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).

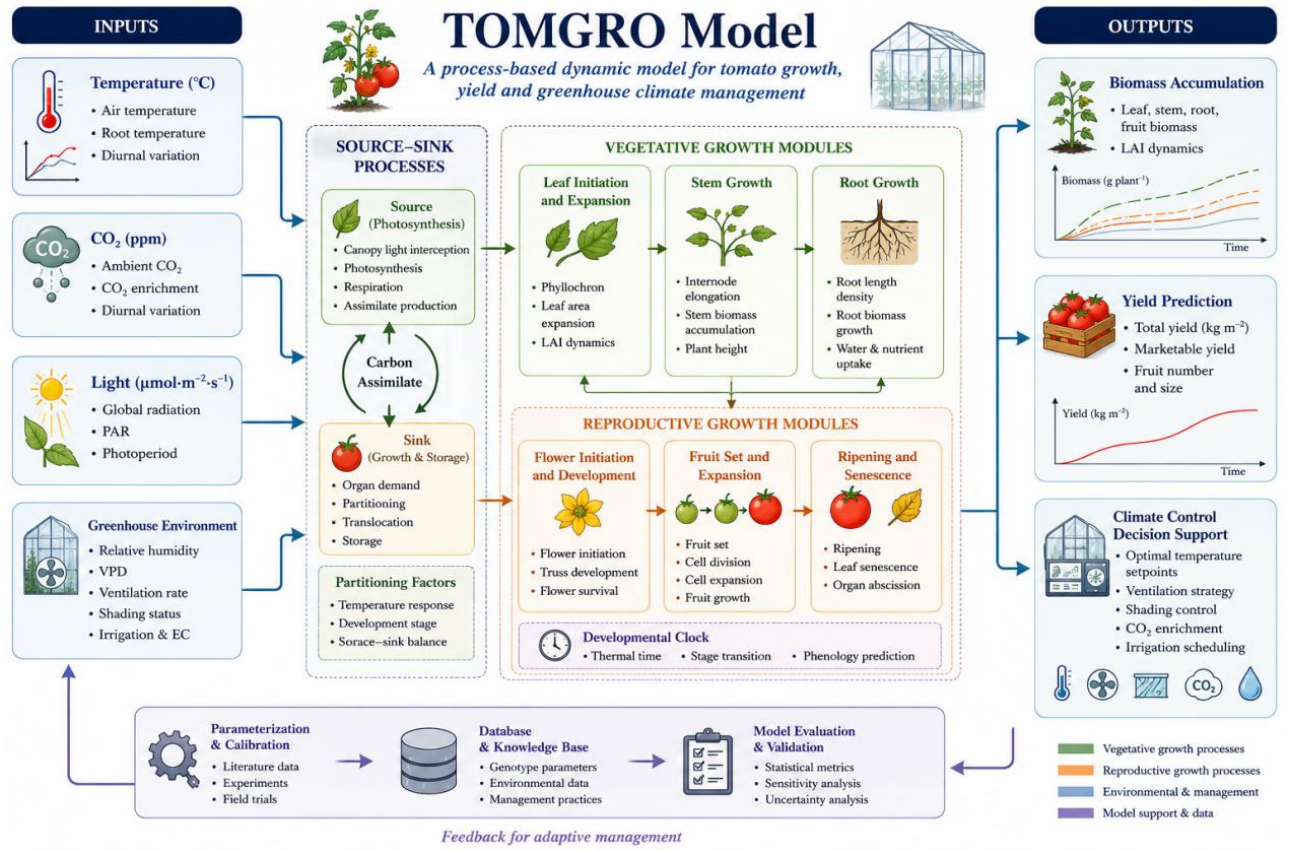

Figure 3 Framework of the TOMGRO process-based tomato growth model showing environmental inputs, source-sink interactions, and yield prediction outputs under greenhouse conditions

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