.

Performance of models

Both empirical and dynamic models were evaluated using a 10% hold-out test split across 11 provinces (Table 3, Fig. 5A). Among the empirical models, R^2 ranged from 0.33 (Yala) to 0.73 (Surat Thani), RMSE from 0.14 to 0.30 t/ha and CV-RMSE from 8.38% (Nakhon Si Thammarat) to 23.75% (Yala), with the MBE never exceeding ± 0.130 t/ha. The CV-RMSE varied markedly across southern Thailand,

with the western provinces (Ranong and Satun) clustered near 15% CV-RMSE, while Nakhon Si Thammarat and Surat Thani in the central zone had the lowest errors. In contrast, higher CV-RMSE values (exceeding 20%) were observed in the southernmost provinces, notably Songkhla and Yala. Most provinces had an intermediate forecast lead time of 6–11 mth, with Trang having the shortest (3 mth) and Yala the longest (20 mth), as shown in Table 2 and Fig. 5B.

Table 3 Test-split performance of empirical models by province.

Province	Number of predictors in each provincial model	R^2	RMSE (t/ha)	CV-RMSE (%)	MBE (t/ha)
Chumphon	6	0.49	0.22	13.74	0
Krabi	7	0.4	0.23	13.03	-0.03
Nakhon Si Thammarat	6	0.67	0.14	8.38	-0.06
Phangnga	5	0.35	0.25	15.14	-0.13
Phatthalung	8	0.65	0.2	14.23	-0.03
Ranong	3	0.41	0.25	15.43	-0.07
Satun	2	0.67	0.16	12.32	0.01
Songkhla	5	0.5	0.3	21.33	-0.06
Surat Thani	5	0.73	0.17	9.95	-0.06
Trang	6	0.68	0.21	13.16	-0.08
Yala	3	0.33	0.2	23.75	-0.01

R^2 = coefficient of determination showing proportion of explained variance (values closer to 1 indicate better model performance); RMSE = average magnitude of prediction error; CV-RMSE = RMSE relative to mean observed yield; MBE = over- or under-estimation due to model bias.

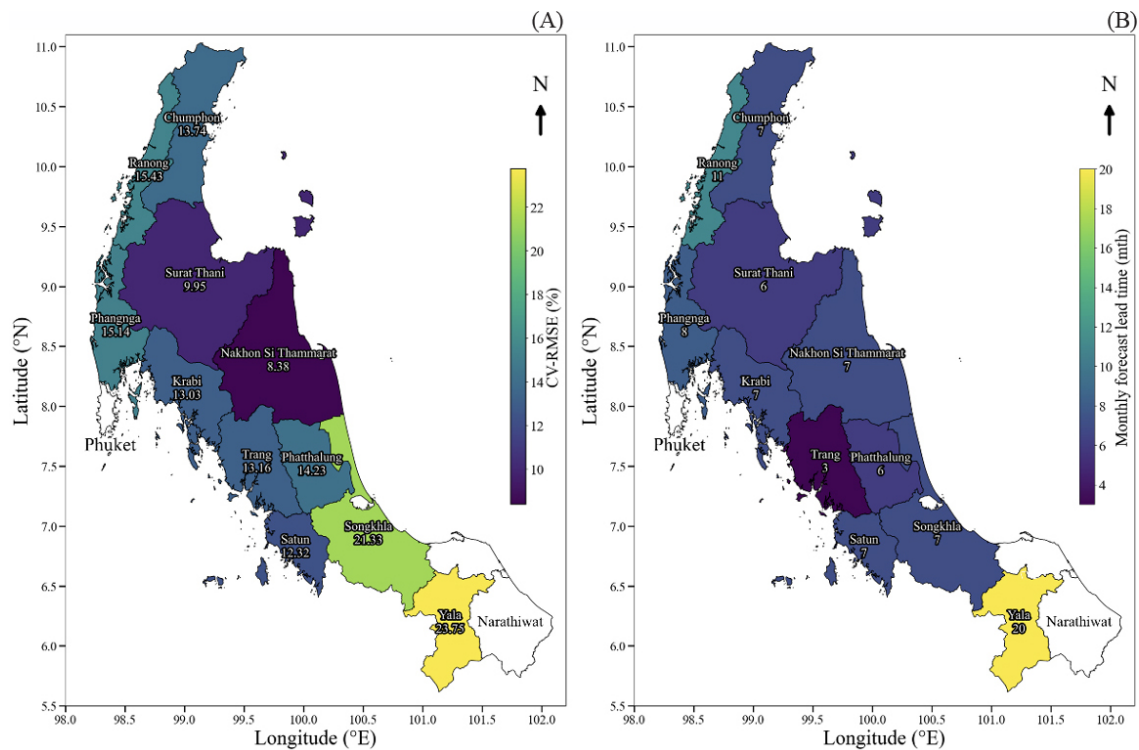


Fig. 5 Spatial distribution from 11 empirical models excluding Phuket and Narathiwat because of unreliable results, showing: (A) the test-split coefficient of variation of root mean square error (CV-RMSE), with the lowest errors observed in Nakhon Si Thammarat and Surat Thani; and (B) seasonal oil palm yield forecast lead time, indicating moderate predictability (6–11 mth) for most provinces.

Discussion

Oil palm yield response to climatic variables

The oil palm yield consists of a complex interplay of moisture, temperature and the crop's developmental stages, manifested through distinct lagged sensitivities over 1–36 mth. Fig. 6 maps these lag windows to key stages of oil palm development: short lags of 3–4 mth align with fruit filling and maturation, lags of approximately 21 mth correspond to reproductive initiation and sex determination and longer lags of 29–36 mth reflect earlier phases of reproductive development occurring roughly 3 yr before harvest. The strongest signals occur at around 21 mth, coinciding with inflorescence primordia formation 2 yr before harvest and at short lags of 3–4 mth, when fruit ripening and final bunch filling depend directly on water availability and thermal conditions for oil accumulation and expansion, consistent with the 4.5–6 mth from pollination to bunch maturity in oil palm (Eyaa Ngombo et al., 2023). Intermediate lags of approximately 8–15 mth capture flower development stages (Fig. 6), consistent with experimental evidence showing suppression of male inflorescence abortion at around 16 mth before bunch maturity under sink-limitation treatments (Legros et al., 2009). Very long lags of 29–36 mth have been linked to initial inflorescence development (Fig. 6; Adam et al. (2011)) and appeared only in a subset of provinces, likely reflecting local microclimatic conditions or differences in management practices (Chow, 1992).

SM-1.6 was a predictor in 10 of the 11 empirical models and also emerged as the most common highest-LMG predictor in 7 of the 11 models (Table 2, Fig. 4). This pattern is biologically plausible because oil palm can develop structural roots that

extend well below the surface in well-drained soils, allowing subsurface moisture to buffer reproductive and vegetative processes during dry periods (Tuan Syaripah et al., 2019); however, effective rooting depth is not uniform and may be restricted where soils are compacted, poorly drained, or seasonally saturated by a shallow water table. Accordingly, SM-1.6 should not be interpreted as direct extraction from exactly 1.6 m at all sites, but rather as an integrative proxy for deeper profile water availability that sustains canopy expansion and reproductive allocation (Tuan Syaripah et al., 2019) and, when deficient, contributes to inflorescence abortion or reduced frond emergence that limits yield potential (Delgado et al., 2024). Machine learning-based yield models likewise identify soil moisture among the top predictors (Khan et al., 2022). This finding highlights the value of reanalysis data, because soil moisture ground stations are largely absent.

Atmospheric moisture and thermal regimes further modulate oil palm yield responses. DT, as a proxy for VPD, influences stomatal conductance and carbon assimilation, since high VPD causes stomatal closure that limits CO₂ uptake for both inflorescence development and fruit filling (Smith, 1989; Dufrene and Saugier, 1993). Temperature interacts with moisture to shape oil palm physiology at multiple scales. Meta-analyses indicate optimum yields occur when mean minimum temperatures are in the range 22–24°C and mean maximums in the range 29–33°C —conditions that favor inflorescence initiation and oil biosynthesis (Corley and Tinker, 2015). During drought, depressed soil moisture can raise VPD by approximately 0.8–2 kPa, which can halve canopy-level photosynthesis (Rötter et al., 2018; Raza et al., 2019; Lesk et al., 2021), underscoring oil palm's sensitivity to water stress (Corley and Tinker, 2015; Oettli et al., 2018; Chiarawipa et al., 2020; Sarkar et al., 2020).

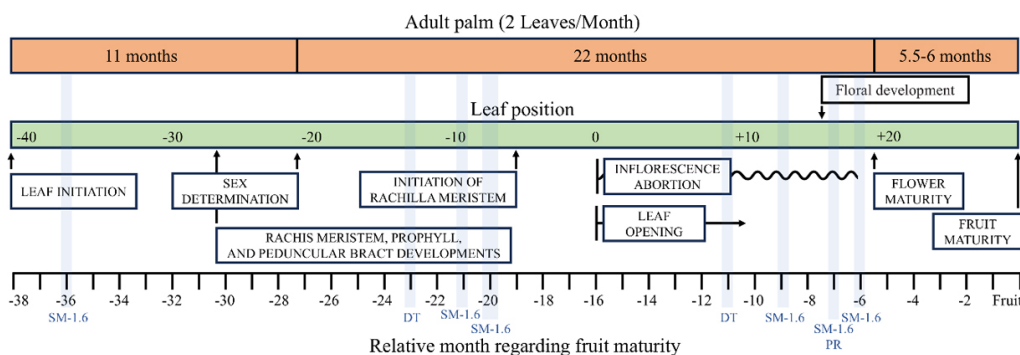


Fig. 6 Oil palm flower development timeline adapted from Adam et al. (2011) with some modification for clarity, illustrating key developmental stages of oil palm and their corresponding timeframes leading up to fruit maturity, leaf position 0 is non-fully opened leaf emerging from the crown (spear leaf), PR = precipitation; DT = dew-point temperature; SM-1.6 = soil moisture at 1.6 m.

Effectiveness of prediction model

Using ERA5 climate variables and CPC soil moisture (SM-1.6), the empirical models achieved R^2 values of 0.50–0.73 in 6 of 11 provinces, consistent with Malaysian studies that explained 60–75% of interannual yield variance (Oettli et al., 2018; Khan et al., 2024), indicating the feasibility of climate-based yield prediction. The partial R^2 analyses indicated that lagged SM-1.6, DT and PR accounted for the majority of predictive power. The expanding-window (walk-forward) cross-validation applied here preserved temporal ordering by training models only on past observations and evaluating performance on subsequent months. This provides a more realistic assessment of operational forecasting skill than random K-fold cross-validation, which can yield biased estimates when applied to temporally dependent climate-yield time series.

The prominence of soil moisture aligned with a machine-learning study in Pahang, Malaysia, where an AdaBoost model ranked soil moisture as the single most influential predictor, surpassing rainfall and temperature (Khan et al., 2022). Observed yield variability across the southern provinces is consistent with prior evidence that interannual oil palm yields are substantially associated with temperature and rainfall, with the dominant climatic control varying by province (Chiarawipa et al., 2020). DT plays a critical role both by directly regulating atmospheric moisture availability and by indirectly modulating evapotranspiration rates that govern plant water use (Oettli et al., 2018). Rising temperatures influence rapid yield recoveries under high humidity, while cooler periods impede recovery and excessive rainfall beyond optimal thresholds depresses yields, likely due to waterlogging, demonstrating that warming and heavy rainfall affect oil-palm output through distinct mechanisms (Go et al., 2024).

Limitations

Despite predictive strength, the empirical models' full potential remains constrained by several factors. First, oil-palm yields respond asymmetrically to temperature spikes, rainfall deficits and extreme wet spells. Estate-level studies in Malaysia and Indonesia have revealed inverted-U temperature relationships and precipitation thresholds that a fixed-lag MLR cannot capture (Roy et al., 2023; Kamil et al., 2024). Omitting these nonlinear dynamics may understate potential production losses during compound hot-dry events and waterlogging periods. Beyond nonlinear mean-state responses,

yield variability may also be sensitive to extreme-event thresholds that are not well represented by monthly averages, such as heatwave episodes, short-duration droughts or intense rainfall events that increase waterlogging risk (Limsakul and Singhruck, 2016; Rivanzha et al., 2025). Studies in southern Thailand have demonstrated that oil palm yields were sensitive to variability in rainfall amount and distribution, with interannual fluctuations strongly influenced by ENSO-related climate anomalies (Chiarawipa et al., 2020). Such threshold-type shocks can cause disproportionate yield losses relative to changes in seasonal means; however, their impacts are likely smoothed when predictors are provincially averaged at monthly resolution and models assume fixed-lag linear effects. A more explicit climate-risk assessment would require event-based indices (such as heatwave days, consecutive dry days, extreme rainfall totals) and nonlinear modeling frameworks, which should be an important direction for future work.

Second, the absence of plantation management data (such as fertilizer application, pruning cycles, replanting intervals) introduces substantial unexplained variance at the provincial scale (Rhebergen et al., 2020; Thoumazeau et al., 2024). Oil palm production in Thailand is dominated by smallholder systems with highly heterogeneous and poorly documented practices, while spatially explicit, temporally consistent provincial-scale proxies for management intensity are not currently available. Field experiments in Ghana and Indonesia have demonstrated substantial yield gains under optimized fertilizer and residue management, highlighting management intensity as a primary driver of smallholder yield gaps (Rhebergen et al., 2020; Gutierrez Al-Khudhairy et al., 2023). This residual variance may partly explain why Songkhla and Yala had higher prediction errors than the other provinces. Based on currently available regional-scale datasets, edaphic type and groundwater table patterns do not appear to show clearly distinct separation across the oil palm cultivated areas of southern Thailand (de Graaf et al., 2015; Hengl et al., 2017; Poggio et al., 2021; Tang et al., 2024). However, formal statistical testing of these mechanisms is difficult because the yield records used here were aggregated from large, cultivated areas into a single provincial observation per month. Notably, Songkhla and Yala tended to have a more dominant older plantation age profile and more spatially dispersed oil palm farmlands, whereas other provinces tended to have more clustered oil palm areas (Descals et al., 2024). Such spatial dispersion may reduce model performance when heterogeneous production environments are combined into a single provincial time series.

Third, remotely sensed indicators of plant health offer untapped predictive power. The normalized difference vegetation index and related canopy-water indices have proven to be effective early predictors of oil palm yield (Ang et al., 2022; Watson-Hernández et al., 2022; Jamshidi et al., 2024). Likewise, incorporating soil-profile texture and moisture data can enhance yield forecasts while promoting sustainable land management (Vhatkar et al., 2025). In addition, the lack of spatially explicit, temporally consistent management records (such as fertilization, pruning, harvest intensity) limits the feature space available for nonlinear modeling at the provincial scale and therefore constrains the performance of machine-learning approaches.

Notably, the MLR models developed in the current study are intended as an interpretable baseline and benchmark rather than an upper bound on predictive skill. The initial trials with machine-learning algorithms, trained solely on climatic variables and yield records, failed to outperform the simpler regression models (data not shown), likely reflecting a narrow feature set that limited the algorithms' ability to learn complex, nonlinear interactions. Expanding the predictor pool to include soil properties, satellite-derived plant-health indices and detailed management records could supply machine-learning methods with the multivariate inputs required to surpass the predictive skill of linear regression.

Practical implications and future work

The developed empirical models offer practical benefits: 1) provincial agriculture offices can issue seasonal forecasts 3–20 mth ahead, with 6–11 mth common before harvest; 2) mill operators can anticipate swings in fresh fruit bunch supply and optimize storage, shipping schedules and contract positions well in advance, reducing demurrage costs; and 3) traders and refiners gain an objective, climate-linked view of forthcoming supply conditions, enabling more informed hedging and spot-purchase decisions.

Looking ahead, enhancing the model with advanced machine learning approaches and integrating seasonal climate forecast datasets represent clear next steps to extend the forecast lead time and improve accuracy. These models can be trained using the same walk-forward cross-validation framework to maintain temporal realism while capturing complex, nonlinear climate-yield relationships. However, for operational deployment, bias correction is a prerequisite rather than an optional refinement, because the statistical distribution of forecast products such as CFSv2 often differs

from the reanalysis data used to train the current models. Consequently, forecast variables must be harmonized to the reanalysis baseline before inference, followed by lead-time alignment and ongoing out-of-sample monitoring under real-time conditions. Once bias-corrected and operationally evaluated, seasonal forecasts can be coupled with the reanalysis-trained framework to create a dynamic system that updates monthly and moves from retrospective analysis toward a fully operational early-warning platform for oil palm yield in southern Thailand.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Acknowledgements

Financial support was provided by the Development and Promotion of Science and Technology Talents Project (DPST). The Office of Agricultural Economics, Thailand provided comprehensive Thai oil palm datasets. Prof Jin-Ho Yoon provided comments and suggestions. This manuscript forms part of the first author's PhD thesis at Kasetsart University, Bangkok, Thailand.

References

- Adam, H., Collin, M., Richaud, F., et al. 2011. Environmental regulation of sex determination in oil palm: Current knowledge and insights from other species. *Ann. Bot.* 108: 1529–1537. doi.org/10.1093/aob/mcr151
- Ang, Y., Shafri, H.Z. M., Lee, Y.P., et al. 2022. A novel ensemble machine learning and time series approach for oil palm yield prediction using Landsat time series imagery based on NDVI. *Geocarto Int.* 37: 9865–9896. doi.org/10.1080/10106049.2022.2025920
- Ang, Y., Shafri, H.Z.M., Lee, Y.P., et al. 2022. Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. *Earth Sci. Inform.* 15: 2349–2367. doi.org/10.1007/s12145-022-00882-9
- Chiarawipa, R., Thongna, K., Sdoodee, S. 2020. Assessing impact of weather variability and changing climate on oil-palm yield in major growing regions of southern Thailand. *J. Agrometeorol.* 22: 274–284.
- Chow, C.S. 1992. The effects of season, rainfall and cycle on oil palm yield in Malaysia. *J. Oil Palm Res.* 4: 32–43.
- Corley, R.H.V., Tinker, P.B.H. 2015. *The Oil Palm*, 5th ed. John Wiley & Sons. Oxford, UK. doi.org/10.1002/9781118953297.
- de Graaf, I.E.M., Sutanudjaja, E.H., van Beek, L.P.H., Bierkens, M.F.P. 2015. A high-resolution global-scale groundwater model. *Hydrol. Earth Syst. Sci.* 19: 823–837. doi.org/10.5194/hess-19-823-2015

- Delgado, T., Ladino, G., Arias, N. 2024. Evaluation of the effect of soil water conditions on the development and water requirements of adult oil palm (*Elaeis guineensis* Jacq.) in the northern region of Colombia. *Agron.* 14: 1976.
- Descals, A., Gaveau, D.L.A., Wich, S., Szantoi, Z., Meijaard, E. 2024. Global mapping of oil palm planting year from 1990 to 2021. *Earth Syst. Sci. Data.* 16: 5111–5129. doi.org/10.5194/essd-16-5111-2024
- Dufrene, E., Saugier, B. 1993. Gas exchange of oil palm in relation to light, vapour pressure deficit, temperature and leaf age. *Funct. Ecol.* 7: 97–104. doi.org/10.2307/2389872
- Eyaa Ngombo, C., Nsimi, M., Ndele, P., et al. 2023. A review of main factors involved in the maturation of oil palm (*Elaeis guineensis* Jacq.) fruit bunches. *Am. J. Plant Sci.* 14: 727–740. doi.org/10.4236/ajps.2023.147049
- FAOSTAT. 2023. Crops and livestock products. <https://www.fao.org/faostat/en/#data/QCL>, 22 September 2023.
- Fox, J., Weisberg, S. 2019. *An R Companion to Applied Regression*. Sage, Thousand Oaks, CA, USA.
- Go, Y.H., Tan, Y.L., Yiew, T.H. 2024. Sensitivity of oil palm yield in Indonesia to climate change: Evidence from threshold cointegration models. *Environ. Dev. Sustain.* doi.org/10.1007/s10668-024-05635-w.
- Groemping, U. 2006. Relative importance for linear regression in R: The package relaimpo. *J. Stat. Softw.* 17: 1–27.
- Gutierrez Al-Khudhairi, S., Howells, T.R., Azlin, B.S., McClean, C.J., Senior, M.J.M., Azmi, R., Benedick, S., Hill, J.K. 2023. Sustainable management practices do not reduce oil palm yields on smallholder farms on Borneo. *Agroecol. Sustain. Food Syst.* 47: 3–24. doi.org/10.1080/21683565.2022.2131691
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., et al. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One.* 12: e0169748. doi.org/10.1371/journal.pone.0169748.
- Hersbach, H., Bell, B., Berrisford, P., et al. 2023. ERA5 monthly averaged data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means?tab=overview>, 10 June 2024.
- Intara, Y.I., Nusantara, A.D., Supanjani, S., Caniago, Z., Ekawita, R. 2018. Oil palm roots architecture in response to soil humidity. *IJOP.* 1: 79–89.
- Jamshidi, E.J., Yusup, Y., Hooy, C.W., et al. 2024. Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models. *Ecol. Inform.* 81: 102595. doi.org/10.1016/j.ecoinf.2024.102595
- Kamil, N.N., Omar, S.F. 2017. The impact of El Niño and La Niña on Malaysian palm oil industry. *OPB.* 7: 1–6.
- Kamil, N.N., Xiao, S., Syed Salleh, S.N., Xu, H., Zhuang, C.C. 2024. Nonlinear impacts of climate anomalies on oil palm productivity. *Heliyon* 10: e35798. doi.org/10.1016/j.heliyon.2024.e35798
- Khan, N., Kamaruddin, M.A., Sheikh, U.U., Bakht, M.P., Mohd, M.N.H., Ab Rahman, A.A.H. 2024. Analysis of weather impacts on oil palm productivity. *AIPConf. Proc.* 3245: 060003. doi.org/10.1063/5.0232251.
- Khan, N., Kamaruddin, M.A., Ullah Sheikh, U., Zawawi, M.H., Yusup, Y., Bakht, M.P., Mohamed Noor, N. 2022. Prediction of oil palm yield using machine learning in the perspective of fluctuating weather and soil moisture conditions: Evaluation of a generic workflow. *Plants* 11: 1697.
- Kuinkel, D., Promchote, P., Upreti, K.R., Wang, S.Y.S., Dahal, N., Pokharel, B. 2024. Projected changes in precipitation extremes in Southern Thailand using CMIP6 models. *Theor. Appl. Climatol.* 155: 8703–8716. doi.org/10.1007/s00704-024-05150-y
- Legros, S., Mialet-Serra, I., Caliman, J.P., Siregar, F.A., Clément-Vidal, A., Dingkuhn, M. 2009. Phenology and growth adjustments of oil palm (*Elaeis guineensis*) to photoperiod and climate variability. *Ann. Bot.* 104: 1171–1182. doi.org/10.1093/aob/mcp214
- Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S.I., Horton, R. 2021. Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields. *Nat. Food.* 2: 683–691. doi.org/10.1038/s43016-021-00341-6
- Limsakul, A., Singhruck, P. 2016. Long-term trends and variability of total and extreme precipitation in Thailand. *Atmos. Res.* 169: 301–317. doi.org/10.1016/j.atmosres.2015.10.015
- Murphy, D. 2014. The future of oil palm as a major global crop: Opportunities and challenges. *J. Oil Palm Res.* 26: 1–24.
- Office of Agricultural Economics. 2023. *Agricultural Statistics of Thailand 2023*. Office of Agricultural Economics, Bangkok, Thailand.
- Oettli, P., Behera, S.K., Yamagata, T. 2018. Climate-based predictability of oil palm tree yield in Malaysia. *Sci. Rep.* 8: 2271. doi.org/10.1038/s41598-018-20298-0
- Paudel, D., Boogaard, H., de Wit, A., et al. 2022. Machine learning for regional crop yield forecasting in Europe. *Field Crops Res.* 276: 108377. doi.org/10.1016/j.fcr.2021.108377
- Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL.* 7: 217–240. doi.org/10.5194/soil-7-217-2021
- Promchote, P., Pokharel, B., Deng, L., Wang, S.Y.S., Yoon, J.H., Kittipadakul, P. 2023. Boosting Thailand's palm oil yield with advanced seasonal predictions. *Environ. Res. Lett.* 18: 071004. doi.org/10.1088/1748-9326/ace06c
- R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rahman, A., Abdullah, R., Nambiappan, B., Shariff, F. 2013. The Impact of La Niña and El Niño Events on Crude Palm Oil Prices: An Econometric Analysis. *OPIEJ.* 13: 38–51.
- Raza, A., Razaq, A., Mehmood, S.S., Zou, X., Zhang, X., Lv, Y., Xu, J. 2019. Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. *Plants* 8: 34.
- Rhebergen, T., Zingore, S., Giller, K.E., et al. 2020. Closing yield gaps in oil palm production systems in Ghana through best management practices. *Eur. J. Agron.* 115: 126011. doi.org/10.1016/j.eja.2020.126011
- Rivanzha, A.T., Klongvessa, P., Wetchayont, P. 2025. Spatial variability of extreme rainfall trends in southern Thailand. In: *9th International Conference on Energy and Environmental Science*. Cham, Switzerland, pp. 307–321.
- Rodell, M., Houser, P.R., Jambor, U., et al. 2004. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* 85: 381–394. doi.org/10.1175/BAMS-85-3-381
- Rötter, R.P., Appiah, M., Fichtler, E., Kersebaum, K.C., Trnka, M., Hoffmann, M.P. 2018. Linking modelling and experimentation to better capture crop impacts of agroclimatic extremes—A review. *Field Crops Res.* 221: 142–156. doi.org/10.1016/j.fcr.2018.02.023

- Roy, W.L.K., Chuen-Khee, P., Hui-Boon, T. 2023. Non-linear effect of temperature on oil palm yield: A note based on panel threshold analysis. *Int. J. Bus. Soc.* 24: 1359–1367. doi.org/10.33736/ijbs.6425.2023
- Sarkar, M.S.K., Begum, R.A., Pereira, J.J. 2020. Impacts of climate change on oil palm production in Malaysia. *Environ. Sci. Pollut. Res.* 27: 9760–9770. doi.org/10.1007/s11356-020-07601-1
- Smith, B.G. 1989. The effects of soil water and atmospheric vapour pressure deficit on stomatal behaviour and photosynthesis in the oil palm. *J. Exp. Bot.* 40: 647–651. doi.org/10.1093/jxb/40.6.647
- Tang, F.H.M., Nguyen, T.H., Conchedda, G., Casse, L., Tubiello, F.N., Maggi, F. 2024. CROPGRIDS: A global geo-referenced dataset of 173 crops. *Sci. Data.* 11: 413. doi.org/10.1038/s41597-024-03247-7.
- Thoumazeau, A., Mettauer, R., Turinah, Junedi, H., Baron, V., Chéron-Bessou, C., Ollivier, J. 2024. Effects of fertilization practices and understory on soil health and oil palm performances in smallholdings: An Indonesian case study. *Agric. Syst.* 213: 103802. doi.org/10.1016/j.agsy.2023.103802
- Tuan Syaripah, N., Mohd Hafiz, I., A'Fifah Abd, R., Rosimah, N., Puteri Edaroyati Megat, W. 2019. Effects of water stress on the growth, physiology and biochemical properties of oil palm seedlings. *AIMS Agric. Food* 4: 854–868. doi.org/10.3934/agrfood.2019.4.854
- van den Dool, H., Huang, J., Fan, Y. 2003. Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *J. Geophys. Res.* 108: 8617. doi.org/10.1029/2002JD003114
- Vhatkar, K.N., Koparde, S.A., Kothari, S., Sarwade, J., Sakur, K. 2025. Enhancing prediction of crop yield and soil health assessment for sustainable agriculture using machine learning approach. *MethodsX* 14: 103418. doi.org/10.1016/j.mex.2025.103418
- Watson-Hernández, F., Gómez-Calderón, N., da Silva, R.P. 2022. Oil palm yield estimation based on vegetation and humidity indices generated from satellite images and machine learning techniques. *AgriEngineering* 4: 279–291.
- Yuan, W., Zheng, Y., Piao, S., et al. 2019. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* 5: eaax1396. doi.org/10.1126/sciadv.aax1396
- Zuur, A.F., Ieno, E.N., Elphick, C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* 1: 3–14. doi.org/10.1111/j.2041-210X.2009.00001.x