# A hybrid mechanistic potato growth model integrating knowledge-guided machine learning and physiological-admissibility diagnostics

by Mhango, K.J., Lewis, M.J., Roberts, J.M. and Harris, W.E.

**Copyright, publisher and additional information:** Publishers' version distributed under the terms of the Creative Commons Attribution License

DOI link to the version of record on the publisher's site

Mhango, K.J., Lewis, M.J., Roberts, J.M. and Harris, W.E. (2027) 'A hybrid mechanistic potato growth model integrating knowledge-guided machine learning and physiological-admissibility diagnostics', *Field Crops Research,* 351, article 110806.

Field Crops Research 351 (2027) 110806

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

# A hybrid mechanistic potato growth model integrating knowledge-guided machine learning and physiological-admissibility diagnostics

Kanthu Joseph Mhango [a,*], Megan Jane Lewis [a], Joe M. Roberts [b], W. Edwin Harris [a]

[a] Centre for Agricultural Data Science, Agriculture and Environment Department, Harper Adams University, Newport, UK
[b] Centre for Crop and Environmental Science, Agriculture and Environment Department, Harper Adams University, Newport, UK

ARTICLE INFO

*Keywords:*
Potato
Crop modelling
Mechanistic modelling
Phenotyping
Breeding

ABSTRACT

*Background:* Potato yield formation is well understood in terms of light capture, radiation conversion, biomass accumulation and partitioning, but mechanistic crop models often rely on fixed or externally calibrated physiological parameters. This limits their use for population-specific inference, where the practical question is not only which line yielded most, but which physiological route produced that yield. We developed a trainable mechanistic potato dry-matter model implemented as a knowledge-guided neural network (KGNN), where the established crop-growth equations are retained, while small neural submodels infer their physiological parameter values from observations.

*Methods:* A multi-source potato panel was assembled with the purpose of constructing and calibrating a KGNN for a Dutch field experiment scenario. The basic modelling unit was a crop sequence or crop time series that represented the longitudinal canopy record for one retained plot, treatment replicate or treatment group. The panel contained 247 sequences: 196 for model fitting and 51 randomly held-out Dutch sequences, stratified by two Dutch environments, for within-population validation. Observed leaf area index (LAI), days after emergence, cumulative photosynthetically active radiation and environment indicators were used to infer effective light capture, radiation-use efficiency (RUE) and a bounded partitioning term. Tuber dry matter was then calculated through the physiological graph. Performance and physiological admissibility were assessed using prediction error, external proxy alignment, graph-consistency checks, structural replays and ±10–20% input perturbations.

*Results:* On the 51 held-out sequences, the KGNN achieved a mean absolute error of 1.41 t $ha^{-1}$ (8.9% of the observed mean), a root mean square error of 1.77 t $ha^{-1}$ and a relative root mean square error of 11.2%, with near-zero mean bias. An apparent modelled RUE diagnostic aligned strongly with the observed apparent RUE proxy (Spearman $r = 0.88$). The learned structure followed the imposed production graph, including associations between intercepted radiation and total dry matter ($r = 0.94$), total dry matter and tuber dry matter ($r = 0.99$), and total dry matter and the model-conditioned partitioning term ($r = -0.54$). Input perturbations produced smooth, monotonic responses; prediction changes were larger for cumulative radiation than for LAI. The partitioning latent was the weakest external anchor ($r = 0.27$) and therefore interpreted cautiously. Consistency in the structural replays and input or latent perturbations provided evidence for physiological-admissibility and knowledge-alignment of the final model despite end-to-end training.

*Conclusions:* A mechanistic potato dry-matter graph can be made locally trainable by replacing fixed coefficients with neural parameter-generating functions. The present validation supports physiological inference for unseen records within the target Dutch experimental population, while a new population would require retraining or recalibration.

## 1. Introduction

Accurate prediction of crop yield is a foundational aim of agronomy, crop physiology and plant breeding (Shawon et al., 2025). For potato (Solanum tuberosum L.), yield formation is well understood at a high level: canopy development determines light capture, radiation-use efficiency (RUE) describes conversion of intercepted radiation into biomass, and harvest index (HI) describes partitioning of biomass to tubers. This

* Corresponding author.
*E-mail address:* jmhango@harper-adams.ac.uk (K.J. Mhango).

https://doi.org/10.1016/j.fcr.2026.110806
Received 28 June 2026; Received in revised form 31 October 2026; Accepted 14 November 2026
Available online 19 November 2026
0378-4290/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

logic underpins the "school of de Wit" tradition in crop modelling (Bouman et al., 1996) and appears in simple dry-matter formulations and more elaborate process-based systems (Jamieson et al., 1991; Lin et al., 2023). Models such as the Simulation of Underground Bulking Storage Organs potato model (SUBSTOR-Potato), the World Food Studies model configured for potato (WOFOST), the Food and Agriculture Organization's AquaCrop crop-water productivity model, and the Agricultural Production Systems sIMulator (APSIM) extend this graph with crop development, soil water balance, nitrogen stress, phenology, cultivar coefficients and management effects (Holzworth et al., 2014; Jones et al., 2003; Raes et al., 2018; Raymundo et al., 2017; de Wit et al., 2019).

Process-based crop models encode biological theory explicitly, but their practical usefulness often depends on fixed equations with externally calibrated coefficients or cultivar-specific parameter sets (Correa et al., 2025). Genetic coefficients, stress response parameters, harvest-index assumptions and canopy-development parameters are usually estimated before application or tuned for specific environments. Once calibrated, the parameter structure is largely fixed, so additional observations rarely change the way yield is physiologically explained (Wallach et al., 2018). This makes conventional crop models powerful for simulation, but less suited to discovering population-specific physiological routes within bespoke breeding material.

Black-box machine-learning models offer the opposite trade-off, being able to adapt to weather records, remote-sensing indices, imagery, management covariates and genotype descriptors, but their internal representations are not normally organised around named physiological processes (Chang et al., 2023). Post-hoc attribution methods can identify influential predictors, but they do not by themselves recover an explicit crop-growth pathway in which light capture, RUE, biomass accumulation and partitioning remain connected (Yenkikar et al., 2025).

This distinction between process-based and black box empirical models is important in applied breeding, where a universal potato yield model is rarely the immediate need. A breeder often works with a particular genetic population, measurement protocol and set of trial environments (Hajjarpoor et al., 2022). The practical question is not only "which line yielded most?" but "why did it yield, what physiological route did it follow, and where is the remaining improvement potential?" A model trained or recalibrated on a breeder's own modest dataset can therefore be useful if it positions lines physiologically within that population, even if it is not a universal global predictor.

A further limitation of both paradigms is their treatment of observations. Traditional process-based crop models often require state variables such as LAI, phenology or biomass, even when these are expensive to measure at scale (Morisse et al., 2022). Machine-learning models can accept alternative sensor streams, but usually without preserving the meaning of the internal crop state being inferred. Much effort is spent converting drone or multispectral imagery into LAI, even though LAI is itself an observation that informs the more mechanistically relevant quantity of effective canopy light interception (Ninomiya, 2022; Morisse et al., 2022). The observation layer should therefore be distinguished from the physiological mechanism it informs. LAI, ground cover, spectral indices, RGB-derived canopy closure or proximal sensor signals may all contain information about canopy light capture, but they should not be treated as identical to that physiological state.

We therefore propose a trainable mechanistic potato dry-matter model implemented as a knowledge-guided neural network (KGNN). In such a model, the mechanistic crop-growth pathway can be fixed, while bounded neural submodels learn how to infer observation-specific values for its physiological coefficients from the crop time series, radiation and environment context. The graph then calculates tuber dry matter through effective light capture, intercepted radiation, RUE, total dry matter and HI (Kooman and Haverkort, 1995; Haverkort et al., 2015). In interpretation, a KGNN designed as such is distinct from post-hoc explainable-AI approaches, including Shapley additive explanations (SHAP) or Shapley-value-based attribution, where a fitted model is interpreted after training by estimating the marginal influence of input variables (Huang et al., 2020). Here, interpretability is built into the model architecture itself, whereby the intermediate nodes are named physiological quantities, and the final prediction must pass through the imposed crop-growth pathway rather than an unconstrained prediction layer. Instead of SHAP values as the basis for explanation, the model assessment rests on an unseen subset of observations for prediction accuracy and prespecified physiological-admissibility criteria. In this case, we propose semantic anchoring (agreement with an external physiological proxy), structural coherence (behaviour consistent with the imposed mechanistic model) and route-specific structural replay (adherent behaviour upon change to one named mechanism while the remaining fitted state is held fixed). These diagnostics test whether the latent states are physiologically interpretable, without necessarily establishing causal field effects.

The intended application (use case) is breeder-facing physiological inference after training or recalibration on the breeder's own target population. The fitted model could then position new lines, crosses or introductions relative to the reference/training population and ask whether yield variation is associated with light capture, conversion efficiency, biomass accumulation or partitioning. The current study utilises an agronomic-trial panel to demonstrate the modelling framework, short of a completed commercial breeding deployment, which would require larger, more genetically diverse breeding panels to train a generalised model.

## 2. Materials and methods

### 2.1. Physiological graph and neural parameter generation

The model starts from a simple traditional crop-growth model (Fig. 1). First, an observed canopy trajectory (crop sequence) informs the fraction of incoming photosynthetically active radiation (PAR) intercepted during the retained crop window. Second, intercepted PAR multiplied by RUE gives total dry matter. Third, HI partitions this total dry matter to tuber dry matter. Finally, the result is converted from g $m^{-2}$ to t $ha^{-1}$ . Neural submodels do not replace or bypass these steps; they infer bounded, sequence-specific light-capture parameters, RUE and HI values that are inserted into the graph. Thus, final yield is calculated by the physiological equations rather than by an unconstrained neural-network output layer.

The neural submodels replace traditional hard-coded parameter requirements. At prediction time, each crop sequence is passed through the trained model which now functions as a chain of parameter-generating functions. It therefore receives its own inferred parameter values from its observed canopy, radiation and environment context. No parameter is manually selected, copied from a training record or refitted during inference.

### 2.2. Dataset assembly and study sources

A multi-source potato panel was assembled to fit and diagnose the KGNN. The four sources were Dutch field experiments (Ten Den et al., 2024), a global pool of potato growth observations (Raymundo et al., 2018), SUBSTOR-Potato experimental files (Woli et al., 2016), and Bangladesh phenology-growth-yield observations (Talukder et al., 2026). They differed in design and measurement timing and were not treated as exchangeable experiments. Each was mapped to the common information required by the graph, that is an observed canopy trajectory, phenological time, cumulative radiation exposure, source or environment context, and tuber dry matter for its retained target window.

The Dutch observational core comprised experiments conducted in 2019 and 2020 at Lelystad (light clay soil) and Vredepeel (sandy soil) (Ten Den et al., 2024). The split-plot design included three replicate blocks, two water regimes and three nitrogen levels. Three processing

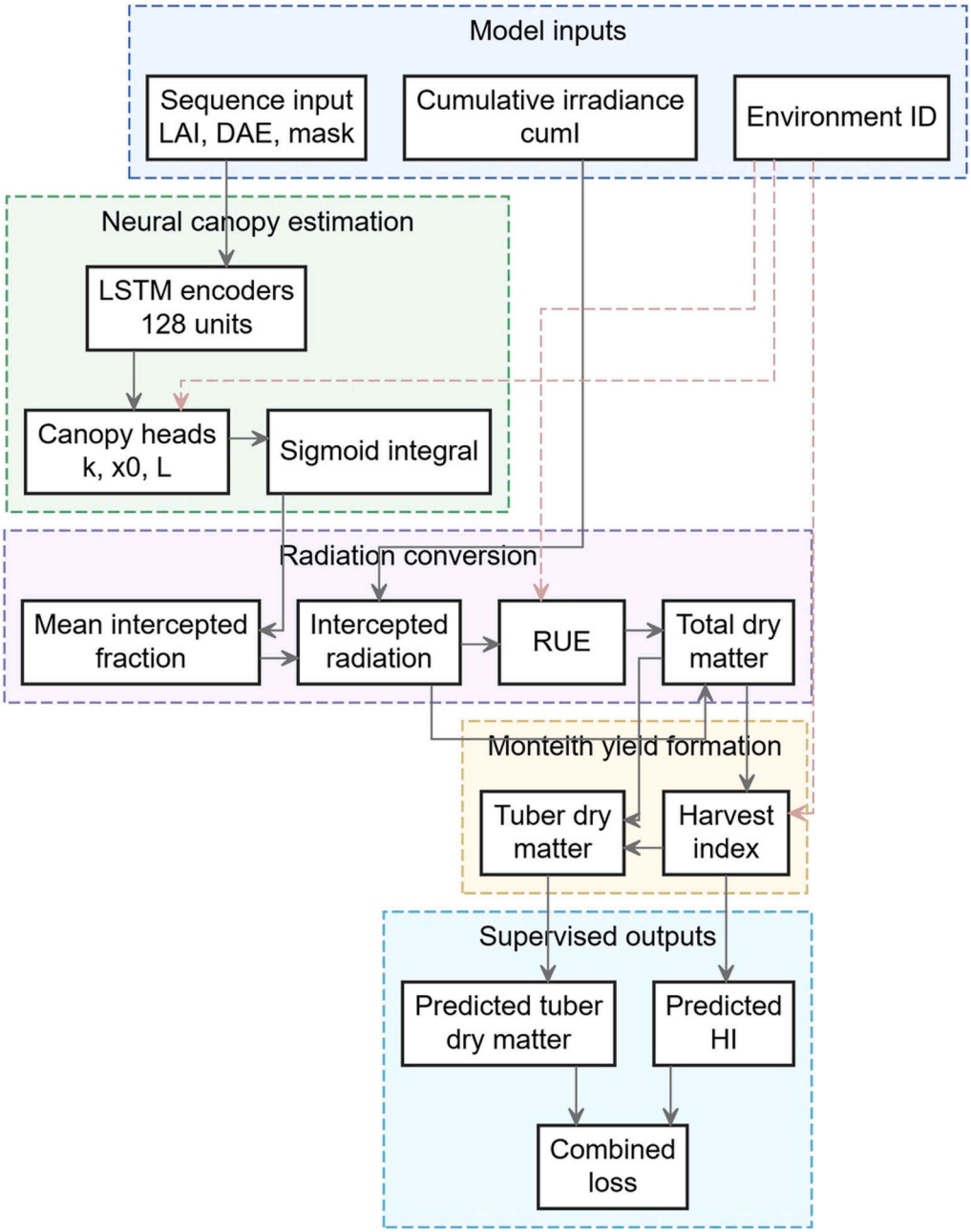


**Fig. 1.** Knowledge-guided neural network computation graph linking observed inputs, neural estimation of bounded canopy parameters, environment-specific mechanism adjustments, the physiological crop-growth calculation and supervised outputs. LAI = leaf area index; DAE = days after emergence; mask = indicator distinguishing measured from appended timesteps; cumI = cumulative incoming photosynthetically active radiation; LSTM = long short-term memory sequence encoder; k = canopy rise-rate parameter; $x_0$ = relative canopy-interception midpoint; L = maximum effective interception potential; RUE = radiation-use efficiency; HI = harvest index.

cultivars were grown at each location: Fontane, Innovator and Markies at Lelystad, and Fontane, Festien and Premiere at Vredepeel. Artificially shaded subplots were excluded. Water and nitrogen were not supplied to the KGNN as explicit coefficients; their effects could enter indirectly through observed canopy development, radiation context, environment and the dry-matter target, and were therefore not assumed to act only through RUE.

The Dutch observations were deliberately the core target population because they provided the most complete and internally consistent field data. The 117 Dutch fitting sequences and 51 held-out Dutch sequences therefore dominated the panel by design. The 18 Raymundo, 37 SUBSTOR-Potato and 24 Bangladesh sequences were fitting support used to widen physiological and environmental coverage, not equivalent external validation domains. The Bangladesh target was measured at 75 days after planting rather than final harvest, so those shorter-window records contributed training information only. Source-wise fitted residuals and latent-state summaries were used to check for source-specific behaviour.

A crop sequence was the longitudinal observations for one retained modelling unit, essentially a plot replicate where plot-level records were available, that could be mapped consistently to the common structure. The potential Dutch factorial comprised 216 unshaded plot-replicate combinations (three cultivars per location × two locations × two water regimes × three nitrogen levels × two years × three replicates).

After shaded plots were excluded and records were required to have usable canopy, phenological, radiation and tuber-dry-matter information, 168 Dutch plot sequences remained: 84 from Lelystad and 84 from Vredepeel. Thus, replicates were retained as separate crop sequences rather than averaged.

### 2.3. Validation design

Only the 168 Dutch sequences were eligible for the prediction-evaluation subset, being the target population for model development and assessment. The additional sources were used to broaden the physiological support available during training, exposing the neural-mechanistic architecture to a wider range of canopy-development, radiation and dry-matter accumulation patterns, but they were not treated as equivalent inference or validation populations because the Bangladesh records represented an earlier growth window rather than final harvest, the DSSAT/SUBSTOR subset was limited in size and temporal scope, and the remaining non-Dutch observational data were too small to support a meaningful held-out target-population test. The intended use is population-specific calibration rather than universal transfer. Compatible data from diverse sources can broaden the physiological support used to fit the defined neural-mechanistic architecture, but the fitted or recalibrated model is evaluated by how well it represents the target breeding population. New varieties or recombinants, once evaluated using comparable canopy and radiation observations, can then be positioned within that population's learned physiological distribution, helping breeders identify candidates with trait values above the reference mean or with complementary improvement routes. Developing a model that transfers unchanged across populations is a different objective requiring larger, more diverse panels and independent validation. The proposed contribution here is the ability to obtain informative within-population inference from comparatively small datasets. Independent cross-source generalisation was not tested because the additional sources differed in target definition, observation period and experimental design.

Using a fixed random seed (42), 30% were held out while stratifying by the Lelystad and Vredepeel environment labels, producing 51 unseen Dutch validation records (26 Lelystad and 25 Vredepeel) and 117 Dutch fitting records. All 79 non-Dutch sequences were assigned to fitting, giving 196 fitting sequences in total. This is within-population validation testing unseen records from the same Dutch experimental network, rather than independent geographic transfer. Performance was also reported separately for the two held-out Dutch environments as a stratified diagnostic. A model used in a new region, experiment or breeding population would require retraining or recalibration.

To test whether the ordinary held-out result was inflated by related observations occurring in both fitting and validation, we also used environment-year-blocked resampling. Each of the four Dutch location $\times$ year environments was withheld in turn, so all cultivars, nitrogen and irrigation treatments, and replicate observations from the withheld environment remained outside model fitting. The non-Dutch datasets remained ancillary training support, and the architecture and hyperparameters were fixed without fold-specific tuning. This grouped analysis tests prediction to unseen environments within the Dutch target population.

Both the ordinary and grouped validation designs were retained, as they are answering different questions. The ordinary split tests prediction for unseen experimental units sampled from the represented target population. The grouped split measures the additional effect of separating related experimental groups. The saved model retains learned neural weights rather than physiological values from individual training observations. At inference, a new experimental unit supplies a compatible canopy sequence, cumulative radiation and a represented environmental context. The parameter-generating functions infer observation-specific light-capture parameters, radiation-use efficiency and harvest index, after which the fixed physiological graph calculates tuber dry matter. Apart from yield prediction, the utility of this model is its ability to infer observation-specific values for otherwise latent traits such as RUE. Application to an unrepresented environmental domain requires an appropriate extension of the environmental encoding and recalibration, while a universally generalised model would require a much larger panel beyond the scope of the current study.

### 2.4. Common panel representation

All data sources were transformed to a common panel of crop time series. Each observed timestep contained LAI, days after emergence (DAE) and an indicator marking it as a real observation. A time series could contain a different number of observations, so artificial rows were appended just to make equal-length arrays for batch processing. These appended rows used the sentinel value −999, and masking instructed the sequence encoder to ignore them. Separate inputs supplied cumulative PAR and a categorical environment indicator; the latter was represented computationally as a one-hot vector, meaning one position identified the relevant source or environment and all others were zero.

The LAI was an observed input from phenotyping. Variation among cultivars, treatments and environments was therefore part of the observed canopy phenotype and the model did not attempt to identify the causes of that LAI variation. LAI informed, but was not equated with, effective light capture. DAE aligned observations across experiments because it was consistently available or derivable in every source. Growing degree days would be more directly thermal, but a harmonised calculation would have required source-specific assumptions about base temperature, daily temperature reconstruction and weather matching. Future single-population applications with consistent weather data could replace or augment DAE with growing degree days. Cumulative PAR represented incoming radiation, and the target was tuber dry matter in t $ha^{-1}$ .

### 2.5. Knowledge-guided neural network architecture

The KGNN was implemented as a mechanistic dry-matter graph with neural parameter estimators, with the full computation graph shown in Fig. 1. Neural submodels estimated effective light-capture parameters, RUE and HI from the observed crop sequence, cumulative radiation and environment context. The supervised output was tuber dry matter in t $ha^{-1}$ rather than fresh yield, in line with conventional mechanistic models (Kooman and Haverkort, 1995; Holzworth et al., 2014). Canopy observations were used to infer an effective light-interception trajectory. Intercepted radiation and RUE were then combined to estimate total dry matter. The HI partitioned total dry matter into tuber dry matter, and a fixed unit conversion transformed the internal dry-matter quantity into t $ha^{-1}$ . Thus, unlike a generic neural-network predictor, the final yield estimate was not produced by an unconstrained dense prediction head but assembled through named physiological states. In this sense, the model generalises calibration from fixed parameter fitting to a reusable parameter-generating function whereby each observation receives mechanism values inferred from its crop trajectory and context by the model.

#### 2.5.1. Model inputs

Each sample comprised the observed LAI–DAE time series, cumulative incoming PAR and an environment identifier, as summarised in Tables 1–3. The time series had three channels: LAI, DAE and the observed-row indicator. Unequal series were extended to the common array length with −999 sentinel rows; a masking layer excluded those rows from the sequence encoders. The original irregular observations, including observed late-season decline where available, were therefore encoded without fitting an external smoothed canopy curve.

The canopy sequence encoded the observed trajectory from which light-interception parameters were inferred. Cumulative PAR supplied the incoming radiation pool available for interception and biomass

**Table 1**
Source-level composition of the retained modelling panel.

| Source | Retained sequences | Training | Validation | Split role |
|---|---|---|---|---|
| Ten Den et al. Netherlands | 168 | 117 | 51 | Train/ validation |
| Raymundo potato observations | 18 | 18 | 0 | Training only |
| DSSAT/SUBSTOR potato | 37 | 37 | 0 | Training only |
| Bangladesh BARI/ Gazipur | 24 | 24 | 0 | Training only |
| Total | 247 | 196 | 51 | - |

Note. DSSAT = Decision Support System for Agrotechnology Transfer; SUBSTOR = Simulation of Underground Bulking Storage Organs potato model. The non-Dutch sources were assigned to training only; validation was reserved for held-out Dutch observational records.

formation. The environment identifier represented the source, location or experiment group and was used only to introduce constrained, parameter-specific perturbations to mechanism values.

### 2.5.2. Sequence encoders and bounded physiological parameters

Separate long short-term memory (LSTM) layers were used as sequence encoders. Each used the available LAI–DAE observations to learn a vector that was subsequently reduced to a scalar. This scalar could be bounded into the plausible range of any of three parameters of a sigmoidal light-interception curve (Hence separate LSTMs for the three parameters). The LSTM did not simulate LAI or biomass recurrently from day to day. The shorter Bangladesh records therefore represented shorter observed target windows rather than truncated final-harvest simulations. The inferred parameters were the rise rate of effective interception (k), the relative curve midpoint ($x_0$) and the asymptotic interception scale (L). Each raw neural output was transformed to a biologically bounded range to represent a learned physiological coefficient constrained to plausible values.

The same bounded-parameter principle was used for all learned physiological parameters, whose roles and constraint rationale are summarised in Table 4. Each parameter was allowed to vary within physiologically plausible limits, enabling the model to learn population-specific mechanism values while preventing implausible parameter combinations.

### 2.5.3. Environment-specific mechanism shifts

Environment information was incorporated as a constrained mechanism effect rather than a direct yield shortcut. A low-dimensional environment code could shift only the named parameters k, $x_0$, L, RUE and HI. Additive shifts were scaled by 0.02 for k, 0.05 for $x_0$, 0.05 for L, 0.25 g DM $MJ^{-1}$ for RUE and 0.05 for HI; RUE could also receive an approximately 10% multiplicative adjustment. These scales reflect the parameter units and admissible ranges. Values were clipped to those ranges after adjustment. The design allowed small source-, location- or experiment-specific physiological shifts without an unconstrained route for memorising source-specific yield.

This design preserves the interpretation of environment effects as small changes in physiological mechanisms. For example, an environment may shift the timing of effective interception, slightly alter attainable interception, or modify the realised RUE or HI. It cannot bypass the physiological graph and predict yield through a separate black-box pathway.

### 2.5.4. Light interception and intercepted radiation

The canopy parameters learned from the LSTMs defined a logistic curve for the effective fraction of incoming PAR intercepted by the canopy over phenological time. The midpoint parameter was learned as a fraction of the observed crop window and then converted to a day value using the final observation day after emergence (DAE). The area under the logistic curve was integrated analytically from emergence to

**Table 3**
Model inputs supplied to the KGNN.

| Input | Contents | Role in the KGNN |
|---|---|---|
| Canopy sequence | LAI, days after emergence and observation mask | Encodes the observed canopy trajectory used to infer light interception |
| Cumulative PAR | Total incoming PAR accumulated over the retained season window | Supplies the radiation pool available for interception and biomass production |
| Environment identifier | Source, location or experiment group | Allows small source- or environment-specific shifts in mechanism parameters |

Note. KGNN = knowledge-guided neural network; LAI = leaf area index; PAR = photosynthetically active radiation.

**Table 4**
Learned physiological parameters and their constraint rationale.

| Latent state | Admissible range | Biological role | Constraint rationale |
|---|---|---|---|
| k | 0.01 – 0.6 | Rate at which effective interception rises with phenological time | Prevents unrealistically abrupt or flat interception trajectories |
| $x_0$ | 0–1 | Relative timing of the interception-curve midpoint | Constrains timing to the observed crop window |
| L | 0.3–1 | Maximum effective interception potential | Represents an interception asymptote rather than raw peak LAI |
| RUE | 0.5–5.0 g DM $MJ^{-1}$ | Conversion of intercepted radiation into dry matter | Keeps conversion efficiency positive and physiologically bounded |
| HI | 0.4 – 0.95 | Partitioning of total dry matter to tuber dry matter | Allows partitioning variation while preventing impossible fractions |

Note. RUE = radiation-use efficiency; HI = harvest index. The parameters k, $x_0$ and L describe the rise rate, midpoint and upper asymptote of the learned effective light-interception curve, respectively.

**Table 2**
Common model-facing variables used by the KGNN.

| Variable | Code variable | Unit or scale | Role in KGNN |
|---|---|---|---|
| Leaf area index | LAI | $m^2$ leaf $m^{-2}$ ground | Canopy observation used to infer effective light interception |
| Days after emergence | DAE | Days | Phenological time input |
| Observation mask | observed_mask | 1 for observed rows; 0 or pad value for padded rows | Separates measured observations from sequence padding |
| Cumulative PAR | X_train_cumI, cumI_sequence | MJ $m^{-2}$ PAR | Incoming radiation for interception and RUE mechanisms |
| Environment label | environment_onehot | One-hot categorical | Allows source/environment effects on mechanism parameters |
| Tuber dry-matter target | y_train, final_tuber_dry_matter | t $ha^{-1}$ dry matter | Supervised target for structural yield output |
| Harvest-index proxy | HI | Fraction | Auxiliary diagnostic and latent semantic check |
| Dry-matter concentration | DMC | % | Quality diagnostic; not always available |

Note. KGNN = knowledge-guided neural network; LAI = leaf area index; DAE = days after emergence; PAR = photosynthetically active radiation; RUE = radiation-use efficiency; HI = harvest index, calculated as tuber dry matter at the retained harvest time divided by reported total dry matter; DMC = dry-matter concentration.

the final observation DAE. Dividing this area by the crop duration produced an average intercepted fraction for the retained season window.

The single-sigmoid interception layer was used as an effective seasonal light-capture approximation rather than as a full canopy life-cycle model. This choice follows the radiation-use-efficiency tradition, in which dry-matter production is organised around the accumulated interception of radiation and its conversion into biomass, and is consistent with potato models such as LINTUL-POTATO, where light interception by green leaves drives dry-matter production while subsequent partitioning to tubers governs crop duration and the end of the growing period (Kooman and Haverkort, 1995). Irreversible senescence was therefore treated not as another equally weighted canopy-construction phase, but as a terminal transition in which the crop has reduced investment in maintaining photosynthetically active leaf area and source–sink dynamics increasingly favour remobilisation and storage-organ filling (Chang and Zhu, 2017). In the present architecture, the LSTM encoder received the full observed LAI trajectory, including decline where present, and environment-specific mechanism shifts allowed source- or environment-governed differences in canopy duration and decline to nudge the effective interception parameters. This allowed senescence to modify the learned seasonal light-capture estimate empirically, without forcing a separately parameterised senescence limb onto a mixed panel in which full-season records coexist with earlier target-window observations in the Talukder dataset (Talukder et al., 2026). The average intercepted fraction was multiplied by cumulative incoming PAR to estimate intercepted radiation as $I_{\mathrm{int}} = f_{\mathrm{int}} \times I_{\mathrm{cum}}$, where $I_{\mathrm{int}}$ is intercepted radiation, $f_{\mathrm{int}}$ is the learned average intercepted fraction, and $I_{\mathrm{cum}}$ is cumulative incoming PAR.

#### 2.5.5. RUE and dry-matter formation

RUE was learned as the slope of the regression of total dry matter to intercepted radiation from the model predictions. This follows the conventional crop-physiology treatment of RUE as the regression slope of dry matter against intercepted radiation (Robles-Zazueta et al., 2021), but generalises it from a single population-level coefficient to a context-dependent slope. The RUE branch received intercepted radiation as the main production predictor, together with cumulative incoming PAR, final DAE and the learned canopy-interception parameters as covariates describing the radiation environment, crop duration and shape of the light-capture trajectory. A compact feed-forward subnetwork therefore served as a nonlinear regression function for estimating this slope, and the bounded scaling layer constrained the resulting RUE value to remain physiologically plausible before environment-specific mechanism adjustment.

Total dry matter was calculated structurally as $DM = RUE \times I_{\mathrm{int}}$. This equation enforces the Monteith-style production logic that dry-matter accumulation is a function of intercepted radiation and radiation-use efficiency (Monteith, 1977). A second diagnostic quantity, whole-season apparent neural RUE, was also calculated. This was defined as the learned sequence-level RUE scalar multiplied by the average intercepted fraction. This diagnostic is equivalent to the modelled total dry matter divided by cumulative incoming PAR, $RUE_{\mathrm{app}} = DM/I_{\mathrm{cum}}$. It therefore represents the model's effective whole-season conversion of incoming PAR into total dry matter, after accounting for the fraction of radiation intercepted by the canopy. This diagnostic was retained to evaluate whether the combined interception-RUE pathway produced plausible whole-season apparent RUE, and to compare it with the observed apparent RUE proxy for tuber dry matter.

#### 2.5.6. Learning HI and tuber dry matter

HI was learned using a compact feed-forward subnetwork applied to a physiological feature set containing total dry matter, final DAE, intercepted radiation, and the whole-season apparent RUE diagnostic. These inputs were chosen because HI represents the partitioning of accumulated biomass into tubers rather than a direct canopy measurement (Kooman and Haverkort, 1995). Total dry matter provides the biomass pool available for partitioning; intercepted radiation and RUE summarise whether that biomass arose from strong light capture, high conversion efficiency, or both; and final DAE provides a developmental-time context for partitioning. The fully connected layer therefore allowed the model to learn a bounded, context-dependent partitioning coefficient from the inferred source and conversion state of the crop. The resulting HI latent was bounded and then adjusted by a small environment-specific shift. Tuber dry matter was calculated structurally as $DM_{\mathrm{tuber}} = DM \times HI$.

The graph calculated dry matter internally in g $m^{-2}$. Because 1 g $m^{-2}$ equals 0.01 t $ha^{-1}$ , tuber dry matter was divided by 100 to report t $ha^{-1}$ . No trainable layer was applied after this conversion, so the reported yield remained the tuber dry matter calculated by the physiological graph.

#### 2.5.7. Loss function and optimisation

The model had final tuber dry matter and HI as supervised outputs. Final tuber dry matter was the primary supervised target while HI was used as an auxiliary target where observed or derived HI values were available. Because HI was not available for all data sources, the auxiliary HI loss was masked so that samples with missing HI did not contribute to that component of the objective.

The total loss was defined as $L = \mathrm{MSE}\left(Y_{\mathrm{dry}}\right) + 3 \times \mathrm{maskedMSE}(HI)$. The HI term was multiplied by 3 because HI is a fractional auxiliary target, whereas the yield loss is expressed on the t $ha^{-1}$ dry-matter scale. Without up-weighting, the HI component would make a relatively small contribution to the total objective and the partitioning node could become weakly anchored even when observed or derived HI values were available. The weighting therefore encouraged the partitioning node to remain semantically connected to HI while still allowing datasets without HI observations to contribute through the final dry-matter target. The model was implemented in Python using Keras with TensorFlow and optimised using Adam with gradient clipping, an initial learning rate of 0.001, a batch size of 16 and 500 epochs chosen after an initial early-stopping run. Learning-rate reduction and early stopping were based on validation loss, and the best validation checkpoint was restored for final evaluation. The multiplier was selected to balance the scales of the two losses rather than through a systematic comparison of alternative loss weights or constraints; such comparisons remain future work.

### 2.6. Physiological-admissibility diagnostics

The model was evaluated by predictive error and three physiological-admissibility diagnostics. A diagnostic was a falsifiable expectation for whether a named intermediate quantity behaved like a crop-growth mechanism rather than an arbitrary hidden feature. The tests examined external proxy alignment, consistency with the imposed model structure and model-conditioned structural replays, without staking a claim to identification of unique causal mechanisms.

#### 2.6.1. Semantic anchoring diagnostic

The first diagnostic asked whether a named latent state was semantically anchored, that is, directionally aligned with independent physiological information where such a proxy existed. Learned RUE-related quantities, HI and canopy timing were compared with observed or independently derived summaries. Exact equality was not expected because several proxies were incomplete or described a related, rather than identical, physiological quantity.

These comparisons were interpreted cautiously because some external proxies are imperfect. In particular, several apparent RUE proxies were calculated from tuber dry matter rather than total biomass, making them ecological anchors rather than strict one-to-one targets. The key requirement was therefore not exact equality between latent and proxy, but directional and biologically interpretable alignment.

#### 2.6.2. Structural coherence diagnostic

The second diagnostic asked whether intermediate quantities were structurally coherent with the imposed model. For example, effective light interception should determine intercepted radiation; intercepted radiation and RUE should determine total dry matter; and total dry matter and HI should determine tuber dry matter. These relationships partly follow algebraically from the graph and are therefore consistency checks rather than independent biological evidence.

Structural coherence was checked in two ways. First, latent covariance described the realised fitted relationships and trade-offs. Second, one learned mechanism at a time was changed while the remaining fitted state was held fixed. An example of expected behaviour from this perturbation was that increasing RUE or light capture should increase dry matter and yield, while increasing HI should change tuber dry matter without changing total dry matter. These were tested on the trained model and tested internal monotonicity rather than causal field effects.

#### 2.6.3. Route-specific structural-replay diagnostic

The third diagnostic asked whether similar yield could arise through different modelled physiological routes and whether effecting a change to the bottleneck of that route produced a sensible model response. These structural replays only meant changing one named fitted mechanism and recomputing downstream quantities while holding the remaining fitted state fixed. They were a model-conditioned diagnostic meant to provide evidence that the trained model was knowledge-aligned, rather than an estimate of the causal effect of a field treatment.

Near-equal predicted yields were searched for among cases with contrasting total-dry-matter and HI combinations as produces by the model. We simulated partitioning replays that targeted cases that combined high-biomass with low-HI. Furthermore light-capture replays were also simulated, targeting cases that were far from interception saturation. Proportional RUE replays were also done to serve mainly as monotonicity checks. These calculations test whether the fitted representation exposes different improvement hypotheses within this population.

Like neural networks and mechanistic crop models, the KGNN can have multiple parameter settings with similar predictive performance. The fitted decomposition is therefore not claimed to be unique or causally identified. The narrower conclusion is that the saved fitted solution achieved within-population prediction while meeting diagnostic expectations for proxy alignment, graph consistency and route-specific structural behaviour.

## 3. Results

### 3.1. A broad retained panel

The retained panel of data spanned a broad range of canopy, phenological and yield conditions, summarised in Table 5. This breadth was important because the model was being tested as a trainable mechanistic parameterisation rather than as a fixed-coefficient simulator. Across all retained sequences, observed LAI values ranged from 1.11 to 7.50, with a median of 4.00. Tuber dry matter ranged from 2.03 to 30.94 t $ha^{-1}$ , with a median of 15.05 t $ha^{-1}$ . Days after emergence ranged from 43 to 179 days, indicating that the panel included both short-season and longer-season crop trajectories. Harvest index, where available, ranged from 0.45 to 0.89, with a median of 0.82. Thus, the model was not fitted only to a narrow Dutch validation domain, but was trained on a deliberately broader physiological support that included lower-yielding, lower-radiation and shorter-duration material from the additional sources.

The additional sources broadened the fitting support in complementary ways. Raymundo et al. (2018) sequences spanned cumulative PAR of approximately 447–1462 MJ $m^{-2}$ and tuber dry matter of 2.0–30.94 t $ha^{-1}$ . SUBSTOR-Potato sequences spanned 428–1364 MJ $m^{-2}$ and 2.03–26.41 t $ha^{-1}$ . Bangladesh sequences covered approximately 276–352 MJ $m^{-2}$, a maximum retained DAE of 64 and 4.4–8.3 t $ha^{-1}$ tuber dry matter. These ranges describe fitting support; they are not external validation results.

### 3.2. Source-wise fitted diagnostics and record audit

Ancillary fitted-row residual summaries showed MAE values of 0.48 t $ha^{-1}$ for Raymundo (n = 18), 0.55 t $ha^{-1}$ for SUBSTOR-Potato (n = 37), 0.57 t $ha^{-1}$ for Bangladesh (n = 24) and 0.90 t $ha^{-1}$ for the Dutch fitting records (n = 117), compared with 1.41 t $ha^{-1}$ on the Dutch validation records. These fitted residuals functioned as descriptive checks rather than independent validation. Source-average latent values remained within the imposed bounds: effective interception ranged from 0.57 to 0.81, RUE from 2.61 to 2.95 g DM $MJ^{-1}$ and HI from 0.76 to 0.80. The summaries therefore revealed source shifts but no source-level collapse to an implausible latent boundary. The Raymundo et al. (2018) dataset included an observation with tuber dry matter yield up to 30.94 t $ha^{-1}$ . Its reported dry-matter fraction was 0.2265, implying 136.6 t $ha^{-1}$ fresh yield. It was a fitting-only record, so the held-out validation metrics were unaffected. Its independent influence on trained weights was negligible so it was retained in the training dataset.

### 3.3. Held-out prediction of tuber dry matter

The model was fitted using 196 sequences and evaluated on 51 unseen Dutch sequences selected by the prespecified 30% environment-stratified split. Fig. 2 is therefore a validation plot for records held out from the target Dutch experimental population, not a calibration plot or external geographic validation. MAE was 1.41 t $ha^{-1}$ , RMSE was 1.77 t $ha^{-1}$ , relative MAE was 8.9% and relative RMSE (RRMSE) was 11.2%. The observed and predicted means were 15.78 and 15.75 t $ha^{-1}$ , respectively. Observed yield had a standard deviation of 2.61 t $ha^{-1}$ , compared with 1.98 t $ha^{-1}$ for predictions, indicating compression of the validation range. Reported separately by the two stratification environments, held-out RMSE was consistent with modest differences at 1.55 t $ha^{-1}$ for Lelystad (n = 26) and 1.97 t $ha^{-1}$ for Vredepeel (n = 25).

The ordinary and environment-year-blocked validation results are compared in Table 6.

Across the 168 pooled out-of-fold Dutch predictions, grouped environment-year-blocked validation produced an MAE of 1.49 t $ha^{-1}$ , an RMSE of 1.63 t $ha^{-1}$ , an RRMSE of 10.42%. This compared with an

**Table 5**
Distribution of key panel variables used by the KGNN.

| Item | Unit | n | Minimum | Maximum | Median | Mean | Stdev |
|---|---|---|---|---|---|---|---|
| Cumulative PAR | MJ $m^{-2}$ PAR | 247 | 275.68 | 1462.19 | 1181.09 | 1064.30 | 293.52 |
| Final DAE | days | 247 | 43.00 | 179.00 | 155.00 | 131.27 | 35.05 |
| Harvest index | fraction | 192 | 0.45 | 0.89 | 0.82 | 0.79 | 0.09 |
| LAI | $m^2$ leaf $m^{-2}$ ground | 247 | 1.11 | 7.50 | 4.00 | 3.92 | 1.35 |
| Tuber dry matter | t $ha^{-1}$ | 247 | 2.03 | 30.94 | 15.05 | 14.11 | 4.80 |

Note. LAI = leaf area index; DAE = days after emergence; PAR = photosynthetically active radiation. LAI and final DAE summarise the retained input sequence for each sample; cumulative PAR is the incoming radiation exposure used by the structural dry-matter pathway.

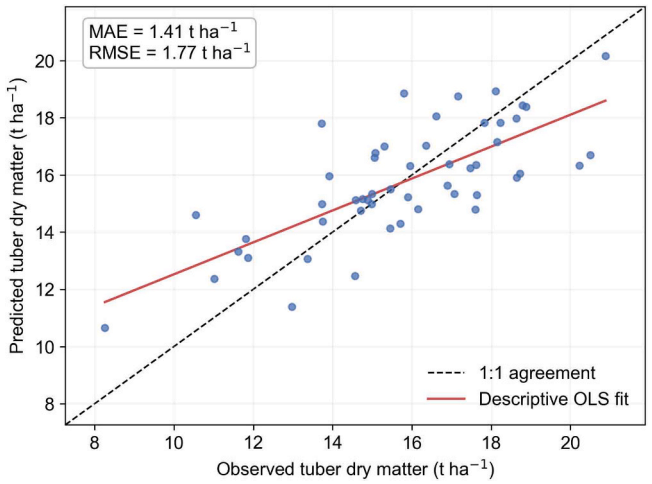


**Fig. 2.** Observed versus predicted tuber dry matter for the 51 Dutch sequences held out from model fitting. This is within-population validation, not calibration or independent geographic validation. MAE = mean absolute error; RMSE = root mean square error; RRMSE = relative root mean square error. The dashed black line shows 1:1 agreement, and the solid red line shows the full-data ordinary-least-squares fit. The regression is descriptive only and was not used to adjust the predictions.

**Table 6**
Ordinary and grouped/blocked validation within the Dutch target population.

| Evaluation measure | Ordinary held-out validation | Grouped environment-year-blocked validation |
|---|---|---|
| Splitting unit | Individual crop sequence | location × year environment |
| Outer validation design | Single 30% split, stratified by location | Four-fold leave-one-location-year-out |
| Dutch out-of-sample predictions, n | 51 | 168 pooled out-of-fold predictions |
| MAE (t ha$^{-1}$) | 1.41 | 1.49 |
| RMSE (t ha$^{-1}$) | 1.77 | 1.63 |
| RRMSE (%) | 11.2 | 10.42 |

Note. Ordinary validation used the existing sequence-level split. In the grouped analysis, all cultivars, treatments and replicates belonging to that environment-year were withheld together, preventing their distribution across fitting and validation. The grouped estimates are based on pooled out-of-fold predictions. MAE = mean absolute error; RMSE = root mean square error; RRMSE = relative root mean square error

MAE of 1.41 t ha$^{-1}$ , an RMSE of 1.77 t ha$^{-1}$ and an RRMSE of 11.2% under the ordinary sequence-level split. Blocking related observations therefore showed that there was no evidence of substantial optimistic bias from the ordinary split.

### 3.4. Learned physiological parameters

The first test was whether the learned mechanism values occupied plausible physiological ranges rather than collapsing to arbitrary boundaries. The fitted model produced latent states for the 196 training samples, summarised in Table 7. The learned sequence-level RUE scalar had a mean of 2.83 g DM MJ$^{-1}$ intercepted PAR and ranged from 0.92 to 3.78 g DM MJ$^{-1}$ intercepted PAR. Learned harvest index had a mean of 0.78 kg kg$^{-1}$ and ranged from 0.66 to 0.85 kg kg$^{-1}$ . The effective light-interception growth-rate parameter ranged from 0.03 to 0.34 d$^{-1}$ , the relative midpoint ranged from 0.04 to 0.55 of the observed crop window, and the maximum effective interception potential ranged from 0.60 to 1.01 as a fraction of incoming PAR. The whole-season apparent neural RUE diagnostic had a mean of 1.72 g DM MJ$^{-1}$ incoming PAR and ranged from 0.25 to 3.45 g DM MJ$^{-1}$ incoming PAR.

**Table 7**
Summary statistics for learned physiological latent states.

| Latent | Count | Mean | Std | Min | Median | Max |
|---|---|---|---|---|---|---|
| Interception rate k (d$^{-1}$) | 196 | 0.16 | 0.12 | 0.03 | 0.15 | 0.34 |
| Interception midpoint $x_0$ | 196 | 0.27 | 0.14 | 0.04 | 0.27 | 0.55 |
| Interception asymptote L | 196 | 0.87 | 0.12 | 0.60 | 0.92 | 1.00 |
| RUE (g DM MJ$^{-1}$) | 196 | 2.83 | 0.36 | 0.92 | 2.92 | 3.78 |
| Mean intercepted fraction | 196 | 0.61 | 0.14 | 0.27 | 0.59 | 0.93 |
| Intercepted PAR (MJ m$^{-2}$) | 196 | 603.49 | 207.39 | 139.38 | 636.11 | 1207.70 |
| NN_total_RUE | 196 | 1.72 | 0.47 | 0.25 | 1.68 | 3.45 |
| Total DM (g m$^{-2}$) | 196 | 1739.89 | 702.68 | 331.88 | 1859.95 | 4566.75 |
| Tuber DM (g m$^{-2}$) | 196 | 1343.43 | 505.09 | 276.59 | 1451.56 | 3123.38 |
| HI (kg kg$^{-1}$) | 196 | 0.78 | 0.03 | 0.66 | 0.78 | 0.85 |
| Tuber DM (t ha$^{-1}$) | 196 | 13.43 | 5.05 | 2.77 | 14.52 | 31.23 |

Note. RUE = radiation-use efficiency; DM = dry matter; HI = harvest index. Apparent neural RUE is the fitted whole-season diagnostic. Total DM and tuber DM are values calculated by the physiological graph; tuber dry matter is reported in t ha$^{-1}$ after unit conversion. All latent states are learned model quantities, not direct measurements.

These values indicate that the model used the allowed physiological parameter ranges rather than collapsing all samples to a single boundary solution. The HI in particular did not simply reproduce the full observed range. Although observed or derived HI values extended down to approximately 0.45, the learned HI distribution was narrower, ranging from approximately 0.66–0.85. This suggests that the model did not force all observed low-HI variation directly into the HI node. Instead, some variation was apparently absorbed through the broader source, biomass and interception structure. This is important because the model was not designed to make each latent node mimic its proxy exactly; rather, each latent must contribute to a coherent yield-forming decomposition. The learned maximum effective interception potential was generally high, with a median near 0.92. It represents the asymptote of the effective intercepted fraction learned from the LAI trajectory. Similarly, the midpoint parameter represents the timing of the effective interception curve.

### 3.5. Semantic anchoring of the learned latent states

The next interpretability test asked whether observation-specific learned states aligned with independently derived physiological information, as summarised in Table 8. Correlations were interpreted as indicators of semantic-anchoring rather than supervised targets. We calculated the confidence intervals around the correlations using non-

**Table 8**
Semantic anchoring correlations between learned latent states and observed or derived physiological proxies.

| Latent | Proxy | Spearman r | 95% Bootstrap CI |
|---|---|---|---|
| Apparent neural RUE | Observed apparent RUE | 0.88 | [0.83, 0.91] |
| Learned midpoint $x_0$ | Observed peak DAE | 0.48 | [0.37, 0.59] |
| Learned HI | Observed/derived HI | 0.27 | [0.14, 0.39] |

Note. RUE = radiation-use efficiency; HI = harvest index; DAE = days after emergence; CI = confidence interval. Correlations are used as semantic anchors rather than supervised targets. The observed apparent RUE proxy used here is derived from tuber dry matter divided by cumulative irradiance, so it is an ecological anchor for the combined yield-forming pathway rather than a strict total-biomass RUE target.

parametric bootstraps over subsets of the dataset during the inference stage. This diagnostic was strongest for the combined interception–RUE pathway, whose whole-season apparent RUE showed the closest alignment with the external apparent-RUE proxy (Fig. 3).

The learned interception midpoint also showed evidence of semantic anchoring to observed canopy development. This comparison was not intended to show that the learned midpoint and observed peak LAI timing are the same quantity. Observed peak timing is a descriptive feature of the sampled LAI trajectory, whereas the learned midpoint is a parameter of the model's effective light-interception curve. However, crop-growth theory implies that these two quantities should be directionally related: canopies that develop later should tend to have later effective interception midpoints, while earlier-developing canopies should tend to have earlier midpoints (Kooman and Haverkort, 1995).

The positive Pearson and Spearman correlations, approximately 0.47 and 0.48, therefore provide a construct-validity check on the timing latent. They suggest that the learned midpoint captured meaningful information about canopy-development timing, while the moderate magnitude of the association is also biologically and statistically plausible because peak LAI timing and the midpoint of cumulative light interception are related but non-identical summaries of canopy dynamics. The result therefore supports semantic anchoring of the latent state without over-interpreting it as a direct estimate of the observed date of peak LAI.

HI was the weakest external semantic anchor. Its correlations with observed or derived HI were 0.26 (Pearson) and 0.27 (Spearman), and its learned range was narrower than the observed range. The learned value should therefore be read as a bounded, model-conditioned partitioning term inside the fitted graph, not as a direct reconstruction of measured HI. Its graph position, bounds and auxiliary loss prevent it from being an unconstrained hidden feature, but stronger interpretation requires datasets with comparable total-biomass and tuber-partitioning measurements.

### 3.6. The learned covariance structure provided a diagnostic map of the production graph

The covariance heatmap in Fig. 4 provides the broad diagnostic map of the fitted graph. Intercepted radiation versus total dry matter returned $r = 0.94$ (95% CI [0.92, 0.96]); total dry matter versus tuber dry matter returned $r = 0.99$ ([0.99, 1.00]); average intercepted fraction versus apparent neural RUE returned $r = 0.90$ ([0.86, 0.94]); and total dry matter versus learned HI returned $r = -0.54$ ([−0.65, −0.42]). The negative biomass–partitioning association was a more informative description of a tradeoff between gross productivity and partitioning pathways toward yield. The persistence of this biomass-partitioning structure across the training data manifold is visualised in Fig. 5. Despite intercepted radiation showing a positive overall relationship with total dry matter, the harvest index showed a negative trend, while the RUE showed an overall negative but study-wise inconsistent trend.

Table 9 summarises how predicted yield changed when either the observed inputs or the model's fitted internal physiological quantities were altered. For the 51 held-out Dutch sequences, changing observed LAI by 0.80, 0.90, 1.10 and 1.20 changed the mean predicted tuber dry matter by −1.42, −0.71, + 0.68 and + 1.30 t $ha^{-1}$ , respectively. The corresponding changes from scaling cumulative PAR were −4.04, −2.18, + 2.06 and + 3.95 t $ha^{-1}$ . Both input responses were smooth and monotonic, with the larger PAR response consistent with its direct multiplicative role in dry-matter formation. Directly replaying fitted HI, RUE and light interception also changed yield monotonically in the directions imposed by the physiological graph. The observed-input rows in Table 9 show how changing measured LAI or radiation values around their observed levels affects the final yield prediction after passing through the entire fitted model. The internal-structure rows check whether changing a fitted physiological quantity alters yield in the expected direction. Neither analysis involved retraining.

### 3.7. Similar yield could arise through contrasting biomass–partitioning routes

The route-specific diagnostic asked whether similar fitted yield could coexist with different upstream model states. The directed graph organised near-equal yields into contrasting combinations of biomass accumulation and partitioning. These alternatives expose fitted equifinality rather than uniquely identified physiological causes.

We defined equifinality as a similar fitted yield reached through different upstream model states. Pairs were considered yield-equivalent when their predicted yields differed by no more than 5% of the median structural yield, and route-distinct when the summed absolute differences in total dry matter and HI, each standardised by its across-sample standard deviation, were at least 1.5. A summary of the equifinal pairs are shown in Table 10. Their median absolute yield gap was 0.38 t $ha^{-1}$ (2.6% of the 14.86 t $ha^{-1}$ median structural yield), while their median total-dry-matter gap was 154.53 g $m^{-2}$ (8.0% of the median) and median HI gap was 0.06 (7.5% of the median). Standardised route distances ranged from 1.50 to 4.99. These are alternative fitted

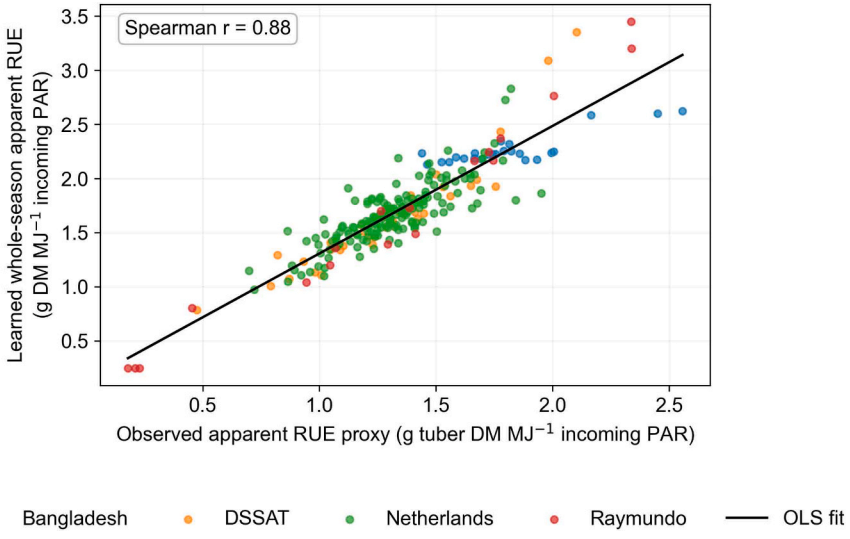


**Fig. 3.** Semantic anchoring of the learned whole-season apparent radiation-use-efficiency (RUE) diagnostic against the observed apparent RUE proxy. Points are coloured by source; the black line is the overall ordinary-least-squares regression fit.

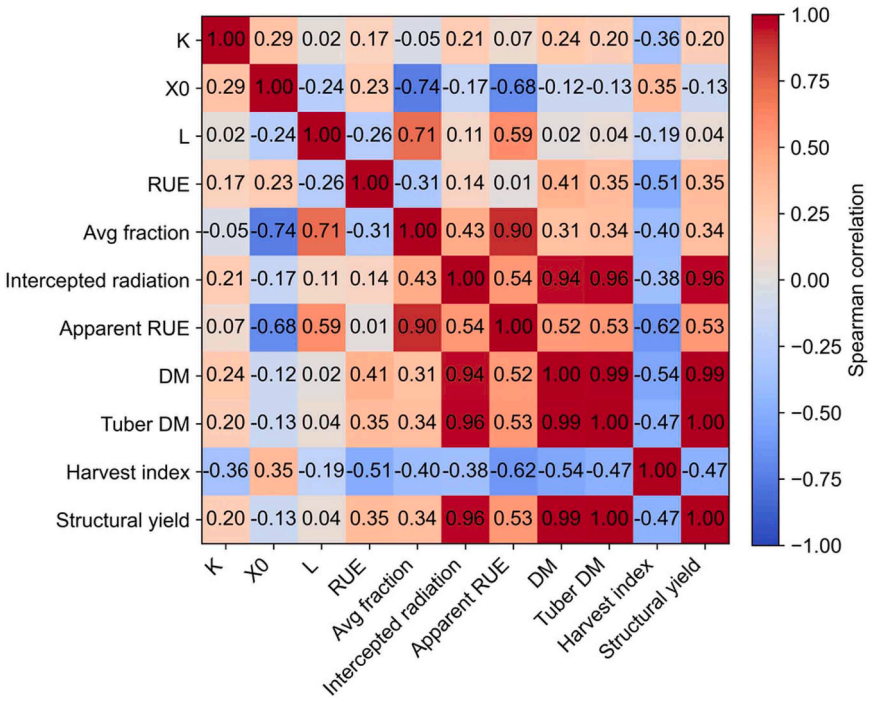


**Fig. 4.** Spearman correlation heatmap for learned latent states and structural yield quantities. Cell labels give correlation coefficients. RUE = radiation-use efficiency; DM = dry matter; HI = harvest index; CI = confidence interval. Correlations between deterministically related variables are shown for transparency but are interpreted as graph-consistency checks, not independent statistical evidence.

explanations within the KGNN contingent on the input data, rather than uniquely identified physiological causes.

Route distance was the sum of standardised absolute gaps in total dry matter and HI, placing the two quantities on comparable scales. Larger values identify pairs with more contrasting fitted biomass–partitioning states despite similar yield. Such pairs can generate testable physiological hypotheses (for example, whether a high-biomass, lower-partitioning line and a lower-biomass, higher-partitioning line should be phenotyped differently) but the latent states should not replace direct measurement.

### 3.8. Bottleneck-aligned interventions produced route-specific responses

Route-specific structural replays were also used to ask whether cases with similar predicted yields, but different fitted physiological profiles would respond differently when their modelled yield-limiting latent features were changed. By separately simulating improvements in partitioning, RUE and light capture while holding the other fitted quantities constant, we tested whether improving a potential yield-limiting factor produced a greater benefit when it was the main factor holding yield back than when yield was more strongly limited elsewhere. Exact interventions and responses are reported in Tables 11 and 12.

For the partitioning replay, HI was moved 20% of the remaining gap toward the numerical ceiling of 0.95. With this scenario, cases that started with the combination of high-dry-matter and low-HI gained 1.06 t $ha^{-1}$, compared with 0.33 t $ha^{-1}$ for low-dry-matter, high-HI cases. This fitted response logically follows from the larger biomass pool and greater remaining distance to the partitioning ceiling in the first route, though it is not a causal estimate of a breeding or field intervention.

Increasing fitted RUE by 10% while holding intercepted radiation fixed gave mean gains of 1.17, 1.43 and 1.57 t $ha^{-1}$ in the low-, mid- and high-RUE backgrounds, respectively. Because the replay is a proportional change in a multiplicative graph, absolute gains primarily reflect baseline yield and downstream partitioning. This replay served as a confirmation that the fitted values were propagated correctly through the specified physiological calculation. Closing 20% of the remaining gap between current intercepted radiation and incoming PAR produced gains of 2.80, 2.13 and 1.01 t $ha^{-1}$ in the low-, mid- and high-interception classes, respectively. The declining response with greater baseline interception is the expected fitted bottleneck pattern because less unrealised radiation remains near saturation.

Taken together, the replays show that the model predicted the greatest benefit when a simulated change addressed the factor currently limiting yield. Improving partitioning produced the largest gains in cases with high biomass but relatively little allocation to tubers, whereas improving light capture was most beneficial in cases intercepting relatively little of the available radiation. By contrast, increasing RUE by the same proportion increased yield proportionally across all groups, so this replay served mainly to verify the model's internal calculation rather than to distinguish among improvement strategies.

## 4. Discussion

### 4.1. A trainable mechanistic crop graph can make local potato datasets physiologically informative

This study tested whether a simple mechanistic potato dry-matter model could be made locally trainable by replacing fixed coefficients with neural parameter-generating functions. The KGNN retained a directed light-capture, RUE, biomass and partitioning graph while inferring observation-specific parameter values from canopy observations, radiation and environment. Unlike a generic black-box predictor, it could not bypass the graph with an arbitrary final yield layer. Unlike conventional fixed-parameter models such as WOFOST, SUBSTOR-Potato, APSIM, AquaCrop and LINTUL-POTATO, its named parameter values varied conditionally among observations (Holzworth et al., 2014; Jones et al., 2003; Raes et al., 2018; Raymundo et al., 2017; de Wit et al., 2019; Kooman and Haverkort, 1995).

The present model learns a conditional parameter-generating function that produces sample-specific physiological states from the crop sequence and its context, in contrast to traditional fixed-parameter

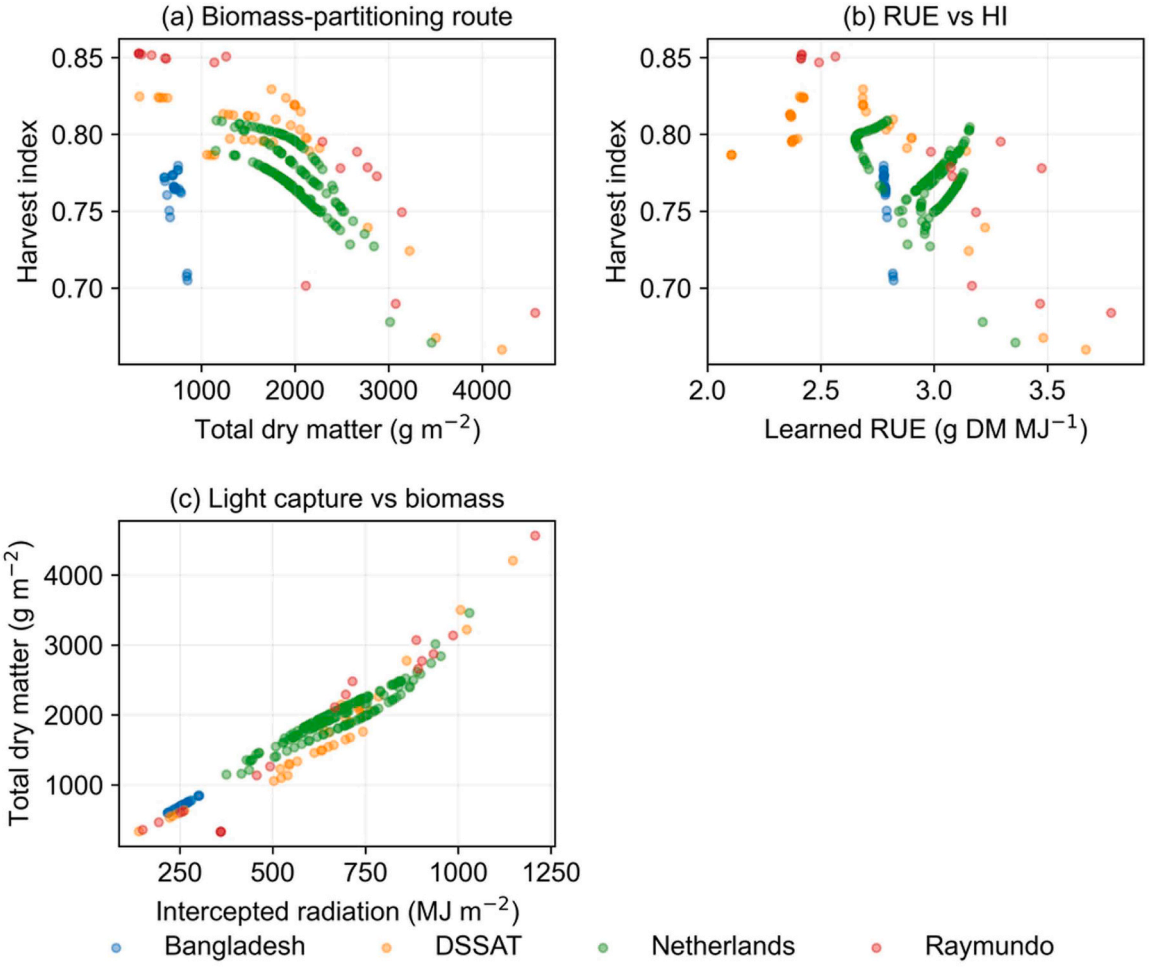


**Fig. 5.** Latent physiological trade-offs by study: (a) biomass-partitioning route, (b) RUE versus HI, and (c) light capture versus biomass. Points show individual sequences grouped by study.

models. It also differs from a mixed-effects parameterisation. Fixed and random effects require the modeller to pre-declare which grouping variables explain variation (Bates et al., 2015). Here, the canopy trajectory, crop duration and radiation context directly inform the inferred mechanism values, while environment effects are restricted to small shifts in named physiological parameters. Detailed datasets containing canopy trajectories, radiation, biomass, harvest index and tuber dry matter are expensive to generate (Morisse et al., 2022; Ninomiya, 2022), so this local, observation-conditioned inference is a practical route for breeder-relevant physiological diagnosis. The physiological graph reduces the hypothesis space available to the neural network, but candidate explanations still need to remain physiologically admissible (Wallach et al., 2018).

### 4.2. *Predictive performance was competitive given the deliberately restricted input space*

Within the held-out Dutch target population, MAE was 1.41 t ha$^{-1}$ (8.9% of the observed mean), RMSE was 1.77 t ha$^{-1}$ and RRMSE was 11.2%, with little mean bias but compressed extremes. These errors are agronomically useful and are comparable with the broader yield-prediction literature, especially given the restricted input set. Kurek et al. (2023) reported strong potato yield prediction using heavily engineered biophysical, satellite and ground-data features (around 5.8% MAPE), but that performance depends on a richer predictor set and does not provide physiological latent states. Lin et al. (2023) reported much larger uncertainty in light-interception-based mechanistic yield modelling, with relative RMSE values that can approach 28%, illustrating the challenge of predicting yield with traditional process models when parameters and scenarios are imperfectly specified.

### 4.3. *The RUE pathway recovered a recognised physiological efficiency signal*

The strongest semantic anchoring occurred for the combined interception–RUE pathway, where the whole-season apparent neural RUE diagnostic showed strong alignment with the observed apparent RUE proxy (Spearman r = 0.88). This is important because RUE was not directly supervised. In classical crop physiology, RUE is commonly estimated as the slope of biomass accumulation against intercepted radiation (Monteith, 1977; Robles-Zazueta et al., 2021). The KGNN retains this interpretation but generalises it from a single population-level coefficient to a context-dependent latent slope learned from intercepted radiation, crop duration, cumulative radiation context and the inferred shape of the light-capture trajectory.

This result is in line with prior potato and crop-modelling work in which RUE is treated as a central determinant of biomass production. Many crop models use fixed or cultivar-specific RUE values, often around 2.5 g MJ$^{-1}$ PAR for potato under favourable conditions, while field studies report substantial variation depending on cultivar, nitrogen, water, temperature, canopy state and measurement protocol (Machakaire et al., 2023; Saldaña-Villota and Cotes-Torres, 2024). The learned RUE values in this study (mean 2.83 g DM MJ$^{-1}$ intercepted PAR) fell within a plausible physiological range and varied contextually,

**Table 9**
Model responses to observed-input perturbations and internal structural replays.

| Perturbation source | Intervention | Mean delta yield (t ha$^{-1}$) | 95% bootstrap CI |
|---|---|---|---|
| Observed input | LAI × 0.80 | -1.42 | [-1.65, −1.17] |
| Observed input | LAI × 0.90 | -0.71 | [-0.84, −0.58] |
| Observed input | LAI × 1.10 | 0.68 | [0.52, 0.86] |
| Observed input | LAI × 1.20 | 1.30 | [0.98, 1.64] |
| Observed input | Cumulative PAR × 0.80 | -4.04 | [-4.16, −3.93] |
| Observed input | Cumulative PAR × 0.90 | -2.18 | [-2.23, −2.13] |
| Observed input | Cumulative PAR × 1.10 | 2.06 | [2.00, 2.12] |
| Observed input | Cumulative PAR × 1.20 | 3.95 | [3.82, 4.09] |
| Internal structure | HI × 0.90 | -1.39 | [-1.45, −1.34] |
| Internal structure | HI × 1.00 | 0.00 | [-0.00, 0.00] |
| Internal structure | HI × 1.10 | 1.39 | [1.33, 1.45] |
| Internal structure | RUE × 0.80 | -2.78 | [-2.89, −2.66] |
| Internal structure | RUE × 1.00 | 0.00 | [-0.00, 0.00] |
| Internal structure | RUE × 1.20 | 2.78 | [2.66, 2.90] |
| Internal structure | Light interception × 0.80 | -2.78 | [-2.90, −2.67] |
| Internal structure | Light interception × 1.00 | 0.00 | [-0.00, 0.00] |
| Internal structure | Light interception × 1.20 | 2.78 | [2.67, 2.89] |

Note. Observed-input perturbations were applied to the 51 held-out Dutch sequences, with confidence intervals bootstrapped across their sequence-level responses. Internal-structure replays were evaluated across the 247 retained fitted sequences. LAI = leaf area index; PAR = photosynthetically active radiation; RUE = radiation-use efficiency; HI = harvest index; CI = confidence interval. All analyses used the saved fitted KGNN without retraining. Observed-input rows assess end-to-end input sensitivity; internal-structure rows replay the fitted graph.

**Table 10**
Near-equal-yield pairs with contrasting modelled biomass–partitioning routes.

| Summary | Value |
|---|---|
| Qualifying comparisons | 586.00 |
| Median absolute yield gap (t ha$^{-1}$) | 0.38 |
| Median absolute DM gap (g m$^{-2}$) | 154.53 |
| Median absolute HI gap | 0.06 |
| Minimum route distance | 1.50 |
| Maximum route distance | 4.99 |

Note. DM = dry matter; HI = harvest index. Route distance is the sum of the standardised absolute DM and HI gaps between two samples with near-equivalent predicted yield.

consistent with known genotype × environment × management effects on conversion efficiency. The finding also agrees with studies showing that radiation conversion efficiency is not a universal constant. Nitrogen status, canopy age, water status and stress can alter efficiency (Zhou et al., 2017; Saldaña-Villota and Cotes-Torres, 2024). Direct RUE measurement requires repeated biomass sampling and radiation interception estimates, which is labour-intensive (Tewes and Schellberg, 2018). The KGNN suggests that a model constrained by the correct production graph can infer a RUE-like state from more accessible canopy and radiation observations.

### *4.4. Canopy timing was semantically anchored, but effective light capture is not equivalent to LAI reconstruction*

The learned interception midpoint showed moderate alignment with observed canopy timing (Spearman r = 0.48), as expected because the two quantities are related but not identical (Kooman and Haverkort, 1995; de Wit et al., 2019). LAI was supplied as an observed phenotype, not simulated from genotype, soil, water, nitrogen or weather. Peak LAI timing is a descriptive statistic of the sampled canopy trajectory, whereas the learned midpoint is a parameter of an effective light-interception curve. The moderate magnitude of the association is therefore biologically sensible.

This distinction is important in relation to prior potato canopy work. Firman and Allen (1989) showed that potato canopy development, ground cover and light interception are closely related, but also that canopy descriptors should not be treated as perfect substitutes for intercepted radiation. The single-sigmoid interception layer was a deliberate seasonal approximation rather than a full canopy life-cycle model, consistent with the radiation-use-efficiency tradition and models such as LINTUL-POTATO (Kooman and Haverkort, 1995). Senescence was implicit in the LSTM encoder and environment shifts without a separate parameter, suiting the mixed dataset where full-season and partial records coexist. Future refinements (e.g., double-sigmoid) should be justified by improvements in prediction, anchoring and postulates rather than visual completeness alone.

**Table 11**
Route-specific structural-replay response summaries.

| Route or background class | Intervention | Mean delta yield (t ha$^{-1}$) | 95% CI for mean delta yield |
|---|---|---|---|
| High-DM / low-HI | HI + 20% of remaining gap to 0.95 | 1.06 | [0.97, 1.17] |
| Low-DM / high-HI | HI + 20% of remaining gap to 0.95 | 0.33 | [0.28, 0.37] |
| High RUE background | RUE + 10%, intercepted radiation fixed | 1.57 | [1.49, 1.66] |
| Low RUE background | RUE + 10%, intercepted radiation fixed | 1.17 | [1.06, 1.28] |
| Mid RUE background | RUE + 10%, intercepted radiation fixed | 1.43 | [1.34, 1.53] |
| High interception | Interception + 20% of remaining incoming PAR gap | 1.01 | [0.88, 1.13] |
| Low interception | Interception + 20% of remaining incoming PAR gap | 2.80 | [2.65, 2.95] |
| Mid interception | Interception + 20% of remaining incoming PAR gap | 2.13 | [2.03, 2.22] |

**Table 12**
Bootstrap contrasts for bottleneck-aligned structural-replay responses.

| Intervention | Contrast | Mean delta-yield difference (t ha$^{-1}$) | 95% Bootstrap CI |
|---|---|---|---|
| HI improvement | High-DM / low-HI minus low-DM / high-HI | 0.74 | [0.63, 0.85] |
| Light-capture improvement | Low interception minus high interception | 1.80 | [1.60, 1.99] |
| Light-capture improvement | Low interception minus mid interception | 0.68 | [0.51, 0.85] |
| Light-capture improvement | Mid interception minus high interception | 1.12 | [0.96, 1.26] |

Note. HI = harvest index; RUE = radiation-use efficiency; CI = confidence interval; KGNN = knowledge-guided neural network. Responses and contrasts are reported in t ha$^{-1}$ . All interventions are model-conditioned structural replays of the saved fitted KGNN, not retrained models or causal field-effect estimates.

### *4.5. Harvest index was only partially anchored, revealing the current limit of partitioning inference*

Learned HI was only weakly aligned with observed or derived values (Spearman r = 0.27) and occupied a narrower range. This is in line with the broader difficulty of modelling partitioning in potato. Harvest index is often treated as a relatively stable or calibrated coefficient in crop models, but in practice it reflects source–sink dynamics, tuber initiation,

tuber bulking, cultivar maturity, stress timing and sampling protocol (Kooman and Haverkort, 1995; Kooman et al., 1996). It is therefore best described as a bounded, model-conditioned partitioning term that closes the fitted dry-matter graph, not as a recovered measurement of harvest index. Its structural role and auxiliary supervision make it more constrained than an arbitrary hidden feature, but direct biomass and partitioning measurements across sources are needed before treating it as a transferable physiological trait.

### 4.6. The biomass–partitioning trade-off agrees with source–sink theory and exposes structured equifinality

The negative fitted association between total dry matter and HI (Spearman $r = -0.54$) made alternative biomass–partitioning routes visible, consistent with source–sink theory (Smith et al., 2018; Reynolds et al., 2021) and empirical potato studies on allocation trade-offs, prolonged canopy duration, and vegetative investment effects (Firman and Allen, 1989; Ávila-Valdés et al., 2020; Li et al., 2016; Chang et al., 2018). However, these outcomes are model-conditioned explanations within one admissible fitted solution. The structural replays translate them into hypotheses for targeted phenotyping (Yenkikar et al., 2025), though they do not establish that manipulating a latent value would cause the predicted field-yield change.

### 4.7. The framework advances knowledge-guided neural networks for crop science

The advance demonstrated here is that a physiologically grounded crop model can be trained end to end while retaining meaningful internal biological structure. Neural subnetworks estimated observation-specific light-capture parameters, RUE and HI, but every yield prediction had to pass through the specified physiological pathway linking intercepted radiation, biomass production and tuber partitioning. The learned whole-season apparent-RUE diagnostic was strongly correlated with the observed apparent-RUE proxy (Spearman ($r = 0.88$)), providing external evidence that the model recovered a meaningful RUE-related signal rather than merely assigning a physiological name to an arbitrary hidden feature. The fitted relationships and structural replays also followed the behaviour required by the physiological graph, confirming that its named mechanisms remained active in producing predictions. Together, these results provide evidence of knowledge alignment within an end-to-end trained neural model. Although they do not establish a unique causal explanation, they go beyond architectural plausibility by directly testing whether the learned internal representation remained consistent with crop-growth knowledge.

### 4.8. Limitations and future directions

The present study represents a feasibility demonstration for a framework whose strongest application is likely to be large, programme-specific breeding and phenotyping populations. Such datasets are often proprietary and not available in the public domain, yet they are precisely where a trainable neural-mechanistic model would be most valuable. The accessible datasets used here are smaller, heterogeneous and partly incomplete, but they support the feasibility of the core concept that a KGNN can retain an explicit crop-growth graph while learning observation-specific latent states from available canopy, radiation and dry-matter information. The results should therefore be read not as a claim that the fitted model is a universal potato predictor, but as evidence for the direction that KGNNs in crop science could take when larger breeding-scale datasets are available. Future work should apply this framework to proprietary or consortium-scale breeding panels, with year-, location- and genotype-blocked evaluations, explicit water and nitrogen covariates, harmonised thermal-time descriptors, and direct biomass and partitioning measurements to strengthen latent-state anchoring. Independent cross-source generalisation was not tested. Alternative auxiliary-HI loss weights and stronger partitioning constraints were not systematically compared and remain priorities for future work on an expanded panel.

### 4.9. Conclusion

This study demonstrates that a physiologically structured potato model can be trained end to end on a comparatively small, heterogeneous dataset while achieving useful held-out prediction within the Dutch target population. The strong correlation between the learned whole-season apparent-RUE diagnostic and the observed apparent-RUE proxy, together with the expected structural responses, provides evidence that the fitted model remained aligned with crop-growth knowledge, most clearly along the light-capture/RUE pathway. The main contribution is the combination of observation-specific neural parameter estimation, a physiological graph that cannot be bypassed when producing yield, and explicit tests of whether the resulting internal states retain their intended meaning. After fitting or recalibration to a target breeding population, the model could help position new lines within its learned physiological distribution and generate hypotheses about their yield-limiting factors and improvement routes. These results do not establish causal field effects or universal transfer of the saved model; evaluation in larger, independent breeding populations is the necessary next step.

## Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## CRediT authorship contribution statement

**Kanthu Joseph Mhango:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Megan Jane Lewis:** Writing – review & editing. **Joe M. Roberts:** Writing – review & editing. **W. Edwin Harris:** Writing – review & editing.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to proofread language and reconstruct prose for readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors declare that no external funding was received for this work and that all contributors are listed as co-authors.

## Data availability

This study utilised publicly available datasets from prior studies. The sources of these datasets have been duly cited in the article for provenance.

## References


Ávila-Valdés, A., Quinet, M., Lutts, S., Martínez, J.P., Lizana, X.C., 2020. Tuber yield and quality responses of potato to moderate temperature increase during tuber bulking under two water availability scenarios. Field Crops Res. 251, 107786. https://doi.org/10.1016/j.fcr.2020.107786.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48.

Bouman, B.A.M., van Keulen, H., van Laar, H.H., Rabbinge, R., 1996. The 'School of de Wit' crop growth simulation models: a pedigree and historical overview. Agric. Syst. 52 (2–3), 171–198. https://doi.org/10.1016/0308-521X(96)00011-X.

Chang, D.C., Jin, Y.I., Nam, J.H., Cheon, C.G., Cho, J.H., Kim, S.J., Yu, H.-S., 2018. Early drought effect on canopy development and tuber growth of potato cultivars with different maturities. Field Crops Res. 215, 156–162. https://doi.org/10.1016/j.fcr.2017.10.008.

Chang, Tian-Gen, Zhu, Xin-Guang, 2017. Source–sink interaction: a century old concept under the light of modern molecular systems biology. J. Exp. Bot. 68 (16), 4417–4431. https://doi.org/10.1093/jxb/erx002.

Chang, Y., Latham, J., Licht, M., 2023. A data-driven crop model for maize yield prediction. Commun. Biol. 6, 439. https://doi.org/10.1038/s42003-023-04833-y.

Correa, E.S., Calderon, F.C., Colorado, J.D., 2025. ML-enhanced mechanistic crop modeling to address noise-induced uncertainty for drought environmental monitoring in rice. Discov. Food 5, 312. https://doi.org/10.1007/s44187-025-00611-3.

de Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D., Supit, I., van der Wijngaart, R., van Diepen, K., 2019. 25 Years of the WOFOST cropping systems model. Agric. Syst. 168, 154–167. https://doi.org/10.1016/j.agsy.2018.06.018.

Firman, D.M., Allen, E.J., 1989. Relationship between light interception, ground cover and leaf area index in potatoes. J. Agric. Sci. 113 (3), 355–359. https://doi.org/10.1017/S0021859600070040.

Hajjarpoor, Amir, Nelson, William C.D., Vadez, Vincent, 2022. How process-based modeling can help plant breeding deal with G x E x M interactions. Field Crops Res. 283, 108554. https://doi.org/10.1016/j.fcr.2022.108554.

Haverkort, A.J., Franke, A.C., Steyn, J.M., Pronk, A.A., Caldiz, D.O., Kooman, P.L., 2015. A Robust potato model: LINTUL-POTATO-DSS. Potato Res. 58 (4), 313–327. https://doi.org/10.1007/s11540-015-9303-7.

Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., et al., 2014. APSIM–evolution towards a new generation of agricultural systems simulation. Environ. Model. & Softw. 62, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009.

Huang, X., Kroening, D., Ruan, W., Sharp, J., Sun, Y., Thamo, E., Wu, M., et al., 2020. A survey of safety and trustworthiness of deep neural networks: verification, testing, adversarial attack and defence, and interpretability. Comput. Sci. Rev. 37, 100270. https://doi.org/10.1016/j.cosrev.2020.100270.

Jamieson, P.D., Porter, J.R., Wilson, D.R., 1991. A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. Field Crops Res. 27 (4), 337–350. https://doi.org/10.1016/0378-4290(91)90040-3.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Kooman, P.L., Fahem, M., Tegera, P., Haverkort, A.J., 1996. Effects of climate on different potato genotypes 2. dry matter allocation and duration of the growth cycle. Eur. J. Agron. 5 (3–4), 207–217. https://doi.org/10.1016/S1161-0301(96)02032-1.

Kooman, P.L., Haverkort, A.J., 1995. Modelling Development and Growth of the Potato Crop Influenced by Temperature and Daylength: LINTUL-POTATO." In *Potato Ecology and Modelling of Crops Under Conditions Limiting Growth* (edited by). In: Haverkort, A.J., MacKerron, D.K.L. (Eds.), Current Issues in Production Ecology, 3. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-0051-9_3.

Kurek, J., Niedbała, G., Wojciechowski, T., Świderski, B., Antoniuk, I., Piekutowska, M., Kruk, M., Bobran, K., 2023. Prediction of potato (Solanum tuberosum l.) yield based on machine learning methods. Agriculture 13 (12), 2259. https://doi.org/10.3390/agriculture13122259.

Li, W., Xiong, B., Wang, S., Deng, X., Yin, L., Li, H., 2016. Regulation effects of water and nitrogen on the source-sink relationship in potato during the tuber bulking stage. PLOS ONE 11 (1), e0146877. https://doi.org/10.1371/journal.pone.0146877.

Lin, Y., Li, S., Duan, S., Ye, Y., Li, B., Li, G., Lyv, D., Jin, L., Bian, C., Liu, J., 2023. Methodological evolution of potato yield prediction: a comprehensive review. Front. Plant. Sci. 14. https://doi.org/10.3389/fpls.2023.1214006.

Machakaire, A.T.B., Steyn, J.M., Franke, A.C., 2023. Photosynthesis rate, radiation and water use efficiencies of irrigated potato in a semi-arid climate using eddy covariance techniques. J. Agric. Sci. 161 (2), 217–229. https://doi.org/10.1017/S0021859623000175.

Monteith, John Lennox, 1977. Climate and the efficiency of crop production in Britain. Philos. Trans. R. Soc. Lond. B Biol. Sci. 281 (980), 277–294. https://doi.org/10.1098/rstb.1977.0140.

Morisse, Merlijn, Wells, Darren M., Millet, Emilie J., Lillemo, Morten, Fahrner, Sven, Cellini, Francesco, Lootens, Peter, et al., 2022. A European perspective on opportunities and demands for field-based crop phenotyping. Field Crops Res. 276, 108371. https://doi.org/10.1016/j.fcr.2021.108371.

Ninomiya, S., 2022. High-throughput field crop phenotyping: current status and challenges. Breed. Sci. 72 (1), 3–18. https://doi.org/10.1270/jsbbs.21069.

Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2018. Reference Manual: AquaCrop Version 6.0–6.1. FAO.

Raymundo, R., Asseng, S., Prassad, R., Kleinwechter, U., Concha, J., Condori, B., Bowen, W., et al., 2017. Performance of the SUBSTOR-potato model across contrasting growing conditions. Field Crops Res. 202, 57–76. https://doi.org/10.1016/j.fcr.2016.04.012.

Raymundo, R., Senthold, A., Rishi, P., Ulrich, K., Bruno, C., Walter, B., Joost, W., et al., 2018. Data from global field experiments for potato simulations. Harv. Dataverse. https://doi.org/10.7910/DVN/4KJ5ME.

Reynolds, M., Owen, K.A., Malcolm, B., Mark, C., Ian, C.D., John, F., Claus, F., et al., 2021. Addressing research bottlenecks to crop productivity. Trends Plant. Sci. 26 (6), 607–630. https://doi.org/10.1016/j.tplants.2021.03.011.

Robles-Zazueta, Carlos A., Molero, Gemma, Pinto, Francisco, Foulkes, M.John, Reynolds, Matthew P., Murchie, Erik H., 2021. Field-based remote sensing models predict radiation use efficiency in wheat. J. Exp. Bot. 72 (10), 3756–3773. https://doi.org/10.1093/jxb/erab115.

Saldaña-Villota, T.M., Cotes-Torres, J.M., 2024. Fluctuations in radiation use efficiency throughout the growth cycle in diploid potato crop. Am. J. Potato Res. 101 (2), 99–113. https://doi.org/10.1007/s12230-023-09938-y.

Shawon, Sarowar Morshed, Ema, Falguny Barua, Mahi, Asura Khanom, Niha, Fahima Lokman, Zubair, H.T., 2025. Crop yield prediction using machine learning: an extensive and systematic literature review. Smart Agric. Technol. 10, 100718. https://doi.org/10.1016/j.atech.2024.100718.

Smith, M.R., Rao, I.M., Merchant, A., 2018. Source–sink relationships in crop plants and their influence on yield development and nutritional quality. Front. Plant. Sci. 9, 1889. https://doi.org/10.3389/fpls.2018.01889.

Talukder, A.H.M.Motiur Rahman, Ahmed, Faruque, Ahsan, A.F.M.Shamim, Mahfuza, S. N., Mustakim, A.A.M.M., Mokarroma, Nadira, Nahar, Lutfun, Khan, Md. Shihab Uddine, Alam, Zakaria, 2026. Dataset on potato phenology, growth, and yield for DSSAT/APSIM crop model calibration at central region of Bangladesh. Data Brief. 64, 112363. https://doi.org/10.1016/j.dib.2025.112363.

Ten Den, T., Van de Wiel, I., Van Ittersum, M., Van Evert, F., de Wit, A., Reidsma, P., 2024. Agronomic dataset on potato growth and yield in the Netherlands. Open. Data J. Agric. Res. 10, 1–13. https://doi.org/10.18174/odjar.v10i0.18680.

Tewes, A., Schellberg, J., 2018. Towards remote estimation of radiation use efficiency in maize using UAV-based low-cost camera imagery. Agronomy 8 (2), 16. https://doi.org/10.3390/agronomy8020016.

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P.J., van Ittersum, M., et al., 2018. Multimodel ensembles improve predictions of crop–environment–management interactions. Glob. Change Biol. 24 (11), 5072–5083. https://doi.org/10.1111/gcb.14411.

Woli, Prem, Hoogenboom, Gerrit, Alva, Ashok, 2016. Simulation of potato yield, nitrate leaching, and profit margins as influenced by irrigation and nitrogen management in different soils and production regions. Agric. Water Manag. 171, 120–130. https://doi.org/10.1016/j.agwat.2016.04.003.

Yenkikar, A., Mishra, V.P., Bali, M., Ara, T., 2025. An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. MethodsX 15, 103442. https://doi.org/10.1016/j.mex.2025.103442.

Zhou, Z., Plauborg, F., Kristensen, K., Andersen, M.N., 2017. Dry matter production, radiation interception and radiation use efficiency of potato in response to temperature and nitrogen application regimes. Agric. For. Meteorol. 232, 595–605. https://doi.org/10.1016/j.agrformet.2016.10.017.