Journal of Applied Automation Technologies

Research Paper

Long-Sequence Informer Modeling for Crop Yield Prediction from Multi-Source Farmland Sensors

Natálie Králová[1] and Kateřina Svobodová[1, *]

[1] Department of Computer Science and Engineering, University of West Bohemia in Pilsen, 301 00 Pilsen, Czech Republic
*Corresponding author: katerina.s@kiv.zcu.cz

Abstract. Obstacles to crop yield prediction from multiple sources of farmland sensors include asynchronous sampling, long-term seasonal variations, sensor failures and significant field-to-field differences. Develop a quality-aware decomposed Informer in this study that aligns the weather, soil, canopy and management streams on a daily basis and keeps observation age and reliability as learnable evidence. A seasonal-residual decomposition separates the slow-varying phenological context from the event-driven fluctuation, and a stage-conditioned probabilistic sparse attention module selects informative historical states without quadratic cost. A total of 18,720 field-season sequences were divided into 64 production areas for testing maize, wheat and soybean. Compared with the strongest recurrent and Transformer baselines, the proposed model reduced the root-mean-square error to 0.438 t ha^{-1} and mean absolute percentage error to 6.1%, and shortened inference time by 41.8% compared to a full-attention Transformer. Under 30% missingness, the error increase was limited to 11.4%, and cross-zone testing maintained an R^2 of 0.827. Calibration reduced the interval undercoverage from 15.8% to 6.4% by 90%. According to the above experiments, explicit sensor-quality encoding and decomposed sparse attention jointly enhance the accuracy, robustness and computational efficiency of the model. The above framework provides an engineering path for all-season forecasting of heterogeneous precision agriculture networks.

Keywords: *Multisource Sensor Fusion, Crop Yield Prediction, Long-Sequence Modeling, Probabilistic Sparse Attention, Temporal Decomposition*

Received on 28 December 2023, Accepted on 25 June 2024, Published on 06 July 2024

Introduction

Crop yield forecasts can help plan procurement, irrigation, insurance and regional food security for the government; however, farm-level data are unavailable before harvest. Weather stations report atmospheric forcing but omit root-zone heterogeneity [1]. Soil probes show the change in moisture and temperature over a small support volume [2]. Canopy sensors respond to crop conditions but are sensitive to viewing angles and light conditions [3]. Machinery logs record interventions at irregular times and have inconsistent semantics [4]. Satellite observations have expanded the area covered, but there are still cloud obstructions and revisit time gaps [5]. These streams are complementary, but their clocks, scales and error mechanisms rarely agree [6].

Traditional statistical models often condense a season into manually aggregated indicators and thus lose the order and duration of stress events [7]. Recurrent networks keep the order of sequences but may lose information from the beginning of a long observation window [8]. Convolutional temporal models are good at extracting local motifs, but to connect planting conditions with harvest results effectively, their effective receptive fields need to be relatively large [9]. Full-attention Transformers offer direct temporal relationships but have a quadratic increase in memory and computation with the accumulation of daily, hourly and event data [10]. Sparse-attention architectures are more feasible for long sequences [11]. However, the performance will drop if missing values, old measurements and different sensor reliability are not considered [12].

https://doi.org/10.64972/jaat.2024v2.278p23e:320-334

Research on the connection between long-sequence models and the quality structure of operational farmland data. We propose a quality-aware decomposed Informer for crop-yield prediction from weather, soil, canopy and management streams. Align the sources without deleting the acquisition time, separate seasonal variations from sharp deviations, and filter out sparse attention based on crop stage. A calibrated probabilistic head is a point estimate and a prediction interval. The engineering contribution is a unified pipeline that links data reliability with efficient temporal selection, so the model remains accurate in the presence of block missingness and is transferable to production zones.

Related Work

Multi-Source Agricultural Sensing and Yield Modeling

Field-scale yield models now incorporate meteorological series, soil observations, remote-sensing indices and management records more extensively. Multi-source fusion can increase the coverage area by addressing the deficiencies of either atmosphere or root-zone measurement instruments alone [13]. Process-based models are relatively complex, and the necessary parameters for calibration are not readily available in commercial farming [14]. Machine-learning regressors can learn non-linear relations among engineered seasonal descriptors [15]. Deep networks are relatively independent of the manual thresholds and accumulated indices [16]. However, naive concatenation assumes that all measurements are equally recent and reliable [17]. Quality flags, observation age and source availability should instead be included in the predictive representation as explicit evidence [18].

There is also the problem of different support areas for satellite pixels, point probes and harvester swaths. Resample to a common grid for ease of computation, but this may introduce artificial precision near the border. Therefore, a practical field model needs a source-specific encoder before fusion and must keep masks to distinguish between an observed zero and a missing value. This requirement prompts the introduction of the aligned and reliable design later.

Long-Sequence Forecasting for Agricultural Systems

Recurrent neural networks and their gated variations are still widely used for crop and environmental time series because they can model temporal dependencies in a compact form [19]. Temporal Convolutional Networks improve parallelism and can learn multi-scale local patterns [20]. Transformer models do not have recurrence and connect distant observations directly [21]. However, their quadratic attention cost will be relatively large if there are many fine-resolution records from multiple sources in a season [22]. Informer reduces this cost by using probabilistic sparse attention and sequence distillation [23]. Later, long-horizon architectures have been shown to be more stable in forecasts through decomposition and frequency-aware processing [24].

The relative weights of different times in the event schedule for agricultural forecasts vary over time due to phenological variations. A moisture deficiency during flowering is likely to be more serious than the same magnitude of deficiency at maturity. Existing long-sequence models have rarely combined crop-stage conditioning, sensor-quality tokens and uncertainty calibration in a single deployable pipeline. The above is the arrangement of the missing two.

Multi-Source Sequence Construction and Problem Formulation

Multi-Source Sequence Construction

Each field-season is shown as weather, soil, canopy and management streams. Weather includes air temperature, precipitation, water vapour pressure deficit, solar radiation, wind speed and humidity. Soil stations provide volumetric water content and temperature at three depths, electrical conductivity and matric potential. Canopy observations can contain normalised vegetation index, enhanced vegetation index, leaf-area proxy, infrared canopy temperature and chlorophyll proxy. Management events record sowing density, cultivar group, irrigation, fertilisation and pesticide application. All values are mapped to a daily axis of length T, and high-frequency measurements are summarized by robust location, spread, extrema, and exposure duration instead of a simple mean.

Figure 1 shows the end-to-end data construction and quality-control workflow: source-specific ingestion is followed by timestamp harmonisation, physical-range screening, robust outlier suppression, daily aggregation, mask generation and field-season packaging. The workflow keeps the original observation age to avoid confusing forward-filled values with new measurements.

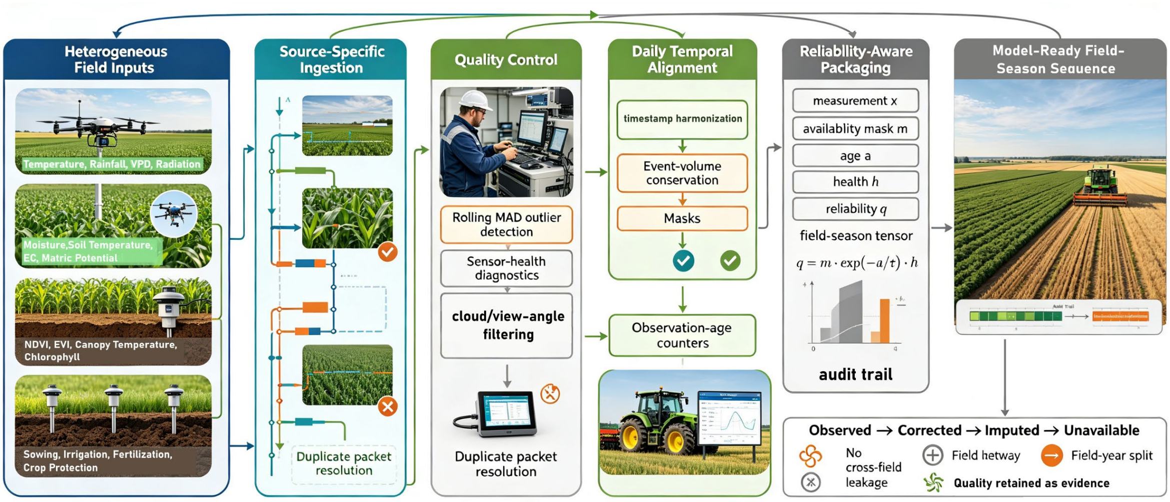

Figure 1. Multi-Source Farmland Sensing, Quality Control, Temporal Alignment, and Field-Season Sequence Construction Workflow.

For source s at day t, the aligned observation vector is $x_t^{(s)}$, its binary availability mask is $m_t^{(s)}$, and its age since acquisition is $a_t^{(s)}$. Reliability combines availability, normalized age, and a sensor-health score $h_t^{(s)}$:

$$q_t^{(s)} = m_t^{(s)} exp\left(-\frac{a_t^{(s)}}{\tau_s}\right) h_t^{(s)}. \qquad \text{Eq.(1)}$$

The source encoder receives measurement, mask, age, and reliability channels. A trainable affine projection maps each source into a common d-dimensional space:

$$e_t^{(s)} = W_s\left[x_t^{(s)} \| m_t^{(s)} \| a_t^{(s)} \| q_t^{(s)}\right] + b_s \qquad \text{Eq.(2)}$$

To prevent a dense source from dominating the fused token, normalized gates are computed across available sources:

$$\alpha_t^{(s)} = \frac{exp\left(g_s\left(e_t^{(s)}, q_t^{(s)}\right)\right)}{\sum_r \, exp\left(g_r\left(e_t^{(r)}, q_t^{(r)}\right)\right)} \qquad \text{Eq.(3)}$$

The daily multisource token is then the gated sum augmented with positional, calendar, and crop-stage embeddings:

$$z_t = \sum_s \alpha_t^{(s)} e_t^{(s)} + p_t + c_t + \phi_t \qquad \text{Eq.(4)}$$

Training windows are divided by field and year, not by individual days, to avoid a partition-boundary leak of adjacent records. Normalise continuous variables based on the training set's statistics and use logarithmic compression for management amounts with a highly skewed distribution. Yield targets are standardized only for optimization and are restored to $\mathrm{t\ ha^{-1}}$ for reporting.

Source registration uses a two-level key of field boundary and season identifier. A sensor is only accepted if its coordinates are within the buffered field polygon and its active period overlaps with the crop calendar. Duplicate packets are resolved based on acquisition time, and the packet with a higher diagnostic score is kept. Source-specific limits are used for physical checks, and a rolling median absolute deviation is calculated based on crop stage for statistical checks. A flagged value is retained in the audit record but not included in the daily aggregate. Cloud probability and view angle of satellite-derived canopy variables are included in the reliability score. Reliability of management events refers to the accuracy of timestamps and whether the operation was recorded by the machine or entered manually. The above rules distinguish among observed, corrected, imputed and unavailable values for tracking purposes.

Reliability-Preserving Imputation and Augmentation

Short gaps of at most two days are initialized by monotone interpolation, whereas longer gaps use a stage-aware seasonal median from the same production zone. The initialized value is never presented alone: its mask and age channels allow the network to discount it. Training augmentation reproduces operational failures through isolated point dropout, contiguous block dropout, channel shutdown, time jitter, and bounded calibration drift. These transformations are applied after the train-validation split.

A reconstruction objective on deliberately hidden observed entries regularizes the encoders:

$$\mathcal{L}_{rec} = \frac{1}{|\Omega|} \sum_{(t,j)\in\Omega} \rho_\delta\left(x_{tj} - \hat{x}_{tj}\right) \qquad \text{Eq.(5)}$$

where ρ_δ is the Huber penalty and Ω contains only values known before masking. The sampling probability increases for sensors with historically frequent outages, avoiding an unrealistically uniform corruption process.

Temporal augmentation is constrained by agronomy: accumulated rainfall and irrigation volumes are conserved when timestamps are jittered, and management events are never moved across phenological-stage boundaries. This preserves physical meaning while exposing the model to acquisition uncertainty.

Augmentation intensity is sampled once per field-season so that correlated failures extend across related channels. A failed soil node, for example, removes moisture, temperature, conductivity, and potential measurements together instead of producing independent missing points. Calibration drift is generated as a slowly varying multiplicative factor with bounded random-walk increments. The reconstruction head is active only during training and shares the source encoders but not the yield head. Its gradients are down-weighted after the first half of training, allowing early representation learning without forcing the final latent space to reproduce sensor noise. Validation data receive deterministic initialization but no stochastic corruption, and all test corruptions are generated from fixed seed lists.

Prediction Task and Evaluation Protocol

The primary task estimates final harvested yield from observations available through a selected forecast cutoff. Cutoffs at $40\%, 55\%, 70\%, 85\%$, and 100% of the season quantify how accuracy evolves. A secondary robustness protocol removes random points or contiguous blocks at rates from 10% to 40%. Cross-zone evaluation withholds complete production zones, while cross-crop evaluation trains on two crops and fine-tunes a small prediction head on the third.

The joint optimization objective combines yield regression, interval likelihood, and masked reconstruction:

$$\mathcal{L} = \mathcal{L}_y + \lambda_{nll}\mathcal{L}_{nll} + \lambda_{rec}\mathcal{L}_{rec} + \lambda_{gate}\mathcal{L}_{gate} \qquad \text{Eq.(6)}$$

Mean absolute error is defined as

$$MAE = \frac{1}{N}\sum_{i} |y_i - \hat{y}_i| \qquad \text{Eq.(7)}$$

and root-mean-square error is

$$RMSE = \sqrt{\frac{1}{N}\sum_{i} (y_i - \hat{y}_i)^2} \qquad \text{Eq.(8)}$$

Model Selection Uses Validation RMSE and Early Stopping. All the reported test values are combined from the five fieldstratified seeds, and paired bootstrap intervals are computed across field-seasons. The protocol eliminates model variation due to favourable partitioning and makes comparisons under missingness reproducible.

Based on the observed planting and harvesting dates in each area, set a forecast cutoff rather than using a fixed calendar month. If the harvest time is unavailable, maturity will be inferred from the latest management record and a sustained canopy decline, and this case will receive a lower target-quality flag. Yield monitor observations are cleaned by excluding header delays, implausible travel speeds, partial swath widths and abrupt moisture-correction jumps before field aggregation. Scale weights prevent large fields from dominating the optimisation, but every field-season has the same weight in headline metrics. Resampling unit for paired model comparisons is the field, which preserves the correlation among seasons at the same location. Hyperparameter selection uses only the validation set, and the same forecast cutoff will be used for testing. A special protocol is used to keep the corruption mask separate from the model's input and is consistently playable across different architectures.

Quality-Aware Decomposed Informer

Decomposed Quality-Aware Informer

The two parts of the architecture are a slow-changing phenological curve and a short-lived outlier. A moving-average operator with stage-dependent width is used to extract the seasonal component, and the residual contains rainfall pulses, heat shocks, irrigation responses and abrupt canopy changes.

$$s_t = MA_{k(\phi_t)}(z_t), r_t = z_t - s_t \qquad \text{Eq.(9)}$$

Seasonal and residual branches use independent projections before entering the stacked Informer blocks. Only after long-range selection are the outputs combined to avoid suppression of event signatures by large seasonal variations. Sensor reliability also affects the strength of the key, and thus an old observation may be retained in the context without drawing attention.

For query i and key j, the stage-conditioned sparse-attention score is

$$A_{ij} = \frac{Q_i K_j^{\top}}{\sqrt{d}} + u_{\phi_i,\phi_j} + \beta log\left(\varepsilon + q_j\right) \qquad \text{Eq.(10)}$$

where u_{ϕ_i,ϕ_j} is a learned compatibility between crop stages and q_j is the fused reliability. Only the most informative queries are evaluated against the full key set. The retained query count follows

$$u = min(T, \lceil cln\, T \rceil) \qquad \text{Eq.(11)}$$

giving logarithmic growth until the sequence becomes shorter than the sparse-query budget.

Distillation layers reduce the temporal resolution of encoder blocks by half using convolution, gated activation and max pooling. The first few layers are relatively local in shape, and those further away capture relations among growth stages. This hierarchy keeps the active history for a long time and limits memory.

The probabilistic query selector uses the difference between the maximum sampled dot product and the mean of all samples for a query as a measure of informativeness. Reliability does not affect the final score or the sampling distribution, and therefore, low-quality keys are not systematically oversampled. Stage compatibility is initially close to zero and jointly learned without a fixed agronomic prior. Residual blocks are employed for pre-normalisation to stabilise mixed-precision training. Tokens after the forecast cutoff are strictly removed rather than masked with synthetic zeros, and calendar embeddings do not reveal the future season length.

Stage-Conditioned Fusion and Forecast Head

Crop Stage determined by growth-degree accumulation, planting date and canopy trajectory. A probability vector for stage membership is used instead of a hard stage label to avoid discontinuities at the transition time. The conditioning vector for attention compatibility and fusion gates is obtained above.

The branch outputs are pooled with a learned yield token. Seasonal and residual summaries are combined through a gate:

$$\begin{gathered} g = \sigma\left(W_g[h_s \| h_r \| h_m] + b_g\right) \\ h = g \odot h_s + (1 - g) \odot h_r + W_m h_m \end{gathered} \qquad \text{Eq.(12)}$$

where h_m summarizes management events. A heteroscedastic head predicts mean yield μ and positive scale σ_y :

$$\left[\mu, log\, \sigma_y\right] = W_o h + b_o, \sigma_y = softplus\left(log\, \sigma_y\right) + \varepsilon. \qquad \text{Eq.(13)}$$

The Gaussian negative log-likelihood is used as a stable interval-training objective:

$$\mathcal{L}_{nll} = \frac{1}{N}\sum_{i}\left[log\ \sigma_i + \frac{(y_i - \mu_i)^2}{2\sigma_i^2}\right]. \quad \text{Eq.(14)}$$

After training, a scalar calibration factor is fitted on the validation set and applied to σ_y. This operation does not change point predictions. It corrects systematic under- or over-dispersion and produces prediction intervals suitable for risk-aware decisions.

Regularise the change of stage probabilities gradually from day to day, except in the vicinity of management operations that may affect canopy development. Missing canopy observations do not freeze the stage; growing-degree accumulation and planting information continue to be updated in distribution, and uncertainty rises. The yield token attends to both branches but cannot attend directly to the raw values, so all source effects must go through quality-aware encoders. Layer Normalisation statistics are shared across crops, and only the final output bias is crop-specific. During cross-crop adaptation, the encoders and attention blocks are frozen for the first five epochs, and then only the upper two blocks are unfrozen at one-tenth of the prediction-head learning rate. Calibration uses the absolute standardised-residual quantile as the initial scale factor and then reduces the interval score on validation data.

Computational Complexity and Implementation

Figure 2 shows the model topology of quality-aware tokens to decomposed branches, stage-conditioned probabilistic sparse attention, temporal distillation, gated recombination and calibrated outputs. The two-branch design does not share attention parameters or normalisation statistics but keeps the same crop-stage embedding for semantic alignment.

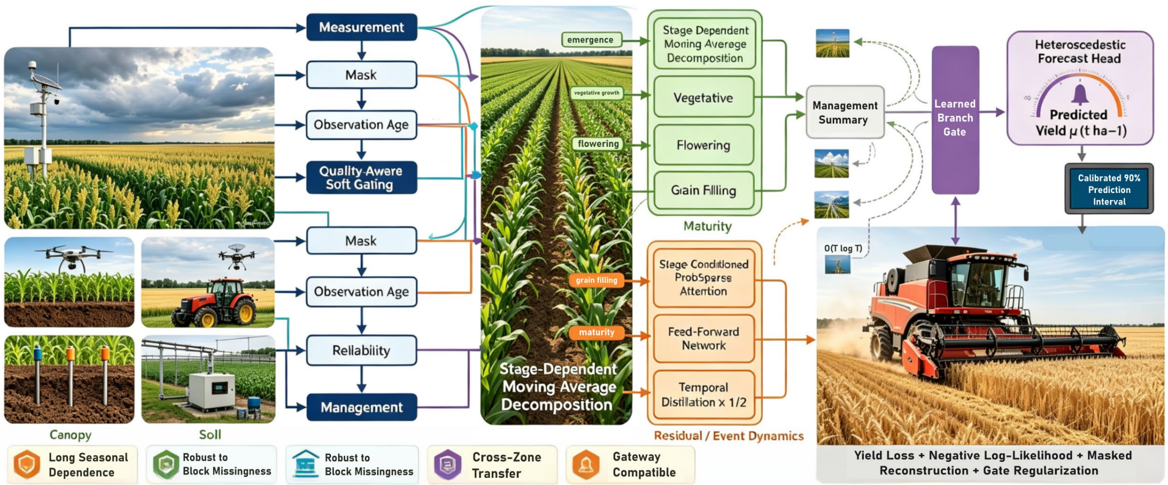

Figure 2. Quality-Aware Decomposed Informer Architecture with Stage-Conditioned Sparse Attention and Calibrated Yield Outputs.

The dominant attention complexity is reduced from quadratic full attention to

$$\mathcal{C}_{attn}(T, d) = \mathcal{O}(Tlog\ T \cdot d) \quad \text{Eq.(15)}$$

and the approximate activation memory becomes

$$\mathcal{M}(T, d, L) \approx \kappa T d \sum_{\ell=0}^{L-1} 2^{-\ell} \quad \text{Eq.(16)}$$

The model uses four encoder blocks, eight heads, $d = 256$, feed-forward width 512, and dropout 0.12. AdamW optimization starts at 3×10^{-4} with cosine decay, a five-epoch warm-up, gradient clipping at 1.0, and batches of 64 field-seasons. Mixed precision is used during training. At inference, source encoders are skipped for absent channels, reducing unnecessary gateway computation without changing learned parameters.

Implementation batches sequences by similar length and adds padding masks before attention and pooling. Gradient accumulation is used only for the 2,048-step scaling experiment. The 40 validation trials of the deep baselines were used to select hyperparameters, and a parameter-count band was employed to prevent bias towards the proposed model. Training stops after 15 unimproved epochs and usually converges in about 78 epochs. Reserve five seeded replicas for the test and calculate an independent calibration factor for each replica.

Experimental Evaluation and Analysis

Experimental Setup and Comparative Performance

The experimental collection contained 18,720 field-season sequences from 64 production zones and three crops: 7,240 maize, 6,180 wheat, and 5,300 soybean records. Seasons ranged from 128 to 246 days. Weather coverage was 96.8%, soil coverage 82.4%, canopy coverage 71.6%, and management-event completeness 88.9%. The field-disjoint split assigned 60% of fields to training, 20% to validation, and 20% to testing. Evaluation included linear regression, random forest, long short-term memory, temporal convolutional network, full-attention Transformer, Autoformer, and the original Informer. Hyperparameters were tuned under the same validation budget. Dataset construction followed the field-level separation principle used in recent agricultural prediction studies [25].

Figure 3(a) uses grouped bars to show RMSE values of 0.421,0.447,0.451, and 0.438 $\mathrm{t\,ha^{-1}}$ for maize, wheat, soybean, and the complete test set; Figure 3(b) uses a lollipop comparison to report corresponding R^2 values of 0.901,0.874, 0.856 , and 0.881; Figure 3 (c) presents an accuracy-latency bubble plot for six models and places the proposed 0.438 $\mathrm{t\,ha^{-1}}$ result at 18.7 ms , compared with 0.612 $\mathrm{t\,ha^{-1}}$ and 32.1 ms for the recurrent baseline. These patterns are consistent with the value of long-range representations reported for crop forecasting [26].

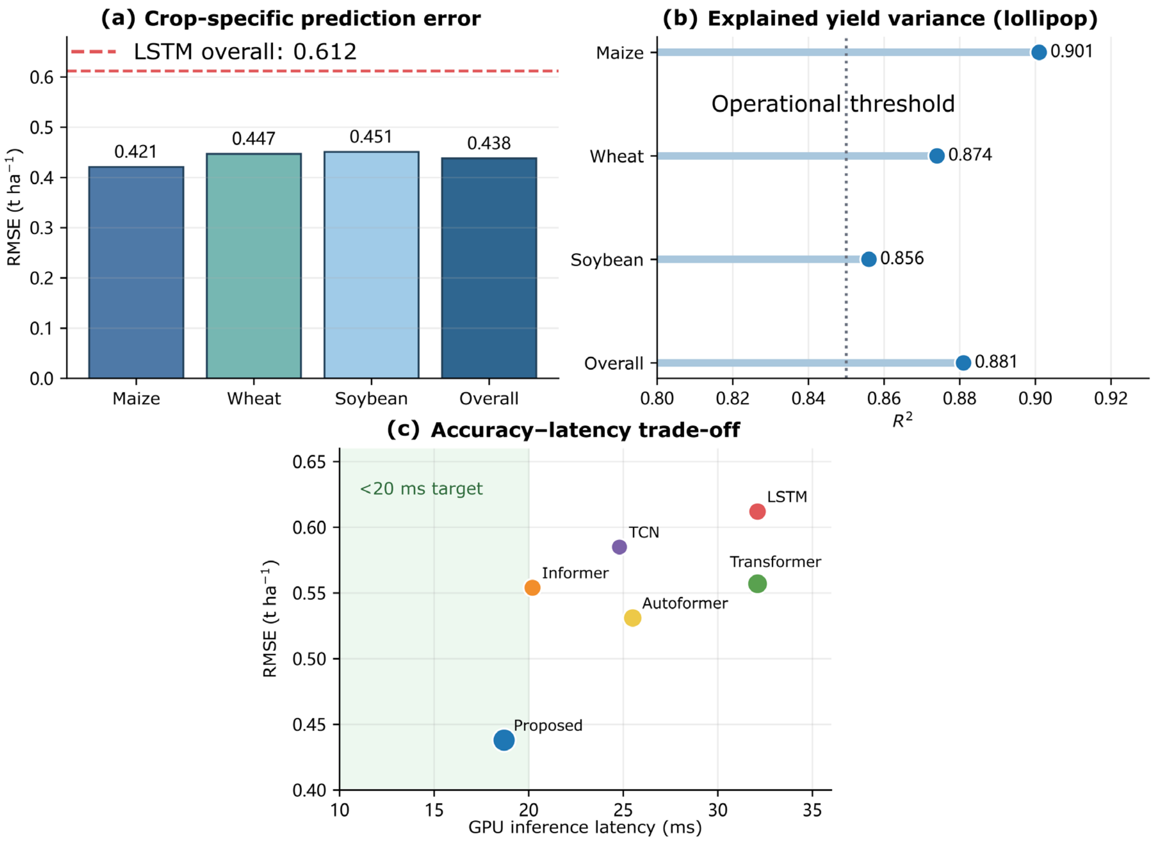

Figure 3. Comparative predictive performance: (a) crop-specific RMSE; (b) crop-specific R^2; (c) accuracy-latency trade-off across baseline models.

Error distributions were examined rather than relying on a single average. The median absolute error was 0.271 $\mathrm{t\,ha^{-1}}$, the 90 th percentile was 0.813 $\mathrm{t\,ha^{-1}}$, and only 2.7% of cases exceeded 1.5 $\mathrm{t\,ha^{-1}}$. Errors were larger in zones with extreme late-season heat, but the proposed model retained a 28.4% RMSE advantage over the full-attention Transformer. This difference indicates that sparse selection was not merely a computational compromise [27].

Baseline behavior varied with crop and sequence density. Linear regression reached 1.104 $\mathrm{t\,ha^{-1}}$ overall RMSE, random forest 0.793, long short-term memory 0.612, temporal convolutional network 0.585, full-attention Transformer 0.557, Autoformer 0.531, and original Informer 0.554. The proposed $0.438\ \mathrm{t\,ha^{-1}}$ result therefore represented reductions of 25.1% against the temporal convolutional network, 21.4% against full attention, 17.5% against Autoformer, and 20.9% against original Informer. Mean absolute errors followed the same ranking at $0.846, 0.603, 0.452, 0.421, 0.397, 0.381$, 0.404, and $0.316\ \mathrm{t\,ha^{-1}}$. Parameter count alone did not explain the ordering: the full-attention Transformer had 16.2 million parameters, whereas the selected proposed configuration had 14.8 million.

Performance was also stratified by yield level and climatic stress. For field-seasons below 4 $\mathrm{t\,ha^{-1}}$, between 4 and 8 $\mathrm{t\,ha^{-1}}$, and above 8 $\mathrm{t\,ha^{-1}}$, the proposed model obtained RMSE values of $0.391, 0.426$, and 0.502 $\mathrm{t\,ha^{-1}}$. In the hottest decile of reproductive-stage temperature, RMSE rose to $0.571\ \mathrm{t\,ha^{-1}}$ but remained below the original Informer value of 0.724. In the driest decile, quality-aware soil encoding produced a $0.137\ \mathrm{t\,ha^{-1}}$ advantage. A residual regression against field area, sequence length, and sensor completeness yielded no meaningful field-size trend, while completeness retained a modest negative coefficient. This pattern supports reporting reliability-conditioned uncertainty rather than presenting a uniform confidence range.

To test whether the advantage depended on a favorable metric, rankings were repeated with normalized RMSE, symmetric mean absolute percentage error, concordance correlation, and median absolute error. The proposed model ranked first on all four measures. Its symmetric percentage error was 5.8%, compared with 7.6% for Autoformer and 7.9% for the full-attention Transformer; concordance correlation was $0.934, 0.901$, and 0.892, respectively. Seed-to-seed RMSE standard deviation was $0.009\ \mathrm{t\,ha^{-1}}$, below the 0.017 observed for original Informer. The improvement persisted when the largest production zone was removed and when each crop was scored with equal weight rather than by sample count. A naive ensemble of the five baselines improved over individual baselines to $0.508\ \mathrm{t\,ha^{-1}}$ but remained 0.070 above the proposed model, indicating that the gain was not reproduced by averaging conventional error patterns.

Ablation, Missingness, and Transfer Analysis

The ablation removed quality channels, decomposition, stage conditioning, reconstruction regularization, and calibration one component at a time. Removing quality channels increased RMSE from 0.438 to $0.492\ \mathrm{t\,ha^{-1}}$; removing decomposition increased it to $0.481\ \mathrm{t\,ha^{-1}}$; removing stage conditioning produced $0.474\ \mathrm{t\,ha^{-1}}$. The largest error under block outages occurred when both quality encoding and reconstruction were disabled, reaching $0.553\ \mathrm{t\,ha^{-1}}$. The findings support explicit missingness modeling rather than silent imputation [28].

Figure 4(a) traces three RMSE curves across five point-missingness levels from 0% to 40%, with the proposed model rising from 0.438 to $0.523\ \mathrm{t\,ha^{-1}}$ while long short-term memory rises from 0.612 to $0.801\ \mathrm{t\,ha^{-1}}$; Figure 4(b) uses grouped bars at the same five outage levels and shows that 30% contiguous block removal increased proposed-model RMSE by 11.4% but original-Informer RMSE by 24.9%; Figure 4(c) uses a four-bubble impact display to report channelshutdown penalties of $0.031, 0.064, 0.046$, and $0.025\ \mathrm{t\,ha^{-1}}$ for weather, soil, canopy, and management channels. The soil channel was most influential because its outages overlapped water-stress periods [29].

Figure 5(a) uses a point-range display with confidence intervals for six held-out zone groups and gives R^2 values between 0.792 and 0.851 ; Figure 5(b) uses a stepped adaptation-area curve across zero, $5\%, 10\%, 20\%$, and 40% target labels, where RMSE decreases from 0.637 to $0.612, 0.541, 0.501$, and $0.466\ \mathrm{t\,ha^{-1}}$; Figure 5(c) maps nine source-totarget crop combinations in a 3×3 heat map, including maize-to-wheat and maize-to-soybean errors of 0.574 and $0.603\ \mathrm{t\,ha^{-1}}$. Domain shift remained visible, yet quality-aware representations transferred more reliably than raw concatenation [30].

https://doi.org/10.64972/jaat.2024v2.278p23e:320-334

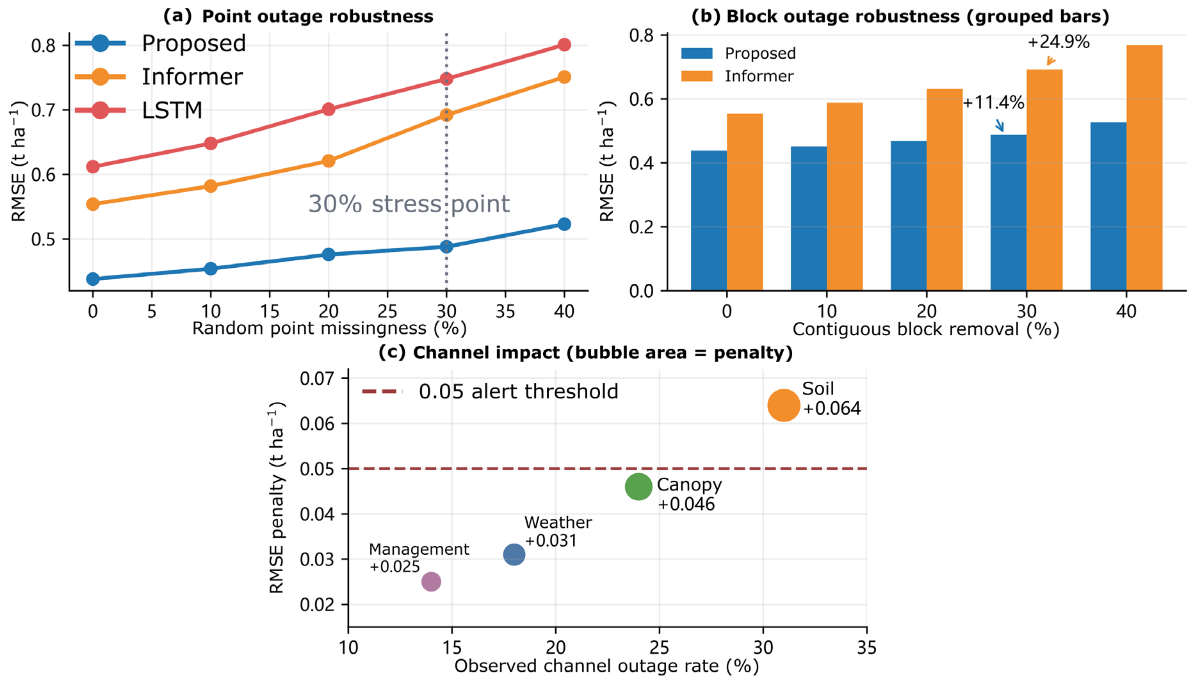

Figure 4. Robustness under sensor degradation: (a) point-missingness response; (b) contiguous block-missingness response; (c) single-channel shutdown penalties.

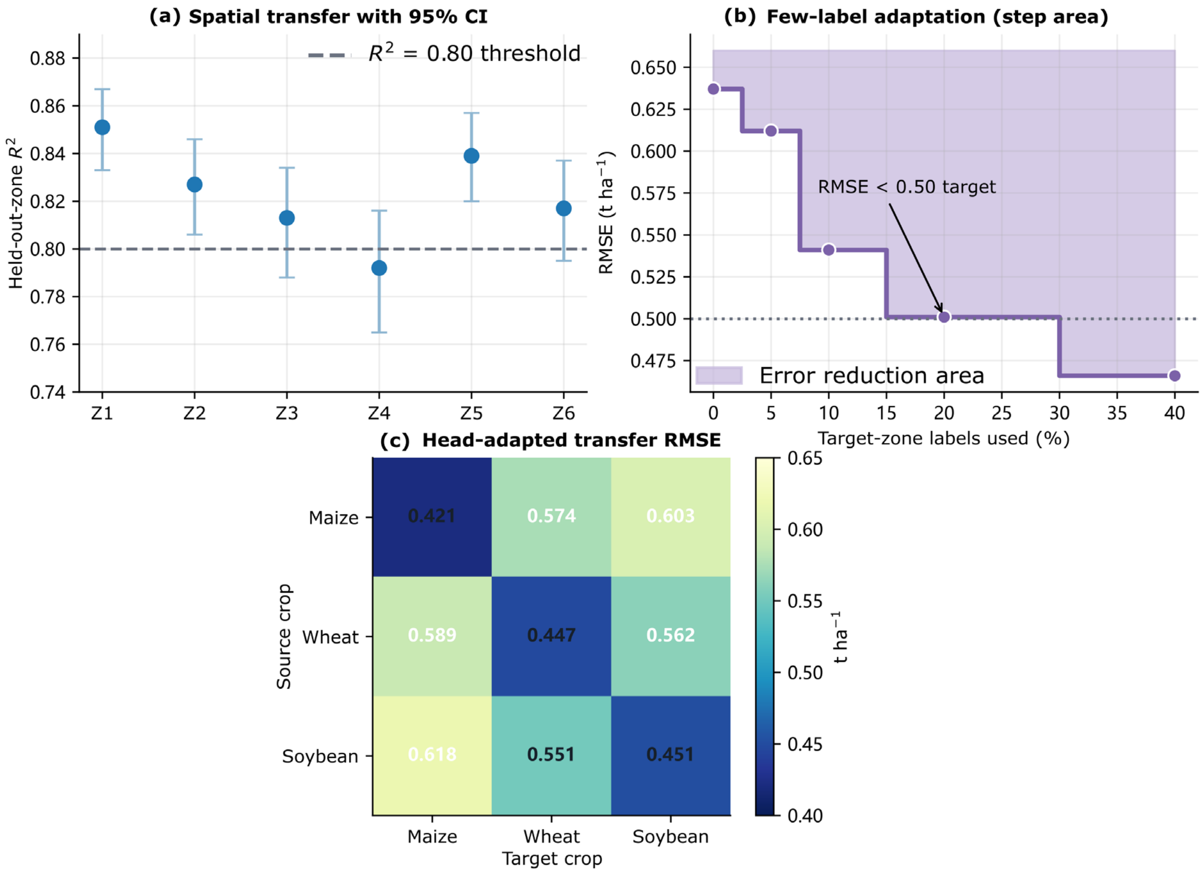

Figure 5. Spatial and crop transfer: (a) held-out zone-group performance; (b) target-label adaptation curve; (c) crop-tocrop transfer error matrix.

The transfer gain was strongest where sensor vendors differed between source and target zones. With no adaptation, vendor shift added $0.083\ \mathrm{t\ ha^{-1}}$ RMSE to the original Informer and $0.049\ \mathrm{t\ ha^{-1}}$ to the proposed model. Reliability channels supplied a common description of observation confidence even when raw calibration distributions changed. This is important because operational networks are rarely homogeneous [31].

Component interactions were non-additive. Quality encoding and reconstruction regularization each improved complete-data RMSE by 0.054 and $0.021\ \mathrm{t\ ha^{-1}}$, but their joint contribution under 30% block missingness reached 0.091 $\mathrm{t\ ha^{-1}}$. Decomposition contributed most in long seasons exceeding 210 days, where its removal added 0.061 $\mathrm{t\ ha^{-1}}$, compared with 0.029 in seasons shorter than 160 days. Stage conditioning primarily affected reproductive cutoffs: removing it changed 70% season RMSE from 0.529 to 0.588 $\mathrm{t\ ha^{-1}}$ but altered full-season RMSE by only 0.036. These interactions match the intended division of labor: reliability channels describe evidence quality, reconstruction improves degraded inputs, decomposition manages long trends, and stage conditioning prioritizes time-sensitive events.

The missingness curves were repeated with outage positions restricted to early, middle, or late season. Early outages were comparatively benign because later canopy and weather evidence remained available. Middle-season outages increased RMSE by $0.073\ \mathrm{t\ ha^{-1}}$, and outages centered on flowering increased it by 0.096. When only canopy data were removed during flowering, wheat suffered the largest R^2 reduction, from 0.874 to 0.812. When soil data were removed during the same interval, maize showed the largest RMSE increase, from 0.421 to $0.503\ \mathrm{t\ ha^{-1}}$. The quality gate shifted average attention mass toward weather and management tokens during these failures instead of amplifying initialized soil values.

Gate behavior provided additional evidence about source substitution. With all channels present, mean gate weights were 0.29 for weather, 0.32 for soil, 0.25 for canopy, and 0.14 for management. During a 30% soil outage, the soil weight fell to 0.18 while weather and canopy rose to 0.36 and 0.30. When canopy images were unavailable for more than 20 days, management weight increased to 0.19, particularly after irrigation and nitrogen events. A control model that renormalized fixed source weights could not reproduce this adaptation and reached 0.536 $\mathrm{t\ ha^{-1}}$ under the same outage. Reliability was not allowed to collapse a gate to zero: a small entropy penalty maintained gradient flow and supported recovery when a source returned. After a simulated 21 -day soil outage, the recovered source regained 90% of its pre-outage average weight within four observed days.

Early Forecasting, Calibration, and Efficiency

Forecast usefulness depends on when a reliable estimate becomes available. At 40% season progress, the model obtained RMSE $0.812\,\mathrm{t\,ha^{-1}}$ and $\mathrm{R}^2 0.612$. At $55\%, 70\%, 85\%$, and full season, RMSE declined to $0.654, 0.529, 0.463$, and $0.438\ \mathrm{t\,ha^{-1}}$. The largest improvement occurred between 55% and 70%, coinciding with reproductive-stage observations. Early prediction behavior agrees with evidence that phenology-aware remote and ground sensing can improve yield estimation [32].

Figure 6(a) presents three cutoff curves at five season-progress levels, where maize improves from 0.791 to $0.421\ \mathrm{t^{-1}}$, wheat from 0.834 to $0.447\,\mathrm{t\,ha^{-1}}$, and soybean from 0.817 to $0.451\,\mathrm{t\,ha^{-1}}$; Figure 6(b) uses a calibration scatter diagram at five nominal coverage levels and shows calibrated 90% coverage of 93.6% rather than the uncalibrated 84.2%; Figure 6(c) uses four box distributions to show median interval widths increasing from $1.08\,\mathrm{t\,ha^{-1}}$ above 90% sensor completeness to $1.71\,\mathrm{t\,ha^{-1}}$ below 60%. The widening is desirable because uncertainty responds to evidence quality instead of remaining constant [33].

Efficiency was measured on one NVIDIA RTX-class GPU and an ARM gateway using batch size one. Full attention required 7.6 GB during training and 32.1 ms per GPU inference, while the proposed model required 4.3 GB and 18.7 ms. On the gateway, latency fell from 284 to 171 ms and peak memory from 1.42 to 0.86 GB. The gain follows the expected long-sequence advantage of sparse attention [34].

Figure 7(a) compares two memory-scaling curves at five sequence lengths from 128 to 2,048 days, where the proposed model grows from 1.2 to 6.8 GB while full attention grows from 1.4 to 18.9 GB; Figure 7(b) uses five latency bars to report 92, 118, 171, 263, and 418 ms on the gateway; Figure 7(c) presents a five-configuration Pareto bubble plot across RMSE, energy, and parameter count, with the selected 14.8 -million-parameter model using 0.74 J per forecast. The configuration remains practical for scheduled field-gateway updates [35].

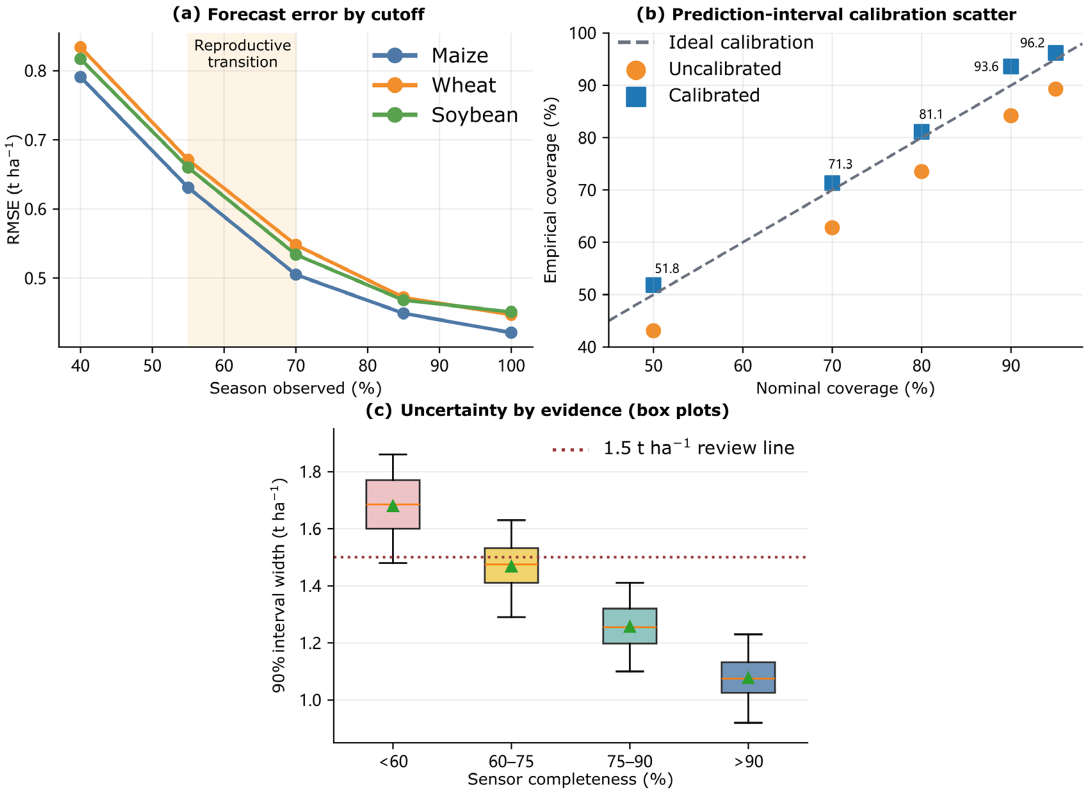

Figure 6. Forecast timing and uncertainty: (a) crop-specific error by season cutoff; (b) nominal versus empirical interval coverage; (c) interval width by sensor completeness.

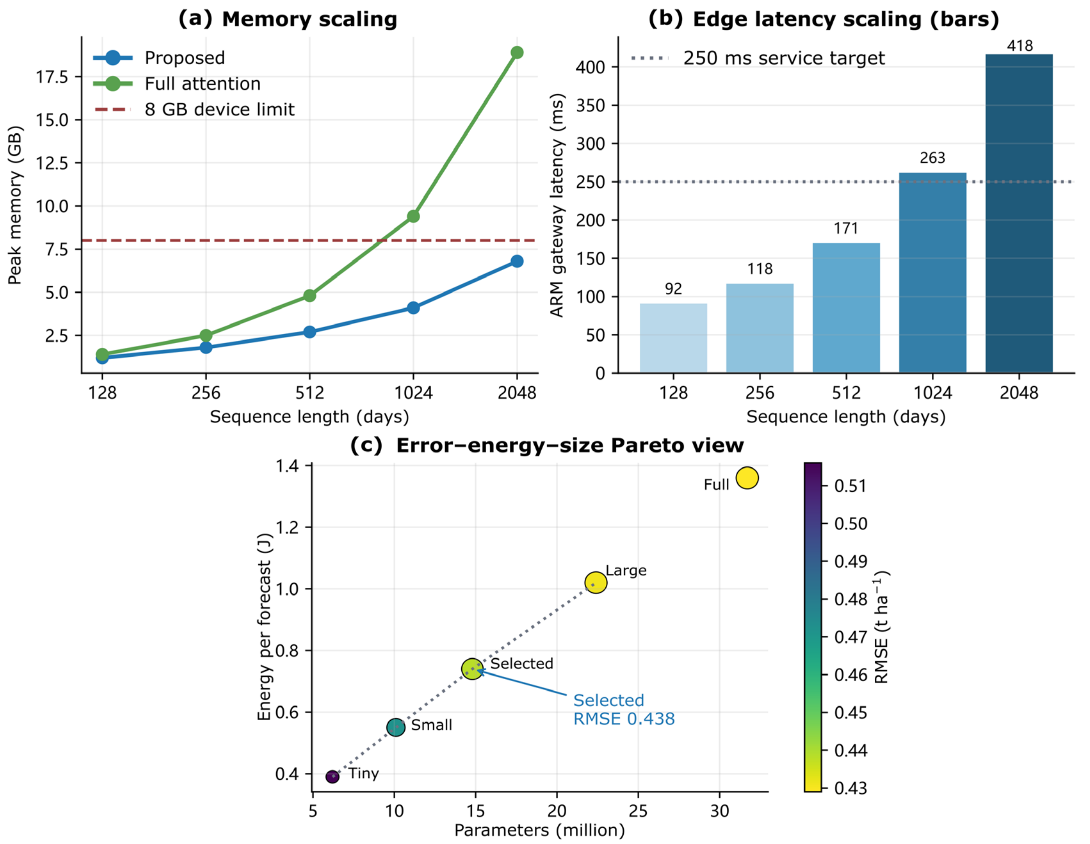

Figure 7. Deployment efficiency: (a) memory scaling; (b) gateway latency scaling; (c) Pareto frontier of error, energy, and parameter count.

A paired bootstrap over 10,000 field resamples placed the RMSE improvement over original Informer at $0.116\ \mathrm{t\,ha}^{-1}$ with a 95% interval of $[0.101, 0.131]$. Improvements remained significant for every crop and cutoff after Holm correction. Residual checks showed mild underprediction above $11\ \mathrm{t\,ha}^{-1}$, suggesting that rare high-yield environments remain underrepresented. The model nevertheless preserved monotonic improvements with added seasonal evidence and avoided unstable error spikes during sensor outages.

Calibration was evaluated with interval score, coverage error, and subgroup reliability. At nominal coverages of 50%, $70\%, 80\%, 90\%$, and 95%, empirical calibrated coverage was $51.8\%, 71.3\%, 81.1\%, 93.6\%$, and 96.2%. The mean interval score at 90% improved from 1.96 to 1.54 after validation scaling. Coverage remained between 91.9% and 94.7% across crops and between 90.8% and 95.1% across sensor-completeness quartiles. The slightly conservative 90% interval was retained because undercoverage is more harmful in procurement and insurance settings. Point RMSE and ranking were unchanged by calibration, confirming that the improvement came from scale correction rather than altered predictions.

A deployment stress test replayed a season through a gateway with intermittent connectivity. Daily ingestion and prediction had a median of 0.81 J for preprocessing and a 95th-percentile end-to-end latency of 236 ms. Caching of static soil and field descriptors reduced encoded input transfer by 38%. Skipping the absent source encoders reduced 9%-17% of the latency for that outage pattern. Quantizing only the source projections and the final head to 16-bit storage reduced the checkpoint size from 59.2 MB to 46.7 MB without an increase in measurable RMSE; more aggressive integer quantization was not employed due to a widened calibrated interval. Based on the above data, a sparse design will improve the training and periodic inference efficiency of the server.

Operational sensitivity was examined by perturbing the forecast cutoff by ± 3 days, changing daily aggregation boundaries by six hours, and replacing zone medians with neighboring-field initialization. The resulting RMSE changes were $0.009, 0.006$, and $0.014\ \mathrm{t\,ha}^{-1}$, respectively. In contrast, suppressing observation-age channels increased RMSE by $0.038\ \mathrm{t\,ha}^{-1}$ when more than one fifth of values had been forward filled. Attention diagnostics placed 31% of retained residual-branch mass on the 20 days surrounding flowering, 22% on major rainfall or irrigation events, and 17% on heat episodes above the crop-specific threshold. Seasonal-branch attention was more diffuse, with no single 20 -day interval exceeding 14%. These diagnostics agree with the architectural intent and provide a useful audit signal, although attention weights are not interpreted as causal effects.

Forecast revisions were also checked for temporal stability. Between consecutive weekly updates, the median absolute change in predicted yield declined from $0.31\ \mathrm{t\,ha}^{-1}$ early in the season to $0.08\ \mathrm{t\,ha}^{-1}$ after the 85% cutoff. Large revisions above $0.75\ \mathrm{t\,ha}^{-1}$ occurred in 3.6% of updates and were concentrated after documented heat, hail, or irrigation anomalies. The original Informer produced 5.9% large revisions, including several during periods with no new valid sensor data. Reliability-aware gating reduced such unsupported movement: when only stale forward-filled values arrived, the median weekly revision was $0.03\ \mathrm{t\,ha}^{-1}$. For decision use, the system records both the new estimate and the evidence change responsible for the update, such as a new canopy acquisition, a rainfall pulse, or restored soil telemetry. This trace does not prove causality, but it gives operators a concrete basis for reviewing abrupt model changes.

Conclusion

Develop a quality-aware decomposed Informer for crop-yield prediction from long, heterogeneous farmland sensor sequences in this paper. Preserve masks, observation age and reliability in alignment; separate seasonal context from event residuals; condition sparse attention on crop stage; and calibrate predictive uncertainty. The resulting pipeline connects the quality of operating data with computationally efficient long-range reasoning in a single model.

The approach is still based on relatively representative field-season histories and carries the uncertainty of yield-map calibration and management records. Rare varieties, extreme compound weather events and sudden sensor replacements are still difficult to show. The current stage estimator does not use jointly learned physiological states for engineered thermal and canopy cues either.

Future work will explore causal management-response modelling, federated adaptation of farms, and physics-guided constraints that preserve water and energy consistency. Continue to calibrate after replacing sensors and crops, improve the stability of deployment, and potentially reduce the amount of labelled yield data needed in new production areas with multimodal foundation representations.

Author Contributions

Kateřina Svobodová contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Natálie Králová contributes to software, validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. Remote Sensing, 14(9), 1990. https://doi.org/10.3390/rs14091990

[2] Nevavuori, P., Narra, N., Linna, P., & Lipping, T. (2020). Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. Remote Sensing, 12(23), 4000. https://doi.org/10.3390/rs12234000

[3] Barzin, R., Lotfi, H., Varco, J. J., & Bora, G. C. (2021). Machine Learning in Evaluating Multispectral Active Canopy Sensor for Prediction of Corn Leaf Nitrogen Concentration and Yield. Remote Sensing, 14(1), 120. https://doi.org/10.3390/rs14010120

[4] Mandal, D., & Rao, Y. S. (2020). SASYA: An integrated framework for crop biophysical parameter retrieval and within-season crop yield prediction with SAR remote sensing data. Remote Sensing Applications: Society and Environment, 20, 100366. https://doi.org/10.1016/j.rsase.2020.100366

[5] Worrall, G., Rangarajan, A., & Judge, J. (2021). Domain-Guided Machine Learning for Remotely Sensed In-Season Crop Growth Estimation. Remote Sensing, 13(22), 4605. https://doi.org/10.3390/rs13224605

[6] Zheng, C., Abd-Elrahman, A., & Whitaker, V. (2021). Remote Sensing and Machine Learning in Crop Phenotyping and Management, with an Emphasis on Applications in Strawberry Farming. Remote Sensing, 13(3), 531. https://doi.org/10.3390/rs13030531

[7] Varela, S., Pederson, T., Bernacchi, C. J., & Leakey, A. D. B. (2021). Understanding Growth Dynamics and Yield Prediction of Sorghum Using High Temporal Resolution UAV Imagery Time Series and Machine Learning. Remote Sensing, 13(9), 1763. https://doi.org/10.3390/rs13091763

[8] Bebie, M., Cavalaris, C., & Kyparissis, A. (2022). Assessing Durum Wheat Yield through Sentinel-2 Imagery: A Machine Learning Approach. Remote Sensing, 14(16), 3880. https://doi.org/10.3390/rs14163880

[9] Zhu, X., Guo, R., Liu, T., & Xu, K. (2021). Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data. Remote Sensing, 13(10), 2016. https://doi.org/10.3390/rs13102016

[10] Nihar, A., Patel, N. R., & Danodia, A. (2022). Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. Journal of the Indian Society of Remote Sensing, 50(8), 1519-1530. https://doi.org/10.1007/s12524-022-01549-0

[11] Oliveira, M. F. d., Ortiz, B. V., Morata, G. T., Jiménez, A. s. F., Rolim, G. d. S., & Silva, R. P. d. (2022). Training Machine Learning Algorithms Using Remote Sensing and Topographic Indices for Corn Yield Prediction. Remote Sensing, 14(23), 6171. https://doi.org/10.3390/rs14236171

[12] Momm, H. G., ElKadiri, R., & Porter, W. (2020). Crop-Type Classification for Long-Term Modeling: An Integrated Remote Sensing and Machine Learning Approach. Remote Sensing, 12(3), 449. https://doi.org/10.3390/rs12030449

[13] Banerjee, B. P., Sharma, V., Spangenberg, G., & Kant, S. (2021). Machine Learning Regression Analysis for Estimation of Crop Emergence Using Multispectral UAV Imagery. Remote Sensing, 13(15), 2918. https://doi.org/10.3390/rs13152918

[14] Evans, F. H., & Shen, J. (2021). Long-Term Hindcasts of Wheat Yield in Fields Using Remotely Sensed Phenology, Climate Data and Machine Learning. Remote Sensing, 13(13), 2435. https://doi.org/10.3390/rs13132435
[15] Aneece, I., & Thenkabail, P. S. (2021). Classifying Crop Types Using Two Generations of Hyperspectral Sensors (Hyperion and DESIS) with Machine Learning on the Cloud. Remote Sensing, 13(22), 4704. https://doi.org/10.3390/rs13224704
[16] Yli-Heikkil?, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., & Klami, A. (2022). Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. Remote Sensing, 14(17), 4193. https://doi.org/10.3390/rs14174193
[17] Sapkota, B., Singh, V., Neely, C., Rajan, N., & Bagavathiannan, M. (2020). Detection of Italian Ryegrass in Wheat and Prediction of Competitive Interactions Using Remote-Sensing and Machine-Learning Techniques. Remote Sensing, 12(18), 2977. https://doi.org/10.3390/rs12182977
[18] Impollonia, G., Croci, M., Ferrarini, A., Brook, J., Martani, E., Blandinières, H., Marcone, A., Awty-Carroll, D., Ashman, C., Kam, J., Kiesel, A., Trindade, L. M., Boschetti, M., Clifton-Brown, J., & Amaducci, S. (2022). UAV Remote Sensing for High-Throughput Phenotyping and for Yield Prediction of Miscanthus by Machine Learning Techniques. Remote Sensing, 14(12), 2927. https://doi.org/10.3390/rs14122927
[19] Canata, T. F., Wei, M. C. F., Maldaner, L. F., & Molin, J. P. (2021). Sugarcane Yield Mapping Using High-Resolution Imagery Data and Machine Learning Technique. Remote Sensing, 13(2), 232. https://doi.org/10.3390/rs13020232
[20] Hara, P., Piekutowska, M., & Niedba?a, G. (2021). Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. Land, 10(6), 609. https://doi.org/10.3390/land10060609
[21] Meraj, G., Kanga, S., Ambadkar, A., Kumar, P., Singh, S. K., Farooq, M., Johnson, B. A., Rai, A., & Sahu, N. (2022). Assessing the Yield of Wheat Using Satellite Remote Sensing-Based Machine Learning Algorithms and Simulation Modeling. Remote Sensing, 14(13), 3005. https://doi.org/10.3390/rs14133005
[22] Bian, C., Shi, H., Wu, S., Zhang, K., Wei, M., Zhao, Y., Sun, Y., Zhuang, H., Zhang, X., & Chen, S. (2022). Prediction of Field-Scale Wheat Yield Using Machine Learning Method and Multi-Spectral UAV Data. Remote Sensing, 14(6), 1474. https://doi.org/10.3390/rs14061474
[23] Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2021). New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: a machine learning approach. International Journal of Remote Sensing, 42(9), 3426-3444. https://doi.org/10.1080/01431161.2020.1871102
[24] Zheng, C., Abd-Elrahman, A., Whitaker, V., & Dalid, C. (2022). Prediction of Strawberry Dry Biomass from UAV Multispectral Imagery Using Multiple Machine Learning Methods. Remote Sensing, 14(18), 4511. https://doi.org/10.3390/rs14184511
[25] Xie, G., & Niculescu, S. (2022). Mapping Crop Types Using Sentinel-2 Data Machine Learning and Monitoring Crop Phenology with Sentinel-1 Backscatter Time Series in Pays de Brest, Brittany, France. Remote Sensing, 14(18), 4437. https://doi.org/10.3390/rs14184437
[26] Marino, S., & Alvino, A. (2021). Vegetation Indices Data Clustering for Dynamic Monitoring and Classification of Wheat Yield Crop Traits. Remote Sensing, 13(4), 541. https://doi.org/10.3390/rs13040541
[27] de Carvalho Alves, M., Sanches, L., Pozza, E. A. l., Pozza, A. l. A. A., & da Silva, F. b. M. (2022). The role of machine learning on Arabica coffee crop yield based on remote sensing and mineral nutrition monitoring. Biosystems Engineering, 221, 81-104. https://doi.org/10.1016/j.biosystemseng.2022.06.014
[28] Lambertini, A., Mandanici, E., Tini, M. A., & Vittuari, L. (2022). Technical Challenges for Multi-Temporal and Multi-Sensor Image Processing Surveyed by UAV for Mapping and Monitoring in Precision Agriculture. Remote Sensing, 14(19), 4954. https://doi.org/10.3390/rs14194954
[29] Xu, R., Li, C., & Bernardes, S. (2021). Development and Testing of a UAV-Based Multi-Sensor System for Plant Phenotyping and Precision Agriculture. Remote Sensing, 13(17), 3517. https://doi.org/10.3390/rs13173517
[30] Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sensing, 12(19), 3136. https://doi.org/10.3390/rs12193136
[31] Ullo, S. L., & Sinha, G. R. (2021). Advances in IoT and Smart Sensors for Remote Sensing and Agriculture Applications. Remote Sensing, 13(13), 2585. https://doi.org/10.3390/rs13132585
[32] Pelta, R., Beeri, O., Tarshish, R., & Shilo, T. (2022). Forecasting seasonal plot-specific crop coefficient (Kc) protocol for processing tomato using remote sensing, meteorology, and artificial intelligence. Precision Agriculture, 23(6), 1983-2000. https://doi.org/10.1007/s11119-022-09910-6

[33] Virnodkar, S. S., Pachghare, V. K., Patil, V. C., & Jha, S. K. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. Precision Agriculture, 21(5), 1121-1155. https://doi.org/10.1007/s11119-020-09711-9
[34] Ma, C., Johansen, K., & McCabe, M. F. (2022). Monitoring Irrigation Events and Crop Dynamics Using Sentinel-1 and Sentinel-2 Time Series. Remote Sensing, 14(5), 1205. https://doi.org/10.3390/rs14051205
[35] Bagha, H., Yavari, A., & Georgakopoulos, D. (2022). Hybrid Sensing Platform for IoT-Based Precision Agriculture. Future Internet, 14(8), 233. https://doi.org/10.3390/fi14080233