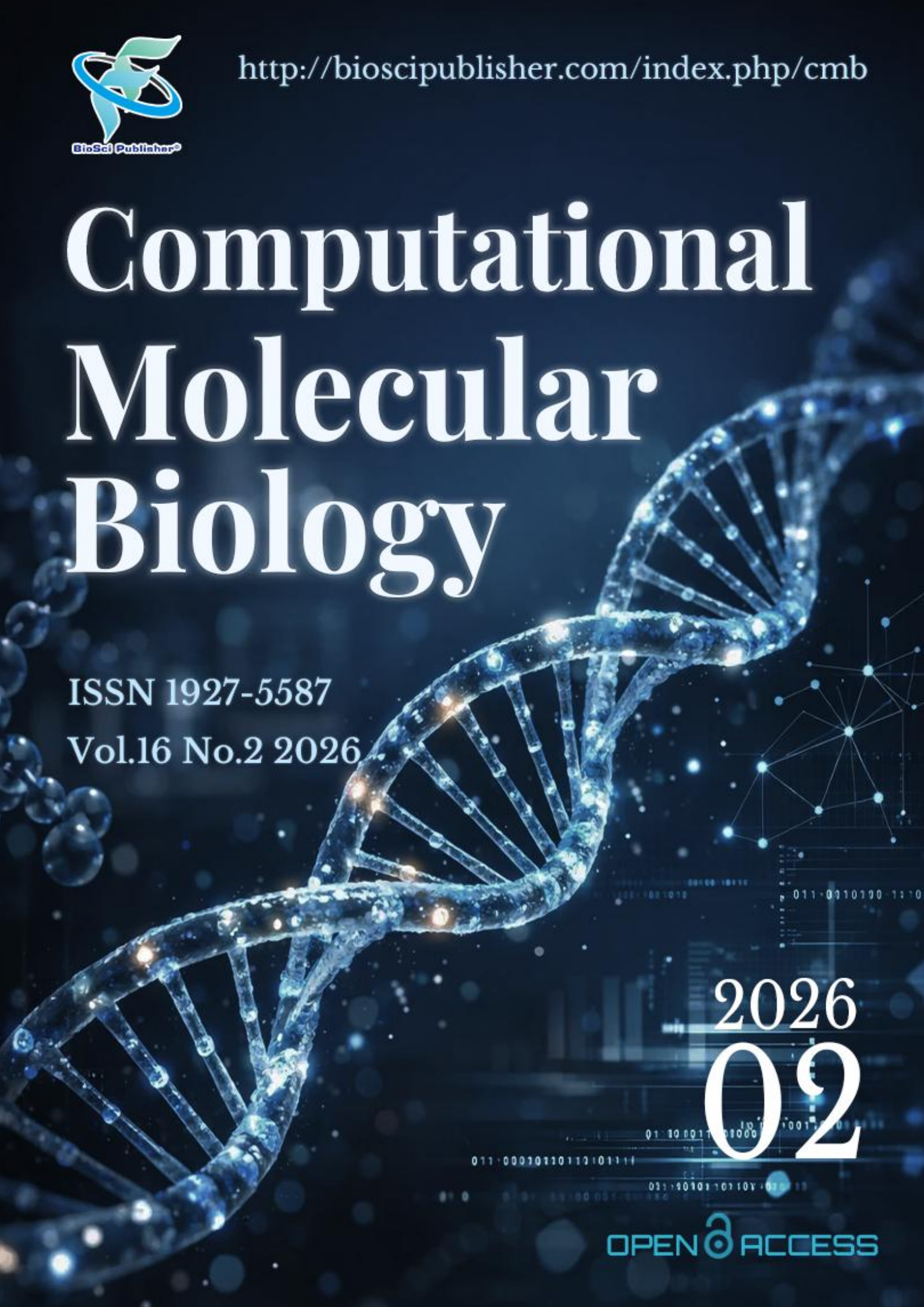
http://bioscipublisher.com/index.php/cmb

Computational Molecular Biology

ISSN 1927-5587
Vol.16 No.2 2026



2026
02

OPEN ACCESS

Publisher
BioSci Publisher

Edited by
Editorial Team of Computational Molecular Biology
Email: edit@cmb.bioscipublisher.com
Website: http://bioscipublisher.com/index.php/cmb

Address:
11388 Stevenston Hwy,
PO Box 96016,
Richmond, V7A 5J5, British Columbia
Canada

Computational Molecular Biology (ISSN 1927-5587) is an open access, peer reviewed journal published online by BioSci Publisher.

The Journal is publishing all the latest and outstanding research articles, letters, methods, and reviews in all areas of computational molecular biology, covering new discoveries in molecular biology, from genes to genomes, using statistical, mathematical, and computational methods as well as new development of computational methods and databases in molecular and genome biology. The papers published in the journal are expected to be of interests to computational scientists, biologists and teachers/students/researchers engaged in biology.

BioSci Publisher is an international Open Access publishing platform that publishes scientific journals in the field of bioscience registered at the publishing platform that is operated by Sophia Publishing Group (SPG), founded in British Columbia of Canada.

Computational Molecular Biology (online), 2026, Vol. 16, No.2
ISSN 1927-5587
http://hortherbpublisher.com/index.php/cmb

Latest Content

Computational Molecular Biology 2026, Vol.16, No.2, 71-84
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Modeling Fruit Weight Formation in Watermelon Based on Environmental Factors

Yuliang Jiang [1,2]
1 Hangzhou Liangyi Agriculture Development Co., Ltd., Hangzhou, 311113, Zhejiang, China
2 Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China
Corresponding author: jyl@hzlyny.net
Computational Molecular Biology, 2026, Vol.16, No.2 doi: 10.5376/cmb.2026.16.0006
Received: 18 Jan., 2026
Accepted: 22 Feb., 2026
Published: 05 Mar., 2026

Preferred citation for this article:
Jiang Y.L., 2026, Modeling fruit weight formation in watermelon based on environmental factors, Computational Molecular Biology, 16(2): 71-84 (doi: 10.5376/cmb.2026.16.0006)

Abstract Watermelon (*Citrullus lanatus*) fruit weight is a crucial indicator for assessing both yield and commercial value, and it is subject to the combined influence of various environmental factors. To elucidate the regulatory mechanisms by which environmental factors govern watermelon fruit weight formation, this study-grounded in the biological processes of fruit development — systematically analyzed the impact patterns of key environmental variables (including temperature, light, water, and soil nutrients) on fruit expansion and dry matter accumulation. Building upon this foundation, field experiments were conducted across multiple environmental settings to acquire data on watermelon fruit weight and related growth parameters; subsequently, utilizing a combination of statistical analysis and modeling techniques, a fruit weight formation model applicable to diverse cultivation conditions was constructed. The model was further employed to simulate and validate watermelon fruit weights under various ecological environments; the results demonstrated that the model possesses high predictive accuracy and stability, effectively capturing the dynamic responses of fruit weight formation to fluctuations in environmental factors. Furthermore, through case studies, the potential applications of the model in irrigation scheduling, fertilization management, and the optimization of controlled-environment cultivation systems were explored. The findings of this study provide a theoretical basis and technical support for achieving high-yield, high-quality watermelon cultivation, while also offering a valuable reference for the broader application and extension of fruit development models within the field of horticultural crops.
Keywords Watermelon (*Citrullus lanatus*); Fruit weight formation; Environmental factors; Model construction; Fruit development

1 Introduction

Watermelon is a major horticultural crop worldwide, valued for its large, fleshy fruits and high consumer demand, with millions of hectares under cultivation and China as the largest producer (Gao et al., 2023). Fruit size and weight are central components of yield and directly influence growers' income and market competitiveness. Industry-oriented studies and cultivar evaluations routinely use average fruit weight, fruit size distribution, and total yield as core performance indices, underlining fruit weight as a pivotal target for breeding, grafting, and agronomic management in commercial production systems (Jordana et al., 2023).

Beyond simple count of fruits, multiple analyses show that increases in total yield often arise mainly from higher average fruit weight rather than fruit number. Systematic review of grafted versus nongrafted watermelon demonstrates that grafting can raise total yield and average fruit weight by more than 10%-20%, highlighting fruit weight as the primary yield driver under diverse production conditions (Jordana et al., 2023). Field trials in different regions and seasons similarly evaluate cultivars and management practices using fruit weight, length, width, and fruit weight per plant, confirming their central role in rating economic performance and recommending cultivars to growers (Kumari et al., 2025). Management strategies such as optimizing flower retention or adjusting nitrogen and boron nutrition are explicitly aimed at maximizing individual fruit size and weight to achieve superior yield and quality (Gülüt, 2021).

At the biological level, fruit weight formation is determined by coordinated phases of cell division and cell expansion during early and mid-development. Early work on watermelon showed that after flowering, pollinated and hormone-induced parthenocarpic fruits undergo active cell proliferation in pericarp and ovule tissues, whereas

unpollinated fruits rapidly cease growth, linking successful fruit set and early cell division to subsequent fresh weight accumulation. Broader studies in model species indicate that fruit development proceeds through fruit set, a growth phase dominated by cell division and then expansion, and maturation/ripening, all tightly regulated by phytohormones such as auxin and gibberellins (Fenn and Giovannoni, 2020). High-resolution analyses highlight that fertilization and seed-derived signals trigger a transition to cell expansion that drives post-fertilization fruit growth and final size, emphasizing that both cellular processes and hormonal regulation underpin fruit weight . In cucurbits and other fleshy fruits, early developmental windows are considered “critical periods” in which disturbances can irreversibly constrain final fruit size and yield potential (Gao et al., 2023).

Environmental factors strongly modulate these developmental processes, making research on their regulatory roles essential for understanding and predicting watermelon fruit weight. Light conditions, for example, have been shown to markedly alter fruit expansion: low-light or shading during 0-15 days after pollination reduces fruit size, soluble sugars, and amino acids, and affects expression of thousands of genes related to metabolism and transcriptional regulation in developing watermelon fruit. Greenhouse orientation and internal microclimate alter solar radiation, temperature, transpiration, and leaf gas exchange around the fruiting zone, which in turn influence fruit volume increase in seedless watermelon (Woo et al., 2022). Supplementary LED lighting and optimized temperature or nutrient regimes around the fruit set region significantly increase fruit mass, size, flesh thickness, and overall yield in plastic-house and winter or early-spring crops, highlighting light, temperature, and mineral availability as key environmental levers for fruit weight formation (Chamchum et al., 2023). Weather studies at field scale further show that precipitation and temperature patterns across seasons drive differences in yield and fruit quality, with drier seasons often associated with higher productivity (Bai et al., 2020).

Despite this growing body of work on environmental impacts, quantitative models that explicitly link dynamic environmental conditions to fruit weight formation in watermelon remain scarce. In other fruit and crop systems, nonlinear growth functions, Bayesian sigmoidal models, and process-based simulation models (e.g., modified WOFOST, SIMBA, or radial basis function neural networks) have been successfully used to describe fruit growth dynamics and predict yields based on time, physiological status, and environmental drivers. Such models capture characteristic sigmoidal or multi-phase growth curves and can integrate factors like temperature, radiation, and plant age to forecast final fruit weight or total yield with high accuracy. However, comparable modeling efforts tailored to watermelon, particularly under controlled-environment or protected cultivation where light, temperature, and supplemental energy inputs are actively managed, are largely absent. Given the demonstrated sensitivity of early fruit expansion and final fruit weight to radiation, temperature, and nutrient conditions in watermelon, there is a clear need to develop and validate models that mechanistically and quantitatively link environmental variables to fruit weight formation. Such models could support decision-making for greenhouse design and orientation, supplemental lighting strategies, temperature control regimes, and fertilization programs, enabling producers to optimize resource use while stabilizing or increasing yield and fruit quality under variable climatic and market conditions.

2 Biological Basis of Watermelon Fruit Weight Formation

2.1 Stages of watermelon fruit development

Watermelon fruit development proceeds through fruit set, a rapid expansion phase, and maturation, each characterized by distinct physiological and molecular changes. Early after pollination, rapid cell division in the young ovary and fruitlets establishes the basic cell number and tissue pattern, a phase tightly coordinated with ethylene- and hormone-related gene expression and high ethylene evolution in fruitlets (Anees et al., 2023). Following this, fruit growth switches to a prolonged cell expansion phase in which vacuolated parenchyma cells enlarge and accumulate sugars, pigments, and other solutes that drive osmotic water uptake and volume increase, forming the bulk of fruit flesh mass.

The expansion and maturation stages are marked by coordinated changes in gene expression and metabolites that define size, texture, color, and sweetness. Transcriptome and digital expression profiling across key developmental stages show thousands of differentially expressed genes related to cell wall metabolism, sugar

accumulation, carotenoid biosynthesis, and stress and hormone signaling, all modulated as fruits pass from immature white to fully ripe red or over-ripe stages (Yu et al., 2022). Non-destructive and physiological measurements further indicate that ripening involves progressive pigment and carotenoid changes at the surface and within the flesh, along with hormone shifts typical of non-climacteric, ABA-and ethylene-modulated maturation, which together stabilize final fruit size and weight (Dhanani et al., 2022).

2.2 Mechanisms by which watermelon flesh cell division and expansion contribute to fruit weight

Fruit weight is fundamentally determined by the final number and size of flesh cells. In watermelon, cell division predominates during only the first several days after anthesis, followed by a long expansion phase in which existing cells enlarge dramatically through vacuolation and wall remodeling. Anatomical studies comparing pollinated, auxin-induced parthenocarpic, and unpollinated fruits show that early fruit growth depends on active cell division in pericarp and ovule tissues; unpollinated ovaries, which lack sustained division, rapidly cease growth, illustrating the centrality of early proliferative activity for subsequent fruit mass potential. After this brief proliferative window, increases in pulp weight from about one week onward are largely ascribed to cellular expansion, consistent with microscopic evidence that watermelon flesh cells become large and visually apparent as fruits enlarge (Kojima et al., 2020).

Cell expansion is driven by coordinated changes in cell wall architecture, hormonal signals, and metabolic status. Studies of firmness and texture reveal that differences in protopectin, cellulose, and hemicellulose contents, together with cell number, packing, and wall thickness, underlie variation in tissue density and mechanical support for expanding cells (Sun et al., 2020; Mashilo et al., 2022). Auxin- and Aux/IAA-mediated pathways modulate cell enlargement and the balance between cell size and number: high expression or allelic variants of Aux/IAA are linked to increased cell number, smaller cell size, and higher firmness, whereas reduced Aux/IAA activity is associated with larger cells and softer flesh, indicating that auxin signaling tunes the cellular composition that ultimately contributes to fruit volume and weight (Anees et al., 2023).

2.3 The impact of source-sink relationships in watermelon plants on fruit weight accumulation

Fruit growth depends on assimilate and water supply from vegetative organs, with the developing fruit acting as a strong sink whose demand changes across development. Dynamic sap-flow monitoring along fruit stalks across successive developmental stages shows that diurnal water distribution between leaves and fruit shifts markedly as fruits expand, slow growth, and reach maturity. During early expansion, nighttime inflow to the fruit dominates, correlating with rapid daily mass increase, whereas at later stages midday transpiration demand in leaves competes more strongly, shortening net inflow periods and reducing net fruit growth, before inflow and outflow balance at maturity when phenotype and weight stabilize (Zhang et al., 2024). Under low-light conditions, reduced photosynthesis and altered carbon and nitrogen metabolism associate with smaller fruits, lower soluble sugar and amino acid contents, and extensive transcriptional reprogramming, demonstrating that source capacity strongly constrains sink development during the critical 0-15 days after pollination expansion window (Gao et al., 2023).

Water supply and photosynthate partitioning also interact with environmental water availability and irrigation management to determine final fruit size and yield. Experiments manipulating drip irrigation at different fractions of crop water requirement across growth stages show that severe water deficits reduce total and marketable yield, average fruit weight, and fruit number, with vegetative growth and active fruit development phases more sensitive than the ripening phase. Deficit irrigation applied only during ripening has a comparatively smaller impact on marketable fruit weight and improves water-use efficiency, implying that the sink strength of fruits is highest and most vulnerable to water limitation during early and mid-development when rapid mass accumulation occurs. Together, these findings indicate that fruit weight formation in watermelon emerges from an integrated source-sink system, in which environmental factors such as light and water modulate assimilate production and transport to the fruit, thereby shaping the trajectory of fruit growth and final weight.

3 Key Environmental Factors Influencing Watermelon Fruit Weight

Watermelon fruit weight is shaped by a complex interaction of temperature, light, water, and nutrient status that

together regulate cell division, cell expansion, and carbohydrate supply to developing fruits. Studies in protected and open-field systems show that modifying microclimate or resource availability around the fruiting zone can substantially alter fruit volume, sugars, and final yield per plant. Temperature and light, in particular, determine photosynthetic capacity and assimilate partitioning, while water and soil factors influence canopy function, root activity, and stress responses that indirectly affect fruit growth rates (Woo et al., 2022). Understanding these key environmental drivers is essential for building predictive models of fruit weight formation and for designing precise cultivation strategies in climate-vulnerable production regions (Barros et al., 2024).

3.1 The impact of temperature variations on watermelon fruit expansion rate and final fruit weight

Targeted heating around the fruiting region clearly demonstrates that temperature during early enlargement can accelerate fruit growth and increase final fruit weight. Raising the minimum temperature around fruit-bearing shoots to 18 °C in early spring plastic-house production significantly increased fresh fruit weight per plant, soluble solids, and fruit set rate, indicating enhanced sink activity and yield potential under suboptimal ambient conditions. High night-time temperature around young fruits (approximately 6 °C above the control) similarly increased fruit length, diameter, and weight by 16 days after anthesis through accelerated cell enlargement, even though final size at harvest later converged with the control (Chamchum et al., 2023).

The timing of temperature elevation also influences internal quality and sugar accumulation associated with fruit weight formation. Heating the fruit and nearby shoots during the early cell enlargement stage (5-20 days after anthesis) increased sucrose phosphate synthase activity and led to higher sucrose content at maturity, particularly in outer flesh tissues, suggesting that warm conditions can promote both structural growth and assimilate storage capacity. When high-temperature treatments were imposed between 10 and 20 days after anthesis, cell enlargement in the central region was stimulated and sugar contents at harvest generally increased; by contrast, high temperatures applied only after 20 days enhanced cell expansion in peripheral tissues but reduced sugar levels overall, indicating that late heat can impair optimal sugar accumulation while still altering fruit morphology.

3.2 The role of light conditions in the accumulation of watermelon photosynthetic products and fruit weight formation

Light intensity and spectral composition strongly control photosynthetic production and distribution of assimilates to fruits, thereby determining fruit size. In vertically trained, high-density systems, increasing shading with planting density reduced per-plant solar radiation interception and whole-plant photosynthetic production, which was closely correlated with reduced fruit size despite little change in soluble solids, highlighting light-driven assimilate supply as a primary determinant of fruit weight. Experiments on greenhouse orientation showed that a southeast-northwest single-span greenhouse received higher integrated solar radiation and higher leaf transpiration near fruits than an east-west structure, and these microclimatic differences were associated with increased fruit volume expansion in seedless watermelon (Figure 1) (Woo et al., 2022).

Artificial supplemental lighting further illustrates the role of light in driving both biomass production and carbohydrate accumulation in fruits. In winter-grown watermelon, evening LED lighting at 900 μmol m^{-2} s^{-1} significantly increased chlorophyll content, photosynthetic rate, fruit number, fruit weight, and flesh thickness, leading to a 31% yield increase and higher fruit sugar content relative to natural-light controls (Hossain et al., 2025). Similarly, plastic-house plants receiving 6-12 h of white LED light at night produced heavier fruits with larger dimensions and thicker flesh than non-supplemented plants, confirming that extended photoperiod and elevated photon flux enhance fruit growth by bolstering daily carbon gain and translocation to reproductive sinks (Gao et al., 2023).

3.3 Regulatory mechanisms of water and soil nutrients on watermelon fruit weight formation

Water availability and root-soil interactions regulate canopy temperature, photosynthesis, and stress physiology, which in turn shape fruit development and final weight. Experiments manipulating the wetted soil area under drip irrigation (12%-22% of surface) in semi-arid Brazil showed that average leaf temperature remained below air temperature and that fruit mass and BRIX were statistically similar across treatments, suggesting that a relatively wide range of localized wetting can maintain physiological stability and high yield if overall water supply is

adequate (Barros et al., 2024). Under drought, however, arbuscular mycorrhizal colonization improved root development, protected chloroplast ultrastructure, and maintained higher photosynthetic efficiency, leading to better water status and greater accumulation of soluble sugars and osmolytes, mechanisms that support sustained assimilate delivery to developing fruits under water-limited conditions.

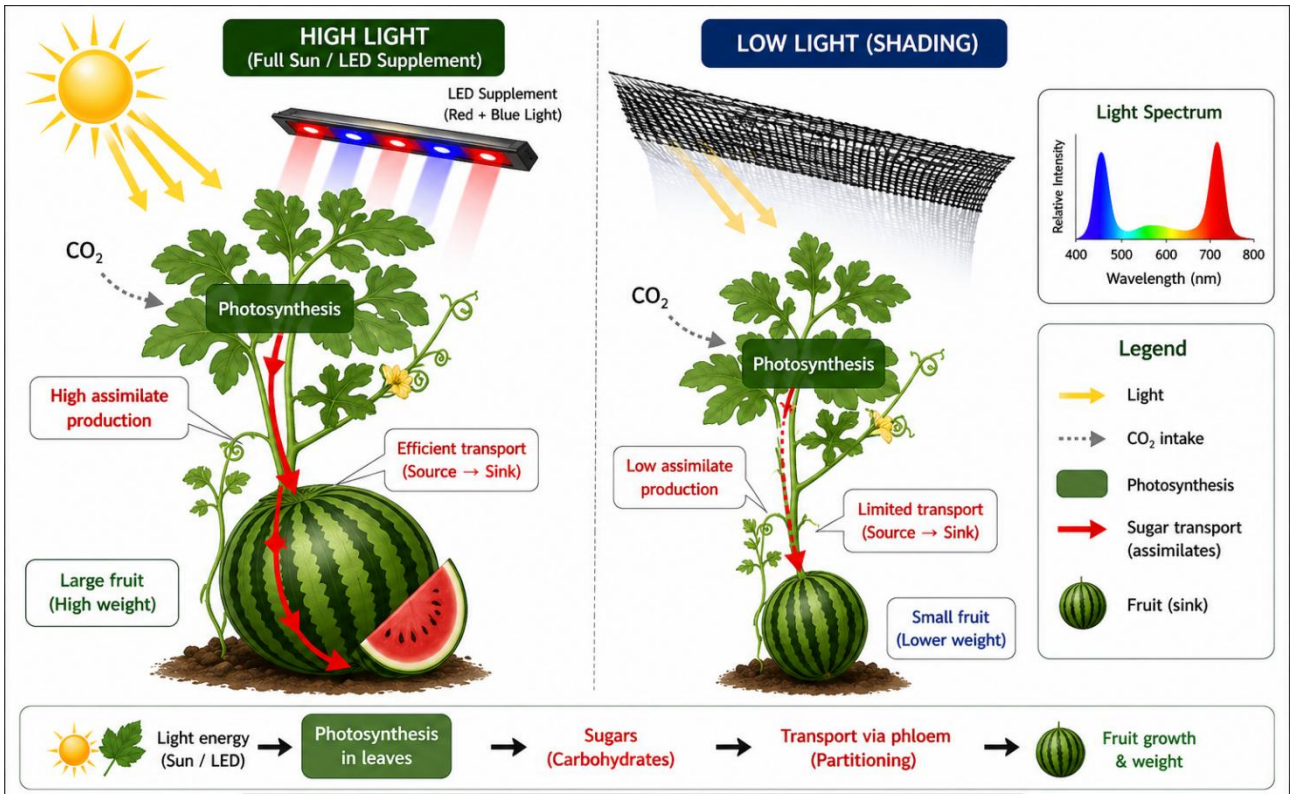


Figure 1 Role of light intensity and spectral quality in regulating photosynthetic carbon assimilation and its allocation to watermelon fruits. Increased light availability enhances assimilate supply and promotes fruit growth and weight formation

Soil nutrient status, particularly calcium and magnesium around the fruit set region, further modulates fruit growth and quality. Heating treatments at 18 °C near the fruiting zone not only increased fruit weight and soluble solids but were accompanied by elevated Ca^{2+} and Mg^{2+} concentrations in leaves adjacent to the fruit set node, implying improved nutrient uptake and transport under optimized temperature, which likely stabilizes cell wall structure and photosynthetic function during critical phases of fruit expansion. Supplemental LED lighting in winter crops likewise increased Ca^{2+} and Mg^{2+} in leaves at the fruit set region, enhancing photosynthetic rates and supporting consistent plant growth, which translated into larger fruit size, thicker flesh, and higher sugars, indicating tight coupling between nutrient status, carbon assimilation, and fruit weight formation under low-light, cool-season conditions (Hossain et al., 2025).

4 Data Acquisition and Experimental Design for Watermelon Fruit Weight Research

4.1 Design of watermelon field experiments

Field experiments on watermelon fruit weight are typically structured as factorial randomized or randomized block designs to evaluate genetic and management factors simultaneously. Representative studies select two or more commercial cultivars differing in fruit size class or adaptation, such as ‘Crimson Sweet’, ‘Sugar Baby’, or locally important hybrids, and test them across multiple locations or seasons to account for environmental variation. Treatments often include mulching materials, fertilizer regimes, or pruning and fruit-thinning levels, arranged with three or more replications to enable analysis of variance and proper error estimation (Deka et al., 2024). This design supports estimation of main and interaction effects on average fruit weight and yield components, while maintaining uniform baseline agronomic practices such as irrigation and pest management across plots (Yismaw et al., 2024).

Planting density is a central experimental factor because it simultaneously affects resource competition, canopy structure, and marketable fruit size. Trials commonly compare several inter- and intra-row spacings (for example, 3.0 × 0.8 m vs. 2.0 × 0.6 m, or 120 × 60 cm vs. denser arrangements) to quantify responses of fruit number, average fruit weight, commercial fruit proportion, and total yield per hectare (Silva et al., 2021). In some studies, plant density is integrated with training system (horizontal vs. vertical), stem number, or fruit-thinning treatments to manipulate source-sink balance and define optimal load per plant for targeted fruit-size categories (Kim et al., 2023). Such multifactor designs allow identification of densities that maximize total yield without excessively shifting the population toward undersized or mini fruits, while preserving desirable quality traits such as soluble solids content (Figure 2) (Tegen et al., 2021).

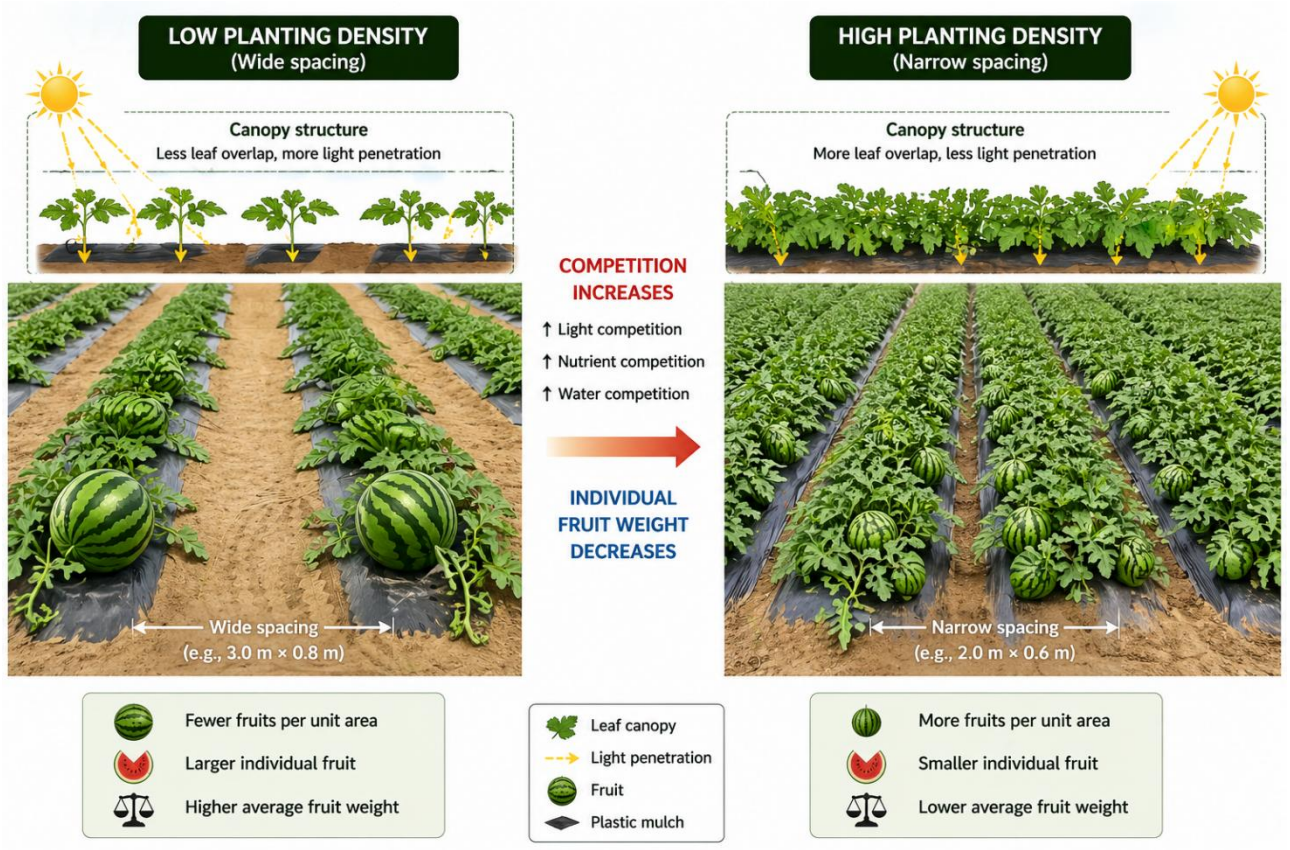


Figure 2 Effect of planting density on the trade-off between fruit number and individual fruit weight in watermelon production systems

4.2 Collection of data on watermelon fruit weight and related growth indicators

Fruit weight is usually recorded at harvest as single-fruit weight and/or average fruit weight per plant or per plot, along with fruit number to derive total and marketable yields. Detailed datasets often include fruit length, fruit diameter, average fruit weight, and total fruit yield, and distinguish between marketable and unmarketable fruits to assess economic performance (Yismaw et al., 2024). Additional yield-related variables such as fruit set, fruit retention, and yield per plant or per hectare are measured to characterize treatment effects on both productivity and fruit-size distribution (Bora et al., 2024). When treatments involve fruit thinning or fruit load per stem, measurements of fruit weight and number at different node positions or fruit counts per plant help quantify how source-sink manipulation alters final fruit mass and commercial yield.

Quality-related indicators, especially soluble solids content (Brix), are commonly assessed to link fruit weight responses with eating quality. Studies generally sample representative fruits from each plot, measure Brix with a refractometer, and record traits such as rind firmness, pulp firmness, juice content, and sugar fractions (reducing, non-reducing, and total sugars). Some trials analyze Brix alongside fruit length, diameter, and average weight to identify trade-offs between density or fruit load and sweetness, while others focus on cultivar × mulch or growth

regulator combinations to detect treatments that raise both yield and Brix (Correa et al., 2020; Raj et al., 2022). Biomass-related traits such as shoot and fruit dry mass, harvest index, and partitioning between vegetative and reproductive organs are collected in source-sink experiments to describe how pruning and fruit number modify photoassimilate allocation and fruit size (Deka et al., 2024).

4.3 Processing and quality control of environmental monitoring data for watermelon

Environmental monitoring is crucial for interpreting treatment effects on fruit weight and for parameterizing environmental-response models. In field and protected experiments, weather data (air temperature, radiation, and sometimes humidity) are commonly obtained from nearby meteorological services or on-site stations to describe seasonal conditions and compare sowing dates or density treatments under similar macroclimates (Gao et al., 2023). Soil moisture is monitored directly in mulching or irrigation studies, where mulched plots generally maintain higher moisture and reduced weed competition than bare-soil controls, supporting higher single-fruit weight and total yield (Yismaw et al., 2024). In subsurface fertigation systems, spacing relative to the irrigation source is explicitly tested, and treatment differences in plant growth and fruit weight are interpreted in light of soil type and the distance from clay pot emitters (Sutarno et al., 2022).

Quality control of environmental data focuses on ensuring consistency, representativeness, and correct linkage to plot-level observations. Experiments using mulches or subsurface irrigation typically collect repeated measurements of soil moisture and sometimes weed biomass, enabling cross-checks between moisture trends and yield responses across treatments (Sutarno et al., 2022; Yismaw et al., 2024). When studying seasonal or sowing-date effects, growth and yield measurements at 60-120 days after sowing are evaluated together with environmental records to identify the sowing window that aligns with favorable temperature and radiation, resulting in superior fruit set, fruit weight, yield per hectare, and Brix (Kim et al., 2023; Bora et al., 2024). Such careful integration of environmental and yield datasets underpins robust inferences about how temperature, light, and moisture regimes drive variation in watermelon fruit weight and related quality traits.

5 Methods for Constructing Watermelon Fruit Weight Models

5.1 Model types applicable to watermelon fruit weight prediction

Empirical models predict fruit weight directly from observable traits or management factors without explicitly representing underlying physiology. For watermelon and other spherical fruits, non-destructive image features (segmented area, bounding box ratios) have been coupled with machine-learning regressors such as Random Forest and Decision Trees to predict individual fruit weight with high accuracy, demonstrating the power of purely data-driven approaches when sufficient labeled images are available (Koç and Kayra, 2024). In agronomic optimization, multiple linear or polynomial regression has been used to relate watermelon fruit weight to input factors such as poultry, cow, and goat manure rates within a Simplex Lattice Design framework, yielding statistically significant quadratic response surfaces for fruit weight and fruit number per plant (Sabouri et al., 2025).

Mechanistic and process-based models, by contrast, attempt to represent fruit growth as the outcome of carbon and water transport, cell expansion, and environmental drivers over time. Biophysical models of fruit such as the virtual-fruit framework describe water and dry-matter flows via xylem and phloem, osmotic and turgor pressures, and cell wall extension, and are capable of simulating seasonal and diurnal dynamics of fruit fresh and dry mass under varying crop load and water status. More recent integrative models explicitly couple carbon and water fluxes with hormonal regulation (e.g., abscisic acid) to simulate fruit mass and its response to heat, cold, and drought, illustrating how mechanistic structures can capture environmental regulation and stress-induced delays in growth in a way that empirical models cannot (Chung et al., 2025).

5.2 Selection of variables influencing watermelon fruit weight

A critical step in model construction is selecting environmental variables that strongly influence watermelon fruit growth. For empirical prediction in other cucurbits, fruit age, harvest date, plant height, fruit length and width, flesh thickness, cavity diameter, branch number, and leaf number have been used as ANN inputs, achieving high determination coefficients for fruit weight, which suggests that morphological and phenological descriptors can

serve as practical predictors when direct physiological measurements are not available (Erniati et al., 2023; Koç and Kayra, 2024). In watermelon, light interception, total solar radiation per plant, and photosynthetic production are closely correlated with fruit weight in vertically trained systems, indicating that radiation-related variables (incident radiation, intercepted PAR, or canopy light-use indices) are key environmental drivers to incorporate in fruit weight models (Gao et al., 2023).

Process-based models require additional internal state variables to link environment to growth. Biophysical fruit models commonly track water content, dry matter, sugar concentrations in fruit and phloem, turgor pressure, transpiration and respiration rates, and xylem-phloem flows, using hourly atmospheric inputs (temperature, humidity) as boundary conditions. Integrated plant-fruit models further include indicators of source-sink balance, such as sucrose concentration in the phloem, stem water potential, and measures of water and nitrogen status, linking these to fruit mass via functions that modify assimilate supply or hydraulic conductance under stress (Zhou et al., 2025).

5.3 Establishment and mathematical expression of dynamic models for watermelon fruit weight

Dynamic modeling of watermelon fruit weight can draw from established formulations in fruit growth modeling. Biophysical approaches treat the fruit as a compartment with state variables for water (w) and dry matter (s), and describe fluxes of water and sugar between fruit, plant, and atmosphere using mass-balance differential equations; sugar uptake is partitioned among mass flow, passive diffusion, and active transport, while cell wall expansion is driven by turgor according to irreversible growth equations at the tissue scale. Such models express fresh mass as the sum of water and dry matter, driven by environmental inputs (temperature, humidity) and plant water status, thereby enabling simulation of diurnal swelling-shrinkage cycles and seasonal growth trajectories that could be adapted to watermelon fruits.

More recent frameworks integrate cellular processes (cell division and expansion), resource limitation, and hormone signaling into compact mathematical structures. A minimal cell-expansion-division model represents temporal changes in cell number and mean cell mass under constraints of carbon and water supply, producing emergent dynamics of total fruit mass and cell size distributions that match observations across genotypes and environments (Miele et al., 2025). Likewise, process-based models that link carbon and water fluxes to endogenous ABA include sub-models for sugar uptake, respiration, hydraulic conductance, and transpiration modulated by ABA concentration, allowing simulation of fruit mass under variable temperature and water availability (Chung et al., 2025). By calibrating such differential-equation systems with watermelon-specific environmental data and fruit growth measurements, dynamic models can be formulated that quantitatively relate environmental factors and physiological indicators to the time course of watermelon fruit weight.

6 Validation and Evaluation of Watermelon Fruit Weight Models

6.1 Evaluation of the fitting accuracy of watermelon fruit weight prediction models

Assessing model accuracy is central to evaluating watermelon fruit weight prediction, and most recent work relies on statistical indices such as root mean squared error (RMSE) and coefficient of determination (R^2). In a non-destructive image-based system for spherical fruits, including watermelon, U-Net segmentation extracted geometric ratios from images and several regression models were trained; performance was evaluated using MSE, MAE, RMSE, and R^2, allowing direct comparison of model fits across algorithms. For watermelon, Random Forest and Decision Tree models showed the highest training success, achieving an R^2 of 0.9112 in the best case, whereas linear and SGD models performed poorly, illustrating the value of non-linear models when fruit appearance and weight relationships are complex (Koç and Kayra, 2024).

Similar criteria are widely adopted in other fruit weight modeling studies and provide a benchmark for what constitutes an acceptable fit. For example, a machine-learning framework for non-destructive plum fruit weight estimation compared SVR, MLR, MLP, and Decision Tree models using RMSE and R^2 in both training and testing, selecting the optimal structure based on lowest RMSE and highest R^2. The best SVR model reached training R^2 of 0.9369 with RMSE 0.4850 g and test R^2 of 0.9267, confirming that accurate fresh-weight models can be obtained when evaluation is rigorously based on these metrics and when training-testing separation is respected to avoid overfitting (Sabouri et al., 2025).

6.2 Verification of the applicability of watermelon fruit weight models across different ecological environments

Beyond goodness-of-fit within a single dataset, watermelon weight models must be evaluated for applicability across environments, especially when environmental factors are explicit inputs. Work on the adaptive potential of large watermelon collections illustrates that genotypes differ strongly in environmental plasticity for average marketable fruit weight, quantified using regression coefficients of genotype response (bi) and stability parameters such as Sgi and general adaptive capacity across multiple years and sites. These parameters effectively characterize how robust fruit weight performance is to fluctuating conditions, providing a conceptual analogue for assessing the environmental robustness of predictive models that incorporate fruit weight as an output trait (Serhiienko et al., 2023).

Model applicability across environments can also be approached through genotype×environment or drought-environment interaction analyses using multivariate tools. In muskmelon, AMMI and GGE biplot models were applied to fruit weight data across irrigation regimes, treating each drought level as an environment and identifying genotypes with stable fruit weight under mild to severe soil water depletion. The GGE biplot classified irrigation regimes into a single mega-environment and distinguished genotypes with wide adaptability and stability, demonstrating that statistical modeling of performance across contrasting water regimes can objectively test whether a predictive framework (or genotype response surface) remains valid under varying ecological conditions (Rad and Bakhshi, 2020).

6.3 Sensitivity and stability analysis of watermelon fruit weight models

Sensitivity and stability analyses clarify how strongly fruit weight predictions depend on specific inputs or environmental drivers, and whether the model behaves reliably under stress or management changes. In a simplex lattice design modeling watermelon fruit weight as a function of poultry, cow, and goat manures, second-order mixture models were fitted and evaluated using analysis of variance; significant F-values and low p-values indicated that the quadratic models were adequate for prediction and captured the effects and interactions of nutrient components on fruit weight. Examining the estimated coefficients and interaction terms provided practical insight into which manure sources most strongly influenced model outputs and under what combinations the model predicted maximum fruit weight per plant.

Stability in the face of environmental variability is also addressed indirectly in studies that quantify genotype stability and plasticity for average fruit weight across multi-year, multi-environment trials. Using parameters such as genotype stability (Sgi), specific adaptive capacity, and plasticity coefficient (bi), large watermelon collections were partitioned into intensive, medium, and highly plastic groups with respect to total and marketable yield and average fruit weight, effectively ranking genotypes by how consistently they express fruit weight under changing conditions. While these analyses focus on biological responses rather than explicit prediction models, the same stability statistics and multi-environment frameworks can be incorporated into model validation protocols to test whether watermelon fruit weight prediction models maintain performance across diverse ecological and management scenarios (Serhiienko et al., 2023).

7 Case Study: Application of Watermelon Fruit Weight Prediction Based on Multiple Environmental Conditions

7.1 Data sources from typical watermelon cultivation regions

Case studies on watermelon fruit weight modeling can draw on diverse cultivation systems, from rain-shelter or greenhouse production to fully open-field systems. Under rain-shelter structures, fertigation trials with water-soluble NPK generated detailed records of leaf traits, fruit weight and quality, enabling regression analyses that identified 125% of the conventional NPK rate as optimal for fruit weight in protected conditions (Figure 3) (Hafiz et al., 2024). In contrast, open-field experiments with soil-moisture-sensor-based drip irrigation under different mulches in North Dakota produced multi-year datasets combining average fruit weight, diameter, and quality traits with rainfall and irrigation records, providing a basis for environment-response modeling in a cool, continental climate (Vaddevolu et al., 2021).

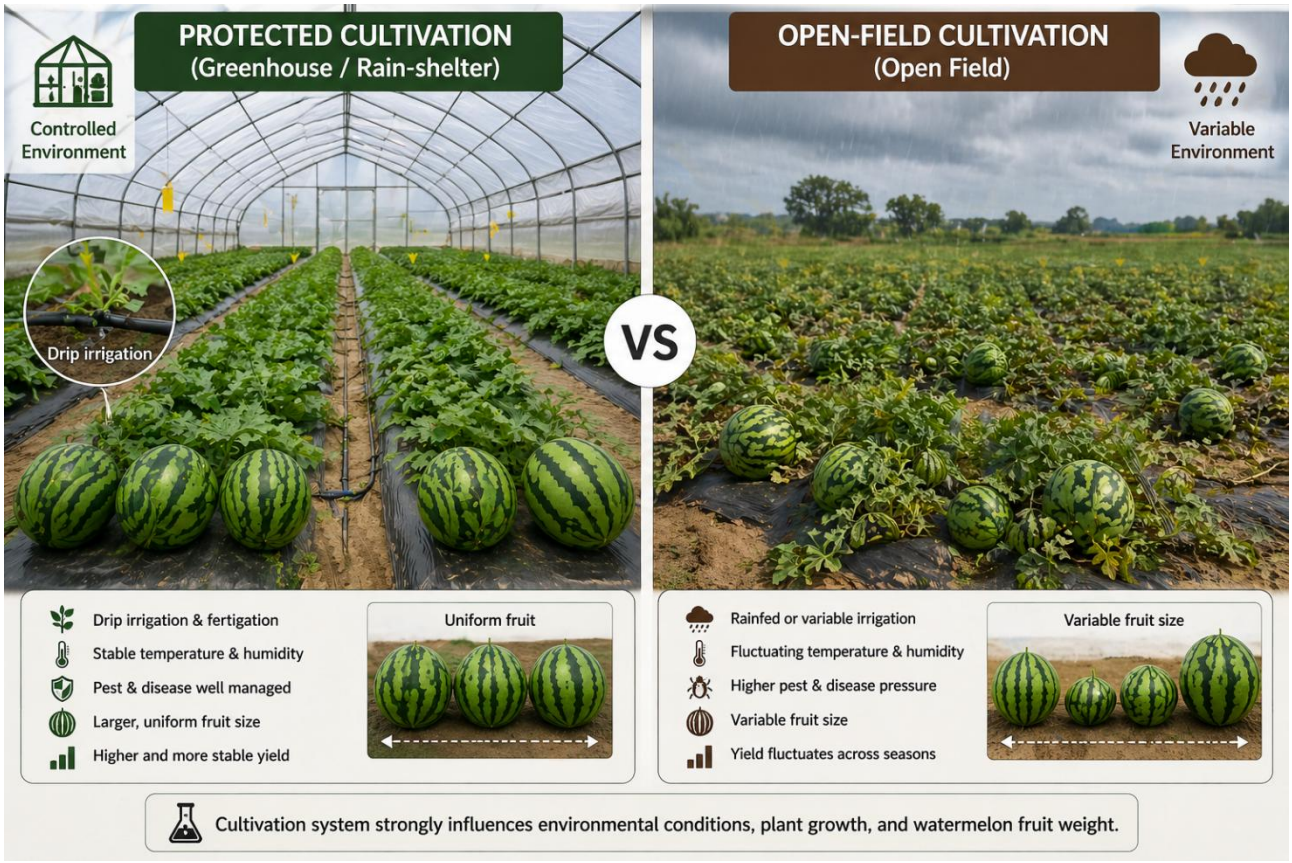


Figure 3 Comparison of protected cultivation and open-field systems used as data sources for watermelon fruit weight modeling under contrasting environmental conditions

Additional environmental contrasts arise from agroforestry versus sole-cropping and from seasonal variation in tropical open fields. In a semi-arid Chinese apple-watermelon agroforestry system, three irrigation quotas were combined with two planting patterns, generating three-year time series for soil water content, photosynthesis, fruit weight, and total yield under systematically different light and water regimes (Qiang et al., 2024). In tropical Tanzania, cultivar trials across dry and wet seasons recorded vine growth and fruit weight under markedly different rainfall and temperature conditions, showing that environmental seasonality significantly modifies yield traits and thus should be reflected in regional fruit weight prediction datasets.

7.2 Analysis of prediction results from watermelon fruit weight models under different environmental scenarios

Non-destructive, image-based models illustrate how fruit weight can be predicted across environments where direct weighing is impractical. A U-Net segmentation plus machine-learning pipeline used simple geometric ratios from fruit images to predict watermelon weight, achieving best training performance with Random Forest and Decision Tree models ($R^2 \approx 0.91$), and highlighting that larger variation in fruit size and image occupation can reduce accuracy relative to other fruits (Koç and Kayra, 2024). An IoT system integrating soil moisture, temperature, humidity, light, and CNN-based image analysis in melon-watermelon cultivation further reported very high sensor reliability and fruit weight prediction accuracies from 99.25% to 99.93%, demonstrating the potential of combining environmental sensing and imaging for real-time prediction under variable field conditions (Sarosa et al., 2024).

Where explicit process-based crop models are used, environmental scenarios such as deficit irrigation and mulching regimes can be evaluated through calibrated simulations. In Ethiopia, an AquaCrop application used soil physical data, climate, and crop records from factorial combinations of water application (50% vs. 100% soil-moisture depletion levels), mulching, and four watermelon varieties to predict yield responses; model performance for mulching-deficit irrigation effects on productivity was acceptable, with RMSE 0.70,

Nash-Sutcliffe efficiency 0.65 and R 0.80. These simulations captured changes in fruit size (diameter, length, and average fruit weight) and yield under water-saving practices, supporting their use as scenario tools for fruit weight formation under constrained irrigation (Gebeyhu and Markos, 2023).

7.3 Model-based optimization of watermelon cultivation management (irrigation, fertilization, and temperature control)

Model outputs linked to irrigation strategies allow optimization of water allocation while maintaining fruit weight. In Mediterranean open fields, staged deficit-irrigation experiments quantified how applying 75%-50% of crop water requirement at vegetative, fruit development, or ripening stages reduces average fruit weight, fruit number, and marketable yield, with ripening identified as the least sensitive stage to water shortage. Similarly, Heliyon-based AquaCrop simulations for Ethiopian mulched systems showed that straw mulching plus 50% deficit irrigation with suitable varieties (e.g., Green Pearl) maximized land and water productivity while maintaining competitive average fruit weights, guiding irrigation scheduling and variety choice under scarcity (Gebeyhu and Markos, 2023).

Fertilization and coupled water-nutrient management can also be optimized using regression and structural approaches informed by model results in related cucurbits. Under rain-shelter watermelon, regression trendlines indicated that a 25% increase above standard NPK (125%) maximized fruit weight and vegetative growth under fertigation, suggesting that fertigation-based models should include fertilizer rate as a continuous decision variable in protected systems (Hafiz et al., 2024). For melon, structural equation modeling under varied water and fertilizer levels identified photosynthetic rate and total dry mass as key intermediates by which water and nutrient inputs control yield and quality, implying that watermelon fruit weight optimization models should similarly treat growth and photosynthesis as mediating variables when evaluating combined irrigation-fertilization strategies across arid and semi-arid environments (Yang et al., 2023).

8 Discussion and Outlook

Most existing watermelon fruit weight models are empirical and developed under narrowly defined conditions, limiting their transferability to other farms and seasons. For example, a tillage-based yield model was fitted using soil physical properties and simple plant traits, achieving very high R^2 (0.98) but relying on linear and weakly nonlinear relationships derived from only two seasons in a semi-arid Nigerian environment. Similarly, mixture models relating poultry, cow, and goat manure rates to fruit weight produced statistically significant quadratic equations, yet they describe only manure effects and do not explicitly account for weather, irrigation scheduling, or pest pressure that commonly constrain yield in practice. Non-destructive image-based models reach high fitting accuracy but also reveal important limitations. A U-Net plus machine-learning approach predicted watermelon weight from image features with $R^2 \approx 0.91$, but training relied on controlled image acquisition at fixed distances and backgrounds, conditions that are rarely met in commercial fields or heterogeneous greenhouses. Furthermore, the authors noted that prediction success is influenced by the diversity and complexity of products in the images, implying that substantial recalibration or retraining would be required when moving from experimental datasets to large-scale, multi-variety production systems.

Watermelon fruit weight responds to interacting environmental drivers rather than isolated factors, which complicates modeling. A three-way factorial experiment in northern Tanzania showed that extra irrigation or fertilizer alone did not increase fruit weight, while pollination strongly affected the probability of setting a second marketable fruit and improved sugar content, with complex higher-order interactions among water, fertilizer, and pollination on fruit initiation. Supplemental hand-pollination across 13 farms increased average fruit weight by 1.3 kg while responses to soil moisture varied with treatment, demonstrating that both biotic (pollinators) and abiotic (soil carbon and water) factors jointly regulate fruit set, abortion, and final weight. Greenhouse experiments further highlight strong water-nitrogen-CO_2 interactions on growth and yield. Under elevated CO_2, increased irrigation improved dry matter accumulation, photosynthesis, and yield, while higher CO_2 partly compensated for low nitrogen, shifting optimal N rates relative to ambient CO_2 scenarios (Hong et al., 2022). The interaction of irrigation and nitrogen significantly affected key physiological indicators such as net photosynthetic rate and

transpiration, and integrated evaluation (TOPSIS) showed that comprehensive growth was positively correlated with yield, implying that fruit weight models must incorporate coupled water-nutrient-CO_2 effects rather than treating each driver independently.

Intelligent irrigation-fertigation systems and IoT platforms offer promising tools to regulate environmental drivers in real time and indirectly stabilize fruit weight. In greenhouse watermelon, an intelligent drip-fertigation system used soil-moisture sensors and IoT-based controllers to adjust irrigation limits and nutrient supply by growth stage, reducing water, N, P_2O_5 and K_2O inputs by 33%-72% without compromising yield or fruit quality. Dry matter accumulation and nutrient uptake followed logistic curves, and improving root traits under intelligent fertigation enhanced water and nutrient acquisition, suggesting that such systems could be coupled with fruit weight models to optimize source-sink balance during the fruit expansion phase. More integrated smart-farming architectures are also emerging for melon and watermelon. An IoT system combining soil moisture, temperature, humidity, and light sensors with CNN-based image analysis automatically regulated watering while predicting fruit weight with accuracies above 99%, and achieved very high reliability of nutrient, pH, and moisture sensors. Broader reviews of AI-IoT in precision agriculture emphasize that fusing remote sensing, high-throughput phenotyping, and machine-learning analytics enables site-specific irrigation and fertilization, automated crop monitoring, and yield forecasting, but also note challenges in data integration, scalability, and real-time decision support that must be addressed before such systems can be widely deployed for watermelon fruit weight regulation.

Acknowledgments

Thanks to the reviewers for providing detailed comments and guidance on the manuscript of this study. The reviewers' keen insights into the issues and attention to detail have greatly benefited the authors.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Anees M., Zhu H., Umer M., Gong C., Yuan P., Lu X., He N., Kaseb M., Yang D., Zhao Y., and Liu W., 2023, Identification of an Aux/IAA regulator for flesh firmness using combined GWAS and bulked segregant RNA-Seq analysis in watermelon, Horticultural Plant Journal, 10(2): 463-476.
https://doi.org/10.1016/j.hpj.2023.05.018

Bai T., Wang T., Zhang N., Chen Y., and Mercatoris B., 2020, Growth simulation and yield prediction for perennial jujube fruit tree by integrating age into the WOFOST model, Journal of Integrative Agriculture, 19(8): 2022-2034.
https://doi.org/10.1016/S2095-3119(19)62753-X

Barros A., Silva R., Neto A., Vellame L., Santos M., and De Oliveira Aguiar Netto A., 2024, Thermal, hydric, and physiological effects on watermelon due to wetted area variation, Revista Ceres, 71(1): 1-10.
https://doi.org/10.1590/0034-737X2024710002

Bora D., Neog M., and Dutta S., 2024, Assessment of morpho-biochemical traits in watermelon (*Citrullus lanatus*) cultivars across staggered sowing intervals under the agro-climatic conditions of Assam, India, Journal of Scientific Research and Reports, 30(10): 480-490.
https://doi.org/10.9734/jsrr/2024/v30i102480

Chamchum W., Glahan S., Kramchote S., Maniwara P., and Suwor P., 2023, Growth and yield of watermelon (*Citrullus lanatus*) in plastic house in response to white LED supplementary lighting, AGRIVITA Journal of Agricultural Science, 45(2): 363-372.
https://doi.org/10.17503/agrivita.v45i2.3967

Chung S., Yun K., and Kim S., 2025, An integrative process-based model of fruit growth as a function of carbon and water fluxes modulated by endogenous abscisic acid in blueberry fruit, Quantitative Plant Biology, 6: e11.
https://doi.org/10.1017/qpb.2025.10011

Correa E., Malla S., Crosby K., and Avila C., 2020, Evaluation of genotypes and association of traits in watermelon across two southern Texas locations, Horticulturae, 6(4): 67.
https://doi.org/10.3390/horticulturae6040067

Dahake L., Sonkamble A., Patil S., and Meshram L., 2020, Influence of spacing on yield, quality and economics of watermelon, International Journal of Chemical Studies, 8(6): 1223-1225.
https://doi.org/10.22271/chemi.2020.v8.i6ab.11052

Deka B., Handique K., Borthakur P., Kotoky U., Saikia A., Kalita P., Gogoi B., Goswami S., Hazarika B., and Hazarika J., 2024, Effect of crop geometry, fruit thinning and nutrient management on growth parameters of watermelon (*Citrullus lanatus* Thumb.), International Journal of Advanced Biochemistry Research, 8(1): 294-299.
https://doi.org/10.33545/26174693.2024.v8.i1sb.294

Deka B., Hazarika J., Borthakur P., Kotoky U., Saikia A., Kalita P., Gogoi B., Goswami S., Hazarika B., and Handique K., 2024, Influence of crop geometry, fruit thinning and nutrient management on yield and yield-related attributes of watermelon (*Citrullus lanatus* Thumb.), International Journal of Environment and Climate Change, 14(1): 854-862.
https://doi.org/10.9734/ijecc/2024/v14i13854

Dhanani T., Dou T., Biradar K., Jifon J., Kurouski D., and Patil B., 2022, Raman spectroscopy detects changes in carotenoids on the surface of watermelon fruits during maturation, Frontiers in Plant Science, 13: 832522.
https://doi.org/10.3389/fpls.2022.832522

Erniati E., Suhardiyanto H., Hasbullah R., and Supriyanto S., 2023, Artificial neural networks to predict melon (Cucumis melo L.) production in tropical greenhouse, Indonesia, Jurnal Keteknikan Pertanian, 11(2): 193-204.
https://doi.org/10.19028/jtep.011.2.193-204

Fenn M., and Giovannoni J., 2020, Phytohormones in fruit development and maturation, The Plant Journal, 105(2): 446-458.
https://doi.org/10.1111/tpj.15112

Gao W., She F., Sun Y., Han B., Wang X., and Xu G., 2023, Transcriptome analysis reveals the genes related to watermelon fruit expansion under low-light stress, Plants, 12(4): 935.
https://doi.org/10.3390/plants12040935

Gebeyhu B., and Markos G., 2023, Assessment of soil mulching field management, and deficit irrigation effect on productivity of watermelon varieties, and AquaCrop model validation, Heliyon, 9(11): e21632.
https://doi.org/10.1016/j.heliyon.2023.e21632

Gerhard D., and Moltchanova E., 2022, A Richards growth model to predict fruit weight, Australian and New Zealand Journal of Statistics, 64(4): 503-520.
https://doi.org/10.1111/anzs.12380

Gülüt K., 2021, Nitrogen and boron nutrition in grafted watermelon I: impact on pomological attributes, yield and fruit quality, PLoS ONE, 16(6): e0252396.
https://doi.org/10.1371/journal.pone.0252396

Hafiz M., Hassan M., Fakhrul M., Omar Z., Abbas H., Mahfuzah W., Ibrahim W., Ghazali N., Hamid M., and Hamid N., 2024, Development of water soluble NPK fertilizer for watermelon cultivation under rain shelter structure, Journal of Agrobiotechnology, 15(S1): 376-385.
https://doi.org/10.37231/jab.2024.15.s1.376

Hossain M., Shibasaki Y., and Goto F., 2025, Enhancement of growth and quality of winter watermelon using LED supplementary lighting, Horticulturae, 11(3): 262.
https://doi.org/10.3390/horticulturae11030262

Jannatizadeh A., Rezaei M., Rohani A., Lawson S., and Fatahi R., 2022, Towards modeling growth of apricot fruit: finding a proper growth model, Horticulture, Environment, and Biotechnology, 64(2): 209-222.
https://doi.org/10.1007/s13580-022-00475-x

Jayasinghe S., Ranawana C., Liyanage I., and Kaliyadasa P., 2022, Growth and yield estimation of banana through mathematical modelling: a systematic review, The Journal of Agricultural Science, 160(2): 152-167.
https://doi.org/10.1017/S0021859622000259

Jordana C., Stapleton S., Colee J., Lee S., Gao Z., Ray Z., Anrecio L., Freed D., and Zhao X., 2023, How does watermelon grafting impact fruit yield and quality? A systematic review, HortScience, 58(10): 1223-1232.
https://doi.org/10.21273/HORTSCI16857-22

Kim E., Jeon Y., Yun G., Noh S., and Lee H., 2023, The characteristics of small-sized watermelons in quality and yield according to planting density, stem number, and node number of fruit-setting, Korean Journal of Horticultural Science and Technology, 41(2): 191-201.
https://doi.org/10.7235/hort.20230023

Koç S., and Kayra H., 2024, Non-destructive weight prediction model of spherical fruits and vegetables using U-Net image segmentation and machine learning methods, Tarım Bilimleri Dergisi, 30(3): 580-592.
https://doi.org/10.15832/ankutbd.1434767

Kojima K., Andou D., and Ito M., 2020, Plant hormone changes in growing small watermelon fruit, The Horticulture Journal, 89(6): 732-739.
https://doi.org/10.2503/hortj.UTD-209

Kumari M., Deltsidis A., Luo X., McAvoy C., and McAvoy T., 2025, Assessment of triploid watermelon cultivars grown in Georgia for yield and quality parameters, HortTechnology, 35(1): 52-61.
https://doi.org/10.21273/HORTTECH05561-24

Mashilo J., Shimelis H., Ngwepe R., and Thungo Z., 2022, Genetic analysis of fruit quality traits in sweet watermelon (*Citrullus lanatus* var. lanatus): a review, Frontiers in Plant Science, 13: 834696.
https://doi.org/10.3389/fpls.2022.834696

Miele L., Roques L., Constantinescu D., Génard M., and Bertin N., 2025, Cell expansion-division under resource limitation: a novel framework for modeling fruit growth dynamics, bioRxiv, preprint: 2024.05.30.596571.
https://doi.org/10.1101/2024.05.30.596571

Qiang X., Sun Z., Li X., Li S., Yu Z., He J., Li Q., Han L., and He L., 2024, The impacts of planting patterns combined with irrigation management practices on watermelon growth, photosynthesis, and yield, Plants, 13(10): 1402.
https://doi.org/10.3390/plants13101402

Rad M., and Bakhshi B., 2020, GGE biplot tool to identify melon fruit weight stability under different drought conditions, International Journal of Vegetable Science, 27(3): 220-230.
https://doi.org/10.1080/19315260.2020.1805538

Raj S., Singh D., and Deepanshu ., 2022, Effect of foliar application of plant growth regulators on growth, yield and fruit quality of watermelon (*Citrullus lanatus* Thunb.), International Journal of Plant and Soil Science, 34(21): 243-251.
https://doi.org/10.9734/ijpss/2022/v34i2131243

Sabouri A., Bakhshipour A., Poorsalehi M., and Abouzari A., 2025, Machine learning techniques for non-destructive estimation of plum fruit weight, Scientific Reports, 15(1): 85051.
https://doi.org/10.1038/s41598-024-85051-2

Sarosa M., Wirayoga S., Syaifudin Y., and Fiernaningsih N., 2024, Internet of Things system for melon/watermelon plant growth with image-based fruit weight prediction, 2024 International Conference of Adisutjipto on Aerospace Electrical Engineering and Informatics (ICAAEEI), pp.1-6.
https://doi.org/10.1109/ICAAEEI63658.2024.10899142

Serhiienko O., Shabetia O., Linnik Z., Serhiienko M., and Povlin I., 2023, Selection of watermelon starting material by adaptability for breeding for suitability for intensive and organic growing technologies, Plant Breeding and Seed Production, 123: 88-99.
https://doi.org/10.30835/2413-7510.2023.293879

Silva A., Da Silva C., Gonçalves C., Filho M., De Sousa Pereira C., Andrade M., and Pessoa W., 2021, Productive potential of watermelon under different plant spacings in the semi-arid region of Brazil, Australian Journal of Crop Science, 15(2): 238-243.
https://doi.org/10.21475/ajcs.21.15.02.p2796

Sun L., Zhang Y., Cui H., Zhang L., Sha T., Wang C., Fan C., Luan F., and Wang X., 2020, Linkage mapping and comparative transcriptome analysis of firmness in watermelon (*Citrullus lanatus*), Frontiers in Plant Science, 11: 831.
https://doi.org/10.3389/fpls.2020.00831

Tegen H., Alemayehu M., Alemayehu G., Abate E., and Amare T., 2021, Response of watermelon growth, yield, and quality to plant density and variety in Northwest Ethiopia, Open Agriculture, 6(1): 655-672.
https://doi.org/10.1515/opag-2021-0037

Vaddevolu U., Lester J., Jia X., Scherer T., and Lee C., 2021, Tomato and watermelon production with mulches and automatic drip irrigation in North Dakota, Water, 13(14): 1991.
https://doi.org/10.3390/w13141991

Woo S., Kim G., Lim J., Jeong J., Cho S., Ahn B., Lee E., Bae J., and Kim H., 2022, Growth environmental factors and fruit enlargement of seedless watermelon according to directions of single-span greenhouse, Korean Journal of Horticultural Science and Technology, 40(4): 447-458.
https://doi.org/10.7235/hort.20220047

Yang Z., Kong T., Xie J., Yang T., Jiang Y., Feng Z., and Zhang Z., 2023, Appropriate water and fertilizer supply can increase yield by promoting growth while ensuring the soil ecological environment in melon production, Agricultural Water Management, 289: 108561. https://doi.org/10.1016/j.agwat.2023.108561

Yismaw G., Fantaw S., and Ayalew A., 2024, Data on effect of mulches on growth and fruit yield of watermelon (*Citrullus lanatus* Thunb.) varieties in west Dembia district, central Gondar zone, Ethiopia, Data in Brief, 53: 110071.
https://doi.org/10.1016/j.dib.2024.110071

Yu Y., Guo S., Ren Y., Zhang J., Li M., Tian S., Wang J., Sun H., Zuo Y., Chen Y., Gong G., Zhang H., and Xu Y., 2022, Quantitative transcriptomic and proteomic analysis of fruit development and ripening in watermelon (*Citrullus lanatus*), Frontiers in Plant Science, 13: 818392.
https://doi.org/10.3389/fpls.2022.818392

Zhang R., Chai Y., Liang X., Liu X., Wang X., and Hu Z., 2024, A new plant-wearable sap flow sensor reveals the dynamic water distribution during watermelon fruit development, Horticulturae, 10(6): 649.
https://doi.org/10.3390/horticulturae10060649

Zhou H., Chen J., and Kang S., 2025, Model-assisted analysis on the response of tomato fruit growth to source-sink ratio regulated by water and nitrogen, Agricultural Water Management, 305: 109222.
https://doi.org/10.1016/j.agwat.2024.109222

Computational Molecular Biology 2026, Vol.16, No.2, 85-97
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Modeling Grain Yield Formation in Rice Based on Temperature and Water Management

Guifang Li [1,2]
1 Jiande Qingrun Modern Agriculture Development Co.,Ltd., Jiande, 311600, Zhejiang, China
2 Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China
Corresponding author: 18179387545@163.com
Computational Molecular Biology, 2026, Vol.16, No.2 doi: 10.5376/cmb.2026.16.0007
Received: 25 Jan., 2026
Accepted: 28 Feb., 2026
Published: 13 Mar., 2026

Preferred citation for this article:
Li G.F., 2026, Modeling grain yield formation in rice based on temperature and water management, Computational Molecular Biology, 16(2): 85-97 (doi: 10.5376/cmb.2026.16.0007)

Abstract Rice yield formation is jointly influenced by temperature and water conditions; these two factors not only determine the progression of rice growth and development but also directly impact photosynthesis, dry matter accumulation, and grain-filling efficiency. With the intensification of global climate change and the increasingly prominent issue of agricultural water scarcity, the development of rice yield formation models-based on the management of temperature and water-holds significant importance for enhancing rice production efficiency and ensuring food security. This paper systematically reviews the physiological and ecological foundations of rice yield formation, with a particular focus on analyzing the mechanisms by which temperature, water, and their interactions influence rice growth and yield components. Furthermore, it compares and summarizes empirical statistical models, process-based mechanistic models, and AI-driven predictive models, while exploring the application of model parameterization, calibration, and validation methods in yield forecasting. Additionally, by incorporating typical management strategies-such as alternate wetting and drying (AWD) irrigation-the paper analyzes rice yield simulation results under various hydrothermal conditions and evaluates their practical value in agricultural applications. The findings indicate that rational temperature regulation and water management can significantly enhance water-use efficiency and yield stability, and that the fusion of multi-source data coupled with intelligent modeling will constitute a key direction for future research on rice yield modeling. This paper serves as a theoretical reference and provides technical support for precision agriculture, the optimal allocation of water resources, and the management of stable rice production within the context of climate change.
Keywords Rice yield model; Temperature regulation; Water management; Crop simulation; Precision agriculture

1 Introduction

Rice is a cornerstone of global food security, feeding more than half of the world's population and supplying a large share of calories in Asia and many low-income regions (Rezvi et al., 2022). However, climate change is already exerting measurable impacts on rice production through shifts in temperature regimes and altered water availability, contributing to observed yield declines in major producing and food-insecure areas (Algarni et al., 2025). Maintaining and increasing rice yields under these pressures requires a quantitative understanding of how grain yield is formed as a function of temperature and water dynamics across critical growth stages (Shrestha et al., 2022). Process-based modeling that links environmental drivers with physiological processes offers a way to anticipate risks, design adaptive management, and support policy decisions for sustainable rice systems (Farooq et al., 2023).

Rising temperatures threaten rice grain yield through both chronic warming and short, extreme events, especially around reproductive stages. High-temperature stress during booting and flowering increases spikelet sterility and alters yield components, with yield per plant declining sharply as heat degree days accumulate at these stages. Recent work also emphasizes that microclimate and organ temperature, rather than air temperature alone, determine sterility risk, indicating that accurate prediction requires modeling canopy and panicle temperature within the crop-water-atmosphere continuum. At the same time, irrigation water is becoming increasingly scarce, and meta-analyses show that water-saving irrigation strategies such as alternate wetting and drying can substantially reduce irrigation inputs and increase water productivity, although yield responses vary with climate and soil conditions. Integrating these temperature and water processes in yield formation models is therefore crucial for realistic projections under future climates.

Process-based rice models such as ORYZA and APSIM-Oryza have been widely used to simulate phenology, biomass accumulation, and grain yield across diverse environments and management scenarios. These ecophysiological models dynamically represent photosynthesis, development, and soil water balance, and have been extended to include responses to drought and salinity, achieving yield root mean square errors generally within experimental uncertainty across stress gradients (Chang et al., 2023). More recently, mechanistic models of grain filling have been developed that explicitly link leaf-level photosynthesis, whole-plant carbon-nitrogen interactions, and panicle sink dynamics, reproducing observed yield formation under varied environmental and genetic perturbations and identifying stability of grain filling rate as a key determinant of maximum yield. Nonetheless, model evaluations indicate that conventional formulations often underperform when simulating yield responses to low-temperature stress or high-temperature-induced sterility at multiple stages, highlighting the need for improved temperature response functions that are stage-specific and variety-dependent (Shrestha et al., 2022; Shi et al., 2024).

The objective of this study is to develop and evaluate a modeling framework for grain yield formation in rice that explicitly couples temperature stress responses with water management effects across key developmental stages. Building on established process-based models, the approach refines algorithms for spikelet fertility, grain number per panicle, and grain filling under low and high temperatures, while incorporating contrasting irrigation regimes representative of traditional flooding and water-saving practices. The scope of the study includes calibration and validation using multi-environment experimental datasets, sensitivity analysis to identify dominant climatic and management drivers, and scenario simulations to quantify yield risks and opportunities under projected warming and alternative water regimes. By integrating temperature and water processes at the level of yield components, the work aims to provide a more reliable tool for assessing adaptation strategies-such as adjusted planting dates, stress-tolerant varieties, and optimized irrigation-for sustaining rice production and water productivity in a changing climate.

2 Physiological Basis of Rice Yield Formation

2.1 Growth stages and yield components of rice

The rice growth cycle is commonly divided into vegetative, reproductive, and grain-filling (ripening) stages, each with characteristic organs and yield-related processes. During vegetative growth, plant height, root development, leaf area, and especially tillering determine the potential panicle number per unit area and thus set the primary framework for yield. In the reproductive stage, panicle development, booting, and flowering occur; here the number of spikelets per panicle is defined, and this stage is the most sensitive to biotic and abiotic stresses, including temperature extremes.

Grain filling and ripening determine spikelet weight through endosperm development and carbohydrate deposition, with asynchronous filling between superior and inferior spikelets often limiting full yield potential (Liu et al., 2025). Yield analyses across diverse varieties show that total spikelet number (a function of panicle number and spikelets per panicle) correlates strongly and positively with grain yield, while higher spikelet numbers can trade off with filled grain percentage and grain weight if sink capacity exceeds source supply (Liu et al., 2024). Thus, yield modeling must capture how growth stages sequentially define panicle number, spikelet number, spikelet fertility, and grain weight.

2.2 Effects of temperature on rice physiology

Rice is highly sensitive to temperature, particularly during reproductive and grain-filling stages, where both high day temperatures and high night temperatures reduce yield through impaired reproductive development and altered carbon balance (El-Mageed et al., 2022; Shrestha et al., 2022). Heat stress at panicle initiation diminishes spikelet number by attenuating secondary branch and floret differentiation and enhancing degradation, while later heat episodes mainly affect spikelet fertility and grain weight, emphasizing the need for stage-resolved temperature response functions in models.

At flowering and grain filling, high temperatures increase spikelet sterility and reduce grain weight via multiple morpho-physiological pathways, including distortion of floral organs, reduced pollen viability, impaired anther

dehiscence, and shortened grain-filling duration (Shrestha et al., 2022). Night-time warming further elevates respiration, accelerates senescence, and contributes to yield penalties estimated at several percent per °C increase above critical thresholds, with reported yield declines of about 4-5% per 1 °C rise beyond 27 °C and up to 41% reduction by 2100 under projected high night temperatures (El-Mageed et al., 2022).

2.3 Effects of water management on rice growth

Water availability and irrigation strategies strongly influence both biomass production and yield components in rice. Under drip irrigation and mulching, progressive water stress at tillering reduces chlorophyll content, leaf photosynthesis, and final tiller number, leading to fewer effective panicles, lower seed-setting rate, and reduced thousand-grain weight, although moderate stress can substantially increase water-use efficiency relative to flooding (Xu et al., 2020). Field experiments with aerobic varieties show that mild water-saving irrigation (~20% less water than conventional) can enhance antioxidant activity, maintain photosynthesis, increase harvest index, and significantly improve grain yield and quality, indicating a non-linear response of growth and yield to water deficit intensity (Gao et al., 2024).

Across broader environments, meta-analysis of water-saving irrigation practices (controlled, intermittent, shallow-wet, AWD) reveals consistent increases in water productivity (4.7%-14.3%) relative to traditional flooding, with variable effects on yield depending on system, soil, and climate. Alternate wetting and drying regimes typically save 17%-34% of irrigation water and can increase or only slightly reduce yield, yielding higher water productivity, while continuous flooding maximizes yield but at the cost of much greater water consumption (Mboyerwa et al., 2021; Roushan et al., 2023). These quantitative relationships between water regime, evapotranspiration, yield components, and water productivity form a critical foundation for modeling grain yield under diverse water management scenarios.

3 Temperature and Water Interactions in Rice Production

3.1 Synergistic effects of temperature and soil moisture

Temperature and soil moisture interact strongly to determine rice yield, with compound extremes often causing larger losses than either stress alone. A panel analysis of rainfed and irrigated rice in India (2000-2018) showed that excessive heat markedly reduced yield, and that losses were greatest when high temperatures coincided with low soil moisture; in contrast, high soil moisture partly offset heat damage, underscoring the importance of managing root-zone water to buffer thermal stress (Mishra et al., 2024). A global empirical assessment similarly found that models using root-zone soil moisture, rather than precipitation, explained much more interannual yield variation and revealed that soil moisture and temperature contribute roughly equally to historical yield fluctuations, highlighting the need to explicitly represent both drivers in yield formation models (Proctor et al., 2022).

These synergistic effects arise because water status controls canopy cooling, stomatal conductance, and thus plant temperature under heat stress. Field experiments with super hybrid rice showed that as water supply was reduced from shallow flooding to mild and severe water stress, canopy relative humidity and plant-atmosphere and soil-atmosphere temperature differences declined, and grain yield fell by up to ~35%, with positive correlations between temperature differentials and yield (Meng et al., 2020). Long-term field data from Taiwan indicated that climate-change-induced increases in water-deficit stress, quantified via crop water status across growth stages, have increasingly constrained rice growth in recent decades, particularly during developmental stages, confirming that water deficits and warming jointly shape yield trajectories over time (Chen et al., 2023).

3.2 Stress responses under extreme climate conditions

Extreme hot-dry or cold-wet events are projected to become more frequent and can sharply disrupt grain yield formation. A global analysis for 1980-2009 showed that co-occurring extremely hot and dry events consistently reduced yields of major crops, including rice, worldwide, with probabilities of such compound extremes increasing over time (Heino et al., 2023). A review of compound heat and moisture extremes reported that hot droughts since about 2000 have been linked to yield losses up to 30% in key breadbasket regions, and that interactions among plant physiology, soil-plant-atmosphere water fluxes, and climate dynamics complicate prediction of net yield impacts under future compound extremes (Lesk et al., 2022).

Rice is particularly vulnerable when drought and temperature stress coincide at sensitive stages such as booting, flowering, and grain filling. Field experiments imposing combined drought and heat during flowering and early grain filling in contrasting cultivars recorded 20%-80% yield reductions, with cultivar-specific differences in which stage was most vulnerable and strong increases in panicle tissue temperature due to reduced panicle conductance under stress. Controlled-environment studies further demonstrated that simultaneous drought and high temperature at early stages (seedling, tillering) can completely prevent panicle formation and thus eliminate yield in susceptible cultivars, while tolerant genotypes maintain some productivity and show distinct patterns in grain quality and health-promoting compounds under combined stress.

3.3 Regional differences in rice production systems

Temperature-water interactions, and thus yield responses, vary widely among regions and production systems. A modeling study for Africa, covering irrigated and rainfed upland and lowland systems under multiple RCP scenarios, projected that without adaptation, higher temperatures shorten crop duration and reduce yields by about 24% under RCP8.5 by 2070; with higher-temperature-sum varieties, some rainfed systems gained modestly, but yields remained constrained by water availability, and irrigated dry-season rice in West Africa still faced large losses driven by photosynthesis reductions at extreme heat. In China, a comprehensive review found that climate change has shifted single and double rice belts northward, altered precipitation patterns, and increased the frequency of droughts and floods, leading to regionally divergent impacts where warming can either increase yields in cooler areas or reduce them in already warm zones through heat damage around flowering (Saud et al., 2022).

Finer-scale analyses highlight that regional differences in climate, water resources, and management produce contrasting yield trajectories and adaptation needs. Regional inequality assessments using ORYZA(v3) combined with climate projections for China showed average yield declines of 3.7-16.4% across regions, with central, eastern, and northwestern China most at risk under both rainfed and irrigated conditions, while northeastern and some southern areas may benefit under low-emission scenarios due to more favorable temperatures and water regimes (Zhan et al., 2023). At the farm scale in Indonesia, qualitative work revealed that upstream irrigated farmers mainly perceive climate impacts through pest outbreaks and heat, whereas downstream farmers, despite nominal irrigation access, experience climate change primarily as water shortages and rising temperatures, leading them to adopt distinct, locally tailored adaptation strategies for managing water scarcity and heat risk (Arifah et al., 2022).

4 Modeling Approaches for Rice Yield Formation

4.1 Empirical and statistical models

Empirical and statistical models relate rice yield directly to weather and agrometeorological indices, providing relatively simple tools for forecasting at regional scales. Panel regression and time-series analysis across 15 major rice-producing countries showed that increases in temperature tend to reduce production, while rainfall volume strongly affects output, highlighting rice's sensitivity to both warming and hydrological variability (Joseph et al., 2023). At subnational scales, climate-index-based regression models using modified Hendrik and Scholl methods successfully linked yields to maximum and minimum temperature, rainfall, humidity and other indices, with good coefficients of determination and accurate forecasts for Maharashtra districts (Sasane, 2023).

Comparisons of alternative statistical formulations emphasize the importance of choosing appropriate regression structures for non-linear climate-yield relationships. In Sri Lanka, multiple linear, power, robust and Gaussian process regressions, together with several machine learning methods, were applied to three decades of climate and yield data; Gaussian process regression achieved the lowest errors and highest correlation between observed and simulated yields (Wickramasinghe et al., 2021). Similar work in Uttarakhand used stepwise linear regression, LASSO, ridge and elastic net on seasonal weather variables, finding that penalized regressions such as LASSO and elastic net generally outperformed ordinary multiple regression, especially when multicollinearity among climate predictors was substantial (Setiya et al., 2023).

4.2 Process-based crop simulation models

Process-based crop models simulate rice growth and yield by representing phenology, biomass accumulation, and soil-water balance, offering mechanistic insight into temperature and water effects. The CERES-Rice model embedded in DSSAT has been extensively evaluated across Asia, typically predicting phenology with normalized RMSE of 1%-5% and grain yield with errors of 2%-5%, though performance often declines under severe water stress (Goswami and Dutta, 2020). Simulations with ORYZA2000 and an empirical energy-equivalent (EEQ) model showed yield declines of about 3.5%-7.6% per °C over a 4 °C range, but also demonstrated that simple regressions on minimum temperature can misattribute yield losses when solar radiation and rainfall covary with temperature (Sheehy et al., 2006).

Recent applications integrate detailed water management and greenhouse gas processes into process-based frameworks. Coupling CSM-CERES-Rice with the DSSAT-GHG module in subtropical Brazil allowed simultaneous evaluation of grain yield and methane emissions under continuous flooding, alternate wetting and drying, and sprinkler irrigation, with grain yield biases below 600 kg/ha and good agreement for daily CH_4 fluxes after calibration of key soil parameters (Figure 1) (Da Silva et al., 2025). In China, a calibrated CERES-Rice model was used with 60-year weather series to compare alternate wetting and drying, controlled drainage, and combined irrigation-drainage schemes; alternate wetting and drying produced the highest yields, while controlled irrigation-drainage treatments maximized irrigation and rainwater use efficiency, guiding optimization of water-saving strategies (Gao et al., 2023).

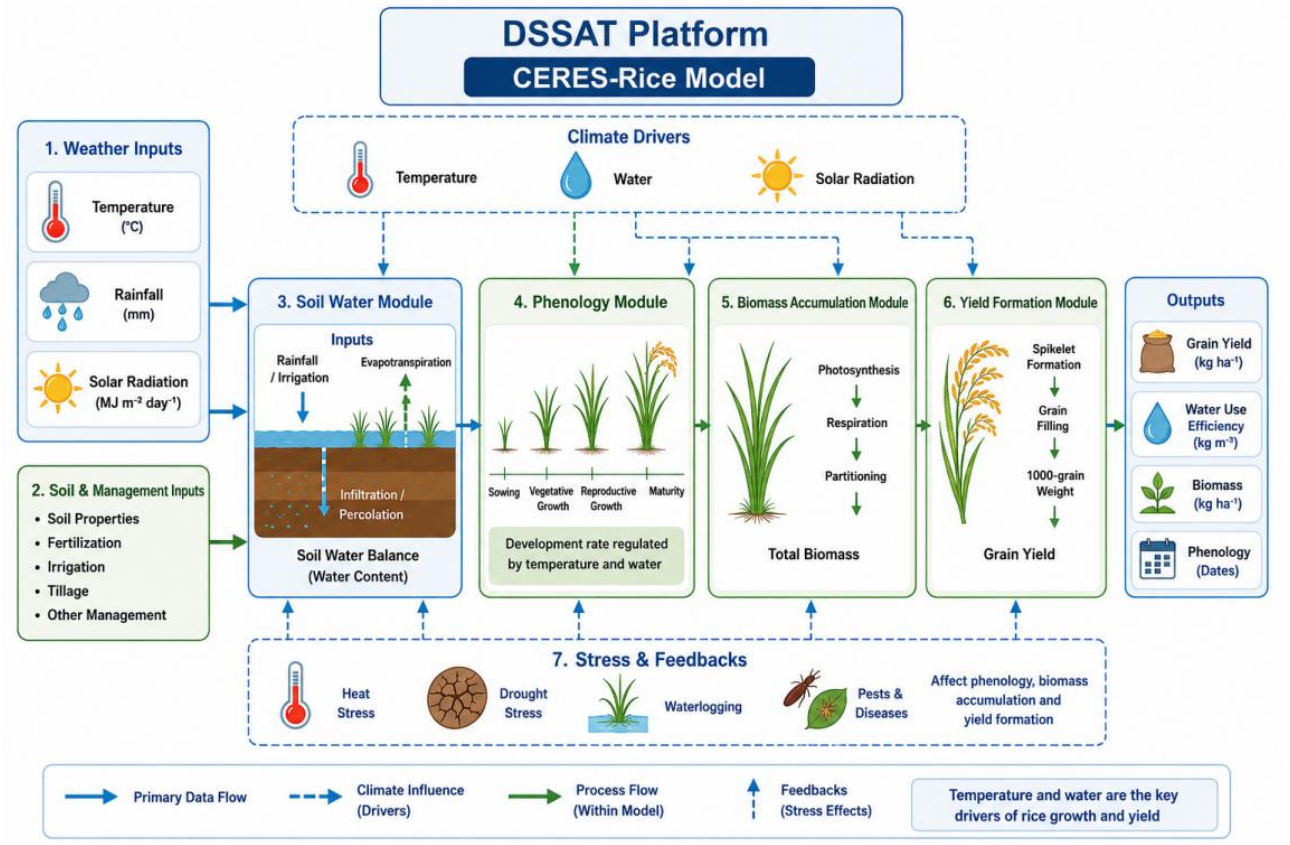


Figure 1 Conceptual framework of the CERES-Rice model embedded in DSSAT, illustrating interactions among weather inputs, soil-water balance, crop phenology, biomass accumulation, and grain yield formation

4.3 Machine learning and artificial intelligence approaches

Machine learning and deep learning methods provide flexible, data-driven alternatives for rice yield prediction that can ingest large climate, soil, and remote-sensing datasets. A comparative study in Chhattisgarh tested stepwise linear regression, penalized regressions, and artificial neural networks with 21 years of district-level yield and weather data; neural networks achieved R^2 values up to 1.0 in calibration and validation for some

districts, and ensemble methods such as random forest further improved performance over single models. Likewise, integrating phenology, growing-season climate, and geographic information in China showed that support vector machines, random forests, and backpropagation neural networks all outperformed multiple linear regression, with phenological variables contributing importance comparable to climatic predictors (Guo et al., 2020).

Deep learning and hybrid architectures have pushed yield prediction towards finer spatial and temporal scales. At the county scale across China, models based on LASSO, random forest, and long short-term memory (LSTM) networks were trained on satellite vegetation indices, meteorological indices, and soil properties; LSTM achieved R^2 values of 0.77-0.87 and lower RMSEs than both random forest and LASSO, and combining solar-induced chlorophyll fluorescence with EVI slightly enhanced performance by capturing drought and heat stress signals (Cao et al., 2021). At pixel scale in South and North Korea, satellite-integrated crop model outputs were used as training labels for a hybrid LSTM-1D-CNN network, which reached $R = 0.859$ and identified water-related indices and maximum temperature (North Korea) and vegetation and geographic variables (South Korea) as key predictors, illustrating the potential of crop model-AI fusion for spatially explicit yield formation under temperature and water variability (Jeong et al., 2021).

5 Key Variables and Parameterization in Yield Models

5.1 Temperature-related parameters

Temperature-related parameters in rice yield models describe how development rate and yield components respond to thermal conditions across growth stages. A foundational approach uses cardinal temperatures-base, optimum, and ceiling-to define a nonlinear response of development rate to temperature; the Beta-function framework derives optimum temperature and maximum development rate from these three temperatures and curvature coefficients, and has been shown to outperform simple thermal-time formulations in predicting flowering time in rice. Empirical and mechanistic simulations further demonstrate that yield declines with warming are moderate when temperature acts alone, but regression-based estimates may be biased if correlated factors such as solar radiation and rainfall are not properly separated, underscoring the importance of mechanistically grounded temperature functions in models.

Recent modeling work has refined temperature sensitivity at the level of yield components. Using a calibrated CERES-Rice model over six climate regions in China, yield sensitivity to temperature was decomposed into panicle number, filled grain number per panicle, and grain weight, revealing that negative yield responses were mainly driven by reductions in filled grains per panicle and were more strongly linked to high-temperature degree days than to growing degree days (Zhou et al., 2025). Other analyses show that conventional rice models often under-represent damage from extreme high or low temperatures, motivating adjustment of base and optimal temperatures or explicit heat-stress modules to improve simulation of growth duration and yield under warm or cold conditions (Figure 2) (Li et al., 2020).

5.2 Water management parameters

Water management parameters in rice models control soil water balance, root-zone moisture, and associated effects on evapotranspiration, biomass, and yield. In water-driven models such as AquaCrop, key parameters include the normalized crop water productivity (WP), stage-specific basal and single crop coefficients (Kc), and water-stress coefficients that reduce transpiration, canopy growth, and harvest index when soil water falls below critical thresholds; for rice, calibrated WP around 19 $g \cdot m^{-2}$ and harvest index near 0.47 have provided good simulations of canopy cover, biomass, yield, and water balance under multiple irrigation regimes in arid and sub-humid environments (Elsadek et al., 2023; Mostafa et al., 2023). A broader review of soil water balance modeling highlights the need for careful parameterization of dual Kc (separating soil evaporation and crop transpiration), soil water holding characteristics, and root-zone depth to derive realistic irrigation requirements and to link water use to yield and water productivity indicators (Pereira et al., 2020).

Process-based rice models with explicit soil modules use management parameters to represent alternative irrigation strategies such as continuous flooding, alternate wetting and drying (AWD), controlled irrigation, and

different ponding depths. For example, a calibrated CERES-Rice model successfully reproduced grain yield, evapotranspiration, irrigation volume, and leaf area index across AWD and controlled irrigation-drainage schemes, then used long-term meteorological scenarios to compare water-saving and yield responses among treatments (Gao et al., 2023). Similarly, AquaCrop-based simulations under drying-wetting cycles in paddy soils and fixed-interval irrigation in direct-seeded rice quantified how changes in irrigation frequency alter evapotranspiration, percolation, and soil moisture dynamics, revealing that current stress coefficients may overestimate water deficit under certain conditions and should be revised for rice-specific hydrology (Elsadek et al., 2023).

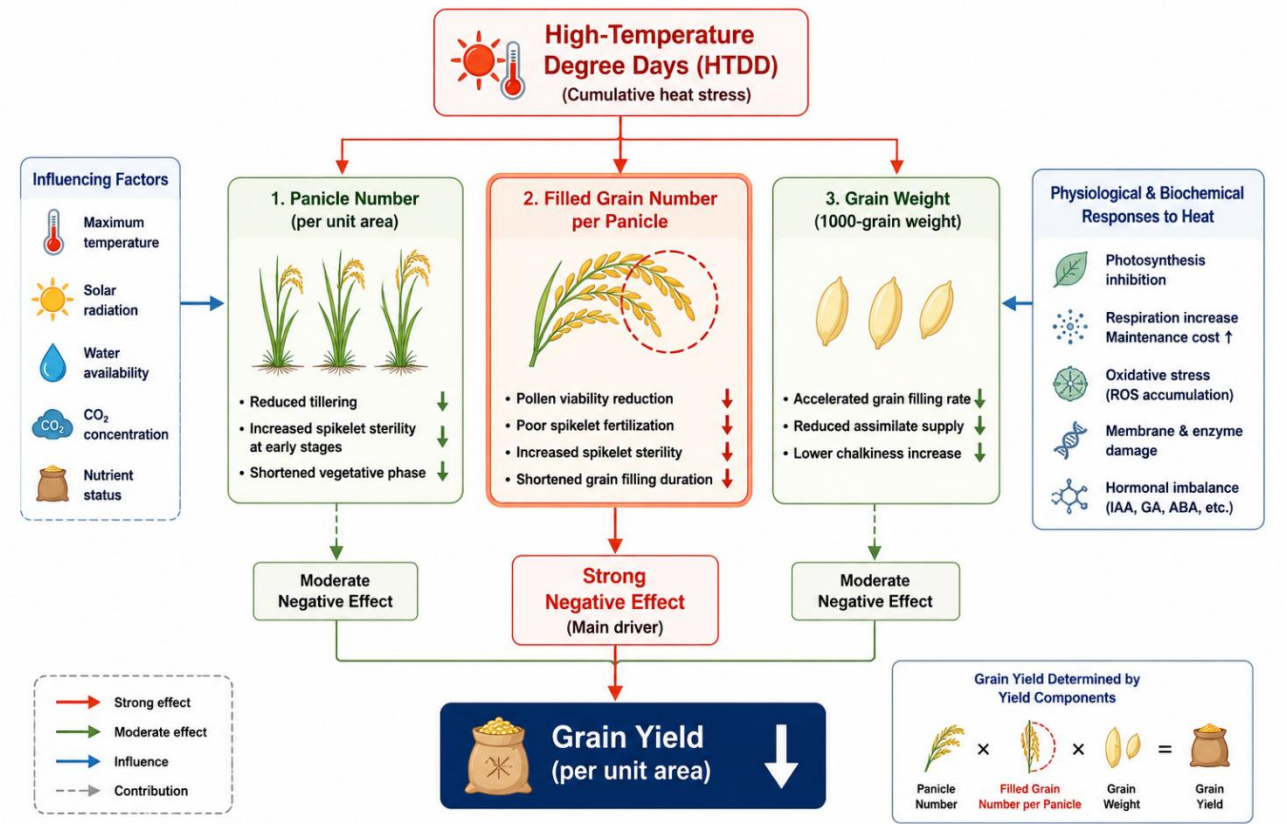


Figure 2 presents the pathways through which high-temperature stress affects rice yield formation. Among the yield components, reductions in filled grain number per panicle contribute most strongly to negative yield responses under warming conditions

5.3 Calibration and validation of models

Accurate representation of temperature and water effects in yield models depends on rigorous calibration and validation of both genetic and environmental/management parameters. In DSSAT-CERES-Rice applications, cultivar-specific coefficients (e.g., phenology, tillering, grain filling, spikelet number, temperature tolerance) are estimated using multi-year field experiments with contrasting genotypes, establishment methods, and nitrogen levels; evaluation against observed yields and phenology has shown low normalized RMSE and realistic sensitivity to ±1 °C temperature changes, confirming that calibrated models can capture both baseline performance and climate sensitivity (Islam and Hasan, 2021). For upland rice, detailed documentation of coefficients such as P2R (photoperiod sensitivity), P5 (grain-filling duration), G1 (spikelet number), G3 (tillering), and G4 (temperature tolerance) illustrates how parameter sets encode cultivar adaptation to different temperature regimes and allow tested models to simulate flowering and maturity across controlled temperature treatments.

Validation must also address parameter uncertainty and model robustness across sites and years. A cross-validation study with ORYZA (v3) generated multiple feasible parameter sets for a high-yielding variety under limited data and showed that several sets produced satisfactory simulations of biomass components and total aboveground biomass when tested with independent datasets, implying that non-uniqueness of calibrated parameters should be explicitly recognized (Nurulhuda et al., 2022). At larger scales, DSSAT-based studies have

calibrated models using gridded weather, management information, and representative genetic coefficients, achieving district-level correlations above 0.7 and relative RMSE below 25% for most major rice-growing districts, and demonstrating reasonable skill in reproducing yield anomalies in out-of-sample years-supporting their use for near-real-time yield estimation and risk assessment (Arumugam et al., 2020).

6 Applications of Rice Yield Formation Models

6.1 Decision support for irrigation and fertilization

Rice yield formation models are increasingly used to optimize coupled water-nitrogen management, balancing grain yield with resource efficiency and environmental impacts. Field experiments combined with regression modeling show that irrigation regime and nitrogen rate jointly determine grain yield, total water productivity, and nitrogen recovery efficiency, but that these objectives cannot be maximized simultaneously, motivating multi-objective decision tools based on water-nitrogen-yield response surfaces (Cao et al., 2020). Multi-objective quadratic models integrating water-nitrogen-yield and water-nitrogen-quality relationships further demonstrate that optimal irrigation and nitrogen combinations differ among management scenarios, and that excessive inputs can become counterproductive for both yield and grain quality.

Model-based seasonal and long-term scenario analyses allow irrigation and fertilization decisions to be tailored to local climate risk. Using CSM-CERES-Rice within DSSAT, one study quantified how early direct seeding, no-tillage, and moderate nitrogen rates simultaneously improved yield, irrigation efficiency, and reduced methane emissions over 35 years, providing concrete guidance on planting date, tillage, and N rate selection (Figure 3) (Darikandeh et al., 2025). Machine-learning decision models that couple ensemble yield prediction (e.g., extremely randomized trees) with swarm-intelligence optimization have also been proposed, enabling site-specific recommendations of N-P-K base fertilizer that increase average rice yields by more than 13% while reducing the need for extensive field experimentation (Gao et al., 2023).

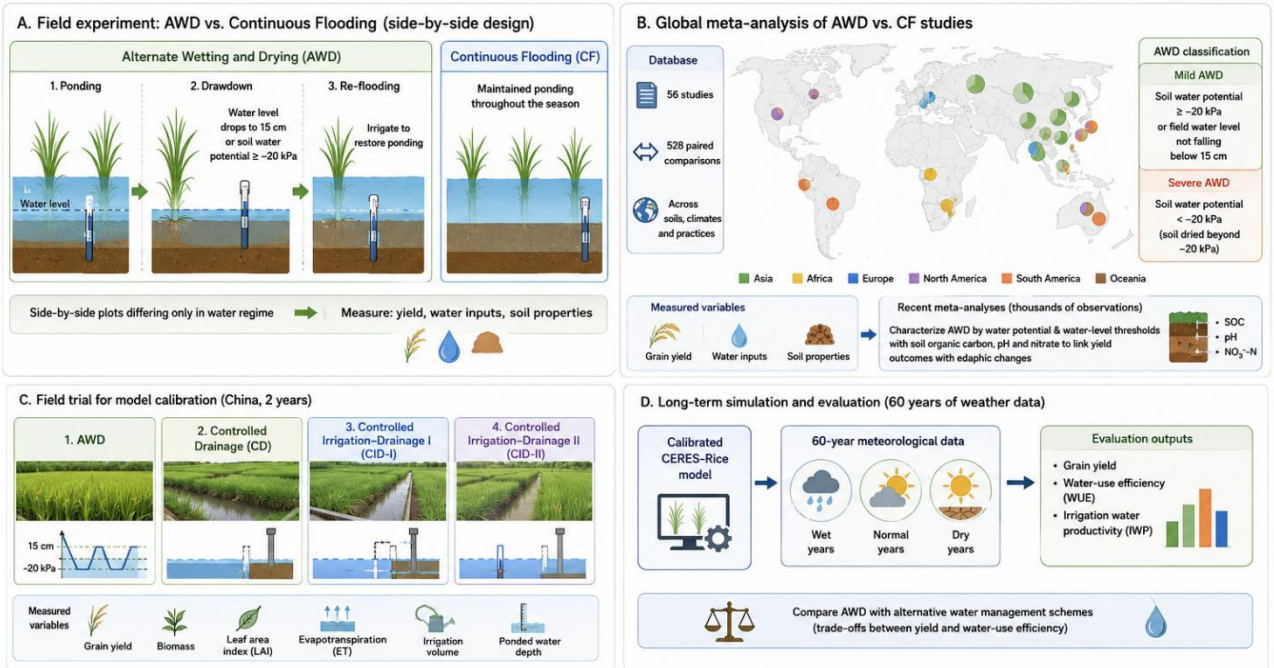


Figure 3 Experimental designs and analytical frameworks used to compare alternate wetting and drying (AWD) with continuous flooding (CF) in rice production systems

6.2 Climate change adaptation and risk assessment

Process-based crop models calibrated for local cultivars are widely applied to assess climate change impacts and identify adaptation levers. In Mediterranean Türkiye, DSSAT-CERES-Rice simulations under multiple GCMs and RCPs showed that irrigated yields could increase slightly in late-century, whereas rainfed yields declined by 15%-25% due to higher temperatures, shorter growth duration, and soil-moisture stress, illustrating how

yield-formation models support evaluation of irrigation as an adaptation option (Baydar, 2026). In Central Java, coupling MarkSim-generated weather with CERES-Rice projected yield decreases in all seasons under RCP2.6-8.5, with up to 11.8% reduction in the second dry season, and pointed to dynamic cropping calendars, irrigation modernization, and integrated nutrient management as priority adaptations (Ansari et al., 2021).

Meta-analytic and multi-model frameworks extend these assessments to global and regional risk profiles. A global meta-model derived from 8,703 process-model simulations showed that, under RCP4.5 without adaptation, major crops including rice face mean yield losses of 6%-21%, but that cultivar choice for rice and irrigation method for maize are among the most influential adaptive strategies, partially offsetting losses as warming intensifies (Abramoff et al., 2023). A separate meta-analysis of 111 climate-rice modeling studies quantified that each 1 °C increase in mean temperature reduces rice yield by 3.85% on average, while elevated CO_2 and adaptive management can compensate some losses, underscoring the role of yield models in probabilistic risk assessment and in designing adaptation portfolios (Li et al., 2024).

6.3 Precision Agriculture and Digital Farming

Yield formation models are increasingly embedded in precision agriculture systems that combine remote sensing, IoT, and AI to support within-field management. An integrated IoT-based framework uses multispectral satellite indices, machine-learning yield prediction (random forest $R^2 \approx 0.96$), and fuzzy-logic irrigation control to recommend suitable crops and fertilizer, while a solar-powered irrigation system achieves about 61% water savings compared with average logic, demonstrating how digital decision layers can operationalize model insights on crop water and nutrient requirements (Saha et al., 2025). At a broader scale, reviews of remote-sensing applications in precision agriculture highlight how satellite and UAV data, linked to crop growth and yield models, are now used operationally for crop monitoring, irrigation scheduling, variable-rate nutrient application, and yield prediction, supported by cloud computing and machine learning workflows (Sishodia et al., 2020).

Recent syntheses of IoT- and AI-enabled sensing technologies emphasize that dense soil-moisture, nutrient, and plant-stress sensor networks, combined with models and ML (e.g., SVMs, CNNs, random forests), underpin real-time optimization of irrigation, fertilization, and pest management across arable systems (Miller et al., 2025). Complementary reviews of precision agriculture for yield prediction stress that hybrid systems merging deep learning (e.g., Bi-LSTM) with multisource remote-sensing inputs can capture the combined effects of temperature, water status, and other stresses on yield, pointing toward digital twins of rice cropping systems where grain yield formation under variable temperature and water regimes is continuously simulated and updated from field data (Saha et al., 2025).

7 Case Study: Modeling Rice Yield Under Alternate Wetting and Drying Irrigation

7.1 Experimental design and data collection

Field experiments comparing alternate wetting and drying (AWD) with continuous flooding (CF) typically use side-by-side plots differing only in water regime, enabling quantification of yield and water responses across soils, climates, and management practices. A global meta-analysis synthesized 56 such studies (528 paired comparisons) and defined mild AWD using thresholds of soil water potential ≥ -20 kPa or field water level not dropping below 15 cm, and severe AWD when soils dried beyond -20 kPa, with associated measurements of yield, water inputs, and basic soil properties. More recent meta-analyses have expanded this database, assembling thousands of observations worldwide and characterizing AWD treatments by water potential and water-level thresholds together with soil organic carbon, pH, and nitrate to link yield outcomes with edaphic changes (Zhou et al., 2025).

Experimental designs used for model calibration and testing often include multiple irrigation schemes and long weather records. In China, a two-year field trial with four irrigation and drainage treatments-AWD, controlled drainage, and two controlled irrigation-drainage regimes-was established to calibrate CERES-Rice using detailed measurements of grain yield, biomass, leaf area index, evapotranspiration, irrigation volume, and ponded water depth. Long-term simulations then combined these calibrated parameters with 60 years of meteorological data classified into wet, normal, and dry years to evaluate yield and water-use efficiency trade-offs among AWD and alternative schemes (Gao et al., 2023).

7.2 Model construction and simulation results

Case-study modeling of AWD typically couples a calibrated crop growth module with an explicit water-balance representation of ponded depth, soil water status, and irrigation triggers. In a DSSAT-CERES-Rice application for central China, cultivar coefficients were first calibrated under observed AWD and non-AWD regimes, achieving normalized RMSE of 3%-10% for yield, biomass, evapotranspiration, irrigation, and leaf area index, indicating reliable capture of growth and water use across treatments. The calibrated model was then driven with historical climate sequences and AWD rules defined by reflooding thresholds, allowing simulation of grain yield and water productivity under contrasting hydrological years and irrigation strategies (Gao et al., 2023).

To better represent AWD hydrology, an improved ORYZA2000 framework integrated a new water-balance simulation tailored to intermittent flooding and drying, along with dynamic root-length growth and revised water-stress algorithms. Applied to paddy fields under CF and AWD in two Chinese regions, the enhanced model substantially improved the simulation of ponded water depth, irrigation and drainage volumes, evapotranspiration, and percolation, with Nash-Sutcliffe efficiencies for ponded depth of 0.82-0.94 and relative errors for total irrigation and drainage mostly within ±10%. Yield prediction remained comparable to or slightly better than the original version, demonstrating that explicit AWD parameterization can capture both water balance and yield formation with good accuracy.

7.3 Implications for sustainable rice production

Model-supported analyses clarify under which conditions AWD can save water without compromising yield, informing sustainable irrigation guidelines. Meta-analysis of 528 AWD-CF comparisons showed that, overall, AWD reduced yields by 5.4%, but mild AWD regimes did not significantly decrease yield, whereas severe AWD caused average losses of 22.6%, particularly in higher-pH and low-carbon soils. A broader synthesis of 3194 observations from 200 studies confirmed that AWD increased water-use efficiency by about 31% but imposed an average 6% yield penalty, and identified optimal thresholds (soil water potential > -15 kPa, water depth < 18.5 cm) and favorable soil conditions under which AWD can actually raise yields by up to 4-7% when combined with appropriate nitrogen, straw, or biochar management.

Process-based simulations extend these insights to long-term climate variability and regional planning. CERES-Rice modeling over 60 historical weather years showed that AWD often produced the highest yields among several water-saving schemes across wet, normal, and dry years, though other controlled irrigation-drainage strategies sometimes achieved greater irrigation and rainwater-use efficiency, suggesting context-specific trade-offs between yield maximization and water conservation (Gao et al., 2023). Global analyses indicate that implementing soil-water-potential-based AWD on suitable irrigated rice areas can increase water productivity over large fractions of Asia, particularly in India, Bangladesh, and central China, demonstrating that AWD-informed models can underpin strategies for sustainable intensification that jointly address food security and freshwater scarcity (Bo et al., 2024).

8 Challenges, Future Perspectives, and Conclusions

Despite substantial advances, rice yield projections under future climate remain highly uncertain. A meta-analysis of 111 studies showed large variability in simulated yield responses to changes in temperature, precipitation, radiation, and CO_2, reflecting differences in climate models, emission scenarios, and, critically, crop model structure and parameterization. Similarly, a multi-model intercomparison of 13 rice models found that spread among crop models exceeded that from 16 global climate models, and that individual models did not consistently reproduce yields across very cool and very warm sites, indicating structural weaknesses in representing temperature and CO_2 responses. Key physiological processes are still imperfectly captured. Sensitivity analysis of the 13-model ensemble identified biomass formation and harvest index responses to warming and elevated CO_2 as major sources of error, while most simulations assumed ideal water and nutrient management and ignored pests, diseases, and sub-optimal farmer practices. Meta-regression work further demonstrated that yield responses aggregate multiple interacting drivers (temperature, precipitation, CO_2, management), and that the choice of study sites, climate scenarios, and adaptation assumptions introduces additional unexplained variation into projected yield changes.

Future modeling must better integrate climate, hydrology, and crop growth, particularly in data-scarce, climate-vulnerable regions. A review of climate-hydrological-crop modeling for Indonesian rice production highlighted critical gaps in long-term observations, local cultivar data, and systematic calibration/validation, as well as limited use of fully coupled multi-model frameworks. Bayesian multi-model ensemble methods applied to Chinese rice regions showed that statistically combining multiple climate models can reduce bias in temperature, radiation, and wind projections and provide more robust estimates of future yield, evapotranspiration, and irrigation requirements. There is also strong scope for hybrid approaches linking process-based models with machine learning and explainable AI. A recent study in China coupled DSSAT with random forests and SHAP analysis to project rice yields under multiple SSP scenarios and to rank the relative influence of variables such as growing degree days, shallow versus deep soil moisture, and precipitation regimes on yield. Global meta-modeling across 8,703 process-model simulations similarly used machine learning to map yield change as a function of climate and adaptation, revealing that for rice, cultivar choice is a dominant lever for avoiding large losses, and demonstrating how statistical emulators can synthesize complex multi-model ensembles for risk analysis.

Rice yield formation under changing temperature and water regimes is governed by interacting physiological processes and management decisions that are only partially resolved in current models. Meta-analyses show that rising temperature and altered precipitation generally reduce rice yields, but that elevated CO_2 and adaptive practices, including improved management, can offset part of these losses; however, the magnitude and direction of impacts vary widely across models and regions. Probabilistic assessments and global meta-models further indicate that without adaptation, most rice-growing areas face significant yield declines as global mean temperature rises, while adaptation-particularly through cultivar choice and irrigation strategies-substantially narrows projected losses. Going forward, credible prediction and decision support will require to reduce structural and parametric uncertainty in ecophysiological models; embedding them in integrated climate-hydrological-crop frameworks; and exploiting machine learning and ensemble techniques to quantify risk and design robust adaptation portfolios. By explicitly representing temperature and water interactions, and by using improved data and hybrid modeling strategies, next-generation rice yield models can more reliably guide climate-smart water management, cultivar deployment, and policy for sustainable rice production under a warming and water-constrained climate.

Acknowledgments

Thanks to the reviewers for providing detailed comments and guidance on the manuscript of this study. The reviewers' keen insights into the issues and attention to detail have greatly benefited the authors.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abramoff R., Ciais P., Zhu P., Hasegawa T., Wakatsuki H., and Makowski D., 2023, Adaptation strategies strongly reduce the future impacts of climate change on simulated crop yields, Earth's Future, 11(3): e2022EF003190.
https://doi.org/10.1029/2022EF003190

Ansari A., Lin Y., and Lur H., 2021, Evaluating and adapting climate change impacts on rice production in Indonesia: A case study of the Keduang Subwatershed, Central Java, Environments, 8(11): 117.
https://doi.org/10.3390/environments8110117

Arumugam P., Chemura A., Schauberger B., and Gornott C., 2020, Near real-time biophysical rice (*Oryza sativa* L.) yield estimation to support crop insurance implementation in India, Agronomy, 10(11): 1674.
https://doi.org/10.3390/agronomy10111674

Baydar A., 2026, Assessing rice yield responses to climate change scenarios using a crop simulation model, PeerJ, 14: e20965.
https://doi.org/10.7717/peerj.20965

Bo Y., Wang X., Van Groenigen K., Linquist B., Müller C., Li T., Yang J., Jägermeyr J., Qin Y., and Zhou F., 2024, Improved alternate wetting and drying irrigation increases global water productivity, Nature Food, 5(12): 1005-1013.
https://doi.org/10.1038/s43016-024-01081-z

Cao J., Zhang Z., Tao F., Zhang L., Luo Y., Zhang J., Han J., and Xie J., 2021, Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches, Agricultural and Forest Meteorology, 297: 108275.
https://doi.org/10.1016/j.agrformet.2020.108275

Cao X., Wu L., Lu R., Zhu L., Zhang J., and Jin Q., 2020, Irrigation and fertilization management to optimize rice yield, water productivity and nitrogen recovery efficiency, Irrigation Science, 39(2): 235-249.
https://doi.org/10.1007/s00271-020-00700-4

Chang T., Wei Z., Shi Z., Xiao Y., Zhao H., Chang S., Qu M., Song Q., Chen F., Miao F., and Zhu X., 2023, Bridging photosynthesis and crop yield formation with a mechanistic model of whole plant carbon-nitrogen interaction, in silico Plants, 5(1): diad011.
https://doi.org/10.1093/insilicoplants/diad011

Chen H., Wu Y., Cheng C., and Teng C., 2023, Effect of climate change-induced water-deficit stress on long-term rice yield, PLoS ONE, 18(4): e0284290.
https://doi.org/10.1371/journal.pone.0284290

Da Silva E.H., Hoogenboom G., Boote K.J., Cuadra S.V., Porter C.H., Scivittaro W.B., Steinmetz S., and Cerri C.E.P., 2025, Implications of water management on methane emissions and grain yield in paddy rice: A case study under subtropical conditions in Brazil using the CSM-CERES-Rice model, Agricultural Water Management, 305: 109234.
https://doi.org/10.1016/j.agwat.2024.109234

Darikandeh D., Shahnazari A., Khoshravesh M., Yousefian M., Porter C.H., and Hoogenboom G., 2025, Optimizing rice management to reduce methane emissions and maintain yield with the CSM-CERES-Rice model, Agricultural Systems, 226: 104248.
https://doi.org/10.1016/j.agsy.2024.104248

El-Mageed T.A., El-Mageed S.A., El-Saadony M.T., Abdelaziz S.M., and Abdou N.M., 2022, Plant growth-promoting rhizobacteria improve growth, morph-physiological responses, water productivity, and yield of rice plants under full and deficit drip irrigation, Rice, 15(1): 47.
https://doi.org/10.1186/s12284-022-00564-6

Elsadek E., Zhang K., Mousa A., Ezaz G., Tola T., Shaghaleh H., Hamad A., and Hamoud Y., 2023, Study on the in-field water balance of direct-seeded rice with various irrigation regimes under arid climatic conditions in Egypt using the AquaCrop model, Agronomy, 13(2): 609.
https://doi.org/10.3390/agronomy13020609

Farooq A., Farooq N., Akbar H., Hassan Z., and Gheewala S.H., 2023, A critical review of climate change impact at a global scale on cereal crop production, Agronomy, 13(1): 162.
https://doi.org/10.3390/agronomy13010162

Gao C., Lin M., He L., Tang M., Li J., and Sun W., 2024, The impact of water-saving irrigation on rice growth and comprehensive evaluation of irrigation strategies, Agronomy, 14(7): 1363.
https://doi.org/10.3390/agronomy14071363

Gao J., Zeng W., Ren Z., Ao C., Lei G., Gaiser T., and Srivastava A., 2023, A fertilization decision model for maize, rice, and soybean based on machine learning and swarm intelligent search algorithms, Agronomy, 13(5): 1400.
https://doi.org/10.3390/agronomy13051400

Gao S., Gu Q., Gong X., Li Y., Yan S., and Li Y., 2023, Optimizing water-saving irrigation schemes for rice (*Oryza sativa* L.) using DSSAT-CERES-Rice model, International Journal of Agricultural and Biological Engineering, 16(2): 120-129.
https://doi.org/10.25165/j.ijabe.20231602.7361

Goswami P., and Dutta G., 2020, Evaluation of DSSAT model (CERES-rice) on rice production: A review, International Journal of Chemical Studies, 8(5): 404-409.
https://doi.org/10.22271/chemi.2020.v8.i5f.10327

Guo Y., Fu Y., Hao F., Zhang X., Wu W., Jin X., Bryant C., and Senthilnath J., 2020, Integrated phenology and climate in rice yields prediction using machine learning methods, Ecological Indicators, 120: 106935.
https://doi.org/10.1016/j.ecolind.2020.106935

Heino M., Kinnunen P., Anderson W., Ray D.K., Puma M.J., Varis O., Siebert S., and Kummu M., 2023, Increased probability of hot and dry weather extremes during the growing season threatens global crop yields, Scientific Reports, 13(1): 29378.
https://doi.org/10.1038/s41598-023-29378-2

Islam S., and Hasan A., 2021, Determination of upland rice cultivar coefficient specific parameters for DSSAT (Version 4.7)-CERES-Rice crop simulation model and evaluation of the crop model under different temperature treatments conditions, American Journal of Plant Sciences, 12(5): 754-770.
https://doi.org/10.4236/ajps.2021.125054

Jeong S., Ko J., and Yeom J., 2021, Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea, Science of the Total Environment, 802: 149726.
https://doi.org/10.1016/j.scitotenv.2021.149726

Joseph M., Moonsammy S., Davis H., Warner D., Adams A., and Oyedotun T., 2023, Modelling climate variabilities and global rice production: A panel regression and time series analysis, Heliyon, 9(4): e15480.
https://doi.org/10.1016/j.heliyon.2023.e15480

Lesk C., Anderson W., Rigden A., Coast O., Jägermeyr J., McDermid S., Davis K.F., and Konar M., 2022, Compound heat and moisture extreme impacts on global crop yields under climate change, Nature Reviews Earth and Environment, 3(12): 872-889.
https://doi.org/10.1038/s43017-022-00368-8

Li N., Zhao Y., Han J., Yang Q., Liang J., Liu X., Wang Y., and Huang Z., 2024, Impacts of future climate change on rice yield based on crop model simulation: A meta-analysis, Science of the Total Environment, 930: 175038.
https://doi.org/10.1016/j.scitotenv.2024.175038

Li S., Fleisher D., Timlin D., Reddy V.R., Wang Z., and McClung A., 2020, Evaluation of different crop models for simulating rice development and yield in the U.S. Mississippi Delta, Agronomy, 10(12): 1905.
https://doi.org/10.3390/agronomy10121905

Liu B., Meng S., Yang J., Wu J., Peng Y., Zhang J., and Ye N., 2025, Carbohydrate flow during grain filling: Phytohormonal regulation and genetic control in rice (*Oryza sativa*), Journal of Integrative Plant Biology, 67(6): 1086-1104.
https://doi.org/10.1111/jipb.13904

Liu K., Zhang K., Zhang Y., Cui J., Li Z., Huang J., Li S., Zhang J., Deng S., Zhang Y., Huang J., Ren L., Chu Y., Zhao H., and Chen H., 2024, Optimizing the total spikelets increased grain yield in rice, Agronomy, 14(1): 152.
https://doi.org/10.3390/agronomy14010152

Miller J.O., de Barros P.R., Schulenburg A.N., Tully K.L., 2025, Coastal stressors reduce crop yields and alter soil nutrient dynamics in low-elevation farmlands, Discover Agriculture, 3(1): 119.
https://doi.org/10.1007/s44279-025-00303-7

Nurulhuda K., Muharam F.M., Shahar N.A.N., Hashim M.F.C., Ismail M.R., Keesman K.J., Zulkafli Z., 2022, ORYZA (v3) rice crop growth modeling for MR269 under nitrogen treatments: Assessment of cross-validation on parameter variability, Computers and Electronics in Agriculture, 195: 106809.
https://doi.org/10.1016/j.compag.2022.106809

Proctor J., Rigden A., Chan D., Huybers P., 2022, More accurate specification of water supply shows its importance for global crop production, Nature Food, 3(9): 753-763.
https://doi.org/10.1038/s43016-022-00592-x

Pereira L.S., Paredes P., Melton F., Johnson L., Wang T., López-Urrea R., Cancela J.J., Allen R.G., 2020, Prediction of crop coefficients from fraction of ground cover and height: background and validation using ground and remote sensing data, Agricultural Water Management, 241: 106197.
https://doi.org/10.1016/j.agwat.2020.106197

Rezvi H.U.A., Tahjib-Ul-Arif M., Azim M.A., et al., 2022, Rice and food security: Climate change implications and the future prospects for nutritional security, Food and Energy Security, 12(1): e430.
https://doi.org/10.1002/fes3.430

Shrestha S., Giri D., Dhital M., Chaudhary B., Pandey R., Bastakoti B., 2022, Effect of different nitrogen levels on yield and yield attributes of different rice varieties in DDSR condition at Kanchanpur, Nepal, Archives of Agriculture and Environmental Science, 7(3): 310–317.
https://doi.org/10.26832/24566632.2022.070302

Saha S., Chant D., Welham J., 2025, A systematic review of the prevalence of schizophrenia, PLoS Medicine, 22(5): e141.
https://doi.org/10.1371/journal.pmed.0020141

Sishodia R.P., Ray R.L., Singh S.K., 2020, Applications of remote sensing in precision agriculture: A review, Remote Sensing, 12(19): 3136.
https://doi.org/10.3390/rs12193136

Setiya P., Satpathi A., Das B., Nain A.S., Jha P.K., Singh S., 2023, Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India, Sustainability, 15(3): 2786.
https://doi.org/10.3390/su15032786

Sheehy J.E., Mitchell P.L., Allen L.H., Ferrer A.B., 2006, Mathematical consequences of using various empirical expressions of crop yield as a function of temperature, Field Crops Research, 98(2): 216-221.
https://doi.org/10.1016/j.fcr.2006.02.008

Wickramasinghe W.M.D.M., Devasinghe D.A.U.D., Dissanayake D.M.D., Benaragama D.I.D.S., Egodawatta W.C.P., Suriyagoda L.D.B., 2021, Growth physiology and crop yields of direct-seeded rice under diverse input systems in the dry zone of Sri Lanka, Tropical Agricultural Research, 32(3): 325-337.
https://doi.org/10.4038/tar.v32i3.8496

Zhou J., Li J., Zhang Y., Yang Y., Lv Y., Pu Q., Deng X., Tao D., 2025, Introgression among subgroups is an important driving force for genetic improvement and evolution of the Asian cultivated rice (Oryza sativa L.), Frontiers in Plant Science, 16: 1535880.
https://doi.org/10.3389/fpls.2025.1535880

Computational Molecular Biology 2026, Vol.16, No.2, 98-113
http://bioscipublisher.com/index.php/cmb

Research Insight Open Access

Prediction of Maize Yield Based on Soil Nutrients and Climate Variables

Jinhua Cheng, Wei Wang
Institute of Life Sciences, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China
Corresponding author: wei.wang@jicat.org
Computational Molecular Biology, 2026, Vol.16, No.2 doi: 10.5376/cmb.2026.16.0008
Received: 02 Feb., 2026
Accepted: 08 Mar., 2026
Published: 21 Mar., 2026

Preferred citation for this article:
Cheng J.H., and Wang W., 2026, Prediction of maize yield based on soil nutrients and climate variables, Computational Molecular Biology, 16(2): 98-113 (doi: 10.5376/cmb.2026.16.0008)

Abstract Maize yield prediction plays an essential role in ensuring food security and promoting sustainable agricultural management. This study explores a prediction framework based on soil nutrient characteristics and climate variables to improve the accuracy and reliability of maize yield estimation. Key soil indicators, including nitrogen, phosphorus, potassium, organic matter, and pH value, were combined with climate factors such as temperature, precipitation, and accumulated growing degree days. Multiple prediction models, including traditional statistical approaches, machine learning algorithms, and deep learning methods, were constructed and compared. The study further analyzed the interaction effects between soil and climate variables and evaluated model performance using indicators such as RMSE, MAE, and R^2. A regional case study was conducted to verify the applicability and robustness of the proposed framework. The results demonstrate that integrating soil nutrient and climate data can significantly enhance maize yield prediction accuracy and provide valuable support for precision agriculture, crop management, and agricultural decision-making.
Keywords Maize yield prediction; Soil nutrients; Climate variables; Machine learning; Precision agriculture

1 Introduction

Global demand for maize is rising steadily as it underpins food, feed, and industrial supply chains, yet production is increasingly constrained by climate variability and degraded soils. Temperature extremes, altered rainfall, and declining soil fertility jointly threaten yield stability, especially in regions already facing food insecurity. Improving the accuracy of maize yield prediction by explicitly linking soil nutrients with key climate variables is therefore essential for optimizing fertilization, managing risk, and designing climate-smart production systems. Maize yields respond strongly to interactions between climate conditions and soil nutrient status. Studies in sub-Saharan Africa and China show that nitrogen (N), phosphorus (P), and potassium (K) inputs can buffer or amplify the impacts of changing CO_2, temperature, and rainfall on yield, and that soil indigenous nutrients strongly modulate yield losses under warming (Falconnier et al., 2020. Long-term experiments further indicate that soil fertility improvements (e.g., higher total and available N and P) enhance yield stability and sustainability, while climate warming tends to reduce yields where soil fertility is low. At the same time, nutrient management alone is insufficient; integrating soil, climate, and management information is needed to maintain productivity under ongoing climate change (Ocwa et al., 2023). In this context, a predictive framework that couples soil nutrient properties with climate variables can support more precise fertilizer recommendations, reduce environmental risks, and improve resilience of maize-based systems.

Internationally, two main directions have emerged. First, process-based crop models are used to simulate maize yield responses to climate scenarios and N management, revealing strong interactions between N inputs, soil N dynamics, and climate drivers in both low-input and intensive systems (Falconnier et al., 2020). Second, data-driven approaches, especially machine learning (ML) and deep learning (DL), increasingly predict crop yields from large datasets combining soil, climate, and management information. Systematic reviews show that temperature, rainfall, soil type, soil nutrients, and vegetation indices are among the most frequently used predictors, and that algorithms such as Random Forest (RF), Support Vector Machines, Artificial Neural Networks, CNNs, and LSTMs dominate recent work. For maize specifically, RF models trained on multi-year field trials in Ghana identified soil properties (e.g., organic carbon, total N, exchangeable bases) and maximum

temperature as the most important predictors of yield, surpassing purely climatic models and improving understanding of nutrient-climate interactions (Asamoah et al., 2024). Related studies using RF and other ML algorithms have shown that including both soil and weather variables substantially improves prediction of maize yield under zero N fertilization and in drought-stressed environments. These advances highlight the potential of combining soil nutrient information with climate variables in robust predictive frameworks, but also reveal gaps: many models rely on limited nutrient descriptors, treat climate and soil separately, or focus on short time periods and narrow environments.

Building on this progress, the present study focuses on prediction of maize yield based explicitly on soil nutrient status and climate variables, aiming to better capture their joint effects. The main research contents are: (1) construction of a comprehensive feature set describing soil nutrients (e.g., N, P, K, organic matter, pH and related properties) and key climate factors (temperature, precipitation, radiation, humidity) relevant to maize growth; (2) development and comparison of data-driven yield prediction models, with emphasis on ensemble methods such as Random Forest and other ML/DL techniques that have shown strong performance in crop yield prediction; and (3) quantitative analysis of variable importance and interaction patterns between soil nutrients and climate variables, to identify critical drivers of yield variation and potential leverage points for management. The technical route begins with data collection and preprocessing, including quality control and normalization of soil and climate data. Next, the dataset is split into training and testing subsets, and multiple candidate models are trained, tuned, and evaluated using metrics such as coefficient of determination (R^2) and root mean square error (RMSE), following best practices from recent ML yield-prediction studies. Finally, model interpretation techniques (e.g., variable importance analysis and partial response analysis) are applied to quantify how specific combinations of soil nutrients and climate variables influence predicted maize yield, providing both a practical prediction tool and theoretical insight for nutrient management and climate adaptation strategies.

Across diverse environments, maize yield is jointly controlled by soil nutrient status and climate conditions, and their interaction largely determines both productivity and stability. While process-based models and ML/DL approaches have advanced yield prediction, there remains a need for models that explicitly integrate detailed soil nutrient descriptors with key climate variables and provide interpretable guidance for management. This study addresses that gap by constructing and evaluating data-driven maize yield prediction models grounded in soil-climate interactions, aiming to support more precise fertilization, risk management, and climate-smart maize production.

2 Analysis of Factors Influencing Maize Yield

2.1 Mechanism of soil nutrients on maize growth

Maize yield is jointly controlled by soil nutrient supply and climate conditions throughout the growing season. Understanding how these drivers act individually and in combination is essential for reliable yield prediction and targeted management. Adequate N, P, and K fertilization strongly enhances maize growth traits such as plant height, leaf area, cob number, and grain weight, which together raise biomass accumulation and grain yield by large margins compared with unfertilized controls (Kaleri et al., 2026). Long-term NPK application improves key soil properties-including soil organic carbon and available N, P, and K-which in turn explain a larger share of yield variation than phenological factors in the North China Plain (Wang et al., 2024).

Nutrient deficiency, especially of nitrogen and phosphorus, markedly reduces yield and dry matter accumulation in maize-based systems (Sun et al., 2024). Under N, P, or K deficiency, maize root growth and activity are inhibited, and hundreds of genes related to nutrient transport, hormones, and transcription factors are differentially expressed, indicating complex molecular regulation of root adaptation to low nutrient supply (Nana et al., 2020).

2.2 Effects of climate factors on maize yield

Temperature, precipitation, drought, and vapor pressure deficit (VPD) strongly shape maize yield anomalies at regional to global scales. Temperature-related extremes generally show stronger associations with yield deviations than precipitation alone, although irrigation can partially buffer high-temperature damage (Figure 1) (Vogel et al., 2019). In Northeast China, compound drought and heat cause greater yield loss than either stress alone, with

warm-dry years producing the largest reductions and yield loss increasing with temperature and VPD but decreasing with precipitation (Li et al., 2021).

Beyond extremes, the balance between atmospheric evaporative demand and soil moisture is critical. Including interactions between VPD and root-zone soil moisture greatly improves statistical prediction of maize yield anomalies, and estimates that ignore soil moisture can overstate climate-induced yield damage by about a factor of two. Similar work in China shows that maize benefits only when atmospheric moisture demand and soil moisture remain in relative balance; accounting for soil moisture halves projected yield losses compared with using atmospheric demand alone (Zhao et al., 2023).

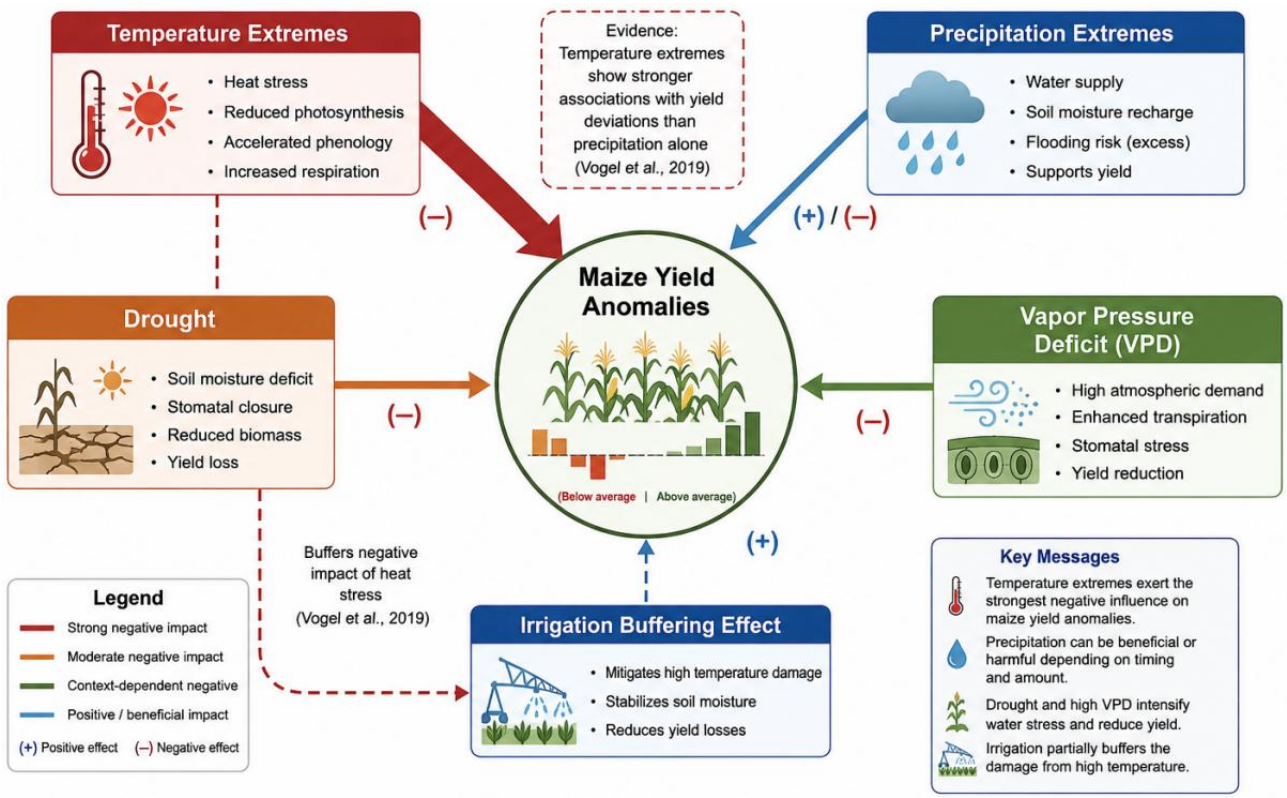


Figure 1 Climate extreme drivers of maize yield anomalies at regional to global scales

2.3 Synergistic mechanism of soil and climate factors

Soil fertility and climate interact to determine both average yield and its stability over time. Long-term experiments show that balanced NPK fertilization not only raises mean maize yield but also improves the stability of relative yield anomalies, while models that combine climate variables with nutrient status explain far more variation in yield anomalies than climate alone (Zhu et al., 2024). In diverse maize systems, soil moisture and temperature jointly drive yield damage, and predictions that include both components outperform those relying only on temperature and precipitation, underscoring the tight soil-climate coupling.

Nitrogen supply particularly modulates maize sensitivity to climate change. In low-input systems, higher N fertilization increases the crop's responsiveness to elevated CO_2, higher temperatures, and altered rainfall, making intensively managed maize more sensitive-and thus more climatically risky-than low-input maize (Falconnier et al., 2020). At larger scales, management intensification (including improved nutrients and technologies) accounts for most historical yield gains, but its benefits are increasingly constrained by warming and drought, meaning that future intensification must explicitly incorporate climate adaptation to sustain yield trends (Medina and Tian, 2023).

Maize yield depends on sufficient N, P, and K to build a productive canopy and reproductive sink, while climate factors-especially temperature extremes, drought, VPD, and soil moisture-govern year-to-year variability. Nutrient

management alters both climate sensitivity and yield stability, so predictive models and management strategies must jointly consider soil fertility and climate interactions rather than treating them in isolation.

3 Data Sources and Overview of the Study Area

3.1 Natural and agricultural conditions of the study area

The major maize-producing regions of northern and northeastern China are characterized by temperate monsoon climates with distinct growing seasons, where temperature, precipitation, and sunshine jointly determine maize climate suitability at different phenological stages (Wang et al., 2024). In the Northeast, relatively cooler temperatures and variable rainfall make precipitation a key limiting factor, while temperature plays a stronger role in the suitability index than in more southerly zones. In contrast, the Huang-Huai-Hai (3H) region has warmer average temperatures and generally higher comprehensive climate suitability, although spatial differences in precipitation and sunshine still create heterogeneous yield potentials. Across China's broader maize belt, temperature variability and climate perturbations can cause substantial yield losses, especially under warming, but these impacts are spatially heterogeneous (Chen et al., 2024).

Soil conditions in the maize belt range from high-soil organic carbon (SOC) soils in parts of the Northeast to more degraded or compacted soils in other regions, and these differences strongly affect yield responses to climate. High SOC, favorable texture, and adequate field capacity enhance buffering capacity against adverse temperature and moisture perturbations, stabilizing yields under climate variability (Feng et al., 2022). In contrast, soils with higher bulk density, coarser texture, or lower water-holding capacity tend to amplify yield losses under warming, underscoring the importance of soil improvement for resilient production. Regional tillage practices, such as deep ploughing or conservation tillage, also interact with local climate: in cooler sites, practices that improve early-season soil temperature and water availability promote maize emergence and growth, whereas in warmer, windier areas, systems that enhance water retention and aeration can be more beneficial (Qian et al., 2025).

3.2 Data sources and acquisition methods

Maize yield data and associated environmental variables can be obtained from long-term field trials, experimental stations, and statistical records, often at plot or county scales. Multi-year experiments in Northeast China and the North China Plain provide detailed measurements of yield, phenology, and management, suitable for evaluating soil-climate interactions and model performance. In some studies, plot-scale experiments under different fertilization or tillage systems supply yield and soil measurements across contrasting climate conditions, enabling analysis of management impacts on yield and soil properties (Meng et al., 2021; Qian et al., 2025). For broader regional coverage, station networks combining agronomic records with local weather observations support large-scale assessments of yield responses to climate variability and soil attributes.

Climate data are typically derived from ground-based automatic weather stations and gridded meteorological datasets, providing variables such as temperature, precipitation, radiation, humidity, and derived indices (e.g., heat degree days, consecutive dry days) during key growth stages (Dandrifosse et al., 2024; Wang et al., 2025). Remote sensing products supply complementary environmental information, including vegetation indices, land surface temperature, and solar-induced fluorescence that capture canopy status over time. Soil data come from field sampling, regional soil surveys, and derived soil property databases, covering SOC, texture, bulk density, water-holding capacity, and nutrient indicators. In advanced yield-prediction frameworks, these multi-source datasets-yield, weather, soil, and remote sensing-are integrated into unified databases for machine learning or crop model applications.

3.3 Data preprocessing and quality control

Prior to model construction, environmental and yield data require systematic preprocessing to ensure completeness and consistency. Weather station data are screened for missing values, range violations, and temporal or spatial inconsistencies, often using automated quality-control algorithms tailored to agricultural decision needs. Such systems flag implausible measurements-e.g., unrealistic temperature sequences, saturated relative humidity at too low values, or anomalous rainfall series-enabling early detection and correction or removal of erroneous records. For gridded or satellite-based climate products, temporal aggregation (e.g., daily to

monthly) and calculation of growing-season indices are performed to match crop growth stages and modeling time steps. Yield and management records are checked for outliers, coding errors, and inconsistent units across years and locations to avoid bias in training datasets (Archontoulis et al., 2020).

Remote sensing and soil datasets also undergo substantial preprocessing. For optical satellite data, procedures include cloud and shadow masking, compositing, and noise reduction to generate consistent vegetation index and land-surface-temperature time series suitable for yield prediction (Li et al., 2022). Novel image-cleaning techniques, such as quartile-based filtering of local pixel neighborhoods, can reduce sensor noise and atmospheric artifacts, improving the signal-to-noise ratio and enhancing model accuracy when combined with deep learning approaches. Soil property and nutrient data from field sampling or databases are harmonized across sources, interpolated or matched to field or grid units, and normalized or standardized for use in machine learning models that combine soil, climate, and management predictors (Diaz-Gonzalez et al., 2022). Overall, rigorous preprocessing and quality control across all data types are essential to ensure robust, interpretable relationships between soil nutrients, climate variables, and maize yield.

4 Construction and Selection of Feature Variables

4.1 Construction of soil nutrient indicator system

A scientific soil nutrient indicator system should reflect both the supply of key macronutrients and the broader edaphic conditions that control maize response. Long-term omission experiments identify available and total N, P, and K, soil organic carbon, C:N and N:P ratios as primary determinants of yield and nutrient use efficiency, showing that edaphic indicators explain more yield variation than phenological factors in maize systems (Wang et al., 2024). Meta-analysis in northern China further supports including soil organic matter, total N, and available P and K as core indicators, because these properties consistently increase under rational fertilization and are closely aligned with yield gains and water use efficiency (Jiang et al., 2024).

For predictive modeling, soil indicators must also capture spatial heterogeneity and nutrient limitations. Maize nutrient omission trials across 324 farmers' fields in the Eastern Indo-Gangetic Plains showed that soil pH was the most critical variable controlling relative N- and P-limited yields, while soil N and Zn strongly influenced Zn-limited yield (Figure 2) (Ahmed et al., 2024). Post-harvest soil test value prediction equations for N, P, and K demonstrate how pre-sowing soil tests, crop uptake, and fertilizer inputs can be combined to estimate dynamic soil nutrient status, supporting targeted fertilizer recommendations for subsequent crops (Abdel-Salam et al., 2024).

4.2 Extraction of climate variable features

Climate feature construction should represent both mean conditions and stress events during sensitive growth stages. Studies that assessed the relevance of climatic attributes for corn yield found that solar radiation, precipitation, vapor pressure, and maximum and minimum temperature are among the most influential variables, with radiation slightly exceeding precipitation in importance in Neotropical environments (Sierra-Forero et al., 2024). Regional analyses that combine multiple climate time series with yield records confirm that temperature- and water-related indicators together explain a large share of yield variability, especially when evaluated over the growing season (Luthra et al., 2024).

Careful temporal aggregation and transformation of climate variables can greatly improve prediction. Monthly vapor pressure deficit and precipitation expressed with spline functions produced the "best climate-only" model for rainfed corn, with high out-of-sample R^2, and adding satellite vegetation indices further enhanced performance (Li et al., 2019). Similar work on climate-driven yield variability uses downscaled temperature, precipitation, and shortwave radiation, plus extreme-climate indices, to quantify how mean growing-season warming, radiation changes, and counts of hot or dry days affect maize yield projections (Chen et al., 2020).

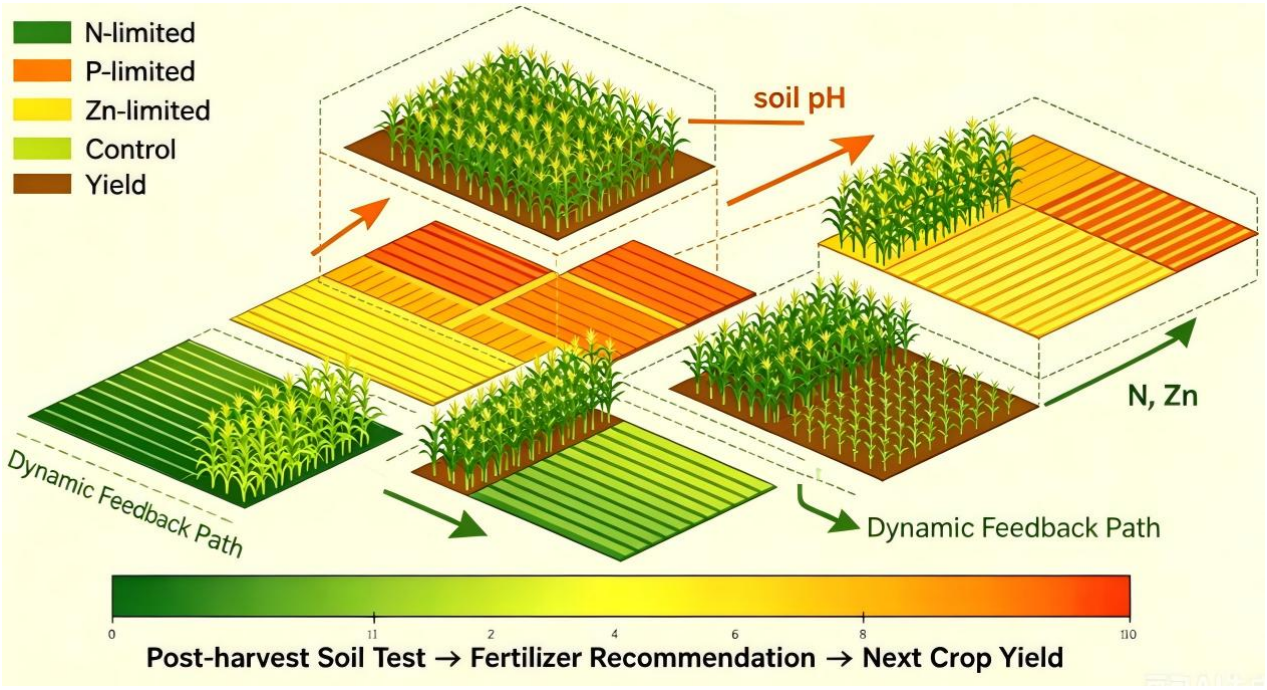


Figure 2 Spatial heterogeneity of soil nutrient limitations and their effects on maize yield

4.3 Feature selection and dimensionality reduction methods

High-dimensional soil-climate datasets require effective feature selection (FS) to avoid overfitting and reduce computational cost. Reviews of machine-learning yield models emphasize that optimal feature sets, obtained by FS, are essential because only a subset of soil, climate, and management variables truly drive prediction accuracy (Hara et al., 2021). In a dedicated framework for yield prediction, a Relief-based FS step was combined with linear discriminant analysis feature extraction, before applying machine-learning classifiers, which markedly improved accuracy over models using all raw variables (Gupta et al., 2022).

Comparative studies of dimensionality reduction for crop yield forecasting show that combining FS and feature extraction (FX) can outperform either alone. In rice yield models based on vegetation and temperature indices, a hybrid approach (FSX) integrating FS with principal component-type FX improved RMSE by up to 60% relative to using all features, and FSX-based models outperformed pure FS or FX in most regions (Pham et al., 2022). More recent works in crop yield prediction apply hybrid FS pipelines (e.g., correlation-based filters, ANOVA, ensemble FS) coupled with advanced learners such as XGBoost or optimized SVR, consistently reporting higher predictive accuracy and lower error once redundant and noisy predictors are removed.

5 Methods for Prediction Model Construction

5.1 Traditional statistical modeling methods

Traditional statistical methods for yield prediction are mainly based on linear or polynomial relationships between yield and a limited set of explanatory variables, often weather indices. Multiple linear regression and its variants have long been used as benchmarks when comparing newer machine learning approaches for maize and other crops, typically using growing-season temperature and precipitation plus a time trend to represent technological progress (Leng and Hall, 2020). Extensions such as quadratic, interaction, and polynomial regression have also been applied to maize and other cereals, and can achieve reasonable accuracy when relationships are approximately linear and the number of predictors is small (Shastry et al., 2017).

More recent work has introduced penalized regression techniques (LASSO, Elastic Net, ridge), which perform variable selection and effectively handle multicollinearity among many weather indices (Vashisth and Aravind, 2026). For maize in semi-arid New Delhi, Elastic Net outperformed stepwise multiple linear regression across vegetative, flowering, and grain-filling stages, with the lowest RMSE and normalized RMSE, highlighting the

value of shrinkage and regularization when many daily weather variables are used. Similar comparisons for rice show that penalized regressions can rival or exceed traditional stepwise regression, though they may still lag behind flexible non-linear models such as neural networks under highly complex climate-yield relationships (Satpathi et al., 2023).

5.2 Machine learning modeling methods

Machine learning (ML) methods such as Random Forest (RF), Support Vector Regression, and boosted trees have become central to crop yield prediction because they capture non-linear responses and interactions between soil, climate, and management variables without strict parametric assumptions. For maize, RF has been shown to outperform multiple linear regression at regional and global scales, reducing RMSE from 14-49% of mean yield with linear models to 6-14% with RF, and better reproducing spatial patterns of yield (Jeong et al., 2016). In the U.S. Midwest, a comparative study using Lasso, Support Vector Regressor, RF, and XGBoost with hundreds of environmental features found that XGBoost was the most accurate and stable algorithm for county-level maize yield prediction (Kang et al., 2020).

In some applications, ML models trained on relatively simple climate inputs also perform strongly. For Irish potato and maize in Rwanda, Random Forest using only rainfall and temperature achieved R^2 values of 0.875 and 0.817, respectively, outperforming polynomial regression and Support Vector Regressor and providing practically useful early-season predictions (Kuradusenge et al., 2023). ML has also been used to model silage maize yields from NDVI time-series; boosted regression trees and RF achieved correlations above 0.87, and were less sensitive to inconsistencies in satellite-derived vegetation profiles than conventional regressions (Aghighi et al., 2018). These studies underline the versatility of ML methods for integrating climate, soil, and remote-sensing predictors in maize yield models.

5.3 Deep learning and ensemble learning methods

Deep learning (DL) extends ML by learning complex, hierarchical representations from large, high-dimensional datasets composed of weather, soil, genotype, and remote sensing inputs. A deep neural network trained on thousands of maize hybrid trials across more than 2,000 locations substantially outperformed Lasso, shallow neural networks, and regression trees, reaching an RMSE close to 11-12% of average yield while also supporting feature selection to reduce input dimensionality with minimal accuracy loss (Khaki and Wang, 2019). However, DL does not always dominate: in a U.S. Midwest maize study, LSTM and CNN architectures did not surpass XGBoost, suggesting that tabular environmental datasets may not always benefit from image- or sequence-oriented deep architectures (Kang et al., 2020).

Ensemble learning combines multiple base learners to improve robustness and accuracy. For corn in the U.S. Corn Belt, CNN-DNN ensembles created via bagging and stacking outperformed ensembles of linear regression, Lasso, RF, XGBoost, and LightGBM, explaining about 77% of spatio-temporal yield variation with an RMSE of 866 kg/ha (Shahhosseini et al., 2021). Hybrid and ensemble DL frameworks that fuse convolutional, recurrent, and fully connected networks have also shown superior performance for crop yield prediction, with CNN-DNN or CNN-RNN-LSTM structures often exceeding single DL or ML models and achieving R^2 values near or above 0.85 in case studies (Oikonomidis et al., 2022). Deep ensemble approaches thus offer a promising route for integrating multi-source soil, climate, and remote-sensing data to achieve robust maize yield prediction under variable environments.

6 Model Training and Evaluation System

6.1 Dataset partitioning and validation strategies

A reasonable partition of the maize yield dataset is the basis for constructing reliable prediction models. In most supervised learning settings, data are divided into training, validation, and test subsets so that model fitting, hyperparameter tuning, and final performance assessment can be clearly separated and avoid information leakage (Bischl et al., 2021). When the number of yearly observations is small, directly reserving an independent test set

becomes difficult, and specialized cross-validation (CV) schemes such as leave-one-out (LOO) or nested CV are recommended to obtain unbiased generalization estimates (Dinh and Aires, 2022).

For crop yield prediction with strong spatial and temporal dependence, the choice of CV strategy affects both apparent skill and interpretability. Studies using simulated or observed yields show that random CV can give overly optimistic accuracy when neighboring samples are highly correlated, while spatial or cluster-based CV provides more realistic estimates on held-out regions (Radočaj et al., 2025). Nested CV or nested leave-two-out schemes further separate inner folds for model selection from outer folds for performance estimation, preventing overly complex models from being chosen and improving transferability across years and locations (Sweet et al., 2023).

6.2 Model parameter optimization methods

Hyperparameters of machine learning models, such as the number of trees in random forests or learning rates in gradient boosting, strongly influence predictive performance and must be tuned systematically rather than by ad-hoc trial-and-error (Bischl et al., 2021). Classical search strategies include grid search and random search, which evaluate candidate configurations on resampling-based performance estimates, but they become inefficient as the hyperparameter space grows.

More advanced approaches treat hyperparameter tuning as a black-box optimization problem and use probabilistic surrogate models. Bayesian optimization with Gaussian processes or related surrogates iteratively proposes promising configurations based on past evaluations and has been shown to find better settings than random search under comparable budgets (Wu et al., 2019). In crop yield estimation, Bayesian optimization frameworks applied to tree-based models such as LightGBM achieve high coefficients of determination and low mean squared error across several agricultural datasets, demonstrating the gains from automated hyperparameter optimization. Random forest-specific tuning via model-based optimization (e.g., tuning mtry, node size, sample size) can further increase accuracy over default settings while controlling runtime (Probst et al., 2018).

6.3 Model evaluation indicator system

Because maize yield prediction is a regression problem, a comprehensive indicator system is needed to evaluate both accuracy and explanatory power. Error-based metrics such as root mean square error (RMSE), mean absolute error (MAE), and related deviations are widely used in crop model evaluation because they directly characterize the magnitude of prediction errors in yield units (Yang et al., 2014). RMSE is particularly sensitive to large errors and is appropriate when error distributions are approximately Gaussian, whereas MAE provides a more robust and interpretable measure of average error and is less influenced by outliers (Chai and Draxler, 2014).

To complement absolute error measures, goodness-of-fit and efficiency statistics assess how much of the observed variance is explained by the model. The coefficient of determination (R^2) is often preferred as a standard metric in regression because it relates performance to the variance of ground-truth yields and is more informative than stand-alone error magnitudes in many applications (Chicco et al., 2021). In process-based crop modeling, additional indices such as modeling efficiency (EF) and the index of agreement (d) are used alongside RMSE and MAE to provide a balanced view of model bias, dispersion, and agreement with observations (Yang et al., 2014). For maize yield prediction models based on soil nutrients and climate variables, combining R^2 (or EF) with RMSE and MAE yields a robust evaluation framework that captures both accuracy and reliability across different environments.

7 Case Study: Empirical Analysis of Regional Maize Yield Prediction

7.1 Study area and sample construction

In many recent maize yield prediction studies, the study area is defined to capture both environmental gradients and management diversity so that models generalize beyond a single field or season. For example, plot-scale work integrates multi-year trials under contrasting fertilizer systems, combining climate, soil, and satellite data to represent heterogeneous growing conditions across years and treatments (Meng et al., 2021). Similar multi-farm designs in Western Australia aggregate yield monitor data from thousands of hectares over several seasons, then

collocate each observation with soil, terrain, and weather variables to form a dense spatio-temporal sample set (Filippi et al., 2019).

Large-area studies, such as county-level maize analyses in the US Midwest or regional work in Northeast China, construct samples by merging official yield statistics with gridded or station-based climate data, soil maps, and multi-source satellite products (Figure 3) (Kang et al., 2020; Li et al., 2022). In Ghana, plot-level samples from hundreds of maize field trials are georeferenced and linked to 0-30 cm soil properties, climate variables during the planting season, and management practices, enabling model training across wide environmental and agronomic ranges (Asamoah et al., 2024).

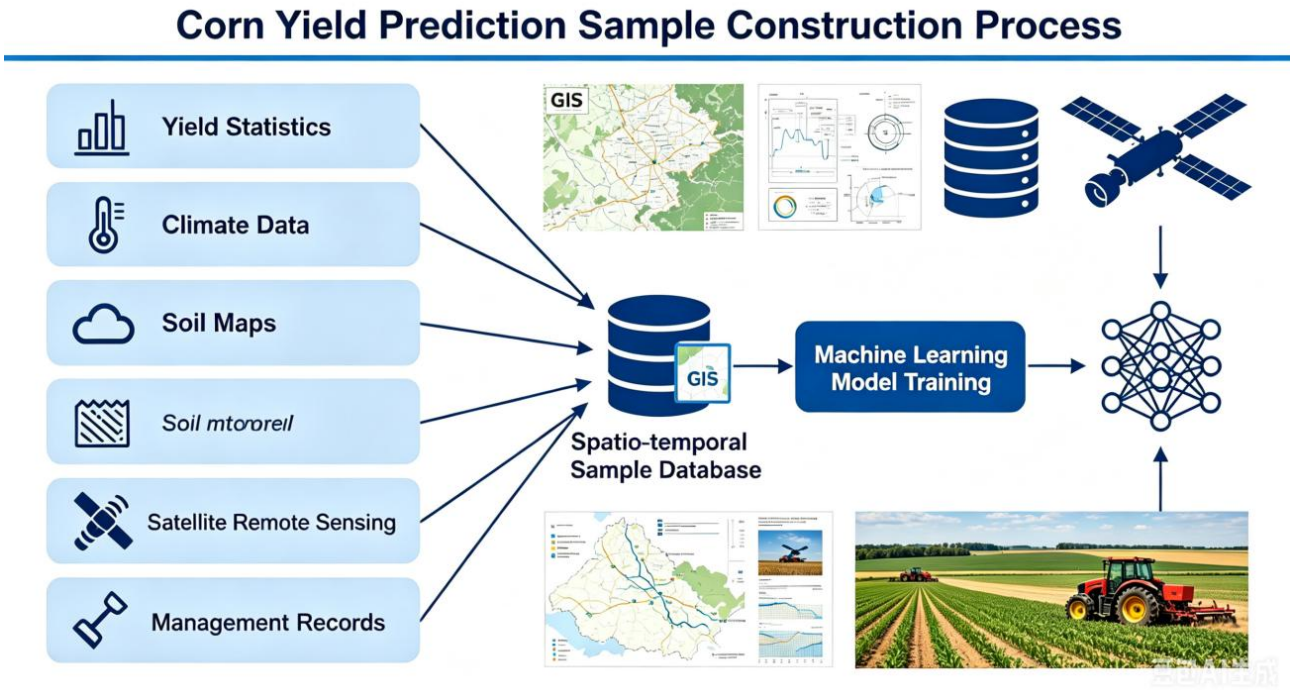


Figure 3 Workflow for integrating multi-source environmental and agricultural datasets into maize yield prediction samples

7.2 Comparative Analysis Of Multi-Model Prediction Results

Comparative studies consistently show that model performance depends strongly on algorithm choice and input richness. At the plot scale, combining vegetation indices, climate, soil, and fertilizer data, Random Forest and Adaptive Boosting clearly outperform linear regression, SVM, GPR, and KNN, with R^2 often above 0.85 and lowest RMSE values (Meng et al., 2021). In a Hungarian field using detailed spatio-temporal soil and micro-relief measurements, XGBoost surpassed neural and kernel methods, reaching test accuracies above 95%, while lattice-based smoothing further improved predictive AUC (Nyéki et al., 2021).

At regional scales, ensemble or tree-based machine learning models generally outperform both traditional regression and deep learning architectures. In the US Midwest, XGBoost provided the most accurate and stable county-level maize forecasts when hundreds of environmental features were used, while LSTM and CNN did not show clear advantages (Kang et al., 2020). Across Northeast China, an ensemble of several ML methods improved yield prediction over individual linear and ML models when integrating environmental and multi-sensor satellite data, explaining more than 70% of maize yield variability (Li et al., 2022).

7.3 Result validation and agricultural application analysis

Robust validation is essential to ensure that multi-model predictions have practical value. Studies highlight that naïve random data splits can substantially overestimate predictive skill, especially when the goal is true forecasting rather than interpolation within a season (Morales and Villalobos, 2023). More rigorous schemes, such as nested k-fold cross-validation across years and fields, or leave-one-field/leave-one-year-out designs, better

reflect operational performance and were used, for instance, in multi-farm machine-learning models and in Ghanaian RF models for maize yield and agronomic efficiency (Filippi et al., 2019; Asamoah et al., 2024).

When rigorously validated, yield prediction models support several agricultural applications. Plot-scale maize models that accurately forecast yield under different fertilizer systems enable assessment of input strategies and refinement of site-specific recommendations before harvest (Meng et al., 2021). Large-area models that integrate climate, soil, and satellite indicators have been used for early-season yield forecasting, outperforming official forecasts and providing actionable information for logistics, market planning, and food-security assessments (Li et al., 2022). Such applications demonstrate how reliable maize yield prediction, grounded in soil-climate interactions, can inform precision fertilization, risk management, and regional policy decisions.

8 Results Analysis and Discussion

8.1 Contribution analysis of soil nutrient variables

Feature-importance and interpretable ML studies highlight that specific soil nutrients can dominate maize yield responses, even in data-rich settings. In a data-intensive farm management trial, Random Forest analysis showed that urea application was consistently the most critical variable for explaining spatial yield variation, with soil phosphorus, pH, clay content, sodium and plant population also among the leading contributors in different seasons (Maseko et al., 2024). This indicates that both applied N and inherent soil fertility properties jointly control yield in high-resolution, within-field prediction. Similar work in precision agriculture, using RF and other models on over 145,000 corn and soybean yield observations, found that soil test P, K, Zn, soil organic matter and cation exchange capacity were key predictors, underscoring the strong explanatory power of nutrient and related soil indicators for yield variation at sub-field scales (Burdett and Wellen, 2022).

Under nutrient-limited conditions, omission trials combined with AutoML provide a more explicit decomposition of nutrient contributions. In 324 nutrient omission plot trials across ten agroecological zones in the Eastern Indo-Gangetic Plains, stack-ensemble and deep learning models predicted relative nutrient-limited yields with low RMSE, and permutation importance identified soil pH as the dominant variable controlling N- and P-limited yields (Ahmed et al., 2024). The same analysis showed that soil N and Zn strongly influenced Zn-limited yield, while spatial trends in K-limited yield emerged along an east-west gradient, revealing distinct fertility constraints for different nutrients. These findings suggest that soil nutrient variables-especially applied N, soil P, Zn, pH and texture-related properties-provide high marginal gains in predictive power and are indispensable components of maize yield models based on soil-climate interactions.

8.2 Influence weight analysis of climate variables

Across diverse modeling frameworks, climate variables frequently emerge as the largest single contributors to interannual maize yield variability. A global meta-analysis using 68 simulation studies for wheat, maize and rice showed that maximum temperature and precipitation significantly affected yield responses, with yields declining by 4.21% per 1 °C increase in maximum temperature but increasing by 0.43% per 1% rise in precipitation (Qin et al., 2023). This quantitative gradient highlights the high negative weight of heat stress and the compensating effect of adequate rainfall in crop-climate response functions. At the global scale, mixed-effects models updating projected yield responses under CMIP6 scenarios indicate that temperature-related stress is a dominant driver of future maize yield losses, with projected global maize declines around 22% by late century under high emissions if adaptation is limited (Li et al., 2025).

Machine-learning-based attribution provides more detailed rankings of individual climate indicators. A hybrid GGCM-Random Forest framework for China's maize belt found that chilling days, drought indicators and crop pests/diseases were the main factors influencing projected maize yield changes, with relative importance quantified via RF partial-dependence analysis (Li et al., 2023). In a separate process-based and ML study on wheat under future climate scenarios, precipitation explained most yield variability in mid-century high-emission conditions, whereas maximum temperature became the dominant limiting factor under later, more strongly warmed scenarios (El-Mahroug et al., 2025). For site-specific maize prediction with spatio-temporal XGBoost models, precipitation during the juvenile growth phase (May) was identified as the single most important factor

over five years, followed by soil pH, clay content, electrical conductivity and NDVI, again emphasizing the high influence weight of water-related variables alongside key soil properties.

8.3 Discussion on model applicability and uncertainty

The applicability of soil-nutrient- and climate-based yield models depends critically on how uncertainty is handled across space, time and scenario conditions. A recent meta-analysis of crop yield responses to projected climate change combined mixed-effects modeling with block bootstrapping to partition uncertainty arising from model structure, climate projections (CMIP6) and emissions pathways, showing that simple pooled OLS tends to underestimate yield losses and under-represent uncertainty ranges (Li et al., 2025). Similarly, a crop-model and ML ensemble for maize and soybean across China demonstrated that coupling GGCMs with Random Forest greatly improved correlation (r up to 0.77 for maize) and reduced normalized RMSE, while variance decomposition revealed that the dominant uncertainty source shifted from crop models in the baseline GGCM runs to global climate models and then scenarios as projections extended further into the century (Li et al., 2023). These results imply that model applicability under future climates requires explicit accounting for structural, climate and scenario uncertainties rather than relying on single-model projections.

Transferability across domains and scales introduces additional uncertainty dimensions for data-driven yield models. Domain-adaptation work on maize in the US Corn Belt, using DANN, KLIEP and RTNN, found that models trained in temperate regions with medium-high growing degree days and moderate vapor pressure deficit generalized well, whereas strong dependence on vegetation indices (GCI) reduced transferability when source and target domains had limited overlap (Priyatikanto et al., 2023). Independent evaluations of cross-validation strategies in UAV-based yield prediction further showed that random CV can substantially overestimate performance when models are applied outside their training spatial domain, whereas spatial or leave-one-field-out CV and simpler, regularized models gave more realistic extrapolation accuracy (Habibi et al., 2024). Together with county-scale ensemble studies that link large prediction errors to low cropland ratios and extreme weather events (Sajid et al., 2022), these findings stress that robust maize yield prediction demands careful validation design, domain-aware training, and transparent uncertainty quantification before models are applied for management or policy decisions in new regions or under novel climate conditions.

9 Conclusions and Future Research Directions

Existing studies confirm that integrating soil nutrients, soil physical properties, and climate variables can explain a substantial share of maize yield variability across diverse agroecological zones. Soil indicators such as nitrogen fertilizer rate, soil organic carbon, pH, bulk density, and exchangeable bases consistently emerge among the most influential predictors, often exceeding the importance of individual climate variables for yield prediction in tropical and semi-arid environments. At the same time, temperature, rainfall, and related weather indices remain key drivers of interannual variation, especially when combined with management and genotype information in large datasets. From a modeling perspective, tree-based and boosting algorithms (Random Forest, XGBoost, Gradient Boosting) generally outperform linear methods and many deep architectures for maize yield prediction using soil-climate feature sets. Meta-modeling of process-based simulations and large empirical trial datasets shows that these methods can achieve relative errors around 10-15% when sufficient training samples and well-designed features are available. Systematic reviews across maize and other crops further indicate that these algorithms are among the most frequently adopted and robust options, particularly when coupled with feature engineering and multimodal data integration.

High-accuracy soil-climate yield models provide actionable information for fertilizer management and nutrient efficiency. In Ghana, Random Forest and XGBoost models trained on long-term maize trials successfully predicted both yield and agronomic efficiency, highlighting nitrogen rate, rainfall, and key soil properties as dominant management levers. Such models support the design of site-specific recommendations that can raise productivity while reducing the environmental costs of blanket fertilizer application. Similar ML-process-model hybrids using APSIM outputs demonstrate that meta-models can rapidly explore genotype-environment-management scenarios for preseason planning. At larger scales, integrating soil maps,

meteorological series, and satellite indicators enables early-season forecasts that outperform conventional statistical baselines and even some official forecasts. County-level yield prediction in the U.S. Midwest has shown that XGBoost models using hundreds of environmental features can provide reliable maize forecasts several months before harvest, improving on models based only on basic weather or historical yields. Reviews of precision agriculture emphasize that such predictive systems contribute to resource optimization, risk management, and food-security planning by linking sensing technologies, big data platforms, and advanced analytics into operational decision support tools.

Despite these advances, several limitations constrain the reliability and transferability of current soil-climate yield models. Studies comparing algorithms against simple baselines show that, under realistic forecasting setups using ordered train-test splits, ML models sometimes offer only modest gains over farm-level average yields, especially when weather forecast errors are ignored. Systematic reviews also highlight persistent challenges with obtaining high-quality, harmonized datasets on soil nutrients, management, and high-resolution yields, which can limit model generalization across regions and seasons. In addition, many models are trained and validated under random data partitioning, leading to over-optimistic performance estimates for true out-of-sample prediction. Future research directions point toward hybrid, transferable, and explainable frameworks. Hybrid models that couple process-based crop simulators with ML or deep learning have improved accuracy and reduced uncertainty in semi-arid maize systems, particularly when fusing remote sensing, climate, and soil information. Domain adaptation and transfer-learning approaches, including partial adversarial networks, are beginning to address domain shifts between ecological zones and could substantially improve cross-regional maize yield prediction. Reviews stress the need for standardized data protocols, interpretable architectures (e.g., SHAP- or XAI-enhanced models), and scalable, crop-agnostic pipelines so that soil nutrient and climate-based yield prediction can be robustly embedded in precision agriculture and sustainability strategies.

Acknowledgments

We would like to thank the anonymous reviewers for their detailed review of the draft. Their specific feedback helped us correct the logical loopholes in our arguments.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdel-Salam M., Kumar N., and Mahajan S., 2024, A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning, Neural Computing and Applications, 36: 20723-20750.
https://doi.org/10.1007/s00521-024-10226-x

Aghighi H., Azadbakht M., Ashourloo D., Shahrabi H., and Radiom S., 2018, Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11: 4563-4577.
https://doi.org/10.1109/jstars.2018.2823361

Ahmed Z., Krupnik T., Timsina J., Islam S., Hossain K., Kurishi A., Emran S., Harun-Ar-Rashid M., McDonald A., and Gathala M., 2024, Prediction of spatial heterogeneity in nutrient-limited sub-tropical maize yield: implications for precision management in the eastern indo-gangetic plains, Artificial Intelligence in Agriculture, 12: 1-15.
https://doi.org/10.1016/j.aiia.2024.08.001

Archontoulis S., Castellano M., Licht M., Nichols V., Baum M., Huber I., Martinez-Feria R., Puntel L., Ordóñez R., Iqbal J., Wright E., Dietzel R., Helmers M., Vanloocke A., Liebman M., Hatfield J., Herzmann D., Córdova S., Edmonds P., Togliatti K., Kessler A., Danalatos G., Pasley H., Pederson C., and Lamkey K., 2020, Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt, Crop Science, 60: 721-738.
https://doi.org/10.1002/csc2.20039

Asamoah E., Heuvelink G., Chairi I., Bindraban P., and Logah V., 2024, Random forest machine learning for maize yield and agronomic efficiency prediction in Ghana, Heliyon, 10: e37065.
https://doi.org/10.1016/j.heliyon.2024.e37065

Bischl B., Binder M., Lang M., Pielok T., Richter J., Coors S., Thomas J., Ullmann T., Becker M., Boulesteix A., Deng D., and Lindauer M., 2021, Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(2): e1484.
https://doi.org/10.1002/widm.1484

Burdett H., and Wellen C., 2022, Statistical and machine learning methods for crop yield prediction in the context of precision agriculture, Precision Agriculture, 23: 1553-1574.
https://doi.org/10.1007/s11119-022-09897-0

Chai T., and Draxler R., 2014, Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature, Geoscientific Model Development, 7: 1247-1250.
https://doi.org/10.5194/gmd-7-1247-2014

Chen F., Xu X., Chen S., Wang Z., Wang B., Zhang Y., Zhang C., Feng P., and Hu K., 2024, Soil buffering capacity enhances maize yield resilience amidst climate perturbations, Agricultural Systems, 222: 103870.
https://doi.org/10.1016/j.agsy.2024.103870

Chen X., Wang L., Niu Z., Zhang M., Li C., and Li J., 2020, The effects of projected climate change and extreme climate on maize and rice in the Yangtze River Basin, China, Agricultural and Forest Meteorology, 282-283: 107867.
https://doi.org/10.1016/j.agrformet.2019.107867

Chicco D., Warrens M., and Jurman G., 2021, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Computer Science, 7: e623.
https://doi.org/10.7717/peerj-cs.623

Dandrifosse S., Jago A., Huart J., Michaud V., Planchon V., and Rosillon D., 2024, Automatic quality control of weather data for timely decisions in agriculture, Smart Agricultural Technology, 8: 100445.
https://doi.org/10.1016/j.atech.2024.100445

Diaz-Gonzalez F., Vuelvas J., Correa C., Vallejo V., and Patiño D., 2022, Machine learning and remote sensing techniques applied to estimate soil indicators - Review, Ecological Indicators, 135: 108517.
https://doi.org/10.1016/j.ecolind.2021.108517

Dinh T., and Aires F., 2022, Nested leave-two-out cross-validation for the optimal crop yield model selection, Geoscientific Model Development, 15: 3519-3536.
https://doi.org/10.5194/gmd-15-3519-2022

El-Mahroug S., Suleiman A., Zoubi M., Al-Omari S., Abu-Afifeh Q., Al-Jawaldeh H., Alta'any Y., Al-Nawaiseh T., Obeidat N., Alsoud S., Alshoshan A., Al-Shibli F., and Ta'any R., 2025, Predictive modeling of climate-driven crop yield variability using DSSAT towards sustainable agriculture, AgriEngineering, 7(5): 156.
https://doi.org/10.3390/agriengineering7050156

Falconnier G., Corbeels M., Boote K., Affholder F., Adam M., MacCarthy D., Ruane A., Nendel C., Whitbread A., Justes É., Ahuja L., Akinseye F., Alou I., Amouzou K., Anapalli S., Baron C., Basso B., Baudron F., Bertuzzi P., Challinor A., Chen Y., Deryng D., Elsayed M., Faye B., Gaiser T., Galdos M., Gayler S., Gérardeaux E., Giner M., Grant B., Hoogenboom G., Ibrahim E., Kamali B., Kersebaum K., Kim S., Laan M., Leroux L., Lizaso J., Maestrini B., Meier E., Mequanint F., Ndoli A., Porter C., Priesack E., Ripoche D., Sida T., Singh U., Smith W., Srivastava A., Sinha S., Tao F., Thorburn P., Timlin D., Traoré B., Twine T., and Webber H., 2020, Modelling climate change impacts on maize yields under low nitrogen input conditions in sub-Saharan Africa, Global Change Biology, 26: 5942-5964.
https://doi.org/10.1111/gcb.15261

Feng P., Wang B., Harrison M., Wang J., Liu K., Huang M., Liu D., Yu Q., and Hu K., 2022, Soil properties resulting in superior maize yields upon climate warming, Agronomy for Sustainable Development, 42(5): 81.
https://doi.org/10.1007/s13593-022-00818-z

Filippi P., Jones E., Wimalathunge N., Somarathna P., Pozza L., Ugbaje S., Jephcott T., Paterson S., Whelan B., and Bishop T., 2019, An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning, Precision Agriculture, 20(5): 1015-1029.
https://doi.org/10.1007/s11119-018-09628-4

Gupta S., Geetha A., Sankaran K., Zamani A., Ritonga M., Raj R., Ray S., and Mohammed H., 2022, Machine learning- and feature selection-enabled framework for accurate crop yield prediction, Journal of Food Quality, 2022: 6293985.
https://doi.org/10.1155/2022/6293985

Habibi L., Matsui T., and Tanaka T., 2024, Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing, Journal of Agriculture and Food Research, 18: 101096.
https://doi.org/10.1016/j.jafr.2024.101096

Hara P., Piekutowska M., and Niedbała G., 2021, Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data, Land, 10(6): 609.
https://doi.org/10.3390/land10060609

Jeong J., Resop J., Mueller N., Fleisher D., Yun K., Butler E., Timlin D., Shim K., Gerber J., Reddy V., and Kim S., 2016, Random forests for global and regional crop yield predictions, PLoS ONE, 11(6): e0156571.
https://doi.org/10.1371/journal.pone.0156571

Jiang M., Dong C., Bian W., Zhang W., and Wang Y., 2024, Effects of different fertilization practices on maize yield, soil nutrients, soil moisture, and water use efficiency in northern China based on a meta-analysis, Scientific Reports, 14: 57031.
https://doi.org/10.1038/s41598-024-57031-z

Kaleri A., Khanzada B., Rajput W., Bijarani A., Shafqat A., Arain A., Mirbahar S., Jokhio N., Majeedano A., and Majeedano S., 2026, Combined effects of nitrogen, phosphorus, and potassium on maize growth, development, and yield, Jammu Kashmir Journal of Agriculture, 5(3): 297-305.
https://doi.org/10.56810/jkjagri.005.03.0297

Kang Y., Ozdogan M., Zhu X., Ye Z., Hain C., and Anderson M., 2020, Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest, Environmental Research Letters, 15(6): 064005.
https://doi.org/10.1088/1748-9326/ab7df9

Khaki S., and Wang L., 2019, Crop yield prediction using deep neural networks, Frontiers in Plant Science, 10: 621.
https://doi.org/10.3389/fpls.2019.00621

Kim K., and Lee B., 2023, Effects of climate change and drought tolerance on maize growth, Plants, 12(20): 3548.
https://doi.org/10.3390/plants12203548

Kuradusenge M., Hitimana E., Hanyurwimfura D., Rukundo P., Mtonga K., Mukasine A., Uwitonze C., Ngabonziza J., and Uwamahoro A., 2023, Crop yield prediction using machine learning models: Case of Irish potato and maize, Agriculture, 13(1): 225.
https://doi.org/10.3390/agriculture13010225

Leng G., and Hall J., 2020, Predicting spatial and temporal variability in crop yields: An inter-comparison of machine learning, regression and process-based models, Environmental Research Letters, 15(4): 044027.
https://doi.org/10.1088/1748-9326/ab7b24

Li C., Camac J., Robinson A., and Kompas T., 2025, Predicting changes in agricultural yields under climate change scenarios and their implications for global food security, Scientific Reports, 15: 87047.
https://doi.org/10.1038/s41598-025-87047-y

Li E., Zhao J., Pullens J., and Yang X., 2021, The compound effects of drought and high temperature stresses will be the main constraints on maize yield in Northeast China, Science of the Total Environment, 812: 152461.
https://doi.org/10.1016/j.scitotenv.2021.152461

Li L., Zhang Y., Wang B., Feng P., He Q., Shi Y., Liu K., Harrison M., Liu D., Yao N., Li Y., He J., Feng H., Siddique K., and Yu Q., 2023, Integrating machine learning and environmental variables to constrain uncertainty in crop yield change projections under climate change, European Journal of Agronomy, 151: 126917.
https://doi.org/10.1016/j.eja.2023.126917

Li Y., Guan K., Yu A., Peng B., Zhao L., Li B., and Peng J., 2019, Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S., Field Crops Research, 234: 55-65.
https://doi.org/10.1016/j.fcr.2019.02.005

Li Z., Ding L., and Xu D., 2022, Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China, Science of the Total Environment, 806: 152880.
https://doi.org/10.1016/j.scitotenv.2021.152880

Luthra N., Srivastava A., Shahi U., Singh V., Dey P., and Singh A., 2024, Prediction of post-harvest soil nutrient status through multiple linear regression for targeted yield of hybrid maize, Indian Journal of Agronomy, 68(4): 547-553.
https://doi.org/10.59797/ija.v68i4.5471

Maseko S., Van Der Laan M., Tesfamariam E., Delport M., and Otterman H., 2024, Evaluating machine learning models and identifying key factors influencing spatial maize yield predictions in data intensive farm management, European Journal of Agronomy, 160: 127193.
https://doi.org/10.1016/j.eja.2024.127193

Matiu M., Ankerst D., and Menzel A., 2017, Interactions between temperature and drought in global and regional crop yield variability during 1961-2014, PLoS One, 12(5): e0178339.
https://doi.org/10.1371/journal.pone.0178339

Medina H., and Tian D., 2023, Synergistic contributions of climate and management intensifications to maize yield trends from 1961 to 2017, Environmental Research Letters, 18(3): 034021.
https://doi.org/10.1088/1748-9326/acb27f

Meng L., Liu H., Ustin S., and Zhang X., 2021, Predicting maize yield at the plot scale of different fertilizer systems by multi-source data and machine learning methods, Remote Sensing, 13(18): 3760.
https://doi.org/10.3390/rs13183760

Morales A., and Villalobos F., 2023, Using machine learning for crop yield prediction in the past or the future, Frontiers in Plant Science, 14: 1128388.
https://doi.org/10.3389/fpls.2023.1128388

Nyéki A., Kerepesi C., Daróczy B., Benczúr A., Milics G., Nagy J., Harsányi E., Kovács A., and Neményi M., 2021, Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods, Precision Agriculture, 22: 1397-1415.
https://doi.org/10.1007/s11119-021-09833-8

Ocwa A., Harsányi E., Széles A., Holb I., Szabó S., Rátonyi T., and Mohammed S., 2023, A bibliographic review of climate change and fertilization as the main drivers of maize yield: Implications for food security, Agriculture and Food Security, 12(1): 19.
https://doi.org/10.1186/s40066-023-00419-3

Oikonomidis A., Catal C., and Kassahun A., 2022, Hybrid deep learning-based models for crop yield prediction, Applied Artificial Intelligence, 36(1): 2031823.
https://doi.org/10.1080/08839514.2022.2031823

Pham H., Awange J., and Kuhn M., 2022, Evaluation of three feature dimension reduction techniques for machine learning-based crop yield prediction models, Sensors, 22(17): 6609.
https://doi.org/10.3390/s22176609

Priyatikanto R., Lu Y., Dash J., and Sheffield J., 2023, Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation, SSRN Electronic Journal, 1: 1-29.
https://doi.org/10.2139/ssrn.4122021

Probst P., Wright M., and Boulesteix A., 2018, Hyperparameters and tuning strategies for random forest, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3): e1301.
https://doi.org/10.1002/widm.1301

Qian Y., Zhang Z., Jiang F., Wang J., Dong F., Liu J., and Peng X., 2025, Impacts of tillage treatments on soil physical properties and maize growth at two sites under different climatic conditions in black soil region of Northeast China, Soil and Tillage Research, 257: 106471.
https://doi.org/10.1016/j.still.2025.106471

Qin M., Zheng E., Hou D., Meng X., Meng F., Gao Y., Chen P., Qi Z., and Xu T., 2023, Response of wheat, maize, and rice to changes in temperature, precipitation, CO2 concentration, and uncertainty based on crop simulation approaches, Plants, 12(14): 2709.
https://doi.org/10.3390/plants12142709

Radočaj D., Plaščak I., and Jurišić M., 2025, A comparative assessment of regular and spatial cross-validation in subfield machine learning prediction of maize yield from Sentinel-2 phenology, Eng, 6(10): 270.
https://doi.org/10.3390/eng6100270

Satpathi A., Setiya P., Das B., Nain A., Jha P., Singh S., and Singh S., 2023, Comparative analysis of statistical and machine learning techniques for rice yield forecasting for Chhattisgarh, India, Sustainability, 15(3): 2786.
https://doi.org/10.3390/su15032786

Shahhosseini M., Hu G., Khaki S., and Archontoulis S., 2021, Corn yield prediction with ensemble CNN-DNN, Frontiers in Plant Science, 12: 709008.
https://doi.org/10.3389/fpls.2021.709008

Shastry A., Sanjay H., and Bhanusree E., 2017, Prediction of crop yield using regression techniques, International Journal of Computing, 6(5): 1-5.

Sierra-Forero B., Barón-Velandia J., and Vanegas-Ayala S., 2024, Assessment of the relevance of features associated with corn crop yield prediction in Colombia, a country in the Neotropical zone, International Journal of Information Technology, 16: 2129-2138.
https://doi.org/10.1007/s41870-024-01762-9

Sun Z., Yang R., Wang J., Zhou P., Gong Y., Gao F., and Wang C., 2024, Effects of nutrient deficiency on crop yield and soil nutrients under winter wheat-summer maize rotation system in the North China Plain, Agronomy, 14(11): 2690.
https://doi.org/10.3390/agronomy14112690

Sweet L., Müller C., Anand M., and Zscheischler J., 2023, Cross-validation strategy impacts the performance and interpretation of machine learning models, Artificial Intelligence for the Earth Systems, 2(4): e230026.
https://doi.org/10.1175/aies-d-23-0026.1

Vashisth A., and Aravind K., 2026, Maize yield estimation at different growth stage using weather variables by LASSO, elastic net and stepwise multiple linear regression techniques, Scientific Reports, 16: 34239.
https://doi.org/10.1038/s41598-025-34239-1

Vogel E., Donat M., Alexander L., Meinshausen M., Ray D., Karoly D., Meinshausen N., and Frieler K., 2019, The effects of climate extremes on global agricultural yields, Environmental Research Letters, 14(5): 054010.
https://doi.org/10.1088/1748-9326/ab154b

Wang N., Ai Z., Zhang Q., Leng P., Qiao Y., Li Z., Tian C., Cheng H., Chen G., and Li F., 2024, Impacts of nitrogen (N), phosphorus (P), and potassium (K) fertilizers on maize yields, nutrient use efficiency, and soil nutrient balance: Insights from a long-term diverse NPK omission experiment in the North China Plain, Field Crops Research, 317: 109616.
https://doi.org/10.1016/j.fcr.2024.109616

Wang X., Li X., Lou Y., You S., and Zhao H., 2024, Refined evaluation of climate suitability of maize at various growth stages in major maize-producing areas in the North of China, Agronomy, 14(2): 344.
https://doi.org/10.3390/agronomy14020344

Wang Y., Shen Y., Yu S., Zhang X., and Xiao D., 2025, Climate extremes are critical to maize yield and will be severer in North China, Climate Risk Management, 47: 100710.
https://doi.org/10.1016/j.crm.2025.100710

Wu J., Chen X., Zhang H., Xiong L., Lei H., and Deng S., 2019, Hyperparameter optimization for machine learning models based on Bayesian optimization, Journal of Electronic Science and Technology, 17(1): 26-40.
https://doi.org/10.11989/jest.1674-862x.80904120

Yang J., Yang J., Liu S., and Hoogenboom G., 2014, An evaluation of the statistical methods for testing the performance of crop models with observed data, Agricultural Systems, 127: 81-89.
https://doi.org/10.1016/j.agsy.2014.01.008

Zhao F., Wang G., Li S., Hagan D., and Ullah W., 2023, The combined effects of VPD and soil moisture on historical maize yield and prediction in China, Frontiers in Environmental Science, 11: 1117184.
https://doi.org/10.3389/fenvs.2023.1117184

Zhu W., Rezaei E., Sun Z., Wang J., and Siebert S., 2024, Soil-climate interactions enhance understanding of long-term crop yield stability, European Journal of Agronomy, 160: 127386.
https://doi.org/10.1016/j.eja.2024.127386

Computational Molecular Biology 2026, Vol.16, No.2, 114-128
http://bioscipublisher.com/index.php/cmb

Feature Review Open Access

Statistical Analysis of Yield Components in Wheat under Different Management Practices

Guoping Yang [1,2]
1 Hangzhou Xiaoshan Daozhihong Family Farm, Hangzhou, 311200, Zhejiang, China
2 Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China
Corresponding author: 869187101@qq.com
Computational Molecular Biology, 2026, Vol.16, No.2 doi: 10.5376/cmb.2026.16.0009
Received: 13 Feb., 2026
Accepted: 20 Mar., 2026
Published: 03 Apr., 2026

Preferred citation for this article:
Yang G.P., 2026, Statistical analysis of yield components in wheat under different management practices, Computational Molecular Biology, 16(2): 114-128 (doi: 10.5376/cmb.2026.16.0009)

Abstract Wheat yield formation is a complex physiological process jointly regulated by genetic traits, environmental conditions, and agricultural management practices. This study systematically investigates the effects of different management strategies on wheat yield components, including spike number, grains per spike, and thousand-grain weight. By integrating multiple management scenarios such as fertilization intensity, irrigation regimes, and planting density adjustments, the responses of yield formation processes were analyzed in terms of growth dynamics, component interactions, and regional variability. The results indicate that fertilization primarily influences spike development and grain setting, while water availability significantly regulates biomass accumulation and yield stability. Planting density further modulates population structure, leading to trade-offs among yield components. Significant coupling relationships were observed among spike number, grain number, and grain weight, suggesting a coordinated but competitive allocation mechanism. Statistical modeling revealed that management practices exert both direct and indirect effects on final yield through yield component mediation. Moreover, regional analysis highlights that climatic and soil conditions amplify or constrain management effectiveness. The findings provide a comprehensive understanding of how integrated agronomic practices shape wheat yield formation and offer theoretical support for optimizing high-yield and stable production systems under diverse agroecological conditions.
Keywords Wheat yield components; Management practices; Fertilization; Irrigation; Yield modeling

1 Introduction

Wheat is a major source of calories and protein worldwide, and further yield gains are essential to meet rising demand for food and feed. Grain yield in wheat is a complex quantitative trait shaped by genotype, environment, and their interaction, making direct selection on yield alone inefficient. A clearer understanding of yield components and how they respond to management practices is therefore critical for designing agronomic strategies and statistical models that improve both yield level and stability. This paper, “Statistical Analysis of Yield Components in Wheat under Different Management Practices,” is positioned within this context. Wheat yield is typically decomposed into spike number per unit area, grain number per spike, and thousand-grain weight, with additional supporting traits such as biological yield and harvest index (Dolijanović et al., 2025). Numerous correlation and path-analysis studies show that grain yield is positively associated with spikes per area, grains per spike, and thousand-grain (or kernel) weight, as well as biological yield and harvest index (Choudhary et al., 2025). Path coefficient analyses repeatedly identify biological yield, grains per spike, and harvest index as having strong direct effects on grain yield, indicating their value as selection or diagnostic traits. Large multi-environment analyses further confirm that grains per unit area, determined by spike number and grains per spike, is the yield component most tightly linked to final yield (Slafer et al., 2022).

Agricultural management practices-especially fertilization, irrigation, tillage, sowing arrangement, and organic amendments-modify resource availability and crop environment, thereby altering yield components rather than yield directly. Meta-analyses and long-term trials show that integrating nitrogen management with irrigation, tillage, and organic inputs can increase wheat yield, nitrogen-use efficiency, and water-use efficiency relative to single-factor optimization. Conventional or well-structured tillage systems often improve grain weight per spike,

thousand-grain weight, grain uniformity, and protein content compared with reduced or no-tillage systems, partly via effects on soil structure, weed pressure, and nutrient dynamics (Ahmadi et al., 2024). Sowing method and organic nutrient management (e.g., raised beds, split farmyard manure with liquid organics) can also enhance spike number, grain yield, and soil biological activity (Sharma et al., 2024). Fertilizer form and biostimulants (e.g., phosphorus, humic acids, mycorrhizae) further influence grains per spike, thousand-grain weight, biological yield, and harvest index, demonstrating multiple pathways from management to yield components and then to total grain yield.

Despite abundant evidence on individual factors, comparatively fewer studies jointly quantify how different management practices reshape the correlation structure and direct versus indirect effects among yield components. This motivates a statistical analysis of yield components in wheat under contrasting management regimes. The central research questions can be framed as: (i) How do key yield components (spikes per area, grains per spike, thousand-grain weight, biological yield, harvest index) respond to different management practices? (ii) How do management-induced changes in these components translate, via correlation and path relationships, into changes in grain yield? (iii) Do specific management practices strengthen the contribution of particular components (e.g., grains per spike or harvest index) to yield?

Based on prior correlation and path-analysis work, the study can hypothesize that: (1) management practices that enhance biological yield, spike number, grains per spike, and harvest index will significantly increase grain yield; (2) the relative importance of numerical components (spikes per area, grains per spike, thousand-grain weight) in determining yield will differ across management regimes; and (3) integrated or optimized management will not only raise yield but also modify the strength and direction of correlations among yield components, revealing management-specific yield-formation pathways. Statistical tools such as correlation, path analysis, and multivariate methods offer an appropriate framework to test these hypotheses and to identify the most responsive and yield-determining components under different management practices.

2 Regulatory Effects of Different Management Practices on Wheat Growth and Development

2.1 Effects of fertilization intensity on vegetative and reproductive growth

Nitrogen fertilization strongly modifies the hierarchy and plasticity of wheat yield components, with early-formed traits such as tiller and spike number responding differently from later traits like grain number and grain weight (Paolo et al., 2022). Increasing N rate up to about 150-300 kg·N·ha^{-1} enhances grain number, grain weight, straw biomass, and plant height, although responses are curvilinear and context dependent (Yokamo et al., 2023). The timing of N also matters: delaying N from early tillering to stem elongation or later tends to reduce total yield but can increase grain weight, reflecting a trade-off between grain number and grain size.

Vegetative photosynthetic capacity and biomass accumulation are also sensitive to N intensity. Moderate to high N rates increase leaf area index, chlorophyll content, and flag-leaf photosynthesis, which in turn support higher total dry matter and reproductive organ biomass (Noor et al., 2023). Under partial shading, appropriate N can compensate for reduced light by boosting photosynthetic efficiency and dry matter transfer to ears, but under heavy shading, the regulatory effect of N on vegetative growth and yield formation becomes limited (Hongzhi et al., 2021). Meta-analysis shows that N use efficiency declines at high N and on fertile soils, implying that excessive fertilization may increase biomass but not proportionally increase grain yield.

2.2 Impacts of water availability on population structure regulation

Water supply around jointing and heading governs tiller survival and spike formation, thereby shaping population structure. Supplemental irrigation that wets the soil to 0-20 cm at jointing reduces tiller mortality, increases productive spike number, and improves leaf photosynthesis of both main stems and tillers, while deeper irrigation layers mainly increase transpiration and reduce water-use efficiency without clear yield gains (Shang et al., 2020). Deficit irrigation schemes that combine moderate water inputs with suitable planting patterns can raise tiller number, spikelets per spike, grains per spike, and radiation use efficiency, particularly when water is applied at both jointing and heading (Zhou et al., 2020).

Water availability also interacts with density and N to determine canopy architecture and uniformity. Under limited irrigation, increasing seeding density can maintain grain yield while markedly improving water productivity by boosting spike number per unit area and canopy apparent photosynthesis in upper layers (Figure 1) (Gao et al., 2021). Optimal combinations of irrigation and N (e.g., irrigation at jointing and anthesis with moderate N) increase spike number, grains per spike, leaf area index, and canopy photosynthesis, while a well-distributed root system supports better extraction of soil water and coordinates root-shoot balance (Wang et al., 2025). Conversely, overly sparse or heterogeneous water distribution in drip systems can create non-uniform subpopulations with reduced leaf area, biomass, and panicle number in disadvantaged rows, lowering overall yield (Jing et al., 2023).

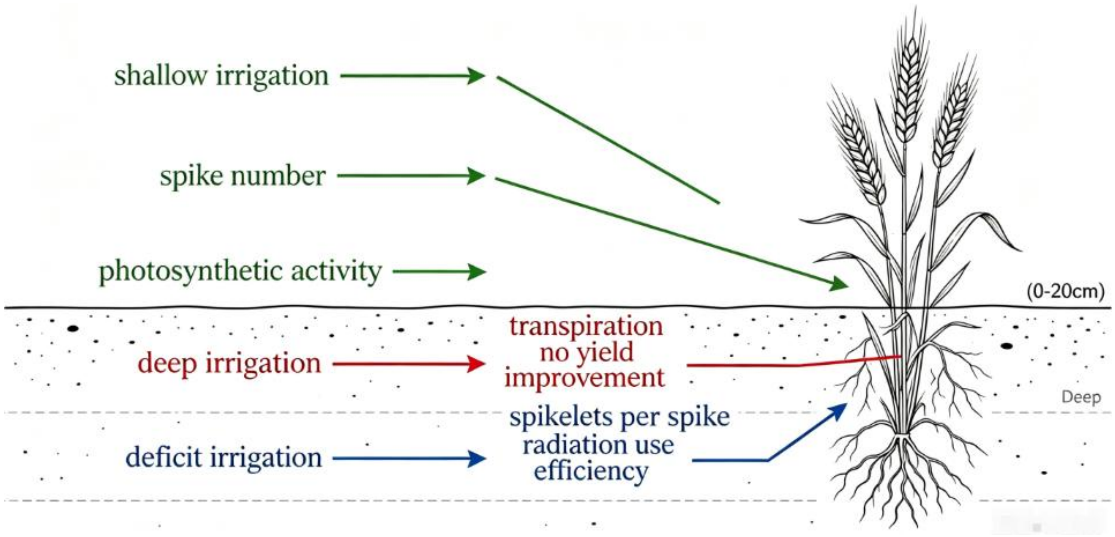


Figure 1 Mechanistic effects of irrigation depth and timing on tiller survival, spike formation, and yield component development in wheat

2.3 Response mechanisms of planting density and crop competition

Planting density regulates the balance between individual plant growth and population yield through intraspecific competition. As density increases, competition intensity rises, leading to reduced tillering, individual biomass, and grain number per plant, but greater spike number per area and, up to an optimum, higher population yield. In high-resource environments, winter wheat can reach near-maximum yields at surprisingly low plant densities, indicating strong compensatory capacity via tillering and spike fertility when competition is relaxed (Lollato et al., 2024). Reviews and field trials show that excessive seeding ultimately depresses yield components and grain yield due to self-thinning and resource limitation, underscoring the need for an optimum rather than maximal density (Arshad et al., 2025).

Competition also drives physiological and structural adjustments. High density often increases shoot elongation while reducing leaf mass per area, tiller number, and per-stem biomass, reflecting shade-avoidance and resource partitioning among competing plants. Breeding trajectories show decreasing shoot competitiveness indices over time, consistent with selecting genotypes that cooperate better at high density by limiting aggressive competitive traits (Manntschke et al., 2025). Experiments combining density with N show that raising density while moderating N can favor superior tillers, optimize spike number, and improve nitrogen-use efficiency, whereas very high N at high density mainly inflates vegetative growth and reduces partial factor productivity (Yang et al., 2019). Together, these mechanisms illustrate that planting density, competition, and genotype jointly determine how yield components respond under intensive management.

3 Dynamic Patterns of Wheat Yield Component Formation

3.1 Temporal characteristics of spike number formation

Spike number in wheat is largely determined by the production and survival of tillers over a defined developmental window from early vegetative stages to anthesis. Studies show that most fertile spikes originate

from tillers initiated early (from three-leaf stage to jointing), whereas late-initiated tillers produce few kernels and contribute minimally to yield (Tilley et al., 2019). The timing of tiller initiation and cessation, together with tiller mortality between jointing and anthesis, governs final spike density and explains large genotypic differences in spike number. Genetic variation in traits such as tillering onset, duration, and survival indicates substantial scope to manipulate these temporal patterns for yield improvement (Xie et al., 2015).

Temporal dynamics of tillering interact with canopy signals and resource competition. Low red:far-red light ratios arising as canopies close tend to accelerate the end of tillering and promote tiller abortion, thereby fixing spike number earlier in genotypes that are more sensitive to shading (Xie et al., 2015). Management and environment modify these dynamics by altering water status, radiation interception, and assimilate availability during the tillering window. Under post-jointing drought, ear-bearing capacity and seed setting of specific tiller positions are highly sensitive to the exact spike developmental stage at which stress occurs, with tillers at intermediate positions showing the greatest reductions in kernels per spike and grain yield per spike (Lin et al., 2020). These findings highlight that both the calendar time and developmental time at which stresses or inputs occur are critical in determining final spike number.

3.2 Limiting factors in grain number formation

Grain number per unit area is jointly determined by spike number and grains per spike, and both components follow a similar pattern of over-production of reproductive structures followed by intense abortion. Detailed observations of tiller and floret primordia show that survival, rather than initiation, is the primary driver of variation in spike number per area and grains per spike, especially during the late reproductive phase when degeneration is most intense (Bicego et al., 2024). Resource availability around stem elongation and heading strongly affects the survival of initiated tillers and florets, making grain number highly plastic in response to shading, thinning, and other changes in assimilate supply.

Within spikes, floret fertility and grain set are constrained by floral degradation and developmental timing. In hexaploid wheat, visible floral degradation from green-anther stage to anthesis strongly influences maximum floret primordia, fertile florets, and final grain number per spikelet, while detillering delays this degradation and increases the number of fertile florets and grains. The spatial gradient along the spike also reflects developmental limitations; basal spikelets are developmentally delayed and therefore exhibit higher floret abortion and more rudimentary spikelets, even when assimilate distribution along the spike is relatively uniform (Backhaus et al., 2023). Ovary size at anthesis emerges as a key integrative trait: larger ovaries, especially in distal florets, are associated with higher floret and grain survival, linking pre-anthesis growth with post-anthesis grain set (Guo et al., 2016).

3.3 Environmental response mechanisms of thousand-grain weight formation

Thousand-grain weight (TGW) is mainly determined during the grain filling period through the interplay between grain filling rate and duration. Field experiments manipulating post-anthesis night temperature show that warming of about 4 °C from 10 days after anthesis to maturity shortens the effective grain filling period, reducing TGW by roughly 3% per degree while leaving grain filling rate largely unchanged (Garcia et al., 2016). Similar work under controlled heat stress indicates that high temperatures hasten physiological maturity and decrease final grain weight, with genotypic differences in heat tolerance closely associated with the capacity to sustain a high grain filling rate (Dias and Lidon, 2009). These responses reflect sink-limited grain growth where accelerated development truncates the time over which potential grain size can be realized.

Broader analyses of climate-related factors confirm that elevated temperature and drought generally reduce grain yield of C3 cereals by depressing TGW. Meta-analysis shows that thousand-grain weight is particularly sensitive to warming, whereas drought and heat together can substantially reduce grain filling and starch accumulation even when grain number is less affected (Mariem et al., 2021). At the crop scale, variation in TGW can be explained by differences in post-anthesis thermal regime and radiation, as well as water availability that maintains photosynthesis and nitrogen metabolism during grain filling (Ru et al., 2023). New remote-sensing approaches, such as UAV-based estimation of grain filling rate and TGW from canopy traits, emphasize that TGW integrates

environmental effects on leaf area, chlorophyll, and biomass during the filling period and can therefore serve as a sensitive indicator of how management practices buffer or amplify climatic stresses.

4 Yield Variations Driven by Management Practices

4.1 Yield responses under different fertilization strategies

Nitrogen rate and timing strongly regulate wheat yield by altering spike number, grain number, and grain weight. Split spring N in winter wheat mainly increased spikes·m^{-2}, with yields peaking when 100 kg·N·ha^{-1} was applied early (BBCH 22-25) and 40 kg·ha^{-1} at stem elongation, reflecting the importance of early N for productive shoot survival (Lachutta and Jankowski, 2024). Across small grains, grain number per unit area is the key driver of yield, and high N (e.g., 100 kg·N) markedly increases grain number and grain yield, although trade-offs with grain weight can occur (Mirosavljević et al., 2025).

Yield responses to N show clear optima rather than linear increases. In semiarid Loess Plateau conditions, higher N rates (e.g., 210 kg·ha^{-1}) maximized yield in wet years, while intermediate N (150 kg·ha^{-1}) was optimal in normal or dry years, indicating precipitation-dependent N demand (Ren et al., 2021). Detailed dose-response studies further show that 210-240 kg·N·ha^{-1} can maximize spike number, grains per ear, thousand-grain weight and grain yield, whereas excessive N (300 kg·ha^{-1}) reduces spike grains and does not improve yield (Qu et al., 2025).

4.2 Yield differences under varying irrigation conditions

Irrigation strategy and amount substantially modify yield level and resource efficiency. Under sprinkler irrigation, full conventional irrigation (CI100) gave the highest grain yield, while a slight reduction (CI75) maintained high yield and increased water use efficiency, indicating that moderate deficit can save water with limited yield loss (Alghory and Yazar, 2018). In drip-fertigated systems, deficit irrigation at 75% ETc combined with moderate N (170 kg·ha^{-2}) produced the highest yields and water- and N-use efficiency, with most of the yield gain attributed to N but a sizeable portion to irrigation (Lu et al., 2021).

Timing of supplemental irrigation is also critical. In North China Plain field trials, irrigation at jointing and anthesis optimized root distribution, post-anthesis dry matter accumulation, and grain filling, increasing grain yield and water use efficiency compared with no irrigation or excessive frequency (Figure 2) (Feng et al., 2023). In Mediterranean durum wheat, two irrigations at flowering and grain filling raised grain yield by about 19-46% and increased thousand-kernel weight relative to rainfed conditions, demonstrating strong benefits of relieving post-flowering drought (Mohammadi, 2024).

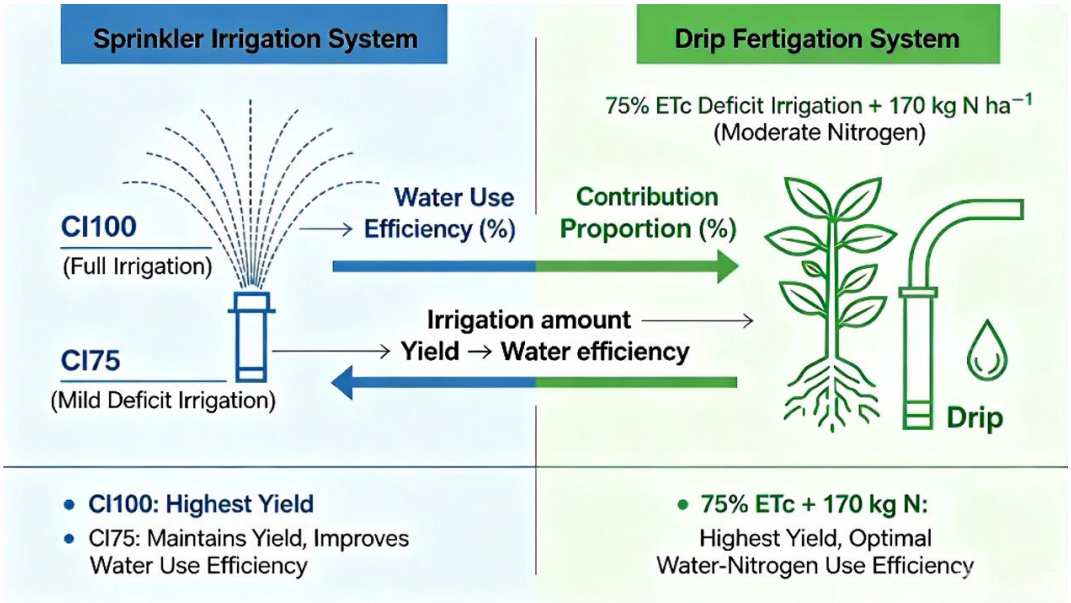


Figure 2 Comparative effects of sprinkler irrigation and drip-fertigated deficit irrigation on wheat grain yield and resource-use efficiency

4.3 Integrated effects of planting density and sowing date adjustment

Planting density and sowing date jointly shape yield by balancing spike number, grains per spike, and grain weight. In northeastern Poland, increasing sowing density from 200 to 400 live grains m^{-2} raised spikes·m^{-2} and grain yield, while delaying sowing by 14-28 days slightly reduced grains per spike but increased thousand-grain weight, with maximum yields achieved under mid-September to early October sowing and higher densities (Lachutta and Jankowski, 2024). A broader multi-genotype study found sowing date had a stronger impact on yield components than plant density, though density strongly correlated with heading time and tillering pattern (Kiss et al., 2018).

Fine-tuning density with N and sowing time can mitigate yield loss in sub-optimal windows. Under rice-wheat rotation, late sowing reduced spikes and kernels, but higher planting densities combined with a 25% N increase (300 kg·N·ha^{-1}) restored yield, with optimal densities depending on whether sowing was delayed 10 or 20 days (Tian et al., 2024). Other semi-winter wheat trials report that sowing around early-mid October at densities near 450 × 10^4 plants ha^{-1} maximizes yield, with higher densities mainly increasing spike number and lower densities favoring thousand-grain weight (Chen et al., 2021).

5 Coupling Relationships among Yield Components

5.1 Trade-off between spike number and grains per spike

The relationship between spike number per unit area and grains per spike is typically antagonistic, reflecting resource limitations during pre-anthesis development. Large multi-environment analyses show that increases in grains per·m^2 are usually achieved either via more spikes per·m^2 or more grains per spike, but simultaneous maximization of both components is rare (Xie and Sparkes, 2021). Compensation between these two routes appears as a hierarchy of plasticity, where spike number acts as a coarse regulator of grains per·m^2 and grains per spike fine-tune the response when resources or genotypes change.

Long-term trial data in winter wheat confirm that the strongest negative correlation among yield components often occurs between spike number and grains per spike, illustrating a robust trade-off in many environments. Yet cultivars differ in how strictly they follow this relationship; some lines deviate positively, combining relatively high spike number with high grains per spike and thus partially escaping compensation (Mandea et al., 2019). Recent work on grain number plasticity further suggests that overlapping developmental periods for tiller and floret mortality can generate feedback between spike number and grains per spike, so that resource shifts during stem elongation alter both components simultaneously (Bicego et al., 2024).

5.2 Synergistic and compensatory effects between grain number and grain weight

Across studies, grain yield is more tightly related to grains per·m^2 than to average grain weight, and negative correlations between grain number components and thousand-grain weight (TGW) are frequently reported. Genome-wide association in tetraploid wheat indicates that many loci for kernel number per spike (KNS) and TGW show opposing phenotypic effects, so gains in one component are often partially offset by losses in the other (Mangini et al., 2018). This genetic antagonism underpins the classic trade-off where increases in grain number reduce average grain size, limiting yield progress.

Nonetheless, both physiological and genetic evidence show that the trade-off is not absolute and can be mitigated. Doubled haploid populations derived from parents contrasting in grain number and weight have produced transgressive segregants with high grain number and relatively large grains, achieving very high yields and reducing the usual compensation. At the molecular level, some manipulations of expansin expression increased grain size without reducing grain number, boosting yield per spike and demonstrating that targeted modification of grain growth can overcome the typical negative association between grain number and grain weight in specific backgrounds (Calderini et al., 2020; Vicentin et al., 2024).

5.3 Mechanisms of multi-factor interactions affecting final yield

Final yield reflects not only pairwise trade-offs among components but also multi-factor interactions among management variables such as nitrogen, water, and density. Drip-fertigation experiments show that irrigation and

nitrogen affect yield mainly through indirect effects on spike density, grains per spike, and TGW, with structural equation modeling confirming that both inputs operate via these components rather than directly on yield (Lu et al., 2021). Optimal combinations (e.g., deficit irrigation with moderate N) maximized grain yield and water-nitrogen use efficiency by increasing spike number and grains per spike while avoiding excessive reductions in grain weight.

More complex factorial trials combining planting pattern, supplementary irrigation, and plant density demonstrate significant three-way interactions on grain yield and resource-use efficiencies. Ridge-furrow planting with plastic mulch, moderate density, and limited supplemental irrigation increased grain yield and water productivity by improving soil water use, effective panicle number, and population N uptake, while maintaining a favorable balance between spike number and spike size (Dai et al., 2023). Other studies highlight that optimal density-nitrogen-water regimes improve spike number in upper and middle canopy layers and population uniformity, thereby enhancing biomass accumulation and yield at the population scale despite reduced per-stem grain weight at very high densities (Gao et al., 2025). These findings indicate that carefully tuned multi-factor management can exploit component plasticity, reduce deleterious trade-offs, and shift yield formation toward more synergistic combinations of grain number and grain weight.

6 Statistical Models and Yield Driving Mechanisms

6.1 Statistical characteristics of relationships between management practices and yield

Across contrasting environments, management practices such as nitrogen rate and irrigation schedule alter wheat yield primarily through effects on spike number, grain number per spike and grain weight. Under drought in Saudi Arabia, multivariate procedures identified spikes·m^{-2}, 100-grain weight, grain weight per spike and biological yield as the most influential variables for grain yield, highlighting how stress conditions sharpen the importance of key yield components (Leilah and Al-Khateeb, 2005). Under semi-arid conditions, variance analysis and regression likewise emphasized number of grains spike^{-1}, spikes·m^{-2} and thousand-kernel weight as main contributors to yield differences among durum genotypes (Frih et al., 2021).

Management-focused meta-analyses and field trials provide quantitative benchmarks for nitrogen-water interactions. A global synthesis showed that nitrogen addition increased wheat grain yield by about 15%, with 100-200 kg·N·ha^{-1} generally optimal for yield, protein and water productivity, and responses modulated by soil texture and climate. In the North China Plain, combined meta-analysis and short-term field experiments indicated that higher nitrogen rates with deficit irrigation improved yield, nitrogen use efficiency and water use efficiency, with an optimal 7:3 inorganic-organic fertilizer ratio under moderate irrigation.

6.2 Regression-based analysis of yield component drivers

Classical and stepwise regression consistently converge on a small set of yield drivers. Under drought, multiple linear and stepwise regressions showed that models including spikes·m^{-2}, grain number spike^{-1} and 100-grain weight explained up to ~93% of grain yield variation, underlining their dominant predictive value (Fouad, 2018). A semi-arid durum study similarly found that grains spike^{-1}, spikes·m^{-2} and thousand-kernel weight significantly explained yield variation in multiple and stepwise regression frameworks (Frih et al., 2021).

Recent work extends regression to farm-level management datasets. Using 22 agronomic and management traits from 90 farms, stepwise regression in Fars province selected nine variables-principally spikes·m^{-2}, grains spike^{-1} and thousand-grain weight, but also spike and awn traits, herbicide use, maturity time and soil salinity-as a parsimonious set, with partial least squares regression achieving $R^2 \approx 0.85$ using only these inputs (Behpouri et al., 2023). Under late-sown conditions, another regression study showed that seven traits (biological yield, harvest index, grain weight per spike, flag leaf length, main spike weight, spikelets per spike and grain appearance) explained ~98% of yield variability, underscoring the value of integrating both structural and physiological traits in yield forecasting (Solanki et al., 2024).

6.3 Multivariate integrated models and contribution decomposition

Multivariate approaches such as path analysis, factor analysis and PCA decompose direct and indirect

contributions of components to yield. In bread wheat, path analysis frequently reveals strong direct effects of spike weight and thousand-kernel weight on grain yield, with grain filling rate and spike number exerting important indirect effects, suggesting these as efficient selection and management targets (Matković-Stojšin et al., 2018; Elmassry and Shal, 2020). Another path-analytic study found that spikelet number and thousand-seed weight, followed by grain size and grain number spike^{-1}, had the largest direct impacts on yield, while multicollinearity diagnostics confirmed that treating all traits as first-order predictors was statistically valid.

Integrated multivariate models have also been used under specific management and climate scenarios. Under drought, combining factor analysis, regression, PCA and clustering confirmed that spikes·m^{-2}, 100-grain weight, grain weight per spike and biological yield form a core determinant set for yield, with these variables loading heavily on principal components associated with productivity (Leilah and Al-Khateeb, 2005). In a Mediterranean irrigation-nitrogen trial, factor analysis grouped variables into yield/water use, yield components and quality factors, while stepwise multivariate regression showed that water footprint indices could be well predicted from NDVI measured at key growth stages, linking spectral signals to integrated yield and resource-use outcomes under different irrigation and nitrogen strategies (Tomaz et al., 2021).

7 Regional Variation in Yield Response Characteristics

7.1 Differences in management responses across climatic zones

Wheat yield responses to management vary strongly among climatic zones because temperature, precipitation, and radiation contribute differently to yield variation across regions. In China, combined changes in mean temperature, precipitation, and solar radiation explain substantial regional differences in yield, with radiation and precipitation often being the dominant drivers and their joint effects exceeding those of any single factor (Han et al., 2023). Similar regional patterns emerge when extreme temperature indicators are used: extreme growing-degree days and other thermal indices cause larger proportional yield losses in northern and spring-wheat regions than in southern winter-wheat zones, even when precipitation increases (Han et al., 2025).

Management practices modify these climate-driven patterns and partially buffer yield gaps. In Mediterranean and MENA agro-ecological zones, simulations show that optimal supplemental irrigation and nitrogen rates, together with adjusted sowing dates, can raise attainable yields by 30-50% and improve water productivity by up to 70%, despite projected 18-30% climate-induced yield declines (Tita et al., 2025). On-farm analyses in the U.S. central Great Plains further reveal that regional clustering by climate is needed, because fertilizer management (N, P, S), fungicide use, and cultivar choice interact with local weather to determine realized yield and yield gaps (Jaenisch et al., 2021).

7.2 Regulatory effects of soil types on yield formation

Soil physical and chemical properties exert strong regulatory effects on wheat yield components and yield gaps. At field scale, small-scale variation in soil texture and organic carbon in both topsoil and subsoil explains nearly half of the spatial variability in grain yield and key components such as tiller density, with higher clay in topsoil enhancing yield but higher clay in subsoil reducing it (Groß et al., 2023). Across arid and semi-arid fields, soil organic carbon, total nitrogen, and available potassium are positively associated with grain number, spike traits, plant height, and both economic and biological yields, indicating that improved physicochemical status narrows yield gaps (Bagheripour et al., 2024).

Management inputs that change soil structure and organic matter further regulate yield formation. A global meta-analysis shows that increasing soil organic carbon up to about 2% is generally associated with higher wheat yields, with diminishing returns beyond this threshold and substantial potential to reduce nitrogen fertilizer needs where SOC is currently low (Oldfield et al., 2018). At plot scale, adding organic residues such as composted bagasse, manure, or straw improves aggregate stability, infiltration, and bulk density, leading to progressive increases in grain and stubble yields as application rates rise (Barzegar et al., 2002).

7.3 Stability comparison under varying hydrothermal conditions

Hydrothermal variability-interacting heat and water conditions-strongly shapes yield stability, with marked

genotype × environment interactions. Multi-environment trials under irrigated, drought, and heat-stress conditions show significant variance for genotype, environment, and their interaction, and AMMI/GGE analyses identify specific genotypes that win in irrigated, drought, or heat environments, as well as a subset that combines high mean yield with broad stability across all stress scenarios (Ram et al., 2020). Similar AMMI/GGE analyses in irrigated versus terminal heat-stress environments indicate that some elite lines are specifically adapted to heat, while others show both above-average yield and high stability across contrasting moisture and temperature regimes.

Beyond individual trials, site conditions determine the magnitude of hydrothermal yield penalties. In Germany, combined heat-drought indices during the reproductive phase have the highest explanatory power for yield loss, with poor sites (low soil quality, lower precipitation) suffering two- to three-fold larger reductions than high-quality, high-rainfall locations under comparable stress (Riedesel et al., 2024). Meta-environment analyses of bread and durum wheat under normal, heat, and drought conditions also reveal complex genotype responses in quality traits and micronutrients, yet identify genotypes that maintain yield and nutritional stability across stress environments.

Across climatic zones, management responses in wheat are tightly conditioned by regional temperature, radiation, and precipitation regimes, with optimized water-nutrient-sowing strategies substantially narrowing yield gaps. Soil type and its physicochemical status regulate yield formation by controlling yield components and mediating the benefits of management inputs and organic amendments. Under varying hydrothermal conditions, both genotype choice and site quality govern yield stability, emphasizing the need to combine statistical G×E analysis with site-specific soil and climate information when designing management practices.

8 Case Study: Comparative Analysis of Typical Management Systems on Yield Structure

8.1 Characteristics of high-input intensive management systems

High-input intensive wheat systems are typically defined by full irrigation and relatively high nitrogen (N) rates, designed to maximize grain yield, grain protein and water productivity. Global and regional meta-analyses show that N addition generally increases grain yield by about 15% and water productivity by around 10%, with optimal responses often at 100-200 kg·N·ha^{-1} under humid or irrigated conditions (Wang et al., 2023). In arid zones with drip or micro-sprinkler irrigation, full evapotranspiration replacement combined with high N rates (e.g., 238 kg·N·ha^{-1}) increased grain yield, biological yield and seed index, and raised crop water productivity by more than 20% relative to lower N inputs (Figure 3) (Abdelrhman et al., 2025).

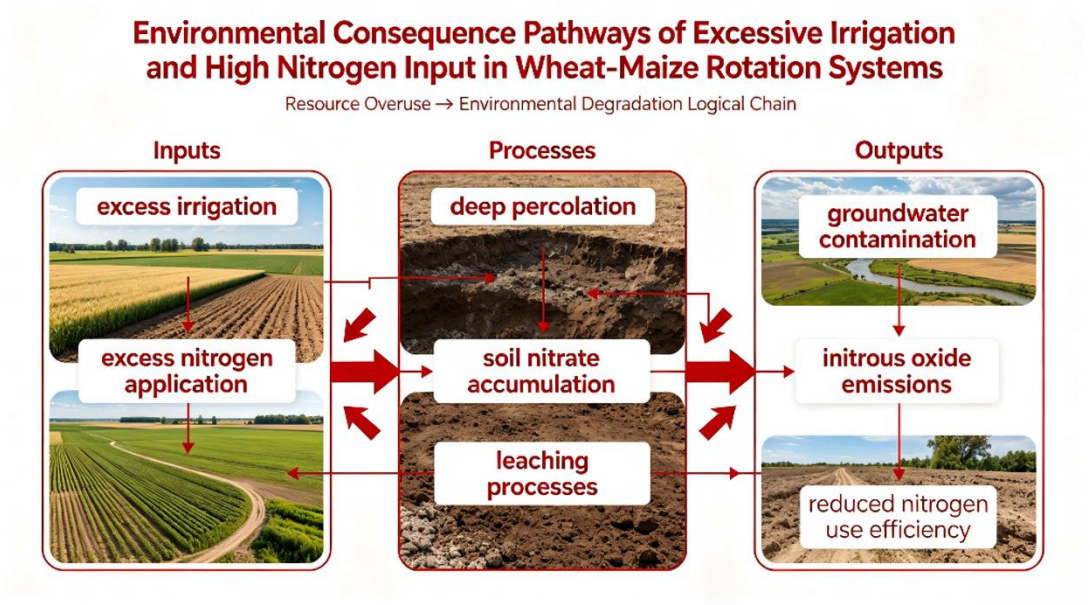


Figure 3 Environmental consequence pathways of excessive irrigation and nitrogen inputs in wheat-based cropping systems, highlighting leaching, groundwater depletion and greenhouse gas emissions

However, intensive management also increases risks of groundwater depletion, nitrate leaching and greenhouse gas emissions if water and N exceed crop demand. In a long-term simulation-experiment framework, excess irrigation and N in wheat-maize rotations led to higher deep percolation and N leaching, requiring careful tuning of irrigation timing and total N to maintain yield while limiting losses (Xu et al., 2020). A global synthesis likewise highlighted that higher N doses, especially beyond 200 kg·ha^{-1}, may reduce nitrogen use efficiency and aggravate environmental burdens, underscoring the need to balance productivity against resource and environmental costs in high-input systems (Wang et al., 2023).

8.2 Comparative analysis of water-saving and reduced-fertilizer production systems

Numerous field trials show that moderate reductions in irrigation and N can sustain or even enhance yield while improving resource use efficiency. Under spring wheat in Northwest China, reducing irrigation from 750 to 600 mm and N from 300 to 225 kg·ha^{-1} maintained dry-matter yield and increased grain yield by 12.9% in one year, while markedly improving water and N use efficiency and economic returns compared with farmers' higher-input practice (Kamran et al., 2023). Similarly, in a drip-irrigated system, deficit irrigation at 75% ETc combined with 170 kg·N·ha^{-2} achieved the highest grain yield and water- and nitrogen-use efficiencies, with structural equation modeling indicating that N contributed over 65% of the yield gain relative to rainfed, unfertilized controls (Lu et al., 2021).

At system scale, alternative water-fertilizer-saving patterns can match conventional yields with lower inputs. In a four-year wheat-maize study in the North China Plain, supplemental drip irrigation at key stages plus 60% of recommended N (272 kg·ha^{-1} yr^{-1}) produced similar double-crop yields and net income to traditional surface irrigation with 453 kg·N·ha^{-1} yr^{-1}, while increasing WUE, N use efficiency and reducing N loss (Li et al., 2023). A regional meta-analysis for the North China Plain further showed that deficit irrigation combined with optimized N splitting and partial organic substitution improved yield, N use efficiency and water use efficiency, with an optimal 7:3 inorganic-organic ratio under moderate irrigation (Wang et al., 2025).

8.3 Regional validation of ecologically adaptive cropping systems

Ecologically adaptive systems integrate crop diversification, conservation practices and input optimization to enhance yield stability and ecosystem services across regions. A review of eco-friendly wheat practices reported that low-input and organic systems generally reduce average yields relative to conventional, yet can deliver competitive and stable production when tailored to cultivar traits, soil agrochemistry and climate, especially under conservation tillage and diversified rotations (Rebouh et al., 2023). Long-term comparisons in Europe and North America illustrate this gradient: conventional systems averaged 6.96 t ha^{-1} with 163 kg mineral N, versus 5.94 t ha^{-1} under low-input and 4.15 t ha^{-1} under organic management, highlighting a clear yield-input trade-off that must be regionally calibrated.

More integrative regional assessments identify "positive deviant" systems that already combine high productivity with reduced environmental footprints. In wheat-maize double cropping on the North China Plain, 16% of surveyed farms were Pareto-optimal across seven sustainability indicators; these systems achieved 17% higher dietary energy output and 49% higher gross margins while lowering groundwater depletion, N loss and greenhouse gas emissions by roughly one-third to one-half compared with other farms (Liang et al., 2022). Key distinguishing practices included lower N in wheat, fewer irrigations, partial manure substitution and reduced pesticide use, demonstrating that regionally validated, ecologically adaptive prototypes can emerge from existing farmer practice rather than purely experimental designs.

9 Optimization Strategies and Regulatory Mechanisms for High-Yield Wheat Management

Studies on nitrogen timing and level show that yield components are highly plastic and can be steered through targeted fertilization regimes. Early and higher N inputs mainly increase grain number via more spikes, spikelets per spike and grains per spikelet, whereas delayed N tends to reduce grain number but increases grain weight, creating a managed trade-off between sink size and grain filling. The very large plasticity of grains per spike and grain number compared with grain weight suggests that management should first secure a high grain number and then avoid excessive late N that only increases grain size at the expense of total yield. Component-level analyses

under different water-N and cultivation modes further clarify thresholds for optimizing spike number and grains per spike. In the Huang-Huai Plain, yields below about 7.5 t ha^{-1} depend on jointly increasing spike number and grains per spike, while higher yields rely mainly on further increasing grains per spike through rapid pre-anthesis dry-matter accumulation. Both spike and non-spike dry matter and nitrogen before anthesis show strong positive relationships with grains per spike, whereas excessive N allocation to spikes reduces grain number, indicating that balanced N distribution between spike and vegetative organs is critical for optimizing the main yield component.

Water-nitrogen combinations create strong synergies in both production and resource-use traits. On the North China Plain, factorial irrigation-N experiments showed that both inputs increased total water use, but also intensified water consumption during grain filling and enhanced soil water use, with a clear positive synergy between crop water productivity and nitrogen-use efficiency across treatments. Decomposing nitrogen-use efficiency revealed that this synergy was driven mainly by higher nitrogen uptake efficiency rather than utilization efficiency, particularly where irrigation increased both pre- and post-anthesis N assimilation into grain. Similar interaction mechanisms appear under more complex management matrices that include planting pattern and density. In a semi-humid but drought-prone region, ridge-furrow planting with plastic film, moderate supplementary irrigation, and medium plant density significantly improved grain yield, water productivity, agronomic N-use efficiency and net income compared with flat planting or sub-optimal densities. These effects arose from interactive gains in soil water consumption, population N uptake and effective panicle number per unit area, demonstrating that coordinated adjustment of canopy structure and soil water capture can synchronize water and N supply with the formation of spikes and grains.

Long-term integrated management strategies illustrate how multiple levers can be combined into robust high-yield systems. In North China, integrated soil and crop system management that delayed sowing, increased seeding rate and optimized fertilization and irrigation achieved yields within about 4-5% of an input-intensive “high-yield” treatment, while markedly increasing nitrogen-use efficiency, water productivity and N balance. A related strategy using higher seeding rate, slightly delayed sowing and re-timed N topdressing similarly produced yields close to the maximum treatment but with much higher NUE and net profit, indicating that coordinated adjustment of sowing date, plant density and N partitioning can support both high yield and economic efficiency. At regional and systems scales, integrated crop management (ICM) and eco-friendly practices provide a broader template for stable high production. Global evidence shows that ICM can raise wheat yields by roughly 16-30% through site-specific nutrient management, conservation tillage, and complementary pest and disease control, while reducing excessive N use and environmental risk. Reviews of eco-friendly wheat systems further emphasize that stable high yields require normative strategies that jointly consider cultivar traits, crop rotation, reduced tillage, biological protection and soil agrochemical status, so that management buffers climatic variability while sustaining yield components across seasons.

Acknowledgments

The author expresses deep gratitude to Professor R.X. Cai from the Zhejiang Agronomist College for his thorough review of the manuscript and constructive suggestions.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdelrhman A., Abdel-Fattah I., Mostafa M., Fadl M., Drosos M., and Scopa A., 2025, Optimizing drip irrigation and nitrogen fertilization for sustainable wheat production in arid soils: Water-nitrogen use efficiency, Water, 17(18): 2708. https://doi.org/10.3390/w17182708

Ahmadi H., Hosseini H., Moshiri F., Alikhani H., and Etesami H., 2024, Impact of varied tillage practices and phosphorus fertilization regimes on wheat yield and grain quality parameters in a five-year corn-wheat rotation system, Scientific Reports, 14: 65784. https://doi.org/10.1038/s41598-024-65784-w

Alghory A., and Yazar A., 2018, Evaluation of crop water stress index and leaf water potential for deficit irrigation management of sprinkler-irrigated wheat, Irrigation Science, 37(1): 61-77.
https://doi.org/10.1007/s00271-018-0603-y

Arshad M., Abbas R., Tonelli F., Baloch R., Haq M., Ahmad A., Zulfiqar U., Djalović I., Prasad P.V.V., and Alshaharni M., 2025, Enhancing wheat yield and quality through optimized weed control timing and seeding density, Archives of Agronomy and Soil Science, 71(9): 1-21.
https://doi.org/10.1080/03650340.2025.2570466

Backhaus A., Griffiths C., Vergara-Cruces Á., Simmonds J., Lee R., Morris R., and Uauy C., 2023, Delayed development of basal spikelets in wheat explains their increased floret abortion and rudimentary nature, Journal of Experimental Botany, 74(16): 5088-5103.
https://doi.org/10.1101/2023.02.17.528935

Bagheripour M., Sharifabad H.H., Mehraban A., and Ganjali H., 2024, Wheat (*Triticum aestivum*) yield gap affected by soil physicochemical properties, Rendiconti Lincei. Scienze Fisiche e Naturali, 35(2): 395-409.
https://doi.org/10.1007/s12210-024-01233-0

Barzegar A.R., Yousefi A., and Daryashenas A., 2002, The effect of addition of different amounts and types of organic materials on soil physical properties and yield of wheat, Plant and Soil, 247(2): 295-301.
https://doi.org/10.1023/A:1021561628045

Behpouri A., Farokhzadeh S., Zinati Z., and Khosravi Z., 2023, Use of multivariate analysis and machine learning methods to characterize traits contributing to wheat yield diversity, Spanish Journal of Agricultural Research, 21(1): e0701.
https://doi.org/10.5424/sjar/2023211-19835

Bicego B., Roxo L., Curci M., Trentin E., Tondelli G., Fontana B., and Rizza A., 2022, Nitrogen fertilization levels and timing affect the plasticity of yield components in bread wheat (*Triticum aestivum* L.), Field Crops Research, 289: 108734.
https://doi.org/10.1016/j.fcr.2022.108734

Bicego B., Savin R., Girousse C., Allard V., and Slafer G.A., 2024, Tillering and floret development dynamics in wheat cultivars of contrasting spike fertility plasticity, Field Crops Research, 317: 109654.
https://doi.org/10.1016/j.fcr.2024.109654

Bicego B., Savin R., Girousse C., Allard V., and Slafer G.A., 2024, Plasticity of grain number and its components in contrasting wheat cultivars, Field Crops Research, 317: 109653.
https://doi.org/10.1016/j.fcr.2024.109653

Calderini D.F., Castillo F.M., Arenas-M A., Molero G., Reynolds M.P., Craze M., Bowden S., Wallington E.J., Dowle A., Gomez L.D., and McQueen-Mason S.J., 2020, Overcoming the trade-off between grain weight and number in wheat by the ectopic expression of expansin in developing seeds leads to increased yield potential, New Phytologist, 230(2): 629-640.
https://doi.org/10.1111/nph.17048

Chen Q., Ou X., Wie W., and Kandyba N., 2021, Influence of the sowing period and density on yield and yield components of three semi-winter wheat varieties, Plant Breeding and Seed Production, 123: 56-65.
https://doi.org/10.30835/2413-7510.2021.251041

Choudhary S., Gaur S., and Kumar S., 2025, Studies on correlation and path coefficient analysis for yield and yield associated traits in wheat (*Triticum aestivum* L.) genotypes, Journal of Advances in Biology and Biotechnology, 28(1): 860-868.
https://doi.org/10.9734/jabb/2025/v28i11860

Dai Y., Liao Z., Lai Z., Bai Z., Zhang F., Li Z., and Fan J., 2023, Interactive effects of planting pattern, supplementary irrigation and planting density on grain yield, water-nitrogen use efficiency and economic benefit of winter wheat in a semi-humid but drought-prone region of northwest China, Agricultural Water Management, 288: 108438.
https://doi.org/10.1016/j.agwat.2023.108438

Dias A.S., and Lidon F.C., 2009, Evaluation of grain filling rate and duration in bread and durum wheat, under heat stress after anthesis, Journal of Agronomy and Crop Science, 195(2): 137-147.
https://doi.org/10.1111/j.1439-037X.2008.00347.x

Dolijanović Ž., Nikolić R., Šeremešić S., Jug D., Biljić M., Pešić S., and Kovačević D., 2025, Effects of conservation tillage and nitrogen management on yield, grain quality, and weed infestation in winter wheat, Agronomy, 15(7): 1742.
https://doi.org/10.3390/agronomy15071742

Elmassry E.L., and Shal M.H., 2020, Estimates of correlation coefficient, path analysis and stepwise regression for some quantitative traits in bread wheat, Menoufia Journal of Plant Production, 5(3): 77-90.
https://doi.org/10.21608/mjppf.2020.138736

Feng S., Ding W., Shi C., Zhu X., Hu T., and Ru Z., 2023, Optimizing the spatial distribution of roots by supplemental irrigation to improve grain yield and water use efficiency of wheat in the North China Plain, Agricultural Water Management, 278: 107989.
https://doi.org/10.1016/j.agwat.2022.107989

Fouad H.M., 2018, Correlation, path and regression analysis in some bread wheat genotypes under normal irrigation and drought conditions, Egyptian Journal of Agronomy, 40(3): 337-350.
https://doi.org/10.21608/agro.2018.3109.1097

Frih B., Oulmi A., Guendouz A., Bendada H., and Selloum S., 2021, Statistical analysis of the relationships between yield and yield components in some durum wheat (*Triticum durum* Desf.) genotypes growing under semi-arid conditions, International Journal of Bio-resource and Stress Management, 12(4): 355-362.
https://doi.org/10.23910/1.2021.2431

Gao Y., Wang Q., Liu Y., He J., Chen W., Xing J., Sun M., Gao Z., Wang Z., Zhang M., and Zhang Y., 2025, Optimal water, nitrogen, and density management increased wheat yield by improving population uniformity, Agricultural Water Management, 311: 109362.
https://doi.org/10.1016/j.agwat.2025.109362

Gao Y., Zhang M., Yao C., Liu Y., Wang Z., and Zhang Y., 2021, Increasing seeding density under limited irrigation improves crop yield and water productivity of winter wheat by constructing a reasonable population architecture, Agricultural Water Management, 254: 106951.
https://doi.org/10.1016/j.agwat.2021.106951

Garcia G.A., Serrago R.A., Dreccer M.F., and Miralles D.J., 2016, Post-anthesis warm nights reduce grain weight in field-grown wheat and barley, Field Crops Research, 195: 50-59.
https://doi.org/10.1016/j.fcr.2016.06.002

Groß J., Gentsch N., Boy J., Heuermann D., Schweneker D., Feuerstein U., Brunner J., von Wirén N., Guggenberger G., and Bauer B., 2023, Influence of small-scale spatial variability of soil properties on yield formation of winter wheat, Plant and Soil, 493(1-2): 79-97.
https://doi.org/10.1007/s11104-023-06212-2

Guo Z., and Schnurbusch T., 2015, Variation of floret fertility in hexaploid wheat revealed by tiller removal, Journal of Experimental Botany, 66(19): 5945-5958.
https://doi.org/10.1093/jxb/erv303

Guo Z., Slafer G.A., and Schnurbusch T., 2016, Genotypic variation in spike fertility traits and ovary size as determinants of floret and grain survival rate in wheat, Journal of Experimental Botany, 67(14): 4221-4230.
https://doi.org/10.1093/jxb/erw200

Han W., Lin X., and Wang D., 2023, Uncovering the primary drivers of regional variability in the impact of climate change on wheat yields in China, Journal of Cleaner Production, 405: 138479.
https://doi.org/10.1016/j.jclepro.2023.138479

Han W., Wang S., Li L., Ali M., Lin X., and Wang D., 2025, Four decades of temperature extremes reshape regional wheat yields and adaptation in China, Journal of Environmental Management, 389: 126271.
https://doi.org/10.1016/j.jenvman.2025.126271

Jaenisch B.R., Munaro L.B., Bastos L.M., Moraes M., Lin X., and Lollato R.P., 2021, On-farm data-rich analysis explains yield and quantifies yield gaps of winter wheat in the U.S. central Great Plains, Field Crops Research, 272: 108287.
https://doi.org/10.1016/j.fcr.2021.108287

Jing J., Li Z., Qian F., Chang X., and Li W., 2023, Effects of different drip irrigation patterns on grain yield and population structure of different water- and fertilizer-demanding wheat (*Triticum aestivum* L.) varieties, Agronomy, 13(12): 3018.
https://doi.org/10.3390/agronomy13123018

Kamran M., Yan Z., Chang S., Ning J., Lou S., Ahmad I., Ghani M., Arif M., Abd El-Sabagh A., and Hou F., 2023, Interactive effects of reduced irrigation and nitrogen fertilization on resource use efficiency, forage nutritive quality, yield, and economic benefits of spring wheat in the arid region of Northwest China, Agricultural Water Management, 281: 108000.
https://doi.org/10.1016/j.agwat.2022.108000

Kiss T., Balla K., Bányai J., Veisz O., and Karsai I., 2018, Associations between plant density and yield components using different sowing times in wheat (*Triticum aestivum* L.), Cereal Research Communications, 46(1): 169-180.
https://doi.org/10.1556/0806.45.2017.069

Lachutta K., and Jankowski K., 2024, An agronomic efficiency analysis of winter wheat at different sowing strategies and nitrogen fertilizer rates: A case study in Northeastern Poland, Agriculture, 14(3): 442.
https://doi.org/10.3390/agriculture14030442

Leilah A.A., and Al-Khateeb S.A., 2005, Statistical analysis of wheat yield under drought conditions, Journal of Arid Environments, 61(3): 483-496.
https://doi.org/10.1016/j.jaridenv.2004.10.011

Li H., Li X., Mei X., Nangia V., Guo R., Hao W., and Wang J., 2023, An alternative water-fertilizer-saving management practice for wheat-maize cropping system in the North China Plain: Based on a 4-year field study, Agricultural Water Management, 282: 108053.
https://doi.org/10.1016/j.agwat.2022.108053

Liang Z., van der Werf W., Xu Z., Cheng J., Wang C., Cong W., Zhang C., Zhang F., and Groot J.C.J., 2022, Identifying exemplary sustainable cropping systems using a positive deviance approach: Wheat-maize double cropping in the North China Plain, Agricultural Systems, 202: 103471.
https://doi.org/10.1016/j.agsy.2022.103471

Lin X., Li P., Shang Y., Liu S., Wang S., Hu X., and Wang D., 2020, Spike formation and seed setting of the main stem and tillers under post-jointing drought in winter wheat, Journal of Agronomy and Crop Science, 206(6): 694-710.
https://doi.org/10.1111/jac.12432

Lollato R.P., Pradella L., Giordano N., Ryan L., Soler J., Simão L., Jaenisch B., and Horton R., 2024, Winter wheat response to plant density in yield contest fields, Crop Science, 64(3): 1460-1474.
https://doi.org/10.1002/csc2.21296

Lu J., Hu T., Geng C., Cui X., Fan J., and Zhang F., 2021, Response of yield, yield components and water-nitrogen use efficiency of winter wheat to different drip fertigation regimes in Northwest China, Agricultural Water Management, 255: 107034.
https://doi.org/10.1016/j.agwat.2021.107034

Mandea V., Mustățea P., Marinciu C., Șerban G., Melucă C., Păunescu G., Isticioaia S., Dragomir C., Bunta G., Filiche E., Voinea L., Lobonțiu I., Domokos Z., Voica M., Ittu G., and Săulescu N.N., 2019, Yield components compensation in winter wheat (*Triticum aestivum* L.) is cultivar dependent, Romanian Agricultural Research, 36: 45-54.
https://doi.org/10.59665/rar3604

Mangini G., Gadaleta A., Colasuonno P., Marcotuli I., Signorile A.M., Simeone R., De Vita P., Mastrangelo A.M., Laidò G., Pecchioni N., and Blanco A., 2018, Genetic dissection of the relationships between grain yield components by genome-wide association mapping in a collection of tetraploid wheats, PLoS ONE, 13(1): e0190162.
https://doi.org/10.1371/journal.pone.0190162

Manntschke A., Hempel L., Temme A., Reumann M., and Chen T.W., 2025, Breeding in winter wheat (*Triticum aestivum* L.) can be further progressed by targeting previously neglected competitive traits, Frontiers in Plant Science, 16: 1490483.
https://doi.org/10.3389/fpls.2025.1490483

Mariem S.B., Soba D., Zhou B., Loladze I., Morales F., and Aranjuelo I., 2021, Climate change, crop yields, and grain quality of C3 cereals: a meta-analysis of [CO2], temperature, and drought effects, Plants, 10(6): 1052.
https://doi.org/10.3390/plants10061052

Matković-Stojšin M., Zečević V., Petrović S., Dimitrijević M., Mićanović D., Banjac B., and Knežević D., 2018, Variability, correlation, path analysis and stepwise regression for yield components of different wheat genotypes, Genetika, 50(3): 817-828.
https://doi.org/10.2298/GENSR1803817M

Mirosavljević M., Momčilović V., Aćin V., Jocković B., Pržulj N., and Jaćimović G., 2025, Yield determination in major small grain crops in response to nitrogen fertilization, Plants, 14(7): 1017.
https://doi.org/10.3390/plants14071017

Mohammadi R., 2024, Effects of post-flowering drought and supplemental irrigation on grain yield and agro-phenological traits in durum wheat, European Journal of Agronomy, 159: 127180.
https://doi.org/10.1016/j.eja.2024.127180

Noor H., Ding P., Ren A., Sun M., and Gao Z., 2023, Effects of nitrogen fertilizer on photosynthetic characteristics and yield, Agronomy, 13(6): 1550.
https://doi.org/10.3390/agronomy13061550

Oldfield E.E., Bradford M.A., and Wood S.A., 2019, Global meta-analysis of the relationship between soil organic matter and crop yields, SOIL, 5(1): 15-32.
https://doi.org/10.5194/soil-5-15-2019

Qu B., Noor H., Feng Y., Di J., Alotaibi M.A., and Noor F., 2025, Nitrogen uptake and water consumption for achieving high yield of winter wheat upon nitrogen addition at different doses, Scientific Reports, 15: 24530.
https://doi.org/10.1038/s41598-025-24530-6

Ram M., Poudel P., Ghimire S., Pandey M., Dhakal K., Thapa D., and Poudel H., 2020, Yield stability analysis of wheat genotypes at irrigated, heat stress and drought condition, Journal of Biology and Today's World, 9(5): 1-10.
https://doi.org/10.35248/2322-3308.20.09.220

Rebouh N.Y., Khugaev C., Utkina A., Isaev K., Mohamed E.S., and Kucher D.E., 2023, Contribution of eco-friendly agricultural practices in improving and stabilizing wheat crop yield: A review, Agronomy, 13(9): 2400.
https://doi.org/10.3390/agronomy13092400

Ren J., Ren A., Lin W., Noor H., Khan S., Dong S., Sun M., and Gao Z., 2021, Nitrogen fertilization and precipitation affected wheat nitrogen use efficiency and yield in the semiarid region of the Loess Plateau in China, Journal of Soil Science and Plant Nutrition, 22(1): 585-596.
https://doi.org/10.1007/s42729-021-00671-1

Riedesel L., Möller M., Piepho H.P., Rentel D., Lichthardt C., Golla B., Kautz T., and Feike T., 2024, Site conditions determine heat and drought induced yield losses in wheat and rye in Germany, Environmental Research Letters, 19(4): 044010.
https://doi.org/10.1088/1748-9326/ad24d0

Ru C., Hu X., Chen D., Wang W., Zhen J., and Song T., 2023, Individual and combined effects of heat and drought and subsequent recovery on winter wheat (*Triticum aestivum* L.) photosynthesis, nitrogen metabolism, cell osmoregulation, and yield formation, Plant Physiology and Biochemistry, 196: 222-235.
https://doi.org/10.1016/j.plaphy.2023.01.038

Shang Y., Lin X., Li P., Gu S., Lei K., Wang S., Hu X., Zhao P., and Wang D., 2020, Effects of supplemental irrigation at the jointing stage on population dynamics, grain yield, and water-use efficiency of two different spike-type wheat cultivars, PLoS ONE, 15(4): e0230484.
https://doi.org/10.1371/journal.pone.0230484

Sharma A., Sharma S., Vyas L., Yadav S., Pramanick B., Naik B.S., Obročník O., Bárek V., Brestic M., Gaber A., Alshehri M.A., and Hossain A., 2024, Innovative organic nutrient management and land arrangements improve soil health and productivity of wheat (*Triticum aestivum* L.) in an organic farming system, Frontiers in Sustainable Food Systems, 8: 1455433.
https://doi.org/10.3389/fsufs.2024.1455433

Slafer G.A., Garcia G.A., Serrago R.A., and Miralles D.J., 2022, Physiological drivers of responses of grains per·m2 to environmental and genetic factors in wheat, Field Crops Research, 286: 108593.
https://doi.org/10.1016/j.fcr.2022.108593

Solanki Y.P.S., Singh V., Rai N.K., and Gangadhar N., 2024, Development of yield forecast model in bread wheat using regression analysis, International Journal of Plant and Soil Science, 36(7): 875-881.
https://doi.org/10.9734/ijpss/2024/v36i74799

Tian Z., Yin Y., Li B., Zhong K., Liu X., Jiang D., Cao W., and Dai T., 2024, Optimizing planting density and nitrogen application to mitigate yield loss and improve grain quality of late-sown wheat under rice-wheat rotation, Journal of Integrative Agriculture, 23(8): 2788-2803.
https://doi.org/10.1016/j.jia.2024.01.032

Tilley M.S., Heiniger R.W., and Crozier C.R., 2019, Tiller initiation and its effects on yield and yield components in winter wheat, Agronomy Journal, 111(3): 1323-1332.
https://doi.org/10.2134/agronj2018.07.0469

Tita D., Mahdi K., Devkota K.P., and Devkota M., 2025, Climate change and agronomic management: Addressing wheat yield gaps and sustainability challenges in the Mediterranean and MENA regions, Agricultural Systems, 219: 104242.
https://doi.org/10.1016/j.agsy.2024.104242

Tomaz A., Palma J.F., Ramos T.B., Costa M.M., Rosa E., Santos M., Boteta L., Dôres J., and Patanita M., 2021, Yield, technological quality and water footprints of wheat under Mediterranean climate conditions: A field experiment to evaluate the effects of irrigation and nitrogen fertilization strategies, Agricultural Water Management, 255: 107214.
https://doi.org/10.1016/j.agwat.2021.107214

Vicentin L., Canales J., and Calderini D.F., 2024, The trade-off between grain weight and grain number in wheat is explained by the overlapping of the key phases determining these major yield components, Frontiers in Plant Science, 15: 1380429.
https://doi.org/10.3389/fpls.2024.1380429

Wang D., Liu S., Guo M., Cheng Y., Shi L., Li J., Yu Y., Wu S., Dong Q., Ge J., and Gong X., 2025, Optimizing nitrogen fertilization and irrigation practices for enhanced winter wheat productivity in the North China Plain: A meta-analysis, Plants, 14(11): 1686.
https://doi.org/10.3390/plants14111686

Wang Y., Peng Y., Lin J., Wang L., Jia Z., and Zhang R., 2023, Optimal nitrogen management to achieve high wheat grain yield, grain protein content, and water productivity: A meta-analysis, Agricultural Water Management, 290: 108587.
https://doi.org/10.1016/j.agwat.2023.108587

Xie Q., and Sparkes D.L., 2021, Dissecting the trade-off of grain number and size in wheat, Planta, 254(3): 57.
https://doi.org/10.1007/s00425-021-03658-5

Xie Q., Mayes S., and Sparkes D.L., 2016, Optimizing tiller production and survival for grain yield improvement in a bread wheat × spelt mapping population, Annals of Botany, 117(1): 51-66.
https://doi.org/10.1093/aob/mcv147

Xu J., Cai H., Wang X., Ma C., Lu Y., Ding Y., Wang X., Chen H., Wang Y., and Saddique Q., 2020, Exploring optimal irrigation and nitrogen fertilization in a winter wheat-summer maize rotation system for improving crop yield and reducing water and nitrogen leaching, Agricultural Water Management, 228: 105904.
https://doi.org/10.1016/j.agwat.2019.105904

Yang D., Cai T., Luo Y., and Wang Z., 2019, Optimizing plant density and nitrogen application to manipulate tiller growth and increase grain yield and nitrogen-use efficiency in winter wheat, Peer J., 7: e6484.
https://doi.org/10.7717/peerj.6484

Yokamo S., Irfan M., Huan W., Wang B., Wang Y., Ishfaq M., Lu D., Chen X., Cai Q., and Wang H., 2023, Global evaluation of key factors influencing nitrogen fertilization efficiency in wheat: A recent meta-analysis (2000-2022), Frontiers in Plant Science, 14: 1272098.
https://doi.org/10.3389/fpls.2023.1272098

Zhang H., Zhao Q., Zhong W., Wang L., Li X., Fan Z., Zhang Y., Li J., Gao X., Shi J., and Chen F., 2021, Effects of nitrogen fertilizer on photosynthetic characteristics, biomass, and yield of wheat under different shading conditions, Agronomy, 11(10): 1989.
https://doi.org/10.3390/agronomy11101989

Zhou X., Yang L., Wang G., Zhao Y., and Wu H., 2021, Effect of deficit irrigation scheduling and planting pattern on leaf water status and radiation use efficiency of winter wheat, Journal of Agronomy and Crop Science, 207(3): 578-593.
https://doi.org/10.1111/jac.12466

Computational Molecular Biology 2026, Vol.16, No.2, 129-145
http://bioscipublisher.com/index.php/cmb

Review Article Open Access

Modeling the Relationship between Temperature and Tomato Yield in Greenhouse Systems

Xingzhu Feng
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding author: xingzhu.feng@hibio.org

Computational Molecular Biology, 2026, Vol.16, No.2 doi: 10.5376/cmb.2026.16.0010
Received: 22 Feb., 2026
Accepted: 30 Mar., 2026
Published:15 Apr., 2026

Preferred citation for this article:
Feng X.Z., 2026, Modeling the relationship between temperature and tomato yield in greenhouse systems, Computational Molecular Biology, 16(2): 129-145 (doi: 10.5376/cmb.2026.16.0010)

Abstract Temperature is one of the most critical environmental factors affecting the growth, development, and productivity of greenhouse tomatoes. This paper systematically reviews and analyzes the relationship between temperature dynamics and tomato yield formation under protected cultivation conditions. The study summarizes the physiological mechanisms through which temperature regulates photosynthesis, respiration, flowering, fruit set, and stress responses during different growth stages. In addition, the characteristics of greenhouse microclimates and the interactions between temperature, humidity, light, and CO_2 are discussed. Various modeling approaches, including statistical regression models, process-based crop models, and machine learning algorithms, are evaluated for their ability to predict tomato yield under variable temperature conditions. The paper also examines methods for model calibration, validation, and performance assessment using multi-season datasets. Several case studies are presented to demonstrate the practical applications of temperature-yield models in greenhouse management and precision agriculture. Finally, the challenges, limitations, and future prospects of intelligent temperature regulation and climate-adaptive modeling are highlighted to support sustainable greenhouse tomato production.

Keywords Greenhouse tomato; Temperature dynamics; Yield prediction; Crop growth model; Precision agriculture

1 Introduction

Greenhouse tomato production has become essential for ensuring stable, year-round supply while making efficient use of land, water, and energy. Within these protected systems, temperature is a primary driver of plant development, resource use, and ultimately yield, and its role is intensifying under climate change and more frequent heat extremes (Kürklü et al., 2025). Elevated temperatures and longer hot seasons already reduce yield, force higher cooling and irrigation demands, and complicate climate control in commercial greenhouses. At the same time, the greenhouse structure creates opportunities to actively manage temperature and to exploit its buffering capacity, provided that quantitative relationships between temperature regimes and tomato yield are well understood (Flores-Velázquez et al., 2022). Developing models that link temperature dynamics to yield is therefore important for climate-resilient design, control, and strategic planning in greenhouse tomato systems.

Over the past decades, numerous experimental and monitoring studies have examined tomato responses to greenhouse temperature. Work in high-tech and plastic houses shows that small but persistent temperature differences within a single compartment (on the order of 3 °C in daily averages) can significantly alter stem growth, fruit growth, and truss mass, even when bulk climate appears uniform (Šalagovič et al., 2024. Experiments manipulating air or soil temperature demonstrate substantial effects on photosynthesis, dry-matter accumulation, quality traits, and yield, with warmer root zones or air generally increasing yield and water productivity up to an optimum, beyond which heat stress causes losses (Efeta et al., 2025). Heat-stress trials further reveal strong genotype-dependent yield declines and quality changes, underscoring the importance of temperature thresholds and exposure duration around flowering and fruit set. Recent climate-change-oriented studies highlight that extreme high temperatures in commercial greenhouses already cause substantial yield losses (often >10%) and sharply increase resource use, confirming temperature as a critical vulnerability factor in modern soilless systems.

On the modeling side, several frameworks explicitly integrate temperature into greenhouse tomato yield prediction. Dynamic crop models such as TOMGRO represent organ initiation and growth via temperature-responsive source-sink processes, calibrated under controlled temperature, CO_2 and light conditions, and have been proposed as tools for environment control decisions (Higashide, 2022). Yield models developed for model-based greenhouse design implement literature-based temperature effects on yield and reproduce responses under both near-optimal and sub-optimal regimes, including extreme diurnal oscillations in contrasting climates. Spatially explicit approaches combining geostatistics with crop growth models show that ignoring intra-greenhouse temperature heterogeneity can bias simulated development rates and yield, particularly between central zones and sidewalls. More recently, integrated climate-and-yield models for specific greenhouse types, such as Chinese solar greenhouses, have been validated against multi-site experiments and used to explore design and operational scenarios, again emphasizing indoor air temperature as a key determinant of predicted yield (Zhou et al., 2025).

Against this background, the present study focuses on modeling the relationship between temperature and tomato yield in greenhouse systems. The first objective is to quantify how different descriptors of temperature regimes-such as daily mean, diurnal amplitude, spatial gradients, and the frequency and duration of supra-optimal events-affect yield and its components under realistic microclimate variability. The second objective is to incorporate these relationships into a modeling framework suitable for coupling with greenhouse climate models and control strategies, building on existing dynamic and design-oriented yield models while simplifying where necessary for operational use. The central hypotheses are that: (i) tomato yield in greenhouses is a non-linear function of both average temperature and exposure to critical heat or cold thresholds at sensitive stages; (ii) explicit representation of intra-greenhouse temperature variation improves yield prediction compared with models using only bulk climate; and (iii) a temperature-focused yield model can support evaluation of design options and climate-control strategies under current and future climate conditions.

2 Physiological Basis of Temperature Effects on Tomato Growth and Yield

2.1 Temperature regulation of photosynthesis and respiration

Tomato photosynthesis operates optimally within a moderate temperature range; deviations in either direction impair carbon gain and growth. Greenhouse and field studies show that high air temperatures, especially above about 38-40 °C, reduce net photosynthetic rate, stomatal conductance and ultimately fruit yield, reflecting damage to both CO_2 assimilation and water relations over time (Figure 1) (Liu et al., 2023). Sub-high temperature combined with high light (35 °C, 1000 μmol·m^{-2}·s^{-1}) sharply decreased net photosynthetic rate, Rubisco activity, PSII and PSI quantum yields, while increasing non-regulated energy dissipation and ROS accumulation, indicating irreversible photoinhibition of both photosystems when thermal and light loads coincide (Talukder et al., 2025).

Lower temperatures also limit photosynthesis by depressing chlorophyll content, electron transport and chlorophyll fluorescence parameters, resulting in reduced dry matter accumulation and yield (Zhang et al., 2023). Under sub-optimal day/night regimes around 15/10 °C, sensitive cultivars show greater reductions in fresh weight, chlorophyll content, Fv/Fm and electron transport rate than tolerant ones, whereas tolerant genotypes maintain higher soluble sugars and proline, supporting photochemistry and osmotic balance. Soil temperature interacts with shoot processes: moderate soil warming to about 26 °C increased leaf assimilation rate, chlorophyll, dry matter and yield in greenhouses, while also stimulating soil respiration and microbial biomass, suggesting coordinated temperature effects on root function and canopy photosynthesis.

Respiration is likewise temperature sensitive, affecting carbon use efficiency and growth. Analyses of tomato under high temperature indicate that respiration rates and growth rates shift together, with elevated temperatures increasing metabolic rates but also reducing metabolic efficiency and substrate carbon use (Alsamir et al., 2020). Nighttime respiratory costs interact with daytime photosynthesis to determine net biomass gain, and high night temperatures have been highlighted as critical constraints in warm greenhouse climates (Sato et al., 2006). Evaluations in solar greenhouses show that water-use efficiency at the leaf level is highest at 20-30 °C, beyond

which heat shock reduces photosynthesis more than transpiration, lowering carbon gain per unit water and contributing to yield losses under hot conditions.

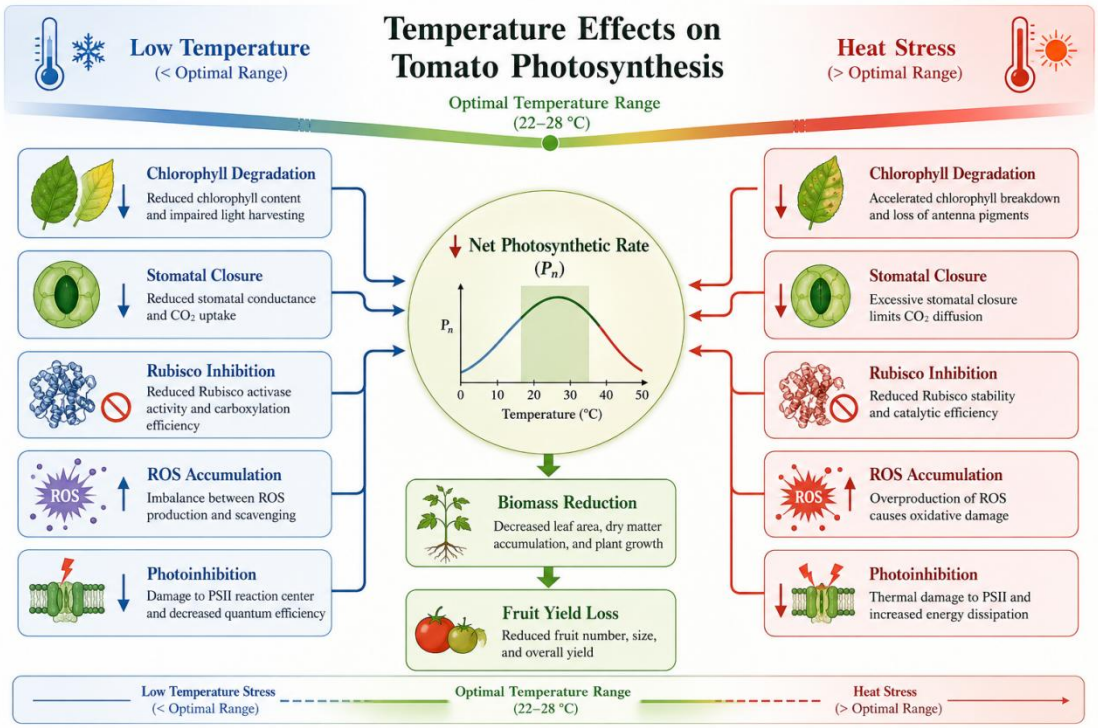


Figure 1 Conceptual diagram illustrating the effects of low and high temperature stress on tomato photosynthesis, photochemical efficiency, and yield formation

2.2 Effects of day/night temperature on flowering and fruit set

Tomato reproductive development is particularly sensitive to relatively small increases in mean day/night temperature. When day/night temperatures rose moderately from 28/22 °C to 32/26 °C, vegetative growth and photosynthesis remained largely unchanged, but fruit set, pollen viability and pollen release declined markedly, demonstrating that reproductive processes fail before canopy carbon assimilation under moderate heat (Sato et al., 2006). In controlled phytotron experiments across 20/24 to 27/37 °C night/day regimes, flowering and fruiting were normal at cooler treatments, but fruit set dropped sharply at 24/32 °C and nearly disappeared at 27/37 °C in most genotypes, underscoring the narrow thermal window for successful fertilization (Yadav et al., 2014).

Night temperature emerges as a key determinant of reproductive success. Work separating day and night effects shows that high night temperature (≥26 °C) at flowering is more detrimental to fruit set than a similar increase in day temperature, even when day temperature is already high. Earlier controlled-environment studies with 26 °C days and 18-26 °C nights reported that total and normal pollen production, seed content, and flower and fruit numbers on the first cluster were all higher at 18-22 °C nights than at 24-26 °C, although pollen germination in vitro could be favored at warmer nights, highlighting a complex trade-off between pollen formation and performance. Under fluctuating ambient day/night conditions in hydroponic summer production, early and late summer regimes with lower mean temperatures produced more flower clusters, fruits and higher yields than mid-summer regimes with warmer nights, again indicating that modest nocturnal warming can substantially depress reproductive efficiency and yield.

2.3 Heat and low-temperature stress mechanisms in greenhouse tomatoes

High temperatures in greenhouses trigger a cascade of morphological, physiological and reproductive disturbances that reduce yield and fruit quality. Reviews of tomato heat stress describe substantial flower abortion, up to about 80 % loss under severe episodes, along with impaired pollen viability and root growth, which together reduce fruit set and marketable yield (Alsamir et al., 2020). Experimental comparisons of high-temperature and

control greenhouses show decreased stem diameter, plant height and fresh weight, elevated electrolyte leakage, lower relative water content, reduced photosynthetic efficiency and increased malondialdehyde, together with accumulation of phenolics, flavonoids and lycopene, consistent with membrane damage and activation of antioxidant defenses under chronic heat (Sellami and Kooli, 2026).

At the reproductive level, moderate but sustained temperature elevation during a critical pre-anthesis window disrupts specific metabolic processes in the androecium. Under 32 °C/26 °C, androecial glucose and fructose decline while sucrose increases, coinciding with reduced acid invertase transcript abundance, altered sugar metabolism and sharply reduced fruit set despite unchanged pollen production. Proline transporter expression on the microspore surface also falls under these conditions, suggesting impaired osmoprotection and turgor regulation in developing pollen. Conversely, screening of contrasting genotypes under high temperature in greenhouses reveals that tolerant cultivars can maintain fruit weight or even improve fruit hardness, whereas susceptible ones show large decreases in fruit size components, highlighting genotypic differences in maintaining reproductive function and fruit quality under heat (Rajametov et al., 2021).

Low temperature stress in greenhouse tomatoes also induces multi-level responses affecting both vegetative and reproductive stages. Reviews of cold responses report delayed flowering, enhanced pollen sterility and strong reductions in fruit set and yield, alongside decreased photosynthetic capacity due to impaired gas exchange, pigment content and chloroplast function (Yadav et al., 2021). Detailed analyses of sub-optimal day/night temperatures show that sensitive cultivars display greater declines in Fv/Fm, photochemical quenching and biomass than tolerant ones, while tolerant genotypes maintain higher levels of osmolytes such as soluble sugars and proline that support ROS scavenging and membrane stability. At the root level, exposure to low root-zone temperature (around 10 °C) reduces root activity, water and nutrient supply to shoots, lowers photosynthesis and chlorophyll fluorescence, and triggers accumulation of hydrogen peroxide, malondialdehyde and proline; only partial recovery occurs after re-warming, indicating lasting damage to the photosynthetic apparatus and growth potential (Zhang et al., 2023). Collectively, these mechanisms explain why both heat and cold episodes in greenhouses can depress tomato yield and underscore the importance of modelling temperature effects across this full stress spectrum.

3 Greenhouse Environmental Characteristics and Temperature Dynamics

3.1 Temperature distribution and microclimate formation in facilities

Air temperature in greenhouses is far from uniform, even when a single central sensor suggests a stable “bulk” climate. Multi-year monitoring in commercial tomato greenhouses has revealed horizontal gradients of up to about 3 °C in daily average temperature and 0.6 kPa in vapour pressure deficit (VPD), driven by structure, airflow patterns, and crop canopy (Šalagovič et al., 2024). These small but persistent differences translated into measurable variability in stem growth, fruit growth rate, and truss mass between locations, indicating that microclimate heterogeneity can meaningfully affect yield. Similar work in single- and multi-span houses reported horizontal temperature differences on the order of 1 °C between center and sides, confirming that assumptions of homogeneous air conditions are unrealistic for modern structures and should be revisited in energy and climate calculations (Ogunlowo et al., 2021).

Vertical stratification further complicates the thermal environment, particularly in tall or large-span facilities. A combined experimental-numerical study in a plastic greenhouse found that air near the roof could be more than 13 °C warmer than air lower in the crop zone at midday, with much smaller differences in the morning (Li et al., 2024). Computational fluid dynamics simulations that explicitly account for crop transpiration and optical effects show that plant canopies can increase temperature standard deviation by more than 30% compared with bare-structure assumptions, and that the hottest air often resides just below the roof where solar gains concentrate (Xu et al., 2022). Field studies in low-automation Mediterranean tomato houses report horizontal differences up to 7 °C-10 °C at certain times, highlighting how limited ventilation and solar load interact to create spatially complex microclimates that challenge single-point monitoring strategies.

3.2 Interaction between temperature, humidity, CO_2, and radiation

Temperature in greenhouses co-varies tightly with relative humidity and VPD, shaping plant responses more than any single variable alone. Microclimate monitoring in commercial tomato systems showed that zones with slightly higher temperatures tended also to exhibit higher VPD, and these combined conditions influenced local growth and fruit development more strongly than either driver considered independently. A detailed sensor-network study using an IoT-based "optimality degree" index quantified how diurnal swings in temperature of nearly 15 °C between day and night were accompanied by large changes in RH and VPD, often driving the climate outside tomato comfort ranges for substantial portions of the day (Rezvani et al., 2020). In warm seasons, natural ventilation alone was insufficient to prevent thermal inversion, leading to high humidity and sub-optimal VPD at night even as daytime conditions became too hot and dry, underscoring the need to manage temperature and humidity jointly rather than in isolation.

Radiation and CO_2 further modulate how temperature and humidity translate into crop performance. Reviews of greenhouse horticulture under climate change in the Mediterranean emphasize that rising temperature, declining RH, increasing VPD and modified solar radiation typically act together, often pushing microclimates beyond optimal thresholds for photosynthesis, transpiration, and reproductive development (Fanourakis et al., 2025). In such conditions, high radiation loads during heat events can exacerbate canopy temperature and water demand, while CO_2 enrichment or shading and cooling strategies may partially offset stress but also alter energy and water use. Process-based and data-driven crop models increasingly incorporate temperature, humidity (via stomatal conductance or VPD), CO_2 and shortwave radiation as coupled drivers, demonstrating that realistic prediction of biomass or yield requires capturing interactions among all four rather than simple temperature sums alone (Sun et al., 2025).

3.3 Seasonal and regional variations in greenhouse temperature regimes

Seasonal shifts in outside climate strongly reshape greenhouse temperature regimes and their suitability for tomato production. Long-term microclimate monitoring in tomato greenhouses has shown distinct spring-summer-autumn patterns, with larger intra-house gradients and more frequent exceedance of high-temperature thresholds during summer, even when the annual mean appears acceptable. An IoT-based assessment in Iran quantified "optimality degrees" for temperature, RH and VPD and found that winter months achieved higher overall optimality, largely because heating systems maintained conditions near target ranges, whereas in summer the lack of automated cooling produced long periods with daytime temperatures above 34 °C and night temperatures below 17 °C, unsuitable for tomato growth (Figure 2) (Rezvani et al., 2020). Numerical analyses of soilless glasshouses in North Africa similarly demonstrated strong seasonal effects on indoor profiles, with differences in roof-level temperatures between crop and no-crop scenarios narrowing but not eliminating the impact of external seasonal forcing (Abid et al., 2024).

Regional climate also determines baseline greenhouse temperature challenges and thus the design of appropriate control strategies. A systematic review across climatic zones reported that optimal tomato production in Mediterranean and arid regions typically requires carefully controlled ranges around 18-25 °C, with rising ambient temperatures reducing yields by more than half in some simulations when air temperatures approach 35 °C (Nițu et al., 2025). Survey-based evidence from high-tech soilless tomato greenhouses in Türkiye showed that extreme heat events during one season led to yield losses averaging 12.5%, alongside substantial increases in water, fogging, fertilizer and electricity use, and widespread reports of difficulty in climate control (Kürklü et al., 2025). In colder regions, solar or soft-shell greenhouses can increase average indoor temperatures by 10 °C-15 °C above outdoors during winter, greatly reducing low-temperature stress but at the cost of pronounced diurnal swings that must be managed to avoid humidity problems and localized stress hotspots. Across structures, seasons and regions, greenhouse temperature dynamics emerge from interacting structural, radiative and airflow processes, tightly coupled with humidity, CO_2 and radiation. Spatial heterogeneity and seasonal-regional contrasts are large enough to influence tomato growth and yield, indicating that both empirical analysis and modeling of temperature-yield relationships must explicitly represent microclimate patterns rather than rely on single-point or season-average conditions.

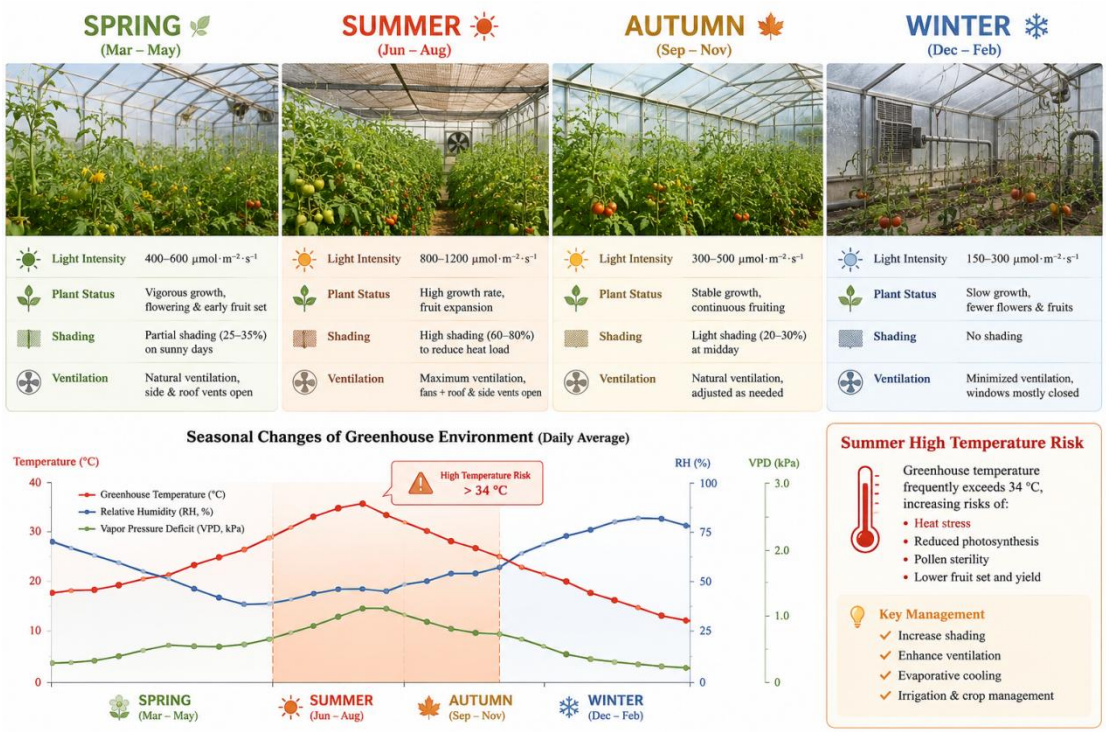


Figure 2 Seasonal variation in greenhouse tomato microclimate conditions, including temperature, relative humidity, and vapor pressure deficit across spring, summer, autumn, and winter periods

4 Tomato Yield Formation and Temperature Sensitivity

4.1 Yield components of greenhouse tomato

Tomato yield in greenhouse systems is determined by the interaction of fruit number, fruit size, and fruit dry matter content, all of which are sensitive to temperature. Yield components such as flower number, fruit set, and number of fruits per plant have been repeatedly proposed as primary markers of performance under heat, because they directly reflect reproductive success under stressful thermal regimes (Ghabileh et al., 2024). In many screening and physiological studies, genotypes that sustain higher fruit set and fruit number under high temperature also maintain higher overall yield, indicating that temperature sensitivity of these components largely controls production potential (Ro et al., 2021).

Fruit size and mass are additional key components shaping final yield. High temperature in greenhouses commonly reduces fruit weight, diameter, and firmness, with reported declines in susceptible cultivars of more than 30% in fruit weight compared with normal temperature conditions (Rajametov et al., 2021). Experimental increases of mean temperature by only a few degrees during fruit development have also been shown to reduce fruit size and alter sugar-acid balance, demonstrating that relatively small thermal shifts can reshape both quantitative and qualitative yield traits. At the same time, modeling and dry-matter studies in greenhouse tomato indicate that high yields are associated with improved total dry-matter production and efficient partitioning to fruits, suggesting that temperature effects on canopy photosynthesis and source-sink balance feed through to both fruit number and fruit size.

4.2 Critical temperature thresholds during different growth stages

Temperature thresholds governing tomato yield are strongly stage-dependent, with the reproductive phase showing the greatest sensitivity. Reviews of heat stress in tomato identify upper threshold temperatures around 30 °C-35 °C as critical for many processes, noting that temperatures above about 35 °C can inhibit seed germination, vegetative growth, flowering time and fruit set (Lee et al., 2022). In commercial protected cultivation in hot Mediterranean summers, mean daily temperatures of 25 °C-26 °C already appear to represent an upper limit for proper fruit set and yield, with even modest reductions of 1 °C-1.5 °C and higher humidity improving pollen viability and fruit set rates (Harel et al., 2014).

Flowering and early fruit set are especially vulnerable to exceedance of these thresholds. Controlled and field experiments consistently show that prolonged daytime or nighttime temperatures above about 32 °C/20 °C (day/night) during the reproductive phase reduce fruit set and fruit weight, leading to significant yield losses (Miller et al., 2021). Studies of pollen performance under episodes of 30 °C-34 °C or short heat shocks around anthesis report sharp declines in pollen viability, germination, and tube growth, which then translate into lower seed set and smaller fruit mass (Zepeda et al., 2026). Conversely, work on optimal ranges and growth-stage specific limits indicates that night temperatures above roughly 21 °C can already discriminate heat-tolerant from heat-sensitive cultivars, emphasizing that even moderate nocturnal warming beyond cultivar-specific thresholds during flowering can markedly depress yield formation.

4.3 Relationship between temperature accumulation and yield stability

Beyond instantaneous thresholds, tomato yield stability in greenhouses reflects the cumulative exposure to supra-optimal or sub-optimal temperatures over the season. Long-term heat stress experiments under greenhouse conditions show that yield loss increases with the duration of exposure: cherry tomato accessions subjected to elevated day set-points for more than 50 days exhibited progressive reductions in harvest index and fruit yield, with some genotypes losing over 40% of yield relative to controls (Park et al., 2023). Recent modeling work combining greenhouse climate projections with morphological yield models similarly demonstrates that future scenarios with 1-8 °C warming and longer hot seasons can decrease yield in heat-sensitive accessions while slightly increasing or stabilizing yield in more heat-resilient ones, highlighting the role of accumulated heat load in determining long-term yield trajectories (Kim et al., 2025).

Temperature accumulation interacts with developmental timing to shape yield stability. A mechanistic model of seed set and fruit mass that incorporates short periods of low (14 °C) and high (30 °C-34 °C) temperature shows that both the level and duration of deviation from the optimum critically affect pollen quality, seed set, and resulting fruit mass, with repeated or longer stress episodes causing cumulative reductions in fruit number and size on a truss (Zepeda et al., 2026). Multi-environment trials comparing performance under optimal field, high-temperature field, and high-temperature greenhouse conditions further confirm that yield stability differs strongly among genotypes: some maintain relatively constant fruit set and yield across environments, while others exhibit steep declines under repeated high-temperature episodes (Ro et al., 2021). Together, these findings support the use of temperature sums or heat-stress indices over sensitive stages as key predictors of yield stability in greenhouse tomato production.

5 Data Acquisition and Experimental Design

5.1 Experimental materials, greenhouse conditions, and cultivation management

Greenhouse tomato studies typically specify cultivar choice, planting density, and structural characteristics to ensure reproducibility and to contextualize yield responses. Experiments on cherry tomato ‘Cheramy F1’ in winter greenhouses used a randomized complete block design with split plots, three rows and three replications, with three plants sampled per row, capturing variability over two consecutive seasons (Arshad et al., 2024). Within this structure, the internal climate ranged from about 8-41 °C across vegetative and fruiting stages, with CO_2 between roughly 386-510 ppm and light intensity from about 95-240 W m^{-2}, providing a broad envelope of temperature regimes for modeling. Other greenhouse trials with large-fruited cultivars have similarly defined plot structure through factorial or split-split plot designs, for example combining cultivar, grafting and plant density (3.5 vs. 5.5 plants m^{-2}) in hydroponic organic systems to test management interactions under hot, humid conditions (Dash et al., 2023).

Representative experimental work also reports greenhouse size, location, and baseline climate. A Ghanaian study used a 270 m^2 greenhouse at a defined latitude, planting tomato ‘Anna F1’ in a 3×3 factorial of spacing and topping treatments, with temperature and relative humidity maintained between 24 °C-32 °C and 63%-80% during the experiment. Orientation and row spacing have been explicitly treated as design variables in Chinese solar greenhouses, where north-south versus east-west orientations and 1.4-1.8 m row spacings were compared to analyze effects on canopy light interception, growth, and yield (Li et al., 2024). Fertilization and soil or substrate

management regimes are usually standardized within each trial; for example, integrated fertilization-tillage experiments in greenhouses applied defined NPK formulations and organic amendments across rotary tillage and plowing systems, and quantified biometric and yield traits under each treatment (Avasiloaiei et al., 2025). Together, such designs provide a template for specifying cultivars, structure, and management in temperature-yield modeling studies.

5.2 Temperature monitoring technologies and sensor deployment

Capturing temperature-yield relationships requires dense, reliable microclimate measurements rather than single-point records. Multi-year monitoring in commercial tomato greenhouses deployed multiple temperature-humidity sensors within the crop canopy, revealing spatial gradients up to about 3 °C in daily mean temperature and 0.6 kPa in vapour pressure deficit between locations, and linking these to local differences in stem and fruit growth (Šalagovič et al., 2024). Sensor networks are increasingly wireless: several studies describe custom wireless nodes or IoT platforms integrating temperature, relative humidity and sometimes CO_2 and light sensors, distributed at multiple horizontal positions and heights to resolve microclimate structure (Kolapkar and Sayyad, 2021). Such systems reduce cabling, facilitate relocation of nodes, and have been shown to detect microclimate layers between lower and upper canopy, as well as climate disturbances near walls or vents.

Recent work has combined distributed sensing with data fusion and model-based indices. In an Iranian commercial tomato greenhouse, a grid of 20 LoRaWAN wireless sensor nodes was installed on two horizontal planes at different heights, while an external weather station recorded outdoor conditions. Sensor calibration and validation were conducted offline in MATLAB/Simulink, and microclimate data were translated into an "optimality degree" index between 0 and 1 for temperature, RH and VPD, enabling direct assessment of how far local conditions deviated from crop comfort zones. Other wireless monitoring systems integrated fruit diameter sensors with 802.15.4-based temperature and radiation nodes and transmitted data via GPRS, achieving mean absolute temperature differences of only about 0.6 °C compared with wired systems, and data loss below 1%. Complementary approaches, such as compliant "plant wearables" measuring temperature and humidity directly on leaf surfaces, illustrate emerging options for ultra-localized microclimate characterization within greenhouse crops (Nassar et al., 2018).

5.3 Yield data collection and statistical preprocessing methods

Tomato yield data in greenhouse experiments are generally collected at plant or area level using standardized protocols, then subjected to statistical analysis and, in modeling studies, further preprocessing. Many agronomic trials quantify number of fruits per plant, individual fruit weight and total yield (e.g., $t \cdot ha^{-1}$ or $g \cdot plant^{-1}$) at one or more harvests, often alongside traits such as fruit size, firmness, soluble solids, and dry matter (Avasiloaiei et al., 2025; Ugbe et al., 2025). In cultivar or spacing-topping trials under greenhouse conditions, randomized or randomized complete block designs with three or more replications are analyzed using analysis of variance, with significance judged at $p < 0.05$ and treatment means separated by least significant difference or similar procedures. Microclimate-growth studies add growth rates of stems and fruits or truss mass at harvest, relating these to local temperature and VPD conditions over defined periods.

For data-driven modeling of temperature-yield relationships, more elaborate preprocessing is required. Yield-prediction studies using artificial neural networks and other machine-learning methods typically assemble datasets that combine environmental descriptors, management variables, and yield as inputs and outputs, then partition data into training and validation sets (Peng et al., 2023). A recent solar-greenhouse study compiled 390 datasets across multiple regions, each including planting density, organic and inorganic N, P, K rates, and effective accumulated temperature, with greenhouse tomato yield as the response; these variables were scaled and classified into different soil fertility levels before being used in neural-network models. In UAV-based yield prediction, ultra-high-resolution imagery was processed into hundreds of plant-level variables (e.g., means and higher-order statistics of vegetation indices), then reduced using feature-selection algorithms before model fitting (Tatsumi et al., 2021). Across these approaches, standard error metrics such as mean squared error, mean absolute error and coefficient of determination are calculated to evaluate predictive performance and to support sensitivity analysis

of which environmental and management factors, including temperature descriptors, most strongly influence modeled greenhouse tomato yield.

6 Modeling Approaches for Temperature-Yield Relationships

6.1 Statistical regression models for yield prediction

Statistical regression remains a fundamental approach to quantifying relationships between temperature variables and tomato yield or its components in controlled environments. In greenhouse cherry tomato under prolonged heat stress, polynomial regression was used to relate external weather (solar radiation, maximum and minimum temperature) to in-house temperature and humidity, forming a climate sub-model that then fed a growth-yield model comparing heat-resilient and heat-sensitive accessions (Kim et al., 2025). This type of regression framework allows explicit estimation of how projected temperature increases of 1 °C-8 °C and longer hot seasons modify yield, and highlights contrasting harvest index responses between genotypes under future climate scenarios.

Regression has also been embedded in broader yield-prediction pipelines as a relatively transparent, data-efficient alternative to complex AI models. In industrial tomato, a platform evaluated multiple algorithms and ultimately selected Ridge regression to predict open-field yield from hybrid and in-season environmental data, achieving prediction errors acceptable to producers and demonstrating that linear penalized models can capture much of the climate-yield signal when sufficient multisite data are available (Kasimatis et al., 2025). Polynomial and multivariate linear regression have similarly been used to approximate nonlinear links between external climate and greenhouse temperature and humidity, with R^2 values above 0.8-0.9 for maximum and minimum temperature, providing statistically robust climate inputs for subsequent tomato yield modeling under both control and heat conditions.

6.2 Process-based crop growth models

Process-based crop models represent temperature effects mechanistically through development rates, photosynthesis, and dry-matter partitioning. The TOMGRO model, a dynamic, source-sink framework based on differential equations, simulates initiation, expansion and senescence of leaves, stems and fruits in response to greenhouse temperature, CO_2 and light; calibration in controlled environments showed that TOMGRO can accurately reproduce observed differences in growth and yield under contrasting temperature regimes, making it suitable for environment-control decisions. Extensions and related models such as TOMSIM and Vanthoor's greenhouse climate-yield model further decompose temperature impacts on processes including truss appearance rate, fruit growth period and dry-matter partitioning, and have been validated across locations with near-optimal and non-optimal temperature and radiation conditions (Figure 3) (Gong et al., 2021).

Cardinal-temperature-driven models refine these process representations by explicitly encoding temperature thresholds for phenology and yield formation. The CROPGRO-Tomato model was improved by updating species coefficients for cardinal temperatures governing pre- and post-anthesis development, leaf appearance, photosynthesis, fruit set and fruit growth, based on recent controlled-temperature experiments (Boote et al., 2012). Recalibration and evaluation against multi-site field data substantially reduced RMSE for leaf area index, fruit number, biomass and fruit dry weight, resulting in Willmott d indices above 0.92 and enabling more reliable prediction of tomato growth and yield responses to temperature change. More recently, an integrated greenhouse yield prediction model combined TOMGRO and Vanthoor structures, using sensitivity analysis and Bayesian optimization to adapt parameters to specific facilities; when tested against four years of greenhouse data, the integrated model produced much lower RMSE than either parent model, indicating that hybridization of process-based schemes can improve robustness under varying temperature regimes (Lin et al., 2019).

6.3 Machine learning and artificial intelligence approaches

Machine learning and AI approaches increasingly complement or replace traditional models for predicting tomato yield or temperature-driven intermediates. A systematic review of tomato-yield ML models found that about two-thirds of best-performing approaches were deep-learning based, with LSTM, generic artificial neural

networks and support vector regression most frequently used when combining climate, soil, plant growth, fertilization and irrigation variables; random forest regression was particularly effective when using image-derived vegetation indices (Odah et al., 2025). In greenhouse applications, an ANN model trained on farm-level energy and input data outperformed multiple linear regression for predicting tomato yield, and sensitivity analysis identified key production factors, illustrating how neural networks can capture nonlinear interactions among management and environment beyond simple temperature terms (Belouz et al., 2022).

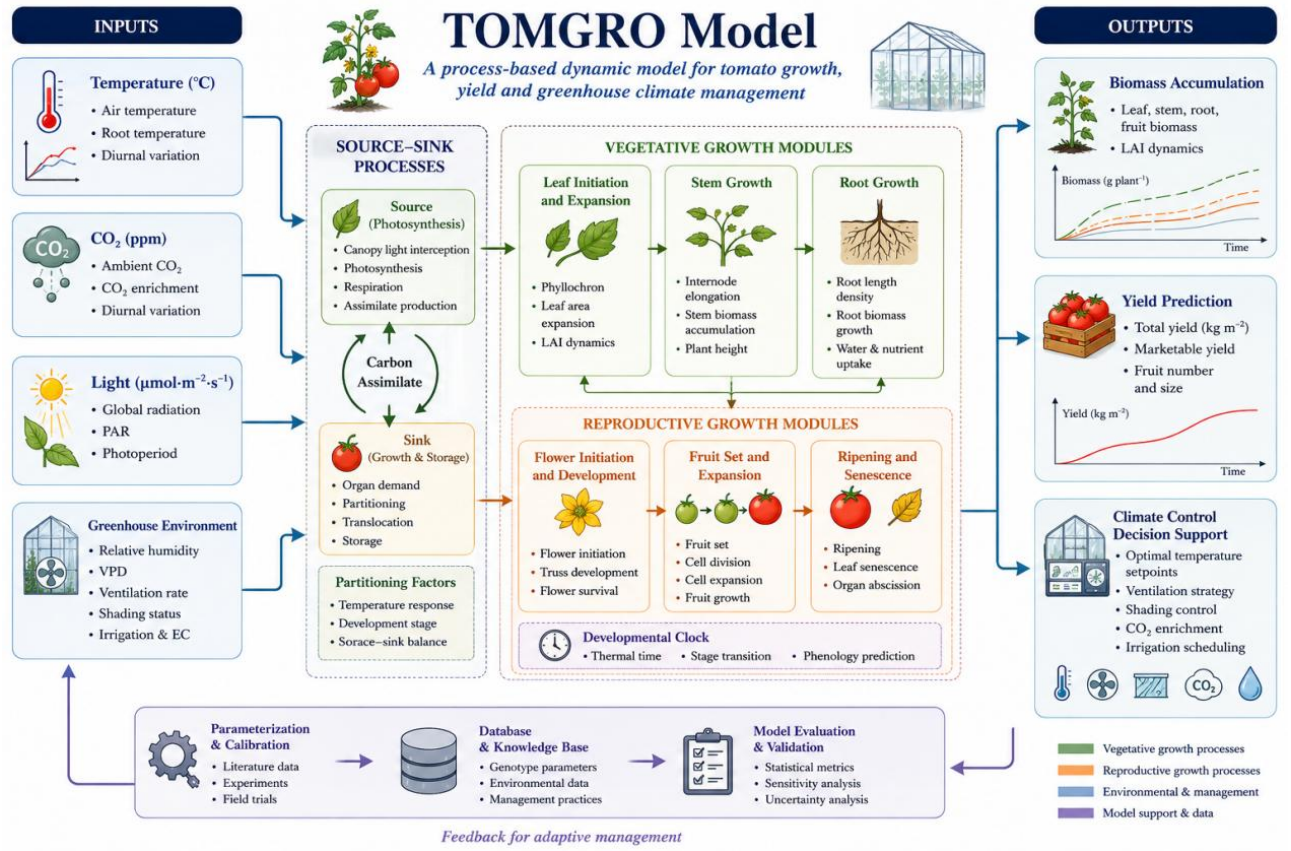


Figure 3 Framework of the TOMGRO process-based tomato growth model showing environmental inputs, source-sink interactions, and yield prediction outputs under greenhouse conditions

Recent studies focus more directly on greenhouse climate-yield linkages and short-term forecasting. A hybrid deep-learning framework combining temporal convolutional networks and recurrent neural networks was developed to predict tomato yield from time series of greenhouse microclimate variables (temperature, CO_2, humidity deficit, radiation) and historical yields, outperforming traditional ML and other deep architectures in RMSE across multiple commercial datasets (Gong et al., 2021). Another RNN-LSTM-based system predicted in-house temperature from external climate, then converted predicted temperatures to growing degree days and drove sigmoid growth models for leaf area index, fruit fresh weight and dry matter, achieving R^2 above 0.80 even when using forecasted rather than observed temperatures (Lin et al., 2024). Model-fusion strategies now integrate biophysical models such as reduced TOMGRO with CNN-RNN predictors, with neural-network fusion delivering higher yield-prediction accuracy than either component alone when applied to multi-year greenhouse temperature and CO_2 records.

7 Model Calibration, Validation, and Performance Evaluation

7.1 Parameter optimization and model calibration

Calibration of tomato temperature-yield models generally focuses on a limited set of influential physiological or empirical parameters, while less sensitive parameters are fixed from literature or prior studies. An integrated greenhouse tomato yield model combining TOMGRO and Vanthoor used an extended Fourier amplitude sensitivity test (EFAST) to classify parameters into optimized, fixed and ignored groups, thereby reducing

dimensionality before calibration (Lin et al., 2019). The remaining parameters were then optimized with Bayesian optimization using multi-year greenhouse data, resulting in an RMSE of 2.60 for yield compared with values above 17 for the individual component models, indicating a much closer match between simulated and observed yields.

Dynamic process-based models follow similar workflows but employ different optimization tools. The DSSAT-CROPGRO-Tomato model calibrated its genetic and management parameters with the GLUE framework, achieving average relative errors around 3%-5% for phenology, plant height and yield dry weight under varying water and nitrogen supplies (Shan et al., 2025). For the HORTSYST model, global sensitivity analysis with Sobol's method first identified nine key parameters controlling photo-thermal time, dry matter production and transpiration; these were then calibrated with a differential evolution algorithm, yielding RMSE values close to zero for leaf area index, nitrogen uptake and dry matter in two greenhouse seasons (Martínez-Ruíz et al., 2021). In reduced TOMGRO, three evolutionary algorithms-genetic algorithm, particle swarm optimization and differential evolution-were compared for calibrating 14 key parameters using multi-year greenhouse datasets; performance was judged from RMSE, relative RMSE and MAE between measured and simulated mature fruit dry matter (Gong et al., 2021).

7.2 Validation using multi-season and multi-region datasets

Robust temperature-yield models must be validated beyond the calibration environment, using independent seasons and, when possible, contrasting regions. The integrated TOMGRO-Vanthoor yield model was verified against four years of greenhouse data, showing consistently high performance across variable environmental conditions, which supports its intended generality under changing greenhouse climates (Lin et al., 2019). AquaCrop was calibrated and validated for greenhouse tomato under full and deficit irrigation; the model reproduced fresh yield, biomass and water productivity for both treatments, and was then used with 30 years of historical weather data to simulate yield responses to external temperature changes, effectively extending validation across multiple climatic years (Locatelli et al., 2024).

Model-based greenhouse design work with the Vanthoor yield model explicitly demonstrated cross-region validity. After implementing temperature effects from a broad literature survey, the model was validated for four temperature regimes in the Netherlands and southern Spain, reproducing yields under both near-optimal and sub-optimal climates with varying light and CO_2. Data-driven yield prediction approaches also rely on multi-environment datasets: a neural-network model for solar greenhouse tomato was trained on 390 experiments from different Chinese regions and soil fertility levels, then evaluated separately within low, medium and high fertility classes to test its generalization across management and climatic gradients (Peng et al., 2023). For dynamic growth models such as HortSyst, autumn-winter and spring-summer greenhouse seasons were simulated, and good agreement for dry matter, nitrogen uptake and transpiration across both seasons indicated that calibrated parameters retained validity under distinct seasonal temperature regimes.

7.3 Evaluation indicators: RMSE, R^2, MAE, and model robustness

Quantitative evaluation of tomato temperature-yield models commonly relies on root mean square error (RMSE), mean (absolute) error, and coefficient of determination (R^2), often complemented by model efficiency or bias. In the integrated yield prediction model, RMSE for yield dropped from above 17 in TOMGRO and Vanthoor to 2.60 in the integrated version, reflecting a substantial improvement in predictive accuracy (Lin et al., 2019). HORTSYST calibration reported RMSE values for leaf area index, nitrogen uptake, dry matter and transpiration that were close to zero, together with high modeling efficiency, indicating that residuals were small relative to observed variability over two crop seasons. In greenhouse AquaCrop applications, RMSE and normalized RMSE were used to evaluate calibration and validation for fresh yield and biomass under full and deficit irrigation, with acceptable errors supporting subsequent use for long-term temperature impact assessment (Locatelli et al., 2024).

Machine-learning models for greenhouse processes and yield frequently add MAE and R^2 to characterize accuracy and robustness. A CatBoost-based model for tomato transpiration achieved $R^2 = 0.92$ over the whole growth stage,

while its RMSE and MAE were reduced by more than 70% relative to a traditional crop-coefficient approach, and further partitioning by growth stage decreased RMSE by up to 97% at night during fruiting, illustrating gains in stability under different regimes (Tong et al., 2023). A neural-network yield model for solar greenhouse tomato across fertility levels reported that an improved particle-swarm-optimized network produced the smallest MSE and MAE and the largest R^2 (up to about 0.94), outperforming baseline networks under low, medium, and high fertility and thus demonstrating robustness across diverse environmental and management contexts (Peng et al., 2023). Broader crop-yield and evapotranspiration studies similarly adopt RMSE, MAE and R^2 as core indicators, emphasizing their usefulness for comparing alternative algorithms and for assessing generalization to unseen seasons or regions.

8 Case Studies of Greenhouse Tomato Temperature-Yield Modeling

8.1 Case study in solar greenhouses under winter cultivation

Winter tomato production in northern China has been used to demonstrate how temperature-yield relationships can be modeled in solar greenhouses. In soft-shell solar greenhouses, dynamic monitoring of light, temperature, and humidity for six cherry and three large-fruited cultivars was combined with yield and quality measurements to build correlation and partial least-squares path models linking microclimate to cluster yield and Brix (Liu et al., 2025). These analyses showed that soft-shell structures raised average daily temperature by 10 °C-15 °C and reduced low-temperature stress duration by 25%, with cherry tomato yield proving more temperature-sensitive than large-fruited types.

A complementary modeling approach integrated a mechanistic climate model with a tomato yield module specifically for Chinese solar greenhouses. This open-source model was calibrated and validated against three experiments, including two commercial winter production greenhouses, and achieved an RMSE of about 1.6 °C for indoor air temperature and 0.61-0.71 $kg \cdot m^{-2}$ for yield, while sensitivity analysis highlighted air-exchange parameters and optimal leaf area index as key determinants of simulated winter yield (Zhou et al., 2025). Active solar heating systems provide another winter case: in paired Canarian greenhouses, an active solar heating installation improved nocturnal thermal conditions and increased total tomato yield by 55% during the cold season, illustrating the strong leverage of improved temperature profiles on winter productivity (Bazgaou et al., 2021).

8.2 Case study in plastic tunnel systems under high-temperature stress

Plastic and walk-in tunnel systems in warm climates offer clear examples of modeling and managing high-temperature impacts on tomato yield. In southern China, daily maximum temperature and mean relative humidity inside plastic greenhouses were simulated using an extreme learning machine to identify high-temperature-high-humidity (HTHH) events, and response surfaces were then used to relate event frequency and return period to tomato physiological losses, showing that flower bud differentiation was the most temperature-sensitive stage (Zhang et al., 2022). The analysis revealed that HTHH events mainly occurred from June to September and that high temperature played a larger role than humidity in reducing growth indicators, providing a quantitative basis for risk assessment and regional layout of plastic-house tomato.

Experimental work in arid regions has focused on modifying tunnel microclimates and quantifying associated yield responses. In late-summer trials comparing a shaded net tunnel, a net tunnel with fogging, and a plastic tunnel with evaporative cooling, all powered by solar energy, cooled tunnels significantly improved leaf area, chlorophyll content, cell membrane stability, and relative water content, while reducing physiological disorders such as sunscald and blossom-end rot (Sharaf-Eldin et al., 2023). These microclimate changes translated into about 31.5% higher marketable yield with evaporative cooling and 28.8% with fogging relative to open field, demonstrating how engineered temperature reductions within plastic systems can be directly linked to yield gains.

8.3 Comparative case study of intelligent greenhouse temperature control strategies

Recent case studies in intelligent greenhouses explicitly couple temperature control strategies with crop and profit models. In the second “Autonomous Greenhouse Challenge”, five AI-supported teams remotely operated high-tech cherry tomato compartments for six months, using sensor data and algorithms to determine temperature,

humidity, and CO_2 setpoints; all AI teams outperformed a reference human grower in net profit, with some teams applying higher early-season and end-season temperatures to accelerate development and ripening while still achieving better heat-use efficiency (Hemming et al., 2020). Analysis with a virtual greenhouse "digital twin" linked these distinct temperature trajectories to differences in yield and resource use, offering a comparative benchmark for data-driven climate strategies.

Model-based optimal control studies provide an additional perspective on intelligent temperature-yield management. A PSO-based model predictive control framework combined a greenhouse climate model with a biophysical yield model and optimized heating, ventilation, and lighting setpoints to maximize yield while minimizing energy costs, outperforming traditional control and genetic-algorithm-based MPC in both yield and energy efficiency in a tomato case study (Gong et al., 2023). At year-round scale, a rule-based MPC using external weather and month-averaged tomato prices to select temperature setpoints was compared with on/off control and open field; simulations for Beijing showed that only strategies jointly optimizing yield and energy cost achieved satisfactory profit, highlighting how economic and biophysical models must be integrated when evaluating intelligent temperature control options (Xu et al., 2024).

9 Discussion, Applications, and Future Perspectives

Existing temperature-yield models for greenhouse tomato are powerful decision-support tools but still face notable limitations and uncertainties. Process-based models often rely on a relatively small set of experiments for parameterization and may not fully capture the variability in low-technology or highly heterogeneous greenhouses, where temperature extremes and spatial microclimate variation are common. Global sensitivity and uncertainty analyses have shown that yield predictions can be highly sensitive to a few temperature-related parameters, such as thresholds for fruit abortion or growth inhibition, meaning that modest parameter errors can translate into large yield errors when conditions move outside the "ideal" range. Many models assume relatively uniform microclimate and well-controlled systems, so their validity can degrade in real commercial settings with imperfect heating and cooling, where optimality degrees for temperature and VPD fluctuate widely between seasons and locations. Furthermore, hybrid approaches that enhance process-based models with deep learning can boost accuracy, but they introduce their own sources of uncertainty related to training data representativeness, sensor noise, and potential overfitting, making model transfer to new greenhouses or future climates less certain unless explicitly tested.

Despite these limitations, temperature-yield models are increasingly embedded in precision agriculture and smart greenhouse platforms to support real-time management. Decision-support systems using crop water productivity models such as AquaCrop already leverage external temperature and long historical weather records to estimate greenhouse tomato yields and optimize irrigation under future climate scenarios. IoT-based microclimate monitoring frameworks quantify "optimality degrees" or comfort ratios for temperature, humidity and VPD, translating dense sensor data into simple indices that link directly to yield risk and guide heating and cooling strategies. Smart greenhouse platforms go further by combining wireless sensor networks, fuzzy or model-based controllers, and cloud dashboards, enabling automated ventilation, shading and irrigation tuned to maintain temperatures within crop-specific ranges. Integrated digital solutions that add machine learning yield predictors, disease recognition, and even fruit expansion analysis use historical temperature-humidity-yield relationships to recommend set-points and interventions, improving resource efficiency and stabilizing production. As these systems mature, temperature-yield models shift from purely research tools to operational components of climate control, irrigation scheduling, and energy optimization in commercial tomato production.

Future work on modeling temperature-yield relationships will likely be driven by convergence of digital twins, dense IoT sensing, and climate-adaptive algorithms. Greenhouse digital twins are beginning to integrate sensor networks, multivariate yield-forecasting models, and edge computing to provide continuous predictions of final yield based on evolving temperature and other climate variables, allowing growers to test "what-if" scenarios for alternative control strategies before applying them in the real house. Broader reviews of agricultural digital twins and smart farming envision virtual replicas that fuse crop models, weather forecasts, soil sensors, and aerial

imagery, creating platforms where temperature-yield modules for tomato become part of a larger cyber-physical production system. Knowledge-based data-driven approaches that couple calibrated process-based models with deep learning show one promising route: process models preserve interpretability and physiological realism, while neural networks and particle filtering correct systematic errors and adapt to new microclimates or management regimes. At the same time, IoT reviews highlight persistent challenges around sensor accuracy, interoperability, and deployment cost, suggesting that future climate-adaptive modeling must explicitly handle data quality, uncertainty propagation, and robust control under extreme events. In this context, next-generation temperature-yield models will need to be both explainable and self-updating, closing the loop between sensing, prediction, and actuation to support resilient, low-carbon greenhouse tomato systems under a changing climate.

Acknowledgments

I would like to thank the anonymous reviewers for their detailed review of the draft. Their specific feedback helped us correct the logical loopholes in our arguments.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abid H., Zghal O., Lajnef M., Ketata A., Zouari S., Gugliuzza G., Mejri M., Arrabito E., and Driss Z., 2024, Analysis of seasonal variations and their impact on the microclimate of soilless glass greenhouses: numerical and experimental investigations, Numerical Heat Transfer Part A: Applications, 86(4): 4576-4600.
https://doi.org/10.1080/10407782.2024.2320829

Alsamir M., Mahmood T., Trethowan R., and Ahmad N., 2020, An overview of heat stress in tomato (*Solanum lycopersicum* L.), Saudi Journal of Biological Sciences, 28(3): 1654-1663.
https://doi.org/10.1016/j.sjbs.2020.11.088

Avasiloaiei D., Calara M., Brezeanu P., Bălăiță C., Brumă I., and Brezeanu C., 2025, Optimizing tomato yield and quality in greenhouse cultivation through fertilization and soil management, Agronomy, 15(9): 2045.
https://doi.org/10.3390/agronomy15092045

Bazgaou A., Fatnassi H., Bouharroud R., Ezzaeri K., Gourdo L., Wifaya A., Demrati H., Elame F., Carreño-Ortega Á., Bekkaoui A., Aharoune A., and Bouirden L., 2021, Effect of active solar heating system on microclimate, development, yield and fruit quality in greenhouse tomato production, Renewable Energy, 165: 237-250.
https://doi.org/10.1016/j.renene.2020.11.007

Belouz K., Nourani A., Zereg S., and Bencheikh A., 2022, Prediction of greenhouse tomato yield using artificial neural networks combined with sensitivity analysis, Scientia Horticulturae, 291: 110666.
https://doi.org/10.1016/j.scienta.2021.110666

Boote K.J., Rybak M.R., Scholberg J.M.S., and Jones J.W., 2012, Improving the CROPGRO-Tomato model for predicting growth and yield response to temperature, HortScience, 47(8): 1038-1049.
https://doi.org/10.21273/HORTSCI.47.8.1038

Dash P., Guo B., and Leskovar D.I., 2023, Optimizing hydroponic management practices for organically grown greenhouse tomato under abiotic stress conditions, HortScience, 58(11): 1378-1386.
https://doi.org/10.21273/HORTSCI17249-23

Efeta B., D'arc U., Claude S., Pancras N., and Jonathan M., 2025, The influence of temperature difference on crop physiological process: Systematic growth analysis of *Solanum lycopersicum* (tomatoes) in both greenhouse and open field, East African Journal of Agriculture and Biotechnology, 8(2): 1-13.
https://doi.org/10.37284/eajab.8.2.3973

Fanourakis D., Tsaniklidis G., Makraki T., Nikoloudakis N., Bartzanas T., Sabatino L., Fatnassi H., and Ntatsi G., 2025, Climate change impacts on greenhouse horticulture in the Mediterranean Basin: Challenges and adaptation strategies, Plants, 14(21): 3390.
https://doi.org/10.3390/plants14213390

Flores-Velázquez J., Rojano F., Aguilar-Rodríguez C., Villagran E., and Villarreal-Guerrero F., 2022, Greenhouse thermal effectiveness to produce tomatoes assessed by a temperature-based index, Agronomy, 12(5): 1158.
https://doi.org/10.3390/agronomy12051158

Ghabileh M., Lotfi M., Aliniaeifard S., and Ramshini H., 2024, Variation in reproductive organ functionality among a population of tomato genotypes reveals the importance of pollen viability and fruit set in response to heat stress, International Journal of Vegetable Science, 30(6): 717-731.
https://doi.org/10.1080/19315260.2024.2429118

Gong L., Yu M., and Kollias S., 2023, Optimizing crop yield and reducing energy consumption in greenhouse control using PSO-MPC algorithm, Algorithms, 16(5): 243.
https://doi.org/10.3390/a16050243

Gong L., Yu M., Jiang S., Cutsuridis V., and Pearson S., 2021, Deep learning based prediction on greenhouse crop yield combined TCN and RNN, Sensors, 21(13): 4537.
https://doi.org/10.3390/s21134537

Harel D., Fadida H., Slepoy A., Gantz S., and Shilo K., 2014, The effect of mean daily temperature and relative humidity on pollen, fruit set and yield of tomato grown in commercial protected cultivation, Agronomy, 4(1): 167-177.
https://doi.org/10.3390/agronomy4010167

Hemming S., Zwart F., Elings A., Petropoulou A., and Righini I., 2020, Cherry tomato production in intelligent greenhouses-Sensors and AI for control of climate, irrigation, crop yield, and quality, Sensors, 20(22): 6430.
https://doi.org/10.3390/s20226430

Higashide T., 2022, Review of dry matter production and growth modelling to improve the yield of greenhouse tomatoes, The Horticulture Journal, 91(2): 143-157.
https://doi.org/10.2503/hortj.UTD-R019

Kasimatis C., Psomakelis E., Katsenios N., Papatheodorou M., Apostolou D., and Efthimiadou A., 2025, Industrial tomato yield prediction using machine learning models, Smart Agricultural Technology, 11: 100920.
https://doi.org/10.1016/j.atech.2025.100920

Kim S., Jeong J., and Kim S., 2025, Morphological analysis-based yield modeling in greenhouse grown cherry tomato (*Solanum lycopersicum*) under prolonged heat stress, Frontiers in Plant Science, 16: 1730694.
https://doi.org/10.3389/fpls.2025.1730694

Kolapkar M.S., and Sayyad S.R., 2021, Greenhouse microclimate study for humidity, temperature and soil moisture using agricultural wireless sensor network system, Advances in Communication and Computational Technology, 668: 278-289.
https://doi.org/10.1007/978-981-16-0493-5_25

Kürklü A., Pearson S., and Felek T., 2025, Climate change impacts on tomato production in high-tech soilless greenhouses in Türkiye, BMC Plant Biology, 25(1): 307.
https://doi.org/10.1186/s12870-025-06307-1

Lee K., Rajametov S., Jeong H., Cho M., Lee O., Kim S., Yang E., and Chae W., 2022, Comprehensive understanding of selecting traits for heat tolerance during vegetative and reproductive growth stages in tomato, Agronomy, 12(4): 834.
https://doi.org/10.3390/agronomy12040834

Li Y., Henke M., Zhang D., Wang C., and Wei M., 2024, Optimized tomato production in Chinese solar greenhouses: The impact of an east-west orientation and wide row spacing, Agronomy, 14(2): 314.
https://doi.org/10.3390/agronomy14020314

Li Y., Jian Y., Wang S., Liu X., Li W., Arıcı M., Zhang L., Li W., and Cao Y., 2024, Spatial temperature distribution and ground thermal storage in the plastic greenhouse: An experimental and modeling study, Journal of Energy Storage, 75: 109938.
https://doi.org/10.1016/j.est.2023.109938

Lin D., Wei R., and Xu L., 2019, An integrated yield prediction model for greenhouse tomato, Agronomy, 9(12): 873.
https://doi.org/10.3390/agronomy9120873

Lin Y., Fang S., Kang L., Chen C., Yao M., and Kuo B., 2024, Combining recurrent neural network and sigmoid growth models for short-term temperature forecasting and tomato growth prediction in a plastic greenhouse, Horticulturae, 10(3): 230.
https://doi.org/10.3390/horticulturae10030230

Liu H., Shao M., and Yang L., 2023, Photosynthesis characteristics of tomato plants and its' responses to microclimate in new solar greenhouse in North China, Horticulturae, 9(2): 197.
https://doi.org/10.3390/horticulturae9020197

Liu H., Zhao H., Liu S., Tian Y., Li W., Wang B., Hu X., Sun D., Wang T., Wu S., Wang F., Zhu N., Tao Y., and Lei X., 2025, When tomatoes hit the winter: A counterattack to overwinter production in soft-shell solar greenhouses in North China, Horticulturae, 11(4): 436.
https://doi.org/10.3390/horticulturae11040436

Locatelli S., Barrera W., Verdi L., Nicoletto C., Marta D., and Maucieri C., 2024, Modelling the response of tomato on deficit irrigation under greenhouse conditions, Scientia Horticulturae, 324: 112770.
https://doi.org/10.1016/j.scienta.2023.112770

Miller G., Beery A., Singh P., Wang F., Zelingher R., Motenko E., and Lieberman-Lazarovich M., 2021, Contrasting processing tomato cultivars unlink yield and pollen viability under heat stress, AoB Plants, 13(4): plab046.
https://doi.org/10.1093/aobpla/plab046

Nassar J.M., Khan S.M., Villalva D.R., Nour M.M., Almuslem A.S., and Hussain M.M., 2018, Compliant plant wearables for localized microclimate and plant growth monitoring, npj Flexible Electronics, 2(1): 24.
https://doi.org/10.1038/s41528-018-0039-8

Lin D., Wei R., and Xu L., 2019, An integrated yield prediction model for greenhouse tomato, Agronomy, 9(12): 873.
https://doi.org/10.3390/agronomy9120873

Lin Y., Fang S., Kang L., Chen C., Yao M., and Kuo B., 2024, Combining recurrent neural network and sigmoid growth models for short-term temperature forecasting and tomato growth prediction in a plastic greenhouse, Horticulturae, 10(3): 230.
https://doi.org/10.3390/horticulturae10030230

Liu H., Shao M., and Yang L., 2023, Photosynthesis characteristics of tomato plants and its' responses to microclimate in new solar greenhouse in North China, Horticulturae, 9(2): 197.
https://doi.org/10.3390/horticulturae9020197

Liu H., Zhao H., Liu S., Tian Y., Li W., Wang B., Hu X., Sun D., Wang T., Wu S., Wang F., Zhu N., Tao Y., and Lei X., 2025, When tomatoes hit the winter: A counterattack to overwinter production in soft-shell solar greenhouses in North China, Horticulturae, 11(4): 436.
https://doi.org/10.3390/horticulturae11040436

Locatelli S., Barrera W., Verdi L., Nicoletto C., Marta D., and Maucieri C., 2024, Modelling the response of tomato on deficit irrigation under greenhouse conditions, Scientia Horticulturae, 325: 112770.
https://doi.org/10.1016/j.scienta.2023.112770

Miller G., Beery A., Singh P., Wang F., Zelingher R., Motenko E., and Lieberman-Lazarovich M., 2021, Contrasting processing tomato cultivars unlink yield and pollen viability under heat stress, AoB Plants, 13(4): plab046.
https://doi.org/10.1093/aobpla/plab046

Nassar J., Khan S., Villalva D., Nour M., Almuslem A., and Hussain M., 2018, Compliant plant wearables for localized microclimate and plant growth monitoring, npj Flexible Electronics, 2(1): 1-12.
https://doi.org/10.1038/s41528-018-0039-8

Nițu O., Ivan E., and Arshad A., 2025, Optimizing microclimatic conditions for lettuce, tomatoes, carrots, and beets: Impacts on growth, physiology, and biochemistry across greenhouse types and climatic zones, International Journal of Plant Biology, 16(3): 100.
https://doi.org/10.3390/ijpb16030100

Odah K., Houetohossou S., Houndji V., and Kakaï R., 2025, Machine learning techniques for tomato yield prediction: A comprehensive analysis, Smart Agricultural Technology, 11: 101067.
https://doi.org/10.1016/j.atech.2025.101067

Ogunlowo Q., Akpenpuun T., Na W., Rabiu A., Adesanya M., Addae K., Kim H., and Lee H., 2021, Analysis of heat and mass distribution in a single- and multi-span greenhouse microclimate, Agriculture, 11(9): 891.
https://doi.org/10.37473/dac/10.3390/agriculture11090891

Park B., Jeong H., Yang E., Kim M., Kim J., Chae W., Lee O., Kim S., and Kim S., 2023, Differential responses of cherry tomatoes (*Solanum lycopersicum*) to long-term heat stress, Horticulturae, 9(3): 343.
https://doi.org/10.3390/horticulturae9030343

Peng X., Yu X., Luo Y., Chang Y., Lu C., and Chen X., 2023, Prediction model of greenhouse tomato yield using data based on different soil fertility conditions, Agronomy, 13(7): 1892.
https://doi.org/10.3390/agronomy13071892

Rajametov S., Yang E., Jeong H., Cho M., Chae S., and Paudel N., 2021, Heat treatment in two tomato cultivars: A study of the effect on physiological and growth recovery, Horticulturae, 7(5): 119.
https://doi.org/10.3390/horticulturae7050119

Rezvani S., Abyane H., Shamshiri R., Balasundram S., Dworak V., Goodarzi M., Sultan M., and Mahns B., 2020, IoT-based sensor data fusion for determining optimality degrees of microclimate parameters in commercial greenhouse production of tomato, Sensors, 20(22): 6474.
https://doi.org/10.3390/s20226474

Ro S., Chea L., Ngoun S., Stewart Z., Roeurn S., Theam P., Lim S., Sor R., Kosal M., Roeun M., Dy K., and Prasad P., 2021, Response of tomato genotypes under different high temperatures in field and greenhouse conditions, Plants, 10(3): 449.
https://doi.org/10.3390/plants10030449

Šalagovič J., Vanhees D., Verboven P., Holsteens K., Verlinden B., Huysmans M., Van De Poel B., and Nicolaï B., 2024, Microclimate monitoring in commercial tomato (*Solanum lycopersicum* L.) greenhouse production and its effect on plant growth, yield and fruit quality, Frontiers in Horticulture, 3: 1425285.
https://doi.org/10.3389/fhort.2024.1425285

Sato S., Kamiyama M., Iwata T., Makita N., Furukawa H., and Ikeda H., 2006, Moderate increase of mean daily temperature adversely affects fruit set of *Lycopersicon esculentum* by disrupting specific physiological processes in male reproductive development, Annals of Botany, 97(5): 731-738.
https://doi.org/10.1093/aob/mcl037

Sellami D., and Kooli S., 2026, Physiological and growth responses of tomato plants to heat stress, Discover Plants, 3(1): 1-15.
https://doi.org/10.1007/s44372-025-00462-3

Shan Z., Chen J., Zhang X., Si Z., Yi R., and Fan H., 2025, Optimizing irrigation and nitrogen application for greenhouse tomato using the DSSAT-CROPGRO-Tomato model, Water, 17(3): 426.
https://doi.org/10.3390/w17030426

Sharaf-Eldin M., Yaseen Z., Elmetwalli A., Elsayed S., Scholz M., Al-Khafaji Z., and Omar G., 2023, Modifying walk-in tunnels through solar energy, fogging, and evaporative cooling to mitigate heat stress on tomato, Horticulturae, 9(1): 77.
https://doi.org/10.3390/horticulturae9010077

Sun W., Coules A., Zhao C., and Lu C., 2025, A lettuce growth model responding to a broad range of greenhouse climates, Biosystems Engineering, 251: 1-16.
https://doi.org/10.1016/j.biosystemseng.2025.01.008

Talukder M., All N., Bappy H., Haque M., Abul M., Molla H., Alam M., Mosharaf M., Limon S., and Quzzaman S., 2025, Fluctuation of ambient day-night temperature influences morphological traits, floral characters, fruit yield and quality of summer tomato genotypes grown in hydroponics, New Zealand Journal of Crop and Horticultural Science, 53(4): 2731-2754.
https://doi.org/10.1080/01140671.2025.2504209

Tatsumi K., Igarashi N., and Xiao M., 2021, Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery, Plant Methods, 17(1): 27.
https://doi.org/10.1186/s13007-021-00761-2

Tong Z., Zhang S., Yu J., Zhang X., Wang B., and Zheng W., 2023, A hybrid prediction model for CatBoost tomato transpiration rate based on feature extraction, Agronomy, 13(9): 2371.
https://doi.org/10.3390/agronomy13092371

Ugbe L., Ushie P., Morebise A., and Akomaye F., 2025, Assessing the impact of climate change on the growth and yield of tomato (*Lycopersicon esculentum*) cultivars in Obudu, northern Cross River State, Nigeria, World Journal of Advanced Research and Reviews, 28(1): 3508.
https://doi.org/10.30574/wjarr.2025.28.1.3508

Xu D., Xu L., Wang S., Wang M., Jin J., and Shi C., 2024, Rule-based year-round model predictive control of greenhouse tomato cultivation: A simulation study, Information Processing in Agriculture, 12(2): 356-370.
https://doi.org/10.1016/j.inpa.2024.11.001

Xu K., Guo X., He J., Yu B., Tan J., and Guo Y., 2022, A study on temperature spatial distribution of a greenhouse under solar load with considering crop transpiration and optical effects, Energy Conversion and Management, 266: 115277.
https://doi.org/10.1016/j.enconman.2022.115277

Yadav D., Meena Y., Bairwa L., Singh U., Bairwa S., Choudhary M., and Singh A., 2021, Morphological, physiological and biochemical response to low temperature stress in tomato (*Solanum lycopersicum* L.): A review, International Journal of Bio-resource and Stress Management, 12(5): 462-471.
https://doi.org/10.23910/1.2021.2480

Yadav R., Kumar R., Kalia P., Jain V., and Varshney R., 2014, Effect of high day and night temperature regimes on tomato (*Solanum lycopersicum*) genotypes, Indian Journal of Agricultural Sciences, 84(2): 228-233.
https://doi.org/10.56093/ijas.v84i2.38052

Zepeda A., Vorage S., Van Mourik S., Heuvelink E., and Marcelis L., 2026, Too cold or too warm? Modelling seed set and fruit mass based on the effect of temperature on pollen quality, AoB Plants, 18(1): plag004.
https://doi.org/10.1093/aobpla/plag004

Zhang H., Sun X., and Song W., 2023, Physiological and growth characteristics of tomato seedlings in response to low root-zone temperature, HortScience, 58(5): 596-603.
https://doi.org/10.21273/hortsci16924-22

Zhang Q., Zhang X., Yang Z., Huang Q., and Qiu R., 2022, Characteristics of plastic greenhouse high-temperature and high-humidity events and their impacts on facility tomatoes growth, Frontiers in Earth Science, 10: 848924.
https://doi.org/10.3389/feart.2022.848924

Zhou B., Lastiri D., Wang N., Yang Q., and Van Henten E., 2025, An opensource indoor climate and yield prediction model for Chinese solar greenhouses, Biosystems Engineering, 250: 244-262.
https://doi.org/10.1016/j.biosystemseng.2024.12.007