in silico Plants, 2026, 8, diag002
https://doi.org/10.1093/insilicoplants/diag002
Published: 5 February 2026
Review

in silico Plants OXFORD

Integrated Modeling Approaches for Agricultural Digital twins: the role of Process Based Models, Agent Based Models, Machine Learning, and Model Coupling

Sathyanarayan Rao [1], Seyed Hamid Ahmadi [2,3], Daniel Leitner [4], Murilo Vianna [4], Felix Maximilian Bauer [4], Sabine Julia Seidel [5], Andrea Schnepf [4*]

[1]Civil Engineering, Indian Institute of Science, Mirinji Marg, Sir C.V. Raman Road, CNR Rao Circle, Malleshwaram, Bengaluru, Karnataka 560012, India
[2]Water Engineering Department, Faculty of Agriculture, Shiraz University, Badjgah Campus, Shiraz, Fars Province 71441-13131, Iran
[3]Drought Research Center, Shiraz University, Badjgah Campus, Shiraz, Fars Province 71441-13131, Iran
[4]Agrosphere Institute, (IBG-3) Institute of Bio- and Geosciences Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich, Germany
[5]Institute of Organic Farming (IFÖL), BOKU University, Gregor-Mendel-Haus Gregor-Mendel-Straße, Vienna 33 1180, Austria

[*]Corresponding author. Agrosphere Institute, (IBG-3) Institute of Bio- and Geosciences Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich, Germany. E-mail: a.schnepf@fz-juelich.de

Abstract

Process-based models (PBMs) provide a mechanistic foundation for simulating complex genotype, environment, and management interactions across multiple scales. Integrating PBMs with agent-based models (ABMs), machine learning (ML), and advanced model coupling strategies (sequential, loose, tight, and framework-based) enables more comprehensive representations of agricultural systems. ABMs capture the adaptive behaviours of individual farmers, while ML techniques efficiently approximate nonlinear physiological processes or act as surrogates for computationally expensive submodules and reveal data-driven patterns, collectively enhancing the simulation of G × E × M interactions under variable climate scenarios. Digital twins (DTs), defined as systems that dynamically synchronize virtual models with real-time sensor data, extend coupled PBM–ABM–ML systems by enabling bidirectional feedback between computational models and farm operations. However, the implementation of DTs remains constrained by challenges such as data integration across heterogeneous sources, computational scalability, and semantic/technical interoperability making them appropriate only where continuous decision support and real-time feedback outweigh complexity and cost. This review evaluates the function of PBMs as the mechanistic core of DT architectures, surveys model integration techniques, and outlines the IT infrastructure required for operationalization. We conclude that digital twins are best deployed when real-time feedback is essential, whereas PBM–ABM–ML couplings suffice for research and policy applications requiring long-term scenario analysis.

Keywords: process-based models, digital twins, agricultural modeling, crop modeling, machine learning

1. Introduction

Agriculture faces urgent challenges as climate variability, resource scarcity, and global population growth heighten the demand for resilient, productive cropping systems capable of adapting to complex genotype-by-environment-by-management (G × E × M) interactions (Yang et al. 2017). Advancing the understanding and prediction of genotype × environment × management (G × E × M) interactions requires modelling frameworks that operate across multiple biological, spatial, and temporal scales (Hernández-Ochoa et al. 2022, Janni et al. 2024).

Process-based models (PBMs) play a central mechanistic role in agricultural research by simulating plant physiological and biophysical processes in response to environmental and management stressors. Their explicit representation of underlying processes from the cellular level (Vos et al. 2010, Schnepf et al. 2018) to fields and regional agroecosystems (Wu et al. 2021) enables quantitative predictions for crop adaptation, water and nutrient use, and the development of evidence-based sustainable practices.

PBMs extend their analytical scope when coupled with agent-based models (ABMs) and machine learning (ML) techniques. ABMs explicitly simulate diverse human decision-making processes, capturing how individual farmers, policies, and social contexts shape agroecosystem behaviours (Troost and Berger 2015). ML complements mechanistic and behavioural approaches by efficiently identifying data-driven patterns, enabling surrogate modelling for computationally expensive simulations, and supporting rapid data assimilation to improve predictions under uncertainty (Droutsas et al. 2022, Zohdi 2024). PBMs, ABMs, and ML thus address distinct computational and representational requirements in agricultural systems analysis.

Digital twins (DTs) are dynamic, virtual replicas of physical agroecosystems that are continuously synchronized with real-time sensor, field, and remote sensing data (Peladarinos et al. 2023, Mitsanis et al. 2024). DTs integrate PBMs, ABMs, and ML into a unified execution environment driven by observational data. In this architecture, PBMs simulate biophysical interactions, ABMs represent adaptive behaviours, and ML accelerates computation through surrogate modelling and data assimilation (Pylianidis et al. 2021, Verdouw et al. 2021).

However, DTs are not without limitations. Despite their transformative potential for real-time scenario analysis and adaptive decision-support, DT deployment faces significant hurdles, including data integration across heterogeneous sources, substantial computational and infrastructural demands, and challenges associated with semantic and technical interoperability between diverse modelling frameworks and IT infrastructures (Wallor et al. 2018, Vereecken et al. 2022, Storm et al. 2024). Consequently, DT adoption is typically justified in tactical contexts requiring continuous feedback, rapid adaptation, or real-time management. In contrast, conventional PBMs, ABMs, or their couplings are often more suitable for long-term analysis when data availability or computational resources are limited.

This review examines the roles and coupling strategies of PBMs, surveys recent advancements in integrating ABMs and ML, and clarifies scenarios where DT implementation is beneficial or where complexity and cost constraints outweigh their advantages. Additionally, we identify major technical barriers and unresolved methodological challenges in developing and operationalizing coupled PBM–ABM–ML systems and agricultural DTs. The intended audience comprises agricultural researchers, modellers, and practitioners interested in multi-scale, interdisciplinary modelling approaches for systems analysis and decision-support. While a full exploration of sensor technologies and IT infrastructures is beyond our scope, we briefly outline essential IT considerations as relevant to model coupling and operational deployments.

The paper is structured as follows. Section 2 introduces the key modelling approaches and their principles. Section 3 compares model coupling strategies across agricultural contexts. Section 4 discusses DTs as an integrative paradigm, including their benefits, limitations, and deployment scenarios. Section 5 outlines current challenges and future perspectives, and Section 6 concludes with actionable recommendations for advancing resilient agroecosystem modelling.

2. Process-based models in agroecosystems

2.1 Definition and role of process-based models in simulating G × E × M interactions

Genotype × Environment × Management (G × E × M) interactions represent the complex interplay between genetic traits, environmental conditions, and management practices that collectively determine crop performance (Cooper et al. 2021, Cooper and Messina 2021). PBMs provide mechanistic frameworks to quantitatively simulate genotype–environment–management interactions by explicitly representing biological and physical processes such as photosynthesis, nutrient uptake, root development, and soil–plant–atmosphere exchanges (Yang et al. 2017, Hajjarpoor et al. 2022).

Unlike empirical statistical models that rely primarily on correlations derived from historical data, PBMs explicitly encode biological processes by integrating genotype-specific physiological parameters with dynamic environmental drivers. This mechanistic foundation enables PBMs to generate reliable predictions of crop performance even under novel climate scenarios or management practices that fall outside historical observations, a critical advantage for climate adaptation research (Li et al. 2021, Rasool et al. 2022, Lopez et al. 2023). For instance, PBMs like CPlantBox mechanistically represent root architecture and function, allowing plant responses to variations in soil moisture or nutrient availability to be predicted across genotypes and management regimes (Schnepf et al. 2018, Giraud et al. 2023).

2.2 Application of process-based models across scales

PBMs have been developed and applied across a wide range of spatial scales, from cellular processes within plant tissues to regional agroecosystem dynamics. The scale distinctions reflect dominant processes and typical modelling objectives at each level, rather than a rigid classification. Figure 1 and Table 1 provide an overview of representative PBMs across scales, summarizing their inputs, core processes, outputs, and typical applications.

At the cellular scale, PBMs focus on resolving fine-scale anatomical and physiological processes that govern water and nutrient transport. Models such as MECHA, GRANAR, and RootSlice explicitly represent root tissue organization and

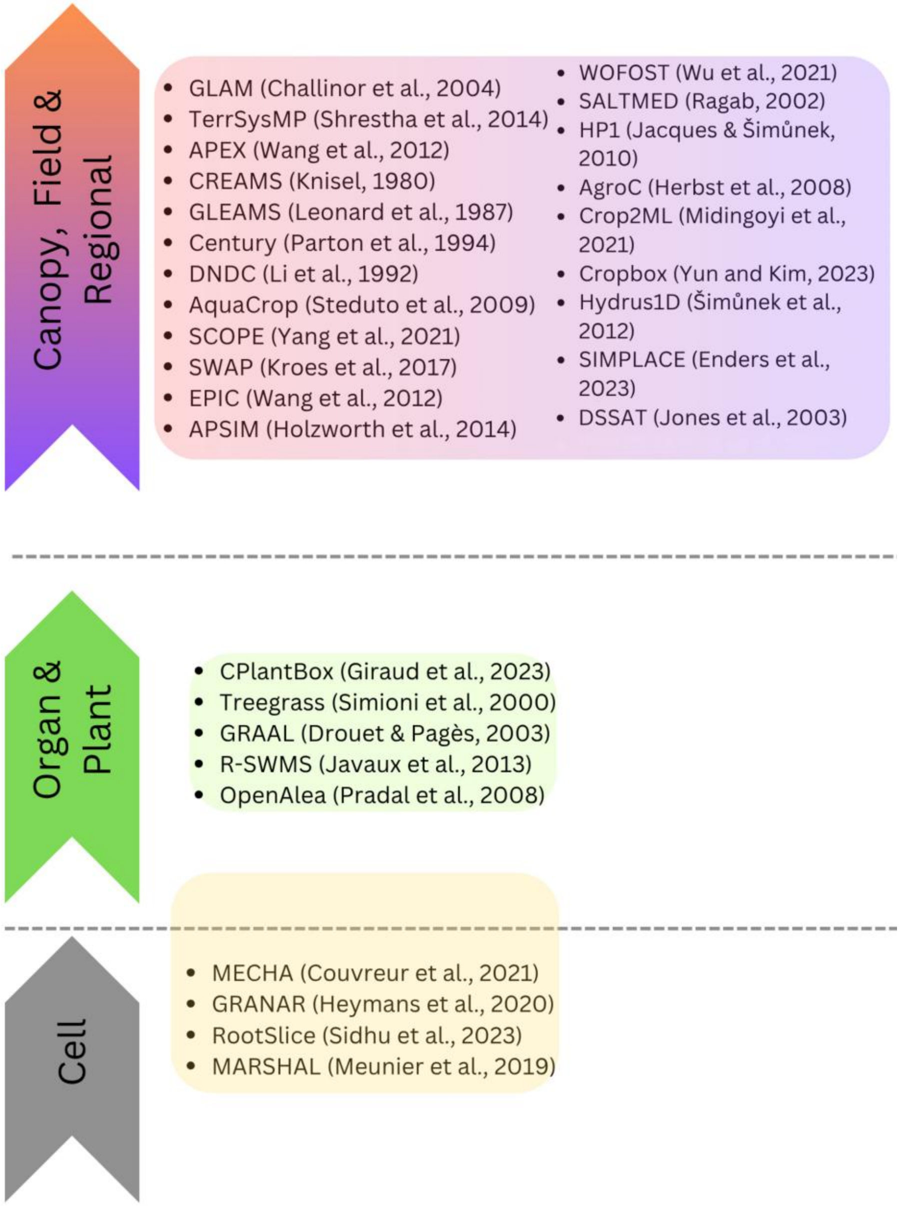

Figure 1 Overview of process-based models (PBMs) frequently applied at different spatial scales in agricultural research, from cellular to regional scales. Although models are grouped according to their typical application scale, some frameworks such as SIMPLACE have been used across scales via modular component assembly; others like Crop2ML supports cross-model component reuse. Thus, this categorization is indicative rather than strictly prescriptive.

cellular pathways, enabling mechanistic analysis of hydraulic conductivity and nutrient uptake efficiency (Heymans et al. 2020, Couvreur et al. 2021, Sidhu et al. 2023). These models link anatomical traits with functional performance, enabling virtual phenotyping and trait-based screening under controlled stress scenarios.

At the organ and whole-plant scale, PBMs increasingly adopt functional structural plant modelling (FSPM) approaches that couple plant architecture with physiological processes. Models such as CPlantBox, R-SWMS, MARSHAL, and GRAAL simulate root and shoot development alongside water and carbon fluxes, providing mechanistic links between genotype-specific traits and plant-level responses to soil and climatic conditions (Drouet and Pagès 2003, Javaux et al. 2008, Schnepf et al. 2018, Meunier et al. 2020). These approaches mark a transition from empirical growth representations towards explicit structural and hydraulic descriptions, increasingly integrated with phenotyping platforms for trait estimation and stress response analysis (Jin et al. 2020).

At the field, canopy, and regional scales, PBMs integrate soil, climate, crop physiology, and management processes to predict yield, water use, and nutrient dynamics under diverse conditions. Widely used models such as APSIM, DSSAT, WOFOST, AquaCrop, and SWAP exemplify this class, supporting applications ranging from field-scale management optimization to regional yield forecasting (Jones et al. 2003, Holzworth et al. 2014, Kroes et al. 2017). Beyond yield prediction, a notable trend at these scales is the coupling of PBMs

Table 1 Process-based models across scales: inputs, core processes, outputs, applications, and websites.

Scale	Representative PBM/Platform	Typical inputs	Core processes (examples)	Typical outputs	Typical applications	Website
Cellular	MECHA	Root anatomy maps, cell properties	Radial/axial root hydraulics	Kr, Kx, tissue conductivities	Drought physiology; trait screening	https://mecharoot.github.io
Cellular	GRANAR	Microscopy-derived cell layers, diameters	Digital root cross-section generation	Annotated anatomy maps	Virtual phenotyping; inputs to hydraulics	https://granar.github.io
Cellular	RootSlice	Root cross-section images/traits	Anatomical modelling, uptake proxies	Anatomy metrics; uptake efficiency proxies	Nutrient uptake at cell scale	https://plantscience.psu.edu/…/rootslice
Organ/Plant	CPlantBox (FSPM)	Genotype params, soil profile, climate	3D root–shoot architecture; water & carbon fluxes	LAI, transpiration, assimilation, RSA	Virtual phenotyping; stress response	https://github.com/Plant-Root-Soil-Interactions-Modelling/CPlantBox
Organ/Plant	R-SWMS	Soil hydraulic props, root geometry	3D soil–root water flow	Root water uptake profiles	Plant–soil hydraulics	https://soil-modeling.org/resources-links/model-portal/r-swms
Organ/Plant	MARSHAL	Root hydraulic/structural traits	Root hydraulics + growth	Hydraulic traits; virtual phenotypes	Drought tolerance screening	https://marshal-root.github.io/
Organ/Plant	GRAAL	Morphology & C allocation params	Organ growth + carbon allocation	Organ biomass, phenology	Systems physiology	https://hal.inrae.fr/hal-02663400
Organ/Plant	OpenAlea (platform)	Component configs, plant–soil data	Component-based FSPM prototyping	Custom model components/graphs	FSPM prototyping & teaching	https://openalea.rtfd.io
Field/Canopy	WOFOST	Daily weather, soil, management, cultivar	Phenology, photosynthesis, water balance	Biomass, yield, ET	Field management; yield prediction	https://www.wur.nl/…/WOFOST.htm
Field/Canopy	AquaCrop	Weather, soil, irrigation, cultivar	Water-driven growth; soil water balance	Yield, WP, ET	Irrigation scheduling; water productivity	https://www.fao.org/aquacrop/en
Field/Canopy	SWAP	Soil hydraulic curve, meteo, cropping	1D soil–water–atmosphere flow	Soil moisture/temp profiles, ET	Coupling with crop models	https://www.swap.alterra.nl/
Field/Canopy	SCOPE	Meteorology, canopy traits (LAI, Cab, etc.)	RTM + energy balance + fluorescence	Fluxes, SIF, Tc	RS–canopy coupling; stress detection	https://scope-model.readthedocs.io/en/master/
Field–Regional	APSIM	Weather, soils, management, genetics	Modular crop/soil/rotation processes	Yield, N fluxes, water balance	Farm systems; regional scenarios	https://www.apsim.info
Field–Regional	DSSAT	Weather, soil, management, cultivar	Multi-crop system simulation	Yield, phenology, resource use	Global crop assessment	https://dssat.net/

(*continued*)

Table 1 Continued

Scale	Representative PBM/Platform	Typical inputs	Core processes (examples)	Typical outputs	Typical applications	Website
Regional	GLAM	Climate, soils, crop calendars	Yield–climate response functions	Regional yield anomalies	Climate risk analysis	https://environment.leeds.ac.uk/climate-change-impacts/doc/general-large-area-model-annual-crops
Regional	TerrSysMP	Atm./land/hydro boundary conds	COSMO–CLM–ParFlow coupling	Land–atmosphere fluxes, states	Fully coupled E–H systems	https://www.terrsysmp.org
Field–Regional	APEX	Weather, soils, management	Watershed hydrology & nutrients	Runoff, N/P, yield	Watershed BMPs; policy	https://epicapex.tamu.edu/apex/
Field–Regional	DNDC	Climate, soil, management	C–N biogeochemistry	GHG emissions, N cycling	Emissions, mitigation	https://www.dndc.sr.unh.edu/
Field–Regional	AgroC	Weather, soil C/N, crop params	Soil C turnover × crop growth	Soil C pools, fluxes, yield	Carbon budgets; coupling with RTMs	https://www.phenorobdaa.de/agroc/
Field	SALTMED	Soil salinity, irrigation quality	Salinity–water–crop growth	Yield, soil salinity dynamics	Salinity management	https://www.ceh.ac.uk/data/software-models/saltmed
Field/Canopy	PROSAIL	Leaf biochemistry (chlorophyll, water, dry matter), canopy structure (LAI, angle), soil reflectance, sun–sensor geometry	Radiative transfer through leaves and canopy	Canopy reflectance spectra, vegetation indices	Estimation of canopy traits, vegetation monitoring, sensor design	https://jbferet.gitlab.io/prosail/index.html
Canopy/Field	SPART	Soil reflectance, canopy traits, atmosphere (aerosols, water vapour), geometry	Radiative transfer across soil, canopy, atmosphere	Surface and TOA reflectance, canopy parameters	Remote sensing inversion, vegetation monitoring	https://github.com/peiqiyang/SPART

with hydrological and radiative transfer models, such as SCOPE and GLAM, to better represent canopy energy balance and large-area crop responses to climate variability (Yang et al. 2021).

Importantly, several modelling frameworks demonstrate cross-scale flexibility, challenging strict scale boundaries. SIMPLACE, for example, enables modular assembly of crop and soil components across spatial domains, while APSIM and DSSAT have been successfully applied beyond field plots to regional assessments (George-Jaeggli et al. 2022, Enders et al. 2023). Similarly, Crop2ML facilitates component reuse across models without prescribing scale-specific performance (Midingoyi et al. 2021). This modularity enables the specific coupling strategies analysed in Section 3.

2.3 Current challenges with process-based models

Despite their strong mechanistic foundations, PBMs face persistent limitations that restrict their accuracy and operational value, particularly in data-scarce or heterogeneous environments. The main limitations arise from two sources: conceptual simplifications in process representation and practical constraints related to computation and scalability.

2.3.1 Representation and realism limits

A primary weakness of PBMs is their simplified treatment of biotic interactions and management realities. While these models excel at simulating potential crop growth, they often lack mechanistic descriptions of pests, diseases, beneficial soil biota, and the nuances of farm-level decision-making. As a result, PBMs often fail to reproduce observed yield variability, a discrepancy that cannot always be attributed to model structure alone.

Yield gaps frequently arise from a combination of uncertainties: farmer-reported management data is often imprecise, soil properties vary spatially, and transient biotic stresses go undocumented (Cooper et al. 2021, George et al. 2024). Even sophisticated PBMs struggle in these contexts because they assume idealized conditions that rarely exist in the field. This limitation underscores the difficulty of capturing agroecosystem dynamics using deterministic equations alone and motivates coupling PBMs with agent-based or data-driven components.

2.3.2 Computational and scalability limits

Beyond representational gaps, PBMs face a steep trade-off between process fidelity and computational feasibility. Detailed models that resolve 3D root architectures or fine-scale soil hydrology are computationally expensive even at the plot scale, yet regional assessments require simulating thousands of such units simultaneously (Marshall-Colon et al. 2017, Enders et al. 2023). Such computational demands limit the scalability of mechanistic models to landscape and regional applications.

The inclusion of hydrological processes further intensifies this challenge, particularly when accounting for preferential flow, hysteresis, or coupled heat and solute transport (Wallor et al. 2018, Vereecken et al. 2022). Similarly, management practices like tillage which fundamentally alter soil physical and biochemical properties are difficult to simulate mechanistically and are often handled through static

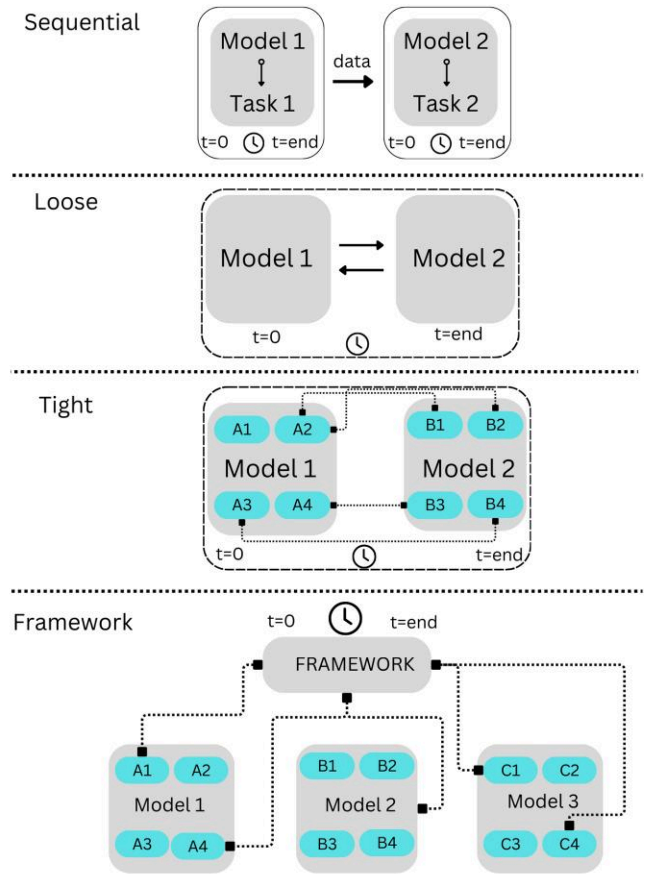

Figure 2 Schematic representation of coupling strategies: Sequential, Loose, Tight, and Framework-based coupling. Sequential coupling involves one-way information flow from an upstream model to a downstream model, whereas Loose and Tight coupling allow bidirectional feedback between interacting models. Models are illustrated with internal submodules (A1–A4, B1–B4, C1–C4), highlighting explicit interaction in Tight and Framework-based coupling.

empirical factors. To mitigate these costs, recent studies have turned to strategic simplification, such as using reduced-order models or surrogate machine-learning emulators. However, the rigid input structures of many legacy PBMs still hinder the dynamic parameter updating required for real-time applications. Addressing these computational and flexibility barriers is a prerequisite for deploying PBMs within advanced digital twin frameworks (Peladarinos et al. 2023, Storm et al. 2024), a topic further explored in Sections 3 and 4.

3. Model coupling strategies

Model coupling integrates PBMs across biological, environmental, and spatial scales, enhancing predictions of agroecosystem dynamics (Herbst et al. 2008). Four principal coupling strategies are commonly implemented: sequential, loose, tight, and framework-based coupling, each characterized by distinct data exchange mechanisms, interdependencies, and computational demands (Jacques and Šimůnek 2010, Li et al. 2012, Midingoyi et al. 2021). Figure 2 illustrates key differences in data flow and model autonomy.

Table 2 Comparison of coupling types.

Attribute	Sequential coupling	Loose coupling	Tight coupling	Framework-based coupling
Dependency (How strongly models depend on each other's outputs)	**One-way data flow**; downstream model uses results from upstream model, but upstream is unaffected.	**Periodic exchange**; models pass information at set intervals, but remain largely independent.	**Mutual dependence**; models share state variables at each timestep and influence each other continuously.	**Configurable dependence**; models integrated within a software environment, with dependency rules defined by the framework.
Flexibility (Ease of modifying or updating one model without breaking the coupling)	**High flexibility**; individual models can be swapped or updated independently, since coupling is minimal.	**Moderate flexibility**; updating one model requires adjustments to synchronization routines.	**Low flexibility**; models are tightly interwoven, making substitution or modification difficult.	**Variable flexibility**; depends on framework design; modular frameworks (e.g. SIMPLACE) enable easier updates.
Scalability (Ease of scaling to large domains or multiple models)	**Limited scalability**; suitable for small case studies or sequential workflows.	**Moderate scalability**; can handle larger datasets but limited by synchronization overhead.	**Challenging scalability**; computationally intensive due to iterative, step-wise exchanges.	**High scalability**; designed for multi-model integration and domain expansion.
Complexity (Implementation and computational cost)	**Low complexity**; simple to implement and computationally light.	**Moderate complexity**; requires synchronization management.	**High complexity**; intensive coding and computation needed.	**Variable complexity**; initial setup effort high, but efficient once framework is established.
Time Synchronization (How and when information is exchanged between models)	**Explicit**; downstream run starts after upstream model completes.	**Explicit at defined intervals**; e.g. daily or hourly exchanges.	**Implicit and iterative**; models exchange information at each timestep.	**Flexible**; both explicit and implicit approaches possible, depending on framework.
Typical Example(s)	HYDRUS-1D → WOFOST (Li et al. 2012)	AgroC ↔ SCOPE (De Cannière et al. 2021)	CPlantBox ↔ DuMux (Giraud et al. 2023)	SIMPLACE integrating phenology, soil water, and N cycling (Enders et al. 2023)

The choice of coupling strategy determines how information is exchanged between models and therefore constrains both scientific scope and practical feasibility. Sequential and loose couplings favour modular experimentation and batch-style analyses, whereas tight and framework-based couplings enable high-frequency feedbacks across interacting domains. These distinctions determine latency and synchronization capabilities, which are critical for DTs and interoperability with external data streams.

Existing reviews typically address coupling strategies either in general environmental modelling or within specific agricultural applications. We compare coupling strategies specifically in the context of agricultural PBMs.

Table 2 and Fig. 2 summarize the main attributes and synchronization schemes. The following Sections 3.1–3.4 illustrate each strategy using representative case studies.

3.1 Sequential coupling: examples

A representative example of sequential coupling at the cellular-to-organ scale is the integration of the GRANAR and MECHA models, designed to mechanistically link root anatomy with root hydraulic function.

Generator of root anatomy in R (GRANAR) is an open-source model that generates realistic digital representations of root cross-sectional anatomy based on minimal experimental input, such as the number of cell layers and their dimensions from root imaging data. Users can parameterize anatomical traits (e.g. cortex thickness, stele diameter, number of xylem poles) from either direct measurement or literature, and GRANAR outputs detailed digital root cross-sections, assigning cell type and geometry for each cell (Heymans et al. 2020). These anatomies are output in formats compatible with downstream hydraulic models (Figs 3, 4).

Model of explicit cross-section hydraulic anatomy (MECHA), in turn, ingests the anatomical maps produced by GRANAR and computes radial and axial hydraulic conductivities (Kr and Kx) by numerically simulating water movement across the multicellular root cross-section. MECHA explicitly accounts for the hydraulic properties of different tissues and cell walls, as well as the spatial arrangement of cell types, suberin, and other barriers (Couvreur et al. 2021). By applying biophysical principles (such as Ohm's law analogy for water flow and considering apoplastic,

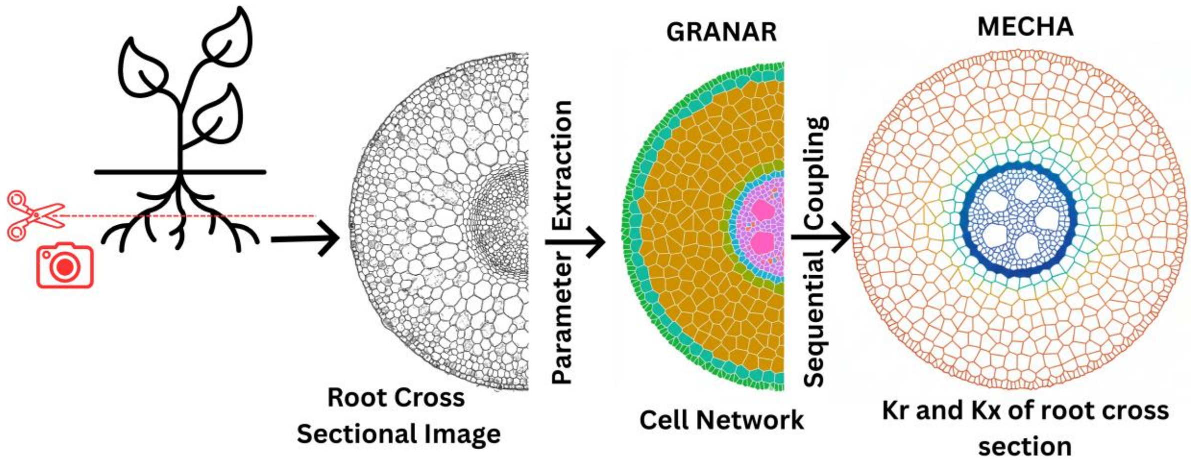

Figure 3 Sequential coupling between GRANAR and MECHA models for root anatomy analysis. Part of this figure generated using MECHA and GRANAR tool (Couvreur et al. 2021).

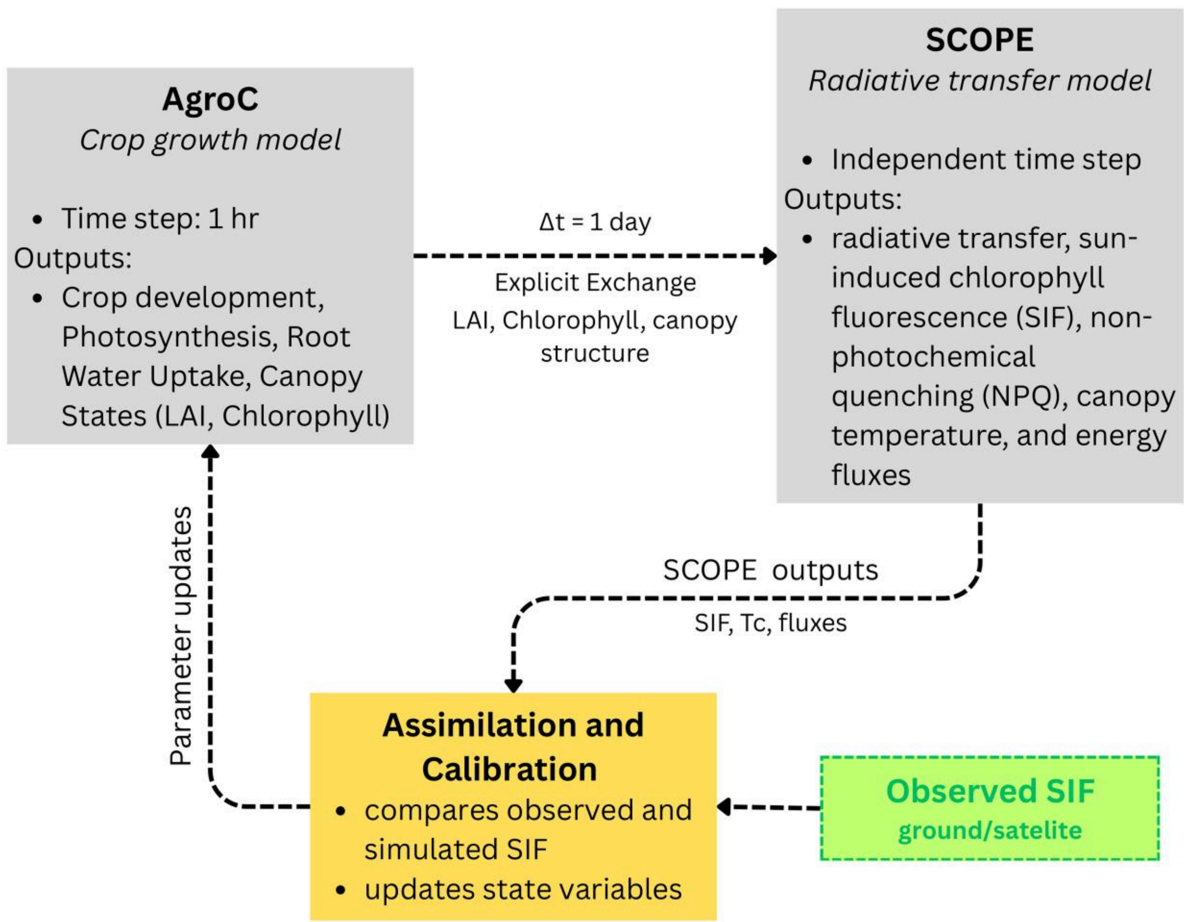

Figure 4 Loose coupling workflow between AgroC (crop growth) and SCOPE (radiative transfer, energy balance, fluorescence). At explicit synchronization intervals ($\Delta t = 1$ day), AgroC provides canopy state (e.g. LAI, chlorophyll, canopy structure) to SCOPE. SCOPE simulates SIF and fluxes, which are compared with observed SIF (ground/satellite). The innovation drives calibration/state updates in AgroC (e.g. water stress function), while both models retain independent internal timesteps. Solid arrows: model-to-model exchange; dashed arrows: observations/ assimilation; dashed loop: update back to AgroC.

symplastic, and transcellular pathways), MECHA estimates how anatomical features influence overall root water uptake efficiency.

The sequential workflow proceeds as follows:

1. GRANAR being used to generate parameterized digital root anatomies, either by sampling traits from microscopy or literature;
2. Exporting these anatomies as input files;

Table 3 Common interface variables in agricultural model coupling, their typical domains, and representative roles.

Exchanged/interface variable	Example coupling cases	Brief description/role
Soil moisture and water fluxes	Soil–plant hydrology couplings (Li et al. 2012)	Link soil water availability with transpiration, growth, and yield under drought/irrigation
Root length density (RLD)	Root architecture ↔ soil transport models (Schnepf et al. 2018)	Quantify soil exploration and uptake efficiency; improve nutrient and water acquisition
Root water uptake and plant hydraulics	Plant hydraulics ↔ soil hydrology (Giraud et al. 2023)	Capture dynamic feedback between soil water potential and plant transpiration under stress
Nutrient concentrations (N, P, etc.)	Soil nutrient cycling ↔ crop growth (Herbst et al. 2008)	Ensure crop growth reflects nutrient availability, uptake efficiency, and management
Evapotranspiration (ET)	Crop growth ↔ atmosphere/canopy models	Integrate water use with climate drivers (temperature, humidity, radiation)
Leaf area index (LAI)	Plant growth ↔ remote sensing/canopy radiative transfer (Yang et al. 2021)	Represent canopy structure for biomass accumulation; connect to vegetation indices
Stomatal conductance	Plant physiology ↔ soil–atmosphere processes (Giraud et al. 2023)	Regulate transpiration, photosynthesis, and water use strategies
Soil organic matter (SOM)	Soil C cycling ↔ plant productivity (Herbst et al. 2008)	Couple soil fertility and carbon turnover with crop growth
Canopy temperature and energy balance	Crop growth ↔ atmosphere/energy balance (Jones and Vaughan 2010)	Link plant stress responses with microclimate and radiative transfer
Soil temperature	Soil heat transfer ↔ crop/respiration (Herbst et al. 2008)	Capture effects on biological activity, growth, and nutrient cycling
Carbon fluxes (CO_2 exchange)	Plant–soil ↔ atmosphere (Herbst et al. 2008)	Couple photosynthesis and respiration for ecosystem-scale C balances

3. MECHA importing these files and performing detailed hydraulic simulations to output axial and radial conductivities, which are key for assessing genotype differences in water uptake efficiency under drought or nutrient stress.

This sequential coupling is strictly unidirectional: changes or variability in anatomy generated by GRANAR directly affect the hydraulic properties computed by MECHA, but not vice versa. Each tool remains modular, facilitating high-throughput virtual phenotyping of root anatomical and physiological traits (Heymans et al. 2020, Couvreur et al. 2021). The GRANAR–MECHA coupling enables quantification of how specific anatomical phenes contribute to water and nutrient acquisition (Table 3).

Other applications. Beyond anatomical modelling, sequential coupling is widely used at the field scale, particularly to enhance soil–crop interactions. A well-known example is the coupling of ‘HYDRUS-1D’ with ‘WOFOST’, where detailed soil moisture profiles simulated by HYDRUS-1D are passed unidirectionally to WOFOST to update transpiration constraints and yield formation under water stress (Li et al. 2012). In such applications, feedback from crop growth to soil hydraulic properties is assumed negligible relative to the study objectives, making sequential coupling a pragmatic and computationally efficient choice. This coupling addresses limitations of simplified soil water representations in crop models by providing physically based soil moisture dynamics under water stress (Li et al. 2012). Traditionally, crop models such as WOFOST employ simplistic ‘tipping bucket’ approaches to soil water flow, limiting their ability to accurately capture water dynamics affecting crops, particularly under water stress. Sequentially coupling HYDRUS-1D, which numerically solves the Richards equation for variably saturated water transport, with a crop model enables a more mechanistic representation of soil moisture stress on plant growth and yield.

The typical sequential workflow involves running HYDRUS-1D first to simulate detailed daily soil moisture profiles across the soil column. These profiles are then passed as input to the WOFOST model at regular, pre-defined intervals usually daily. WOFOST utilizes these soil moisture data to update transpiration, growth, and yield computations. The data flow is strictly unidirectional: soil hydrology (HYDRUS-1D) → crop growth (WOFOST), with no feedback from the crop model to the hydrology model during the same cycle. Exchanged variables usually include soil moisture in each soil layer, actual transpiration, and available water for roots. After these variables are provided to WOFOST, there is no immediate feedback; any adjustment occurs in the subsequent simulation cycle.

3.2 Loose coupling: examples

Loose coupling is illustrated by the integration of the AgroC crop growth model with the SCOPE radiative transfer and fluorescence model (De Cannière et al. 2021). In this approach, the AgroC model simulates crop development, photosynthesis, and root water uptake, while SCOPE models the transfer of energy, photosynthetic activity, and the emission of sun-induced chlorophyll fluorescence (SIF) from the crop canopy. Each model operates independently, using its own time-stepping and internal routines. At pre-defined synchronization intervals, typically daily, AgroC outputs key crop parameters such as leaf area index, chlorophyll content, and canopy structure, which are then provided as input to SCOPE.

SCOPE uses these AgroC-derived parameters to simulate radiative transfer and calculate SIF, as well as physiological quantities like non-photochemical quenching and maximum carboxylation capacity (VCMax) that reflect crop water stress. The resulting SIF time series, simulated by SCOPE, are then compared to observed SIF (from ground-based or satellite sensors) to calibrate AgroC's internal water stress function. This process constrains AgroC's simulation of photosynthesis and evapotranspiration under drought, improving the realism of water and carbon flux predictions. This workflow preserves the modularity and autonomy of both models, exchanging data only at scheduled intervals without requiring continuous or iterative feedback.

Other applications. Loose coupling is a standard strategy for linking distinct process domains without enforcing full numerical integration. It has been widely applied to couple 'HYDRUS-1D' with crop models for irrigation and salinity management (Hao et al. 2015, Kanda et al. 2018). A related example is the 'HP1' framework, which links HYDRUS-1D with 'PHREEQC' to couple water and solute transport with geochemical reactions through periodic variable exchange (Jacques and Simunek 2010). In these systems, explicit but infrequent data exchange allows interaction while maintaining model independence.

3.3 Tight coupling: selected case studies

Tight coupling denotes high-frequency, bidirectional exchange at each computational step with in-memory data sharing and, typically, an implicit (iterative-to-convergence) synchronization. A representative example of tight coupling in soil–plant–atmosphere modelling is provided by the integration of the functional–structural plant model CPlantBox with the DuMux simulation framework (Giraud et al. 2023). In this system, plant and soil domains are not treated as independent modules with periodic data exchange, but as dynamically interacting components within a single simulation environment. The tight coupling is achieved through direct, in-memory exchange of physiological and environmental variables at each computational time step, with plant growth, water flow, photosynthesis, stomatal conductance, and soil hydrodynamics all computed in a unified loop.

At each simulation step, DuMux calculates the three-dimensional soil water potential field, accounting for the effects of plant water uptake and atmospheric demand. These updated soil variables are immediately made available to CPlantBox, which then computes the distribution of water potentials and flows throughout the plant architecture, including detailed xylem and phloem transport, transpiration, and carbon assimilation. The computed root water uptake profiles, as well as changes in plant water status, are then directly communicated back to the soil module, closing the feedback loop without recourse to external file I/O. The two systems iterate within each time step until convergence is achieved, allowing the simulation to capture rapid feedbacks such as the onset of drought stress, recovery after rainfall, or phenotypic plasticity in response to fluctuating environmental conditions. At each coupling step, the exchanged variables include the soil matric potential field ($\psi_s(\mathbf{x})$) transferred to the plant, and the plant's root water uptake $S(\mathbf{x})$, xylem/phloem flows, and leaf water potential (ψ_l) transferred back to the soil. The synchronization is implicit, iterating to convergence within each global timestep.

This tight coupling allows simulation of coupled soil–plant–atmosphere processes, including coordinated water and carbon fluxes and transient stress responses. For example, shifts in carbon allocation, dynamic root-to-shoot ratio adjustment, and coordinated water and carbon fluxes under transient stress. The model architecture is designed to allow user-defined spatial and temporal resolution, and to accommodate a wide variety of plant types and environmental scenarios, making it well-suited for hypothesis testing, virtual phenotyping, and digital twin development. The high-frequency, bidirectional coupling is supported by a unified codebase with shared data structures and Python bindings, ensuring computational efficiency and robust process integration.

Other applications. Similar tight coupling strategies have been implemented in the SOILCO2–RothC system, which links soil water and heat transport with carbon turnover to simulate heterotrophic respiration under dynamic environmental conditions (Herbst et al. 2008). Likewise, STEMMUS–SCOPE tightly integrates canopy radiative transfer, photosynthesis, and energy balance with soil–root water flow, enabling coupled simulation of carbon, water, and energy fluxes (Wang et al. 2021). While such approaches substantially improve physical realism, they also impose high computational and implementation costs, motivating the use of framework-based coupling in broader applications (Section 3.4).

3.4 Framework-based coupling: selected case studies

Framework-based coupling integrates multiple models within a single software architecture, where components are executed inside a shared runtime environment. In contrast to sequential or loose coupling, which relies on file exchange between separate executables, framework-based approaches embed model components directly within a common computational framework. This enables in-memory interaction, coordinated execution, and systematic management of shared state variables.

A suitable example of this approach is Scientific Impact Assessment and Modelling Platform for Advanced Crop and Ecosystem Management (SIMPLACE; Enders et al. 2023). In SIMPLACE, framework-based coupling is achieved by organizing soil, plant, and atmosphere processes into modular components with explicitly defined interfaces (Fig. 5). Each module encapsulates its internal process representations while exchanging key state variables and fluxes through standardized coupling interfaces. This architecture allows alternative process formulations to be combined and compared without modifying the underlying source code.

During execution, all variables are centrally managed within a shared data container (VarMap). Using a shared runtime and centralized variable management makes it possible to compare alternative process formulations under identical inputs. As a result, SIMPLACE supports both modular

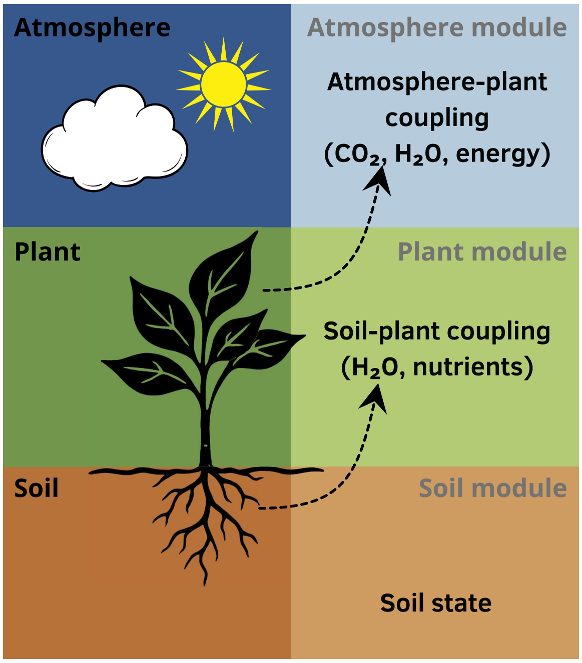

Figure 5 Conceptual representation of modular soil–plant–atmosphere coupling in framework-based crop modelling platforms. The atmosphere, plant, and soil are represented as interacting modules, with explicitly defined coupling interfaces for the exchange of mass and energy. The plant module acts as a central integrator, receiving environmental forcing from the atmosphere and resource supply from the soil, while internal process representations remain encapsulated within individual modules.

experimentation and scalable integration across crop, soil, and management domains.

Other applications. Framework-based coupling is also used outside crop modelling, particularly in remote sensing and Earth system modelling. In remote sensing, the radiative transfer model PROSAIL integrates the leaf optical properties model PROSPECT with the canopy reflectance model SAIL within a unified code structure, enabling coherent simulation of leaf- and canopy-scale radiative processes (Jacquemoud et al. 2009). The SPART model extends this framework by incorporating soil and atmospheric radiative transfer, allowing end-to-end simulation of surface reflectance from the canopy to the sensor (Yang et al. 2017).

At the ecosystem scale, platforms such as TerrSysMP employ a standardized interface layer to integrate atmospheric (COSMO), land-surface (CLM), and hydrological (ParFlow) models within a single execution environment, enabling coordinated land–atmosphere–hydrology simulations (Shrestha et al. 2014).

While typically classified as PBMs, major agricultural platforms such as APSIM and Crop2ML also adopt modular framework architectures internally, using standardized component definitions and shared runtime structures to facilitate process reuse and interoperability. In this review, APSIM and Crop2ML are discussed primarily as PBMs, whereas PROSAIL, SPART, and TerrSysMP are used to illustrate framework-based coupling across distinct model domains (Holzworth et al. 2014, Midingoyi et al. 2021).

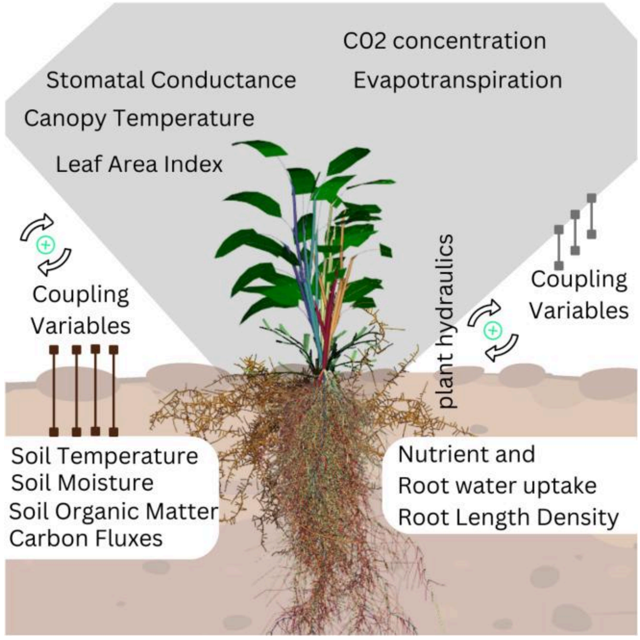

Figure 6 Interface variables between soil–plant and atmosphere in model coupling. This diagram showcases a CPlantBox-generated plant, detailing key interface variables that connect atmospheric and soil models. Variables such as soil moisture, stomatal conductance, and leaf area index are placed in context around a central plant illustration to indicate their relevance in different environmental domains.

3.5 Key variables in model for coupling opportunities

Despite differences in implementation, all coupling strategies rely on the exchange of a core set of interface variables. These variables connect soil, plant, and atmosphere submodels, or serve as training data and calibration targets in hybrid approaches (e.g. PBM–ML integrations). Summarizing these variables helps clarify where meaningful interactions occur across scales and disciplines.

These interface variables constitute the data exchange layer for hybrid couplings and the digital twin architectures discussed in Section 5. Figure 6 provides a visual overview of these interfaces, mapping their roles within the plant–soil–atmosphere continuum and highlighting their importance in model coupling.

4. Integrating machine learning and agent-based models with PBMs

The coupling techniques discussed in Section 3, such as sequential, loose, tight, and framework-based coupling, have primarily been illustrated through examples of integrating different PBMs. These strategies also apply to the integration of ML models and ABMs with PBMs. Similar coupling methodologies can also be applied to integrate other modelling approaches with PBMs, most notably ML models and ABMs.

ML and ABMs address specific computational and behavioural limitations of PBMs. ML primarily addresses computational scalability and pattern extraction, whereas ABMs

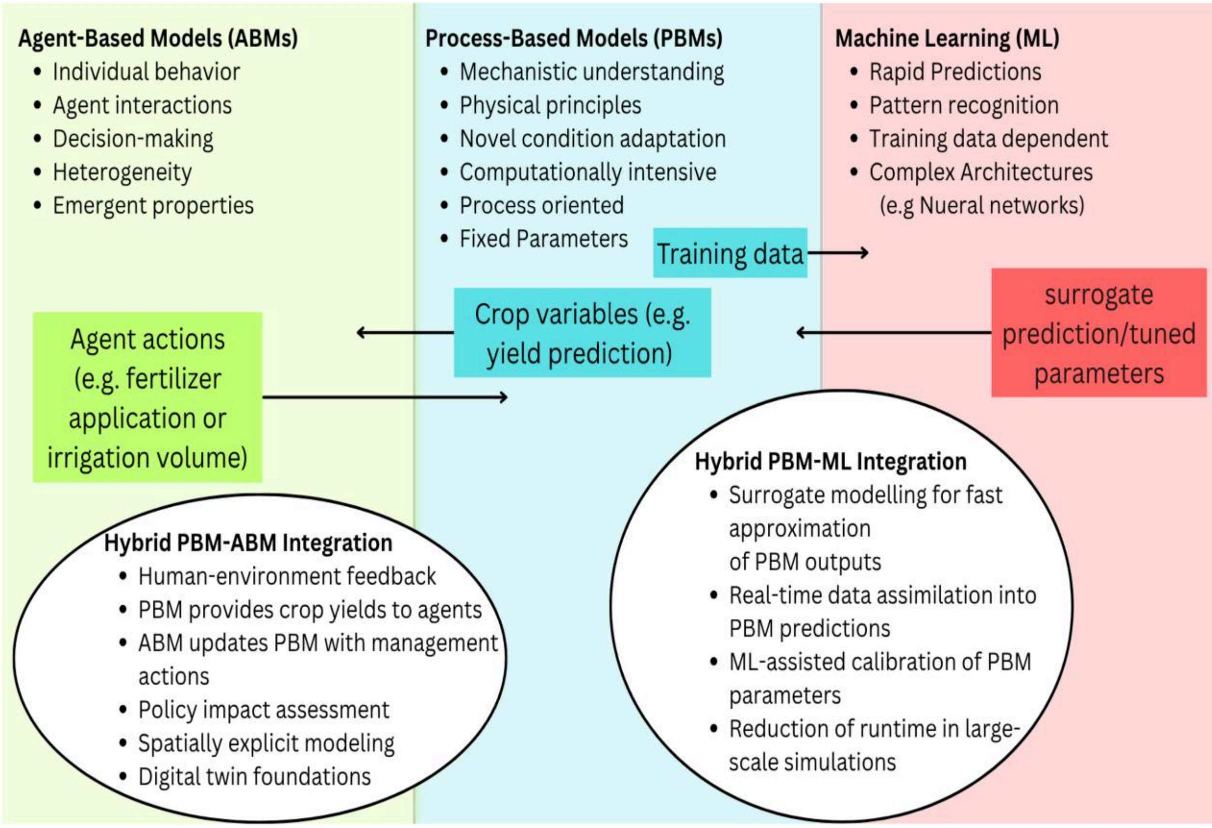

Figure 7 Complementary integration approaches with process-based models. This figure illustrates how process-based models (PBMs) can be integrated with both agent-based models (ABMs) and machine learning (ML) to address different aspects of agricultural system modelling. PBMs provide the mechanistic foundation, while ABM integration enhances human decision-making representation and ML integration improves computational efficiency and prediction accuracy.

introduce behavioural realism and institutional heterogeneity that PBMs alone cannot represent.

ML models are commonly used as surrogate models to approximate computationally expensive process-based simulations. Their primary strength lies in approximating complex biophysical processes through learned input–output relationships, allowing predictions to be generated much more rapidly and with lower computational cost than full mechanistic simulations. For example, ML surrogates have been used in precision agriculture to identify suitable crop choices based on soil conditions and management practices, thereby reducing reliance on costly and time-consuming field trials (Bhat et al. 2023).

At the same time, the performance of surrogate models depends strongly on the quality and representativeness of their training data. To remain accurate and relevant, ML surrogates must be retrained or updated as PBMs evolve or as new experimental insights become available. This need for continual updating underscores that ML models are not replacements for PBMs, but rather accelerators that extend their practical applicability, particularly in contexts where computational or data constraints would otherwise limit analysis.

Conversely, ABMs simulate agricultural systems by representing individual decision-makers, such as farmers, each operating under distinct rules and constraints. By explicitly modelling human choices and interactions, ABMs add a critical layer to PBMs, capturing how individual and collective decisions shape agricultural outcomes (Storm et al. 2024). Rather than assuming uniform behaviour, ABMs rely on decision rules informed by survey data, census information, or behavioural theory, enabling representation of diverse real-world practices such as irrigation scheduling, crop selection, and technology adoption.

As illustrated in Fig. 7, ML and ABMs address different limitations of PBMs rather than duplicating their function. Unlike ML models, which are primarily driven by statistical correlations, ABMs explicitly represent feedback between human decisions and environmental dynamics. This makes them particularly valuable for exploring how policy changes, market conditions, or climate variability influence farm-level decisions and long-term system trajectories.

4.1 Machine learning integration with PBMs

The integration of ML with PBMs is primarily motivated by the need to address persistent limitations in mechanistic modelling, particularly the strong sensitivity of PBMs to site-specific inputs (Wallor et al. 2018) and the substantial uncertainty associated with simulating soil–plant hydrological processes (Vereecken et al. 2022). Hybrid ML–PBM approaches replace selected mechanistic components with data-driven approximations while retaining the overall process-based structure

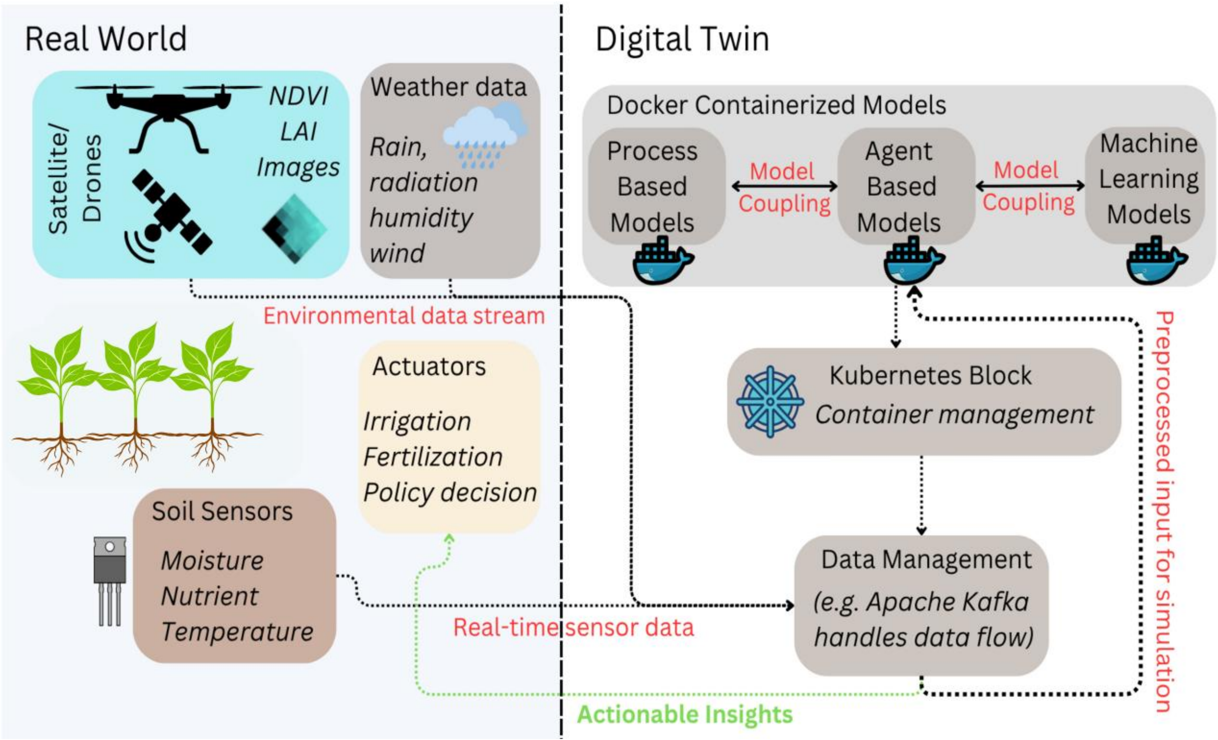

Figure 8 Environmental and crop observations from sensors, weather feeds, and drones stream into a broker that buffers and routes data in real time. Kafka plays that broker role so models receive timely, ordered inputs. Docker packages each model as a portable service so PBMs, ABMs, and ML components can be run, updated, and reproduced consistently. Kubernetes schedules and scales those containers and keeps them healthy. The models exchange state through defined interfaces, and the loop closes when the twin's outputs drive field actions such as irrigation or fertilization.

(Droutsas et al. 2022). Increasingly, these hybrid approaches are viewed as foundational components for adaptive modelling frameworks and DTs (Pylianidis et al. 2021, Verdouw et al. 2021).

Several recent studies illustrate how ML surrogates can reduce computational cost while preserving model behaviour. Shang et al. (2024) showed that neural network surrogates of farm-scale process models can achieve speedups of up to five orders of magnitude while retaining predictive accuracy. Similarly, Droutsas et al. (2022) replaced simplified empirical stress functions within PBMs with ML-based predictors of radiation use efficiency and phenology, leading to more robust simulations under high-temperature stress. Beyond computational acceleration, ML has also enhanced predictive performance in data-rich environments. Sun et al. (2022) coupled the GEPIC crop model with a Random Forest algorithm to improve soybean yield predictions under climate extremes, while Drees et al. (2024) developed a two-stage CWGAN framework linking process-based simulations with image-driven trait estimation, thereby bridging mechanistic modelling and visual phenotyping.

Operational deployment of ML–PBM hybrids remains constrained by maintenance and generalizability requirements. Storm et al. (2024) identify hybrid learning systems that integrate expert knowledge with ML as a priority for next-generation agricultural modeling, but they also emphasize the difficulty of maintaining such systems over time. Unlike PBMs, which evolve through refinement of physical equations, ML components require retraining whenever the underlying data distribution, model structure, or process understanding changes. Their performance remains tightly coupled to the quality and representativeness of training data, raising concerns about robustness when extrapolating to novel climate conditions or management regimes. Consequently, successful ML–PBM integration requires explicit frameworks for updating, validating, and governing ML components alongside evolving mechanistic models (Fig. 8).

4.2 Coupling agent-based models with process-based models

While ML integration primarily addresses computational efficiency and predictive uncertainty, coupling PBMs with ABMs targets a different limitation, namely the absence of explicit human decision-making in most PBMs. ABM–PBM integration extends the modelling scope by representing individual farmers, institutions, or agents operating under heterogeneous constraints and behavioural rules. This enables simulation of adaptive responses to environmental variability and policy interventions that cannot be captured through biophysical processes alone.

The value of this approach was demonstrated by Troost and Berger (2015), who coupled ABMs with PBMs to reproduce regional agricultural supply functions. By explicitly representing crop rotation constraints and heterogeneous farmer decisions, their framework captured emergent market dynamics that are typically missed by aggregate economic models. Such integrations are particularly relevant for assessing adaptation strategies and policy impacts. Streefkerk et al. (2023) applied the ADOPT-AP framework in Kenya by combining spatially explicit hydrological models with behavioural simulations to analyse differentiated drought adaptation strategies among agropastoralists. Similarly, Shang et al. (2024) linked detailed process-based farm models with the AgriPoliS ABM to evaluate agri-environmental policies, allowing individual farm decisions to be traced through to regional environmental outcomes. These studies show that ABM–PBM coupling captures decision heterogeneity and policy responses that are absent from biophysical models alone (Storm et al. 2024).

Despite this potential, integrating ABMs with PBMs presents substantial methodological challenges. As noted by Verdouw et al. (2021) and Pylianidis et al. (2021), ABMs and PBMs typically operate at different spatial and temporal scales and rely on fundamentally different modelling paradigms. Ensuring meaningful feedback between human decisions and biophysical responses therefore requires careful coupling design. In addition, behavioural parameters are often difficult to observe directly, necessitating advanced calibration and validation approaches such as Bayesian inference or pattern-oriented modelling. Nevertheless, when combined with ML–PBM integration, ABM–PBM coupling forms a complementary framework in which ML mitigates computational and parametric uncertainty, while ABMs ensure behavioural and policy relevance. ABM–PBM integrations convert PBMs from explanatory tools into systems capable of simulating adaptive management and policy scenarios.

5. Digital twins in agriculture: model coupling and implementation challenges

5.1 Concept and scope of digital twins in agriculture

DTs are increasingly described as interactive, real-time representations of agricultural systems that connect physical farms with virtual models to enable monitoring, prediction, and decision support (Verdouw et al. 2021, Peladarinos et al. 2023, Mitsanis et al. 2024). Rather than relying on a single modelling approach, DTs integrate process-based, agent-based, and data-driven models that are continuously updated using field and remote sensing observations (Pylianidis et al. 2021). DTs function as composite systems integrating scientific, computational, and data-driven components, rather than acting solely as high-fidelity PBMs.

PBMs often provide the scientific foundation of DTs, since they capture mechanistic crop–soil–climate interactions (Vereecken et al. 2022, Mitsanis et al. 2024). For example, functional structural plant models (FSPMs) reconstruct plant growth in three dimensions, offering highly detailed representations of crop architecture (Mitsanis et al. 2024). At the same time, ABMs contribute the capacity to represent farmer decision-making and collective behaviours at landscape scales (Troost and Berger 2015, Storm et al. 2024). ML complements these by serving as surrogates for complex PBMs, accelerating computations or extracting patterns from sensor data (Droutsas et al. 2022, Zohdi 2024). When integrated within a digital twin, PBMs simulate biophysical responses, ABMs represent management decisions, and ML enables rapid state updating from observations, allowing management actions to be evaluated against current system states.

5.2 Practical illustrations of digital twins

Several recent studies illustrate how these concepts are being trialled in practice, ranging from physics-based calibration to open twin infrastructures. Recent case studies illustrate both the technical feasibility of DTs and the constraints that currently limit their operational use. Zohdi (2024) coupled a physics-based growth model with a genetic algorithm for rapid calibration against canopy height data, achieving efficient but biologically grounded predictions. Storm et al. (2024) proposed the 'Digital Agricultural Avatar', combining agroecosystem models with drone-based sensing to create a nitrogen management twin. Greenhouse horticulture provides further illustration: imaging systems measure tomato plant traits, which are fed into twin models that simulate plant responses to interventions such as pruning or irrigation, allowing growers to test strategies virtually before implementation. Recently, Zeng et al. (2025) advanced the field by developing an open soil–plant twin based on the STEMMUS-SCOPE model, which integrates multi-source observations via data assimilation. This open infrastructure follows FAIR principles, enabling data reuse, reproducibility, and independent extension of the note soil–plant digital twin.

5.3 Challenges in deployment

Despite such advances, significant challenges constrain the deployment of operational agricultural DTs. Most existing digital twin implementations remain at the prototype stage and rely on retrospective datasets rather than live data streams (Peladarinos et al. 2023). Synchronizing simulated state variables (e.g. soil moisture, canopy status) with observed field conditions remains difficult, particularly given the heterogeneity of soil, crop, and management practices (Vereecken et al. 2022). Scaling DTs to farm and regional levels further requires managing large, diverse data streams while ensuring interoperability between models and software tools (Storm et al. 2024).

5.4 Supporting IT infrastructures for digital twin deployment

The computational and scalability limitations of PBMs discussed in Section 2.3.2 have direct implications for the design of digital twin infrastructures. Because PBMs are computationally intensive and often rely on legacy codebases (e.g. Fortran, C + +), they require encapsulation to interface with web-based architectures. As a result, operational DTs require not only scientific model coupling but also IT infrastructures that enable reproducible execution, scalable deployment, and stable long-term operation.

5.4.1 Containerization and orchestration

The first challenge is reproducibility. Many PBMs require specific software environments or libraries that are difficult to replicate across different computers. 'Containerization' technologies, such as Docker, solve this by packaging the model, its mathematical solvers, and all necessary software libraries into a single, self-contained unit (a 'container'). This ensures that a PBM runs exactly the same way on a laptop, a server, or a cloud platform, significantly simplifying the update process.

To manage the computational burden of running these models for thousands of fields simultaneously, orchestration tools like 'Kubernetes' are used. Kubernetes orchestrates container distribution across computing resources, dynamically scaling allocation based on computational workload. This capability is essential for DTs that require near–real-time responsiveness, where running models manually on a single machine would be too slow.

5.4.2 Data streaming and IoT connectivity

The second challenge is moving data from the field to the model in real time. Platforms such as Apache Kafka manage high-throughput data streams, ensuring reliable routeing of weather feeds, sensor readings, and model outputs between components.

However, gathering data from the field often involves battery-powered sensors with limited internet connection. To address this, lightweight protocols like 'MQTT' (Message Queuing Telemetry Transport) are used to transmit data from soil probes and weather stations to central servers efficiently. MQTT transmits telemetry from edge devices, while Kafka manages central data ingestion, enabling the bidirectional data flow required for feedback loops.

5.4.3 Semantic interoperability

Beyond simply moving data, DTs require semantic interoperability to ensure that the data is interpreted correctly. While connectivity protocols (MQTT, Kafka) manage data transport, semantic interoperability ensures consistent data interpretation across heterogeneous modelling components. In agricultural modelling, ambiguity is a frequent source of error. For instance, one model might define 'soil moisture' as volumetric water content (cm^3), while another expects soil water potential (kPa). If a variable is exchanged without an explicit definition of its physical meaning and units (e.g. volumetric water content vs. matric potential), the receiving model may interpret it incorrectly and generate physically invalid results. Semantic standards address this by enforcing explicit variable definitions, units, and metadata so that exchanged quantities are interpreted consistently across models.

Recent developments, such as the open soil–plant digital twin infrastructure of Zeng et al. (2025), demonstrate promising pathways for implementing these architectures. However, robust IT infrastructure alone does not guarantee improved predictive performance. As demonstrated by Wallor et al. (2018), calibrating models against diverse data sources does not automatically guarantee better predictions, and complex hydrological feedbacks often remain underrepresented (Vereecken et al. 2022). Furthermore, widespread adoption depends on socio-economic factors, including data ownership, stakeholder trust, and usability for farmers and advisors (Ma and Rahut 2024). Consequently, effective Digital Twin deployment requires robust IT infrastructure alongside calibrated scientific models and defined data governance protocols.

5.5 Tactical vs. strategic utility: when is a digital twin necessary?

A critical synthesis of the literature suggests that DTs are not universally applicable. The relevance of a digital twin depends on whether the task requires real-time control or long-term scenario analysis.

5.5.1 Tactical applications (real-time control)

DTs are most appropriate in scenarios requiring real-time monitoring and rapid feedback loops. For instance, Mitsanis et al. (2024) demonstrate how plant models integrated into DTs enable real-time pruning and spraying decisions. Similarly, Peladarinos et al. (2023) highlight that in smart farming, DTs act as 'operational copilots', empowering dynamic interventions such as irrigation scheduling by fusing sensor and drone data. In these cases, the high complexity and cost of DT infrastructure are justified by the value of immediate responsiveness (Zohdi 2024).

5.5.2 Strategic applications (long-term planning)

Conversely, when the objective is system understanding, exploring genotype × environment × management (G × E × M) interactions, or supporting policy design, classical model couplings are often more suitable than full DTs. Pylianidis et al. (2021) note that for strategic questions (e.g. climate adaptation or cultivar design), PBMs can provide sufficient insight without the overhead of real-time synchronization. Droutsas et al. (2022) further show that embedding ML within PBMs improves yield prediction without requiring a live digital shadow.

Therefore, we recommend that DTs be prioritized for tactical domains where continuous sensing and rapid actuation are essential such as greenhouse climate control or robotic harvesting. For strategic research questions involving long-term scenario analysis, established model couplings often offer a more scalable and resource-efficient alternative.

In summary, DTs in agriculture differ from conventional model couplings through their continuous feedback between physical systems and virtual representations; they are integration frameworks defined by a continuous feedback loop between physical and virtual systems. While PBMs provide the mechanistic backbone and ABMs capture human decision-making, ML serves as a critical bridge for computational efficiency and data assimilation.

Moving forward, the transition from proof-of-concept to operational reality relies on two pillars:

1. Technological integration: Adopting IT infrastructures (e.g. containerization, data streaming) that enable real-time data flow and model interoperability.

2. Socio-technical trust: Addressing governance, data ownership, and usability to ensure stakeholder adoption (Ma and Rahut 2024).

By respecting these distinctions using DTs for tactical precision and coupled models for strategic insight, agricultural modelling can effectively address the dual challenges of operational efficiency and long-term resilience.

6. Conclusion and outlook

This review shows that agricultural systems analysis increasingly relies on integrating complementary modelling paradigms rather than on any single approach. PBMs provide the essential mechanistic quantification of soil–plant–atmosphere interactions. However, their ability to address complex $G \times E \times M$ challenges is significantly amplified when coupled with ABMs to capture human decision-making and ML to enhance computational efficiency and data assimilation.

Model coupling, whether sequential, loose, tight, or framework-based, represents a substantive methodological advance in agricultural systems modelling. Integrated PBM–ABM–ML frameworks support strategic analysis of climate adaptation, policy impacts, and system sustainability, independent of the real-time synchronization required by DTs. These coupled systems allow researchers to analyse feedback between biophysical processes (e.g. water and carbon fluxes) and socio-economic decisions (e.g. management choices).

DTs represent a specialized, high-fidelity application of these integrated frameworks, distinguished by their continuous feedback loop with the physical world. While DTs offer transformative potential for tactical, operational management (e.g. precision irrigation, robotic intervention), they should not be viewed as a universal replacement for classical modelling approaches. As discussed, the implementation of DTs entails significant costs related to sensor networks, data governance, and IT infrastructure. Therefore, the choice between developing a standard coupled model versus a fully operational digital twin should be driven by the specific decision context: strategic planning favours flexible model coupling, whereas real-time responsiveness justifies the complexity of a digital twin.

A primary remaining challenge is enhancing the interoperability and reusability of these modelling components. Whether building a research-grade coupled model or an operational twin, success depends on shared semantic standards, modular software architectures (such as containerization), and transparent data pipelines. Modular and interoperable combinations of PBM, ABM, and ML components allow agricultural modelling to address both operational decision-making and long-term system analysis.

Acknowledgements

We also wish to extend our gratitude to Prof. Guillaume Lobet for his valuable suggestions and review of this article. We further thank the anonymous reviewers for their constructive comments, which greatly improved the quality of this manuscript.

Funding

The authors gratefully acknowledge funding from the Phenorob project, supported by the German Research Foundation (DFG) under Germany's Excellence Strategy, grant number EXC-2070 – 390732324. Dr. Sathyanarayan Rao also acknowledges funding support from the REWARD project (Project Number: CP/OTHR-22-0037.13), administered through the Department of Civil Engineering, Indian Institute of Science, Bengaluru.

Conflicts of interest

None declared.

Data availability

This study is a review article synthesizing existing research. It does not involve the generation, collection, or analysis of primary data.

References

Bhat SA, Hussain I, Huang N-F. Soil suitability classification for crop selection in precision agriculture using GBRT-based hybrid DNN surrogate models. *Ecol Inform* 2023;**75**:102109. https://doi.org/10.1016/j.ecoinf.2023.102109

Cooper M, Messina CD. Can we harness "enviromics" to accelerate crop improvement by integrating breeding and agronomy?. *Front Plant Sci* 2021;**12**:735143. https://doi.org/10.3389/fpls.2021.735143

Cooper M, Voss-Fels KP, Messina CD *et al.* Tackling $G \times E \times M$ interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *Theor Appl Genet* 2021;**134**:1625–44. https://doi.org/10.1007/s00122-021-03812-3

Couvreur V, Heymans A, Lobet G *et al.* Evidence for a multicellular symplasmic water pumping mechanism across vascular plant roots. BioRxiv 2021-04. 2021.

De Cannière S, Herbst M, Vereecken H *et al.* Constraining water limitation of photosynthesis in a crop growth model with sun-induced chlorophyll fluorescence. *Remote Sens Environ* 2021;**267**:112722. https://doi.org/10.1016/j.rse.2021.112722

Drees L, Demie DT, Paul MR *et al.* Data-driven crop growth simulation on time-varying generated images using multi-conditional generative adversarial networks. *Plant Methods* 2024;**20**:84. https://doi.org/10.1186/s13007-024-01205-3

Drouet J-L, Pagès L. GRAAL: a model of GRowth, architecture and carbon ALlocation during the vegetative phase of the whole maize plant: model description and parameterisation. *Ecol Modell* 2003;**165**:147–73. https://doi.org/10.1016/S0304-3800(03)00072-3

Droutsas I, Challinor AJ, Deva CR *et al.* Integration of machine learning into process-based modelling to improve simulation of complex crop responses. *In Silico Plants* 2022;**4**: diac017. https://doi.org/10.1093/insilicoplants/diac017

Enders A, Vianna M, Gaiser T *et al.* SIMPLACE a versatile modelling and simulation framework for sustainable crops and agroecosystems. *in silico Plants* 2023;**5**:diad006. https://doi.org/10.1093/insilicoplants/diad006

George-Jaeggli B, Wu A, McLean G *et al.* Exploiting the variability and heritability of leaf angle in sorghum to design optimal canopies for different target environments. In *Proceedings of the 20th Agronomy Australia Conference 2022*. 2022.

George TS, Bulgarelli D, Carminati A *et al.* Bottom-up perspective—the role of roots and rhizosphere in climate change adaptation and mitigation in agroecosystems. *Plant Soil* 2024;**500**:297–323. https://doi.org/10.1007/s11104-024-06333-0

Giraud M, Gall SL, Harings M *et al.* CPlantBox: a fully coupled modelling platform for the water and carbon fluxes in the soil–plant–atmosphere continuum. *In Silico Plants* 2023;**5**: diad009. https://doi.org/10.1093/insilicoplants/diad009

Hajjarpoor A, Nelson WCD, Vadez V. How process-based modeling can help plant breeding deal with G × E × M interactions. *Field Crops Res* 2022;**283**:108554. https://doi.org/10.1016/j.fcr.2022.108554

Hao Y, Xu X, Ren D *et al.* Distributed modeling of soil water-salt dynamics and crop yields based on HYDRUS-EPIC model in Hetao Irrigation District. *Trans Chin Soc Agric Eng* 2015;**31**:110–6. https://doi.org/10.11975/j.issn.1002-6819.2015.11.016

Herbst M, Hellebrand HJ, Bauer J *et al.* Multiyear heterotrophic soil respiration: evaluation of a coupled CO2 transport and carbon turnover model. *Ecol Model* 2008;**214**: 271–83. https://doi.org/10.1016/j.ecolmodel.2008.02.007

Hernández-Ochoa IM, Gaiser T, Kersebaum K-C *et al.* Model-based design of crop diversification through new field arrangements in spatially heterogeneous landscapes. A review. *Agron Sustain Dev* 2022;**42**:74. https://doi.org/10.1007/s13593-022-00805-4

Heymans A, Couvreur V, LaRue T *et al.* GRANAR, a computational tool to better understand the functional importance of monocotyledon root anatomy. *Plant Physiol* 2020;**182**: 707–20. https://doi.org/10.1104/pp.19.00617

Holzworth DP, Huth NI, deVoil PG *et al.* APSIM—evolution towards a new generation of agricultural systems simulation. *Environ Model Softw* 2014;**62**:327–50. https://doi.org/10.1016/j.envsoft.2014.07.009

Jacquemoud S, Verhoef W, Baret F *et al.* PROSPECT + SAIL models: a review of use for vegetation characterization. *Remote Sens Environ* 2009;**113**:S56–66. https://doi.org/10.1016/j.rse.2008.01.026

Jacques D, Simunek J. Notes on HP1 a software package for simulating variably-saturated water flow, heat transport, solute transport, and biogeochemistry in porous media. *HP1 Version* 2010;**2**. https://www.pc-progress.com/Documents/BLG-1068.pdf

Janni M, Maestri E, Gullì M *et al.* Plant responses to climate change, how global warming may impact on food security: a critical review. *Front Plant Sci* 2024;**14**:1297569. https://doi.org/10.3389/fpls.2023.1297569

Jin X, Zarco-Tejada PJ, Schmidhalter U *et al.* High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. *IEEE Geosci Remote Sens Mag* 2020;**9**: 200–31. https://doi.org/10.1109/MGRS.2020.2998816

Jones JW, Hoogenboom G, Porter CH *et al.* The DSSAT cropping system model. *Eur J Agron* 2003;**18**:235–65. https://doi.org/10.1016/S1161-0301(02)00107-7

Jones HG, Vaughan RA. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford, UK: Oxford University Press, 2010.

Kanda EK, Mabhaudhi T, Senzanje A. Coupling hydrological and crop models for improved agricultural water management—a review. *Bulg J Agric Sci* 2018;**24**:380–90.

Kroes JG, van Dam JC, Bartholomeus RP *et al.* SWAP version 4: Theory description and user manual, No. 2780. Bulletin: Alterra-Rapport—Wageningen University and Research Centre, 2017.

Li C, Li L, Reynolds MP *et al.* Recognizing the hidden half in wheat: root system attributes associated with drought tolerance. *J Exp Bot* 2021;**72**:5117–33. https://doi.org/10.1093/jxb/erab124

Li Y, Kinzelbach W, Zhou J *et al.* Modelling irrigated maize with a combination of coupled-model simulation and uncertainty analysis, in the northwest of China. *Hydrol Earth Syst Sci* 2012;**16**:1465–80. https://doi.org/10.5194/hess-16-1465-2012

Lopez G, Ahmadi SH, Amelung W *et al.* Nutrient deficiency effects on root architecture and root-to-shoot ratio in arable crops. *Front Plant Sci* 2023;**13**:1067498. https://doi.org/10.3389/fpls.2022.1067498

Ma W, Rahut DB. Climate-smart agriculture: adoption, impacts, and implications for sustainable development. *Mitig Adapt Strateg Glob Chang* 2024;**29**:44. https://doi.org/10.1007/s11027-024-10139-z

Marshall-Colon A, Long SP, Allen DK *et al.* Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. *Front Plant Sci* 2017;**8**:786. https://doi.org/10.3389/fpls.2017.00786

Meunier F, Heymans A, Draye X *et al.* MARSHAL, a novel tool for virtual phenotyping of maize root system hydraulic architectures. *In Silico Plants* 2020;**2**:diz012. https://doi.org/10.1093/insilicoplants/diz012

Midingoyi CA, Pradal C, Enders A *et al.* Crop2ML: an open-source multi-language modeling framework for the exchange and reuse of crop model components. *Environ Model Softw* 2021;**142**:105055. https://doi.org/10.1016/j.envsoft.2021.105055

Mitsanis C, Hurst W, Tekinerdogan B. A 3D functional plant modelling framework for agricultural digital twins. *Comput Methods Appl Mech Eng* 2024;**431**:108733. https://doi.org/10.1016/j.compag.2024.108733

Peladarinos N, Piromalis D, Cheimaras V *et al.* Enhancing smart agriculture by implementing digital twins: a comprehensive review. *Sensors* 2023;**23**:7128. https://doi.org/10.3390/s23167128

Pylianidis C, Osinga S, Athanasiadis IN. Introducing digital twins to agriculture. *Comput Electron Agric* 2021;**184**:105942. https://doi.org/10.1016/j.compag.2020.105942

Rasool F, Khan MR, Schneider M *et al.* Transcriptome unveiled the gene expression patterns of root architecture in drought-tolerant and sensitive wheat genotypes. *Plant Physiol Biochem* 2022;**178**:20–30. https://doi.org/10.1016/j.plaphy.2022.02.025

Schnepf A, Leitner D, Landl M *et al.* CRootBox: a structural–functional modelling framework for root systems. *Ann Bot* 2018;**121**:1033–53. https://doi.org/10.1093/aob/mcx221

Shang L, Wang J, Schäfer D *et al.* Surrogate modelling of a detailed farm-level model using deep learning. *J Agric Econ* 2024;**75**:235–60. https://doi.org/10.1111/1477-9552.12543

Shrestha P, Sulis M, Masbou M *et al.* A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Mon Weather Rev* 2014;**142**:3466–83. https://doi.org/10.1175/MWR-D-14-00029.1

Sidhu JS, Ajmera I, Arya S *et al.* RootSlice A novel functional-structural model for root anatomical phenotypes. *Plant Cell Environ* 2023;**46**:1671–90. https://doi.org/10.1111/pce.14552

Storm H, Seidel SJ, Klingbeil L *et al.* Research priorities to leverage smart digital technologies for sustainable crop production. *Eur J Agron* 2024;**156**:127178. https://doi.org/10.1016/j.eja.2024.127178

Streefkerk IN, de Bruijn J, Haer T *et al.* A coupled agent-based model to analyse human-drought feedbacks for agropastoralists in dryland regions. *Front Water* 2023;**4**:1037971. https://doi.org/10.3389/frwa.2022.1037971

Sun Q, Zhang Y, Che X *et al.* Coupling process-based crop model and extreme climate indicators with machine learning can improve thepredictions and reduce uncertainties of global soybean yields. *Agriculture* 2022;**12**:1791. https://doi.org/10.3390/agriculture12111791

Troost C, Berger T. Process-based simulation of regional agricultural supply functions in Southwestern Germany using farm-level and agent-based models. *Paper presented at the International Conference of Agricultural Economists*, Milan, Italy, August 9–14, 2015.

Verdouw C, Tekinerdogan B, Beulens A *et al.* Digital twins in smart farming. *Agric Syst* 2021;**189**:103046. https://doi.org/10.1016/j.agsy.2020.103046

Vereecken H, Amelung W, Bauke SL *et al.* Soil hydrology in the Earth system. *Nat Rev Earth Environ* 2022;**3**:573–87. https://doi.org/10.1038/s43017-022-00324-6

Vos J, Evers JB, Buck-Sorlin GH *et al.* Functional–structural plant modelling: a new versatile tool in crop science. *J Exp Bot* 2010;**61**:2101–15. https://doi.org/10.1093/jxb/erp345

Wallor E, Kersebaum K-C, Ventrella D *et al.* The response of process-based agro-ecosystem models to within-field variability in site conditions. *Field Crops Res* 2018;**228**:1–19. https://doi.org/10.1016/j.fcr.2018.08.021

Wang Y, Zeng Y, Yu L *et al.* Integrated modeling of canopy photosynthesis, fluorescence, and the transfer of energy, mass, and momentum in the soil–plant–atmosphere continuum (STEMMUS–SCOPE v1. 0.0). *Geosci Model Dev* 2021;**14**:1379–407. https://doi.org/10.5194/gmd-14-1379-2021

Wu S, Yang P, Ren J *et al.* Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm. *Remote Sens Environ* 2021;**255**:112276. https://doi.org/10.1016/j.rse.2020.112276

Yang P, Prikaziuk E, Verhoef W *et al.* SCOPE 2.0: a model to simulate vegetated land surface fluxes and satellite signals. *Geosci Model Dev* 2021;**14**:4697–712. https://doi.org/10.5194/gmd-14-4697-2021

Yang X, Tian Z, Sun L *et al.* The impacts of increased heat stress events on wheat yield under climate change in China. *Clim Change* 2017;**140**:605–20. https://doi.org/10.1007/s10584-016-1866-z

Zeng Y, Verhoef A, Vereecken H *et al.* Monitoring and modeling the soil–plant system toward understanding soil health. *Rev Geophys* 2025;**63**:e2024RG000836. https://doi.org/10.1029/2024RG000836

Zohdi TI. A machine-learning enabled digital-twin framework for next generation precision agriculture and forestry. *Comput Methods Appl Mech Eng* 2024;**431**:117250. https://doi.org/10.1016/j.cma.2024.117250