Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

AI-Assisted Phenotyping and Physiological Modeling for Yield Stability in Future Cropping Systems

Dr. Nadia Jabeen *
Department of Agriculture, Hazara University Mansehra
nadia_khan909090@yahoo.com
nadia.agri@hu.edu.pk
Orcid id : 0000-0001-6617-8301

Ameer Jan
University of Makran
ameerjan@uomp.edu.pk

Bai-Bureh O'Bai Kamara
Sierra Leone Agricultural Research Institute (SLARI).
Agriculture Department, Central University, Mile91 Sierra Leone
kamarabaibureobai@gmail.com
Orcid.org/0009-0003-9705-6210

Ahmid Chernor Jalloh
Youth Action for Relentless Development Organization - Sierra Leone (YARDO-SL)
ahmidjalloh99@gmail.com

ABSTRACT

Climate change has intensified the frequency of drought, heat, and other abiotic stresses, threatening global crop productivity and yield stability. Conventional breeding and process-based crop models have independently contributed to crop improvement; however, both remain constrained by limited phenotypic information, labor-intensive parameterization, and insufficient capacity to capture dynamic genotype × environment × management (G × E × M) interactions. Recent advances in high-throughput phenotyping (HTP), artificial intelligence (AI), machine learning (ML), and remote sensing have created new opportunities for integrating real-time phenotypic observations with physiological crop models. This review examines the convergence of AI-assisted phenotyping and process-based crop modeling as a framework for improving yield prediction and climate resilience. It discusses major sensing technologies,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

including RGB, hyperspectral, thermal, LiDAR, and root imaging systems, together with functional trait extraction methods. The review further summarizes the mathematical foundations of leading crop growth models such as DSSAT, APSIM, WOFOST, AquaCrop, and InfoCrop, followed by recent developments in data assimilation techniques, including Ensemble Kalman Filters, variational methods, and particle swarm optimization. Field-based case studies in wheat, cotton, canola, and spinach demonstrate substantial improvements in crop monitoring and prediction accuracy through the assimilation of high-throughput phenomic data. Finally, the emerging Knowledge- and Data-Driven Modeling (KDDM) paradigm is explored, highlighting parallel, serial, and modular integration strategies that combine mechanistic understanding with data-driven intelligence. The synthesis demonstrates that AI-assisted phenotyping integrated with physiological modeling provides a robust pathway toward digital agriculture, precision crop management, accelerated breeding, and enhanced yield stability under future climate scenarios.

Keywords: Artificial intelligence (AI); High-throughput phenotyping; Crop phenomics; Process-based crop models; Knowledge- and data-driven modeling (KDDM); Machine learning; Data assimilation; Precision agriculture; Yield prediction; Climate-smart agriculture; Remote sensing; Digital agriculture.

Introduction

Modern agricultural production faces the unprecedented challenge of ensuring global food and nutrition security under the accelerating pressures of climate change. Cultivar yield (P) is a highly complex, emergent phenotype resulting from non-linear interactions between the underlying genetic architecture (G), highly variable environmental conditions (E), and strategic agronomic management decisions (M), historically formalized as the G × E × M paradigm (Liang et al., 2025). Under current climate projections, cropping systems are increasingly subjected to combinations of multi-stress factors such as unexpected temperature anomalies, shifts in precipitation patterns, and prolonged drought events. These stressors disrupt critical physiological processes, particularly during reproductive development, causing severe crop yield instability (Ghaffar et al., 2022)

To decompose yield into genetic and environmental factors, crop science has traditionally relied on two distinct methodological approaches. The first is a statistical modeling approach, which

Online ISSN 3007-3197
Print ISSN 3007-3189

seeks to identify superior genotypes by evaluating mean performances across environments and estimating sensitivities through relatively simple linear or non-parametric models (Gauch et al., 2008). While computationally straightforward, these statistical frameworks are limited because they treat the biological system as a static entity, failing to capture the dynamic, temporal feedback loops that govern crop growth. The second approach relies on process-based crop simulation models (PBMs). These mechanistic models integrate principles of plant physiology, micrometeorology, soil physics, and agronomy to simulate crop growth on a daily or hourly time-step (Elias et al., 2016). This allows for the prediction of yield as an emergent end-point driven by environmental variables. However, process-based models are often limited by the labor-intensive physical measurements required to calibrate their highly localized, genotype-specific parameters (Alexander et al., 1984).

Parallel to these modeling paradigms, crop breeding has been historically constrained by the "phenotyping bottleneck". Traditional manual phenotyping methods are slow, destructive, and lack the temporal resolution required to capture rapid physiological stress responses. To overcome this barrier, crop phenomics has emerged as a data-driven discipline (Tharan et al., 2026). High-throughput phenotyping (HTP) platforms utilizing advanced imaging sensors, automated greenhouses, unmanned aerial vehicles (UAVs), and satellite constellations enable the continuous, non-destructive observation of crops under real-world conditions. Despite their obvious synergies, crop phenotyping and biophysical crop modeling have frequently been treated as isolated domains, limiting their combined potential.

The systematic integration of artificial intelligence (AI), computer vision, and machine learning (ML) with process-based models offers a path toward bridging this gap (Li et al., 2021). This combined approach, known as Knowledge- and Data-Driven Modeling (KDDM), simplifies model parameterization, corrects simulation biases using real-time field observations, and enables the unsupervised recovery of latent physiological variables. This review article examines the structural, algorithmic, and biological frameworks underpinning the integration of AI-assisted phenotyping with crop physiological models to optimize yield stability in future cropping environments (Wang et al., 2025).

2. High-Throughput Sensing Technologies and Functional Trait Extraction

The transition from manual crop observations to digital phenomics is driven by non-destructive imaging sensors that capture physical, physiological, and biochemical signals across various organizational scales. Modern high-throughput phenotyping utilizes a multi-sensor array to evaluate crop performance. Visible spectrum (RGB) cameras provide foundational structural

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

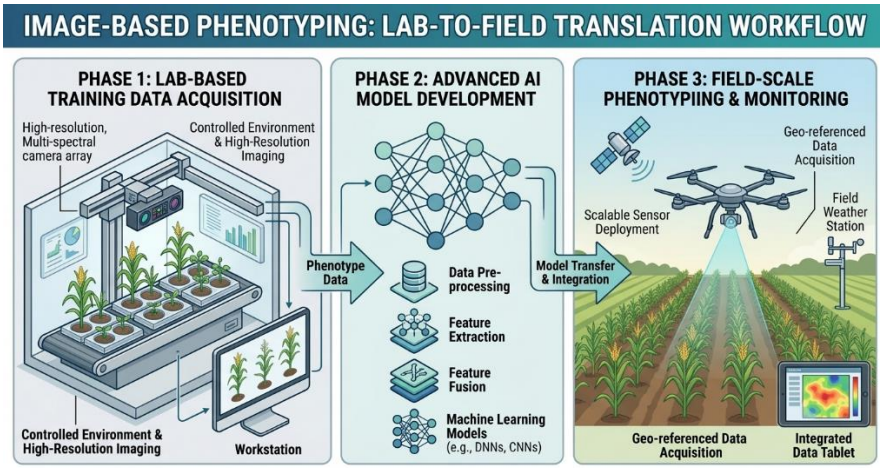
Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

data, allowing for the quantification of canopy cover, plant density, and mechanical lodging indices (Angid et al., 2025). When coupled with photogrammetry or Light Detection and Ranging (LiDAR) sensors, these systems generate detailed 3D point clouds that reconstruct plant architecture, canopy volume, and leaf angle distributions. These structural metrics correlate directly with light interception efficiency and total biomass accumulation (Chauhan & Singh, 2025).

To assess physiological and biochemical function, multi-sensor platforms integrate spectroscopic, thermal, and fluorescence imagers. Multispectral and hyperspectral sensors capture canopy reflectance across visible, near-infrared (NIR), and short-wave infrared (SWIR) spectra (typically ranging from 350 to 2,500 nm). Machine learning models utilize these spectral signatures to derive key biophysical variables, such as Leaf Area Index (LAI) and plant nitrogen concentration (PNC). This supports the early, non-destructive diagnosis of crop nutritional status and photosynthetic capacity (Zandi et al., 2025).

Thermal infrared thermography monitors canopy temperature to provide a real-time proxy for crop water transpiration and stomatal conductance. Under water-limiting conditions, plants close their stomata to prevent water loss, which reduces evaporative cooling and increases canopy temperature. Fusing these structural and functional modalities allows for the detection of "pre-visual" physiological stress before physical symptoms, such as wilting or chlorosis, manifest in the field (Ndlovu et al., 2024).

Figure 1. Methodological workflow for image-based plant phenotyping, illustrating the multi-phase translation from controlled lab environments to advanced AI feature extraction and scalable field-level monitoring.



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Table 1: Primary Sensing Technologies and Associated Physiological Target Traits

Sensor Category	Spectral & Physical Modalities	Primary Target Traits	Crop Physiology Implications	Processing & Extraction Algorithms
RGB Imaging	Visible spectrum (400–700 nm)	Canopy cover, leaf count, organ morphology, plant height, lodging index	Biomass accumulation, phenological progression, structural mechanical integrity	3D Structure-from-Motion (SfM), semantic segmentation, object detection
Multispectral / Hyperspectral	NIR, SWIR, and Visible (350–2,500 nm)	Leaf Area Index (LAI), chlorophyll absorption, plant nitrogen concentration	Photosynthetic capacity, nitrogen nutrition dynamics, grain protein content	Vegetation indices (NDVI, SAVI, GNDVI, MTVI), Partial Least Squares Regression
Thermal Infrared	Long-wave Infrared (8–14 μm)	Canopy temperature, stomatal conductance proxy	Transpiration rates, stomatal closure, early water stress responses	Crop Water Stress Index (CWSI), thermal-to-spatial calibration algorithms
Active LiDAR	Laser ranging (905–1,550 nm)	Canopy height, leaf angle distribution, 3D structures	Light interception efficiency, structural volume,	Voxelization, point cloud semantic segmentation, spatial density

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

			biomass density	clustering
Root Rhizotrons	Infrared transillumination (850 nm)	Main root length, lateral root density, gravitropic angles	Subsoil water extraction potential, nutrient foraging efficiency	nnUNet multi-class semantic segmentation, skeletonization, functional PCA

At the plant-organ scale, capturing below-ground root system architecture (RSA) has historically been a significant bottleneck in plant phenomics. Because root developmental plasticity determines a crop's capacity to forage for water and nutrients, understanding RSA is essential for breeding drought-resilient cultivars. The development of ChronoRoot 2.0, an open AI-powered platform for 2D temporal plant phenotyping, addresses this limitation (Bilkey, 2023).

By utilizing modular, low-cost imaging hardware equipped with Raspberry Pi Noir V2 cameras and an infrared (IR) backlight (850 nm), ChronoRoot 2.0 captures high-resolution imagery independently of diurnal light cycles. To process this data, the platform implements the nnUNet deep learning architecture for multi-class semantic segmentation (Lopez-Ruiz et al., 2017). This allows the platform to simultaneously segment and track six distinct seedling organs over time:

1. Main root
2. Lateral roots
3. Seed
4. Hypocotyl
5. Leaves
6. Petiole

Integrating functional principal component analysis (fPCA) directly into this time-series segmentation pipeline enables the quantification of complex temporal parameters. These parameters include gravitropic response rates and circadian growth oscillations, providing researchers with actionable traits to study physiological adaptation under climate stress (Gaggion et al., 2026).

3. Process-Based Crop Growth Models and Mathematical Formulations

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Process-based crop models use mathematical equations to represent the biophysical processes that drive crop growth, development, and yield formation. By simulating the interactions within the soil-plant-atmosphere continuum, these models help evaluate how different genetic parameters influence crop performance across diverse environments. Prominent modeling frameworks include the Decision Support System for Agrotechnology Transfer (DSSAT), the Agricultural Production Systems Simulator (APSIM), the World Food Studies (WOFOST) model, AquaCrop, and InfoCrop (Teh, 2006).

The core biophysical engine of these models calculates daily biomass accumulation (M) using either a carbon-driven or water-driven formulation. In carbon-driven models such as WOFOST and DSSAT, potential biomass accumulation is a function of the intercepted photosynthetically active radiation (I_{PAR}) and a genotype-dependent radiation use efficiency (RUE) (Salman et al., 2025):

$$\frac{dM(t)}{dt} = \mathrm{RUE} \times I_{\mathrm{PAR}}(t) \times \min\left(f_{\mathrm{water}}(t), f_{\mathrm{temp}}(t), f_{\mathrm{nitrogen}}(t)\right)$$

where $f_{\mathrm{water}}(t)$, $f_{\mathrm{temp}}(t)$, and $f_{\mathrm{nitrogen}}(t)$ are daily stress adjustment factors, bounded between 0 and 1, that represent limitations imposed by soil water deficits, thermal extremes, and nitrogen deficiencies, respectively. In contrast, water-driven models such as AquaCrop calculate daily biomass accumulation using water productivity (WP), which normalizes above-ground dry matter production relative to crop transpiration (Tr) and atmospheric evaporative demand (ET_0) (Deryng et al., 2011):

$$\frac{dM(t)}{dt} = WP \times \left(\frac{Tr(t)}{ET_0(t)}\right)$$

This water-driven formulation is highly effective for simulating biomass and yield dynamics under arid or water-limited conditions.

The simulated biomass is dynamically partitioned into various crop organs including leaves, stems, roots, and storage organs (grain) based on the crop's phenological development stage. Crop phenology is typically modeled as a function of thermal time accumulation, calculated by interpolating daily minimum and maximum temperatures. In the APSIM Plant module, for example, phenological progression is divided into 11 distinct crop stages and 10 growth phases (Brown e al., 2019).

The transition between these phases is governed by the accumulation of growing degree days

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

(GDD), which is adjusted to account for photoperiod sensitivity (including long-day, short-day, or qualitative responses) and vernalization requirements. Once the reproductive stage begins, the allocation of assimilates to the grain is often driven by a cultivar-specific daily harvest index increase (hi_incr), defining the sink strength of the reproductive organs (Anandhi, 2016).

In tropical and sub-tropical environments, the InfoCrop model serves as a user-friendly tool to simulate crop growth. InfoCrop processes edaphic, varietal, meteorological, and crop management inputs to model key agronomic outputs, while also simulating soil water balances, organic carbon dynamics, nitrogen transport, greenhouse gas emissions, and crop-pest interactions (Aggarwal et al., 2006).

Table 2: Functional Comparison of Major Process-Based Crop Models

Crop Model	Primary Biophysical Formulation	Input Data Requirements	Standard Simulated Outputs	Key Physiological Assumptions	Principal Computational Bottlenecks
DSSAT	Carbon-driven radiation use efficiency (RUE)	Daily weather, multi-layer soil properties, management schedules, cultivar coefficients	Leaf area index, phenological stages, biomass partitioning, grain yield	Growth responds linearly to temperature between cardinal thresholds	High sensitivity to genetic parameters that are difficult to measure in the field
APSIM	Modular, process-based carbon and nitrogen resource queues	Dynamic meteorological series, stratified soil physical/chemical profiles, management rotations	Transpiration, nitrogen uptake, root depth distribution, soil organic carbon	Resource uptake is limited by root surface area and soil water potential	High model dimensionality; execution can require minutes per run
WOFOST	Net leaf	Daily	Gross CO_2	Assimilate	Simplified

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

	carbon assimilation minus maintenance respiration	temperature, radiation, relative humidity, wind speed, soil water retention	assimilation, respiration, dry matter accumulation, crop yield	partitioning is a fixed function of phenological development stage	root water uptake mechanics that exclude 3D architecture
AquaCrop	Water-driven crop water productivity (WP)	Daily reference evapotranspiration (ET_0), temperature, rainfall, soil properties	Canopy cover (CC), crop transpiration, biomass, water use efficiency	Biomass accumulation is directly proportional to crop transpiration	Highly dependent on accurate initial soil moisture and hydraulic parameters
InfoCrop	Dynamic dry matter allocation coupled with crop–pest interactions	Edaphic properties, cultivar genetic data, pesticide/fertilizer schedules, weather	Crop duration, dry matter, nitrogen uptake, methane/nitrous oxide emissions	Yield is driven by source–sink relationships under pest and environmental stress	Simplified representation of crop canopy microclimates in diverse systems
SiriusQuality	Process-based wheat development and canopy microclimates	Hourly weather, nitrogen fertilizer schedules, canopy architectural traits	Intercepted radiation, anthesis date, grain protein content (GPC)	Grain protein content is determined by post-anthesis nitrogen translocation	Demands highly localized parameterization of leaf development and senescence

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

4. Algorithmic Data Assimilation and Scale-Mismatch Mitigation

While process-based crop models provide biophysical representations of plant growth, their regional and localized accuracy can be degraded by uncertainties in soil properties, weather inputs, and cultivar-specific parameters. Data assimilation (DA) addresses this challenge by combining real-time phenomic observations with dynamic crop simulation equations (Ramirez-Villegas et al., 2015).

Data assimilation strategies are broadly categorized into three methodologies:

- **Forcing Methods**: Direct replacement of simulated state variables with observed values on the date of observation. This simple approach can introduce physical mass-balance inconsistencies within the model (Divac et al., 2009).
- **Variational Assimilation (e.g., 4DVar)**: Adjusting the crop model's initial conditions and parameters within a defined temporal window to fit simulated growth curves to observed trajectories. While mathematically robust, variational methods are computationally intensive due to the requirement of calculating model adjoints (Wang et al., 2024).
- **Sequential Assimilation**: Dynamically updating model states at each time step when an observation becomes available. This is commonly achieved using the Ensemble Kalman Filter (EnKF), which represents state uncertainties using an ensemble of model runs (Moradkhani et al., 2005).

A major technical challenge in sequential data assimilation is the spatial scale mismatch between remote sensing observations and localized crop models. Crop simulation models typically simulate a single homogeneous site. In contrast, coarse-resolution remote sensing platforms, such as MODIS (which features a 250m to 1km pixel size), capture mixed spatial signals from heterogeneous fields. This mismatch can introduce errors when assimilating raw satellite data (Lahoz & Schneider, 2014).

To mitigate this scale-mismatch, researchers have implemented multi-step data assimilation workflows. For example, high-resolution Landsat TM imagery (30m resolution) has been combined with daily MODIS imagery using a Kalman filter algorithm to generate a synthetic, high-resolution time-series of Leaf Area Index (LAI) (Ruiying et al., 2026).

When this synthetic leaf area index (LAI) series is assimilated into the WOFOST crop growth model using an ensemble Kalman filter (EnKF) strategy, it provides a more realistic

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

representation of crop phenological dynamics. In regional field trials, this approach significantly improved winter wheat yield estimation accuracy ($R^2 = 0.43$, RMSE = 439 kg ha^{-1}) compared to three alternative baselines (Huang et al., 2016):

1. WOFOST without assimilation ($R^2 = 0.14$, RMSE = 647 kg ha^{-1})
2. Direct assimilation of sparse Landsat TM observations ($R^2 = 0.37$, RMSE = 472 kg ha^{-1})
3. Direct assimilation of course, filtered MODIS imagery ($R^2 = 0.49$, but with a very high RMSE = 1,355 kg ha^{-1} due to spatial sub-pixel contamination).

To bridge the gap between canopy-level observations and individual leaf-level parameters, researchers developed the "Digital Plant Phenotyping Platform" (D3P). The D3P workflow couples the Architectural model of DEvelopment based on L-systems (ADEL) wheat model which simulates the 3D canopy architecture of wheat crops with high-throughput image simulators (Sanchez, 2024).

Using this approach, the sequential assimilation of the green fraction derived from RGB images of the crop canopy allows for the estimation of five key genotypic parameters:

- The phyllochron (the thermal time interval between leaf appearances)
- The lamina length of the first leaf
- The rate of leaf lamina elongation
- The number of green leaves present at the start of leaf senescence
- The rate of leaf senescence.

In silico experiments conducted on wheat crops demonstrate that these micro-architectural parameters can be estimated using canopy-level RGB observations acquired at 0° and 60° zenith views with a temporal frequency of 100 °Cd (degree days) (Kundrat et al., 2010).

5. Case Studies in Field-Validated Phenomic Data Assimilation

To evaluate the real-world performance of these data assimilation workflows, several field-validated integration studies have been conducted across diverse environments. These studies demonstrate that combining high-throughput phenotyping with physiological modeling can improve yield prediction accuracy under environmental stress (Ghamkhar et al., 2019).

5.1. Durum Wheat Characterization via the SiriusQuality Model

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

A data assimilation pipeline was developed to estimate 12 key genotypic parameters governing crop phenology, canopy growth, light interception, and biomass accumulation in durum wheat (*Triticum durum*). High-throughput phenotyping was conducted using a ground-based Phenomobile, which captured time-series RGB imagery over eight years of multi-environment field trials involving nine distinct cultivars (Padovan et al., 2020).

The pipeline assimilated three key variables:

1. The fraction of intercepted photosynthetically active radiation (fIPAR)
2. Heading date
3. Final grain yield.

Two data assimilation strategies were compared: a Bayesian DREAM (zs) algorithm and a Lookup Table (LUT) inversion technique. The field validation results demonstrated that the LUT method achieved comparable accuracy to Bayesian inference while requiring significantly less computation (Cai et al., 2025).

Additionally, multi-year field trials showed that collecting data in just two contrasting environments (defined by site and year) was sufficient to characterize these genotypic parameters. This allowed for the reliable prediction of cultivar phenological development and yield stability across a wider range of environments, providing breeders with a tool to streamline genotype characterization (Leggett et al., 2015).

5.2. Precision Water Management and Yield Correction in Cotton

In cotton (*Gossypium hirsutum L.*), water stress during flowering and boll development can cause severe yield loss. To address this challenge, high-resolution Leaf Area Index (LAI) data derived from Unmanned Aerial System (UAS) multispectral imagery was assimilated into the DSSAT crop model (Ghorbanzadeh et al., 2025).

During the early squaring stage, the Normalized Difference Vegetation Index (NDVI) was used to estimate cotton LAI ($R^2 = 0.56$), while the Modified Triangle Vegetation Index (MTVI) and Enhanced Vegetation Index (EVI) provided more stable and accurate estimates during later growth stages (Peng et al., 2026).

The UAS-derived LAI values were assimilated into the DSSAT model using a Particle Swarm Optimization (PSO) algorithm. This workflow corrected spatial simulation biases, resulting in a more accurate representation of canopy development and transpiration under drought stress (Peng et al., 2026).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

A comparison of yield prediction errors before and after assimilation across three distinct irrigation levels is detailed in Table 3.

Table 3: DSSAT Cotton Yield Prediction Error Before and After UAS-LAI Assimilation

Soil Irrigation Level	Unassimilated DSSAT Prediction Error	Assimilated DSSAT Prediction Error	Root-Zone Soil Moisture Representation	Biophysical Correction Mechanism
30% Standard Irrigation	40.0% – 52.0%	3.6%	Corrected to capture rapid drought depletion in top soil layers	Adjusts the transpiration reduction factor, preventing overestimation of biomass under water stress
60% Standard Irrigation	40.0% – 52.0%	4.8%	Dynamic tracking of mild water-stress events	Calibrates daily stomatal resistance to match simulated water loss with observed canopy temperature
90% Standard Irrigation	40.0% – 52.0%	6.3%	Correctly identifies near-saturated soil water states	Constrains Leaf Area Index expansion, preventing the model from simulating excessive vegetative growth

By integrating high-resolution, multi-temporal phenotype parameters, the dynamic crop model corrected structural biases. This approach reduced yield prediction errors to under 7% across all irrigation levels, providing an effective framework for precision irrigation scheduling (Wang et al., 2025).

Online ISSN 3007-3197
Print ISSN 3007-3189

Similar UAS-based data assimilation studies have been validated in other crops. In canola (*Brassica napus*), the assimilation of high-resolution phenotype parameters reduced yield simulation errors from 452 kg/ha to 234 kg/ha (Yang et al., 2023).

In spinach, researchers demonstrated that the assimilation of LAI observations using the Red Edge Soil Adjusted Vegetation Index could accurately predict dry biomass yields, even when utilizing only two observation points in the middle growth stages, more than a month before harvest (Ochiai & Kamada, 2026).

6. Knowledge- and Data-Driven Modeling (KDDM) Topologies

To combine the strengths of both modeling paradigms, researchers have developed Knowledge- and Data-Driven Modeling (KDDM) frameworks. This approach integrates process-based models (PBMs) with machine learning (ML), improving predictive accuracy while preserving biophysical interpretability (Cerone, 2021). KDDM systems are designed using three primary topologies:

- **Parallel Structures**: The PBM and ML models receive the same input variables. The final KDDM output is calculated as either the sum or multiplication product of their respective predictions. In a sum configuration, the ML model is trained on the residual errors between real-world observations and the output of the biophysical model. For example, combining the 3D-architecture-based GreenLab crop model with an artificial neural network to predict tomato organ dry weights significantly reduced systematic simulation errors compared to using the PBM alone (Zhang et al 2023).

- **Serial Structures**: One model's output serves directly as the input or parameter value for the other. In a primary configuration, an ML model converts complex, unstructured sensor data into physiological parameters that are subsequently passed into a process-based model. This simplifies parameterization in cases where raw sensor signals cannot be directly processed by the crop model (Solle et al., 2017). In an alternative configuration, simulated biophysical variables from a process model are used as input features for a downstream machine learning predictor. For instance, coupling simulated soil moisture and crop water-stress variables from the APSIM model with a machine learning model significantly outperformed purely data-driven algorithms for regional yield prediction in the US Corn Belt (Feng et al., 2020).

- **Modular Structures**: Individual biophysical modules within a process-based model (such as light interception, transpiration, or root water uptake) are substituted with

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

localized ML models. This allows researchers to utilize high-performing data-driven algorithms for specific, complex processes, while retaining the overall biophysical framework of the crop model (Rao et al., 2026).

Figure 2. Conceptual framework of knowledge- and data-driven modeling (KDDM) under multifactorial abiotic stress, highlighting systemic data constraints, artificial intelligence (AI) interpretability limitations, and the biological realism gap in predicting emergent genotype-by-environment-by-management ($G \times E \times M$) phenotypes.

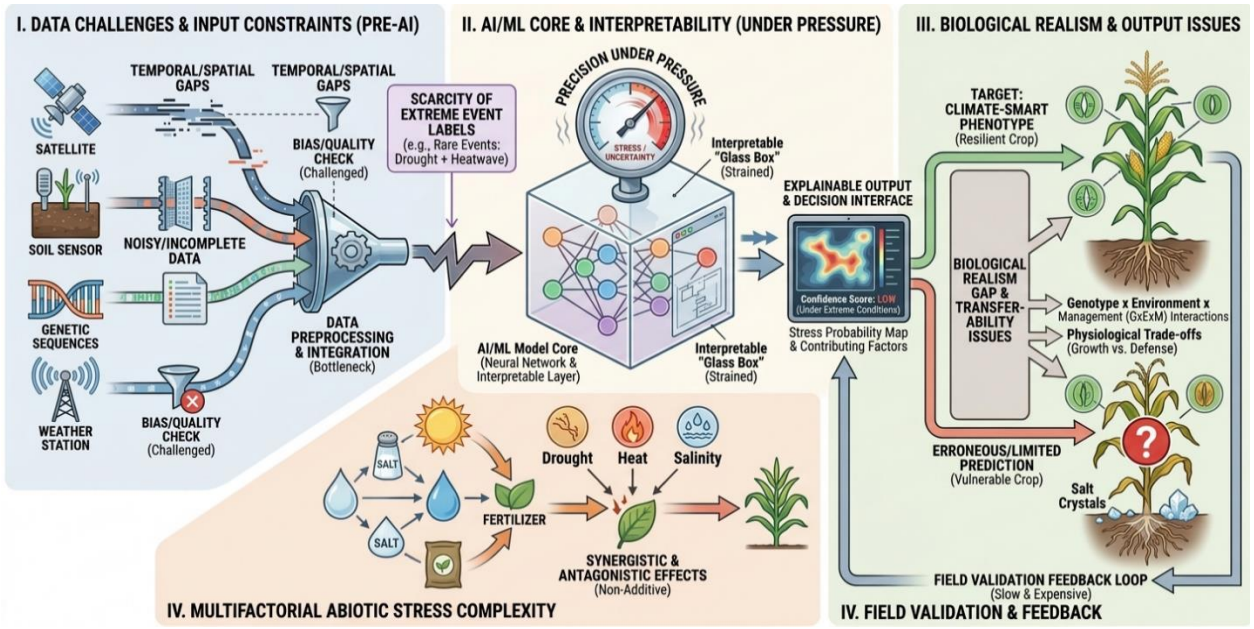


Table 4: Analytical Overview of KDDM Topologies and Physics-Informed ML Models

Modeling Strategy	Structural Architecture	Principal Physiological Applications	Primary Strengths	Identified Weaknesses
Parallel KDDM	PBM and ML run in tandem; final output is a combined function *Y*PBM ±*Y*ML	Correcting systematic biomass estimation errors in complex greenhouse	Reduces systematic model bias without altering underlying code	Does not resolve the parameterization bottlenecks of the process model

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

		tomato environments		
Serial KDDM	One model's output serves as the input or parameter value for the other	Mapping raw hyperspectral sensor data to crop parameters, which are then passed to the PBM	Enables the integration of unstructured sensor data into biophysical models	Vulnerable to error propagation if the upstream model introduces systematic biases
Modular KDDM	Replaces specific biophysical modules of a PBM with targeted ML algorithms	Replacing empirical root-zone soil water uptake routines with an LSTM module	Enhances simulation accuracy for highly complex or poorly characterized processes	Demands highly standardized interfaces to ensure mass and energy conservation
AgriPINN (Process-Informed)	Biophysical ODEs are embedded as differentiable constraints in deep learning	Regional above-ground biomass (AGB) and water-stress prediction in winter wheat	Unsupervised recovery of latent variables (LAI, RUE); 8x faster than LINTUL5	Demands differentiability of the underlying biophysical equations

KDDM frameworks are also applied to construct deep learning emulators of complex biophysical models. While process-based models like APSIM capture detailed crop-soil-climate interactions, they are computationally intensive. A single APSIM simulation run can require approximately two minutes, making large-scale spatial simulations or multi-generation breeding evaluations computationally expensive (Shen et al., 2023).

To address this limitation, researchers have developed probabilistic neural network emulators. In a representative study, an APSIM emulator was trained on two million simulations generated using a Sobol low-discrepancy sequence across 22 genetic, soil, and management parameters (Lueckmann et al., 2019).

The resulting deep learning emulator predicted 13 key crop growth and yield outputs with high

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

accuracy ($R^2 = 0.93$). By deploying this emulator on high-performance computing systems, such as the Frontera supercomputer, simulation times were reduced by several orders of magnitude. This allows researchers to rapidly evaluate crop performance across millions of genetic and environmental combinations (Proctor, 2021).

Conclusion

The integration of AI-assisted phenotyping with process-based crop physiological models represents a transformative advancement in modern agricultural research and precision farming. High-throughput sensing technologies now enable continuous, non-destructive measurement of structural, physiological, and biochemical crop traits across multiple spatial and temporal scales, substantially reducing the long-standing phenotyping bottleneck. Simultaneously, process-based crop models provide mechanistic insight into plant growth, resource allocation, and environmental responses, while advanced data assimilation techniques bridge the gap between observed phenotypes and simulated crop dynamics. The emergence of Knowledge- and Data-Driven Modeling (KDDM) further enhances predictive performance by combining the biological interpretability of mechanistic models with the adaptive learning capabilities of artificial intelligence.

Evidence from recent field applications demonstrates that integrating phenomic observations into crop models significantly improves the accuracy of biomass estimation, yield prediction, irrigation management, and stress detection across diverse crops and environments. These integrated systems also accelerate genotype evaluation and facilitate climate-smart breeding by improving the characterization of genotype × environment × management interactions.

Despite these advances, several challenges remain, including standardized phenotypic data acquisition, interoperability among sensing platforms and crop models, computational complexity, model transferability across environments, and the interpretability of deep learning algorithms. Addressing these limitations will require interdisciplinary collaboration among plant scientists, agronomists, computer scientists, remote sensing specialists, and data engineers.

Overall, AI-assisted phenotyping coupled with physiological crop modeling provides a powerful foundation for next-generation digital agriculture. Continued advances in sensor technologies, explainable artificial intelligence, digital twins, and physics-informed machine learning are expected to further enhance crop resilience, optimize resource-use efficiency, and support sustainable food production under increasingly variable climatic conditions.

Online ISSN 3007-3197
Print ISSN 3007-3189

7.Refrences

Aggarwal, P. K., Kalra, N., Chander, S., & Pathak, H. (2006). InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. I. Model description. *Agricultural systems*, *89*(1), 1-25.

Alexander, C., Barber, D., & Hutchison, R. (1984). Chondrule rims and interchondrule matrix in UO C's. *Meteoritics, Vol. 19, p. 184*, *19*, 184.

Anandhi, A. (2016). Growing degree days–Ecosystem indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. *Ecological Indicators*, *61*, 149-158.

Angidi, S., Madankar, K., Tehseen, M. M., & Bhatla, A. (2025). Advanced high-throughput phenotyping techniques for managing abiotic stress in agricultural crops A comprehensive review. *Crops*, *5*(2), 8.

Bilkey, N. (2023). *Mechanisms and Consequences of Microtubule-Based Symmetry Breaking in Plant Roots* (Doctoral dissertation, Washington University in St. Louis).

Brown, H. E., Huth, N. I., Holzworth, D. P., Teixeira, E. I., Wang, E., Zyskowski, R. F., & Zheng, B. (2019). A generic approach to modelling, allocation and redistribution of biomass to and from plant organs. *in silico Plants*, *1*(1), diy004.

Cai, D., Zhu, C., Wang, X., Manceau, L., Jezequel, S., Marguerie, M., ... & Martre, P. (2025). Predictions of wheat phenotypic variability by integrating high-throughput phenotyping observations into a crop growth model. *Plant Phenomics*, 100149.

Cerone, A. (2021, December). Ten years of DataMod: The synergy of data-driven and model-based approaches. In *International Symposium: From Data to Models and Back* (pp. 7-24). Cham: Springer International Publishing.

Chauhan, S., & Singh, A. (2025). High-throughput phenotyping and phenomics for exploration of plant growth and development. In *Plant high-throughput phenotyping and functional phenomics* (pp. 155-174). CRC Press.

Deryng, D., Sacks, W. J., Barford, C. C., & Ramankutty, N. (2011). Simulating the effects of climate and agricultural management practices on global crop yield. *Global biogeochemical cycles*, *25*(2).

Divac, D., Milivojević, N., Grujović, N., Stojanović, B., & Simić, Z. (2009). A procedure for state updating of SWAT-based distributed hydrological model for operational runoff

Online ISSN 3007-3197
Print ISSN 3007-3189

forecasting. *Journal of the Serbian Society for Computational Mechanics*, *3*(1), 298-326.

Elias, A. A., Robbins, K. R., Doerge, R. W., & Tuinstra, M. R. (2016). Half a century of studying genotype× environment interactions in plant breeding experiments. *Crop Science*, *56*(5), 2090-2105.

Feng, P., Wang, B., Li Liu, D., Waters, C., Xiao, D., Shi, L., & Yu, Q. (2020). Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agricultural and Forest Meteorology*, *285*, 107922.

Gaggion, N., Boccardo, N. A., Bonazzola, R., Legascue, M. F., Mammarella, M. F., Rodriguez, F. S., ... & Ferrante, E. (2026). ChronoRoot 2.0: an open AI-powered platform for 2D temporal plant phenotyping. *GigaScience*, *15*, giag018.

Ghaffar, A., Rahman, M. H. U., Ahmed, S., Haider, G., Ahmad, I., Khan, M. A., ... & Ahmed, A. (2022). Adaptations in cropping system and pattern for sustainable crops production under climate change scenarios. In *Improvement of plant production in the era of climate change* (pp. 1-34). CRC Press.

Ghamkhar, K., Irie, K., Hagedorn, M., Hsiao, J., Fourie, J., Gebbie, S., ... & Barrett, B. (2019). Real-time, non-destructive and in-field foliage yield and growth rate measurement in perennial ryegrass (Lolium perenne L.). *Plant Methods*, *15*(1), 72.

Ghorbanzadeh, Z., Panahi, B., Purhang, L., Hossein Panahi, Z., Zeinalabedini, M., Mardi, M., ... & Ghaffari, M. R. (2025). Integrative Genomics and Precision Breeding for Stress-Resilient Cotton: Recent Advances and Prospects. *Agronomy*, *15*(10), 2393.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., ... & Wu, W. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agricultural and Forest Meteorology*, *216*, 188-202.

Kundrat, M. E. (2010). *A Comprehensive Series for Predicting Bone Dynamics: Forecasting Osseous Tissue Formation Using the Molecular Structure of a Biomaterial* (Doctoral dissertation, University of Dayton).

Lahoz, W. A., & Schneider, P. (2014). Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, *2*, 16.

Leggett, M., Newlands, N. K., Greenshields, D., West, L., Inman, S., & Koivunen, M. E. (2015). Maize yield response to a phosphorus-solubilizing microbial inoculant in field trials. *The Journal of agricultural science*, *153*(8), 1464-1478.

Online ISSN 3007-3197
Print ISSN 3007-3189

Li, D., Quan, C., Song, Z., Li, X., Yu, G., Li, C., & Muhammad, A. (2021). High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Frontiers in bioengineering and biotechnology*, *8*, 623705.

Liang, X., Yu, S., Ju, Y., Wang, Y., & Yin, D. (2025). Multi-scale remote-sensing phenomics integrated with multi-omics: Advances in crop drought–heat stress tolerance mechanisms and perspectives for climate-smart agriculture. *Plants*, *14*(18), 2829.

Lopez-Ruiz, N., Granados-Ortega, F., Carvajal, M. A., & Martinez-Olmos, A. (2017). Portable multispectral imaging system based on Raspberry Pi. *Sensor Review*, *37*(3), 322-329.

Lueckmann, J. M., Bassetto, G., Karaletsos, T., & Macke, J. H. (2019, January). Likelihood-free inference with emulator networks. In *Symposium on advances in approximate Bayesian inference* (pp. 32-53). PMLR.

Moradkhani, H., Hsu, K. L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water resources research*, *41*(5).

Ndlovu, H. S., Odindi, J., Sibanda, M., & Mutanga, O. (2024). A systematic review on the application of UAV-based thermal remote sensing for assessing and monitoring crop water status in crop farming systems. *International Journal of Remote Sensing*, *45*(15), 4923-4960.

Ochiai, S., & Kamada, E. (2026). Assimilating Unmanned Aerial Vehicle-based Leaf Area Index Observations into a Spinach (Spinacia oleracea) Dry Yield Prediction Model for Processing Applications. *The Horticulture Journal*, *95*(1), 70-79.

Padovan, G., Martre, P., Semenov, M. A., Masoni, A., Bregaglio, S., Ventrella, D., ... & Dibari, C. (2020). Understanding effects of genotype× environment× sowing window interactions for durum wheat in the Mediterranean basin. *Field Crops Research*, *259*, 107969.

Peng, H., Esirige, Gu, H., Gao, R., Zhou, Y., Men, X., & Wang, Z. (2026). Estimation of Cotton LAI and Yield Through Assimilation of the DSSAT Model and Unmanned Aerial System Images. *Drones*, *10*(1), 27.

Peng, H., Esirige, Gu, H., Gao, R., Zhou, Y., Men, X., & Wang, Z. (2026). Estimation of Cotton LAI and Yield Through Assimilation of the DSSAT Model and Unmanned Aerial System Images. *Drones*, *10*(1), 27.

Proctor, K. W. (2021). Development of a Statistical Emulator of US Crop Yields via a Deep Neural Network Approach.

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Ramirez-Villegas, J., Watson, J., & Challinor, A. J. (2015). Identifying traits for genotypic adaptation using crop models. *Journal of Experimental Botany*, *66*(12), 3451-3462.

Rao, S., Ahmadi, S. H., Leitner, D., Vianna, M., Bauer, F. M., Seidel, S. J., & Schnepf, A. (2026). Integrated Modeling Approaches for Agricultural Digital Twins: The Role of PBM, ABM, ML, and Model Coupling. *in silico Plants*, diag002.

Ruiying, R., Zhang, K., Liang, Q., Maocheng, Z., Weijun, X., Zhou, C., & Li, M. (2026). Monitoring Carbon Stock Change at the Individual-Plant Scale: A Methodological Review and Integrative Framework. *Forests*, *17*(5), 563.

Salman, M., García-Vila, M., Fereres, E., Raes, D., & Steduto, P. (2021). *The AquaCrop model–Enhancing crop water productivity: Ten years of development, dissemination and implementation 2009–2019* (Vol. 47). Food & Agriculture Org..

Sanchez, J. R. (2024). *Automated unmanned systems and data analytics for In-field digital crop phenotyping* (Doctoral dissertation, University of Georgia).

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., ... & Lawson, K. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552-567.

Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., ... & Steckenreiter, T. (2017). Between the poles of data-driven and mechanistic modeling for process operation. *Chemie Ingenieur Technik*, *89*(5), 542-561.

Teh, C. B. (2006). *Introduction to mathematical modeling of crop growth: How the equations are derived and assembled into a computer model*. Dissertation. com.

Tharan, S. T., Chandan, M., Vidyarani, J. S., Swetha, M., Srinivasamurthy, S., Kumar, P. K., & Prasath, K. R. (2026). Accelerating Climate Resilience in Vegetables: The Synergistic Role of Genomics and Phenomics. *International Journal of Environment and Climate Change*, *16*(4), 674-689.

Wang, J., Wang, Y., & Qi, Z. (2024). Remote sensing data assimilation in crop growth modeling from an agricultural perspective: new insights on challenges and prospects. *Agronomy*, *14*(9), 1920.

Wang, P., Zou, S., Li, J., Ju, H., & Zhang, J. (2025). Advancing Water Quality Management: Harnessing the Synergy of Remote Sensing, Process-Based Models, and Machine Learning to Enhance Monitoring and Prediction. *Remote Sensing*, *17*(18), 3157.

Wang, T., Tong, R., Xu, T., Li, Y., & Chen, Y. (2025). Artificial intelligence in plant science: from image-based phenotyping to yield and trait prediction. *Frontiers in Plant*

Online ISSN 3007-3197
Print ISSN 3007-3189

Science, *16*, 1732979.

Yang, Y., Wei, X., Wang, J., Zhou, G., Wang, J., Jiang, Z., ... & Ren, Y. (2023). Prediction of seedling oilseed rape crop phenotype by drone-derived multimodal data. *Remote sensing*, *15*(16), 3951.

Zandi, A., Hosseinirad, S., Kashani Zadeh, H., Tavakolian, K., Cho, B. K., Vasefi, F., ... & Tavakolian, P. (2025). A systematic review of multi-mode analytics for enhanced plant stress evaluation. *Frontiers in Plant Science*, *16*, 1545025.

Zhang, N., Zhou, X., Kang, M., Hu, B. G., Heuvelink, E., & Marcelis, L. F. (2023). Machine learning versus crop growth models: an ally, not a rival. *AoB Plants*, *15*(2), plac061.