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