

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-Ruiz 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 CO₂. 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, R², 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 (R²), 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 R² to characterize accuracy and robustness. A CatBoost-based model for tomato transpiration achieved R² = 0.92 over the whole growth stage,