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