

historical range. They also explain association better than mechanism. So they are useful for local forecasting and first-pass diagnosis, but they rarely suffice for scenario analysis on their own (Tolosa et al., 2023).

4.2 Process-based crop simulation models

Process-based crop models are the backbone of much sorghum climate-impact research because they translate weather into plant development through explicit biological rules. APSIM and DSSAT-CERES-Sorghum are the most widely used in the literature reviewed here, with AquaCrop often used for irrigation and water-productivity questions. Their main attraction is interpretability: they can represent thermal-time driven phenology, soil-water balance, biomass production, and yield formation in a way that makes adaptation experiments possible. That is why they are especially useful for testing cultivar maturity, sowing date, fertilizer response, supplemental irrigation, and trait ideotypes under future climates. APSIM-based studies in Ethiopia and Mali, for example, have been used to characterize drought patterns, explore genotype \times environment \times management interactions, and identify where trait changes or sowing shifts might reduce risk. DSSAT-based studies in Ethiopia have simulated both future yield decline and the performance of adaptation packages under SSP scenarios. The cost of this power is data demand and calibration effort. A crop model can formalize biology beautifully and still perform poorly if soils, varieties, or management are mis-specified (Tirfessa et al., 2023; Diancoumba et al., 2024; Gardi et al., 2025; Ali and Kothari, 2026).

A general process-based expression of sorghum yield prediction can be written as:

$$Y=f(T,R,S,G,M,\epsilon)$$

where Y is yield, T is the temperature regime, R is rainfall and soil water supply, S is soil condition, G is genotype, M is management, and ϵ captures unobserved variation. In empirical models, f is usually a fitted statistical relation. In crop simulation models, f is decomposed into linked sub-processes such as phenology, transpiration, biomass accumulation, and partitioning.

4.3 Remote sensing applications

Remote sensing expands sorghum yield modeling by observing the crop directly across space rather than relying only on weather and field samples. The main value of remote sensing lies in its ability to capture canopy status, vegetation indices, spatial heterogeneity, and sometimes stage-specific crop responses that are difficult to measure manually at large scale. Recent sorghum studies show that multispectral imagery from satellites or UAVs can support reasonably strong yield prediction when aligned with key phenological stages. In tropical environments, artificial neural network models built from vegetation indices and soil elevation data reached strong performance in estimating sorghum grain yield, while arid-region UAV studies found that integrating multispectral and meteorological data can predict yield with high accuracy and reveal which growth stage contributes most to predictive skill. A recurring theme is that timing matters: the best observation date is not necessarily the latest one, because some stages carry more information about final yield formation than others. Remote sensing therefore works best when it is phenology-aware, not just image-rich (Ferraz et al., 2024; Deng et al., 2025).

4.4 Machine learning and artificial intelligence approaches

Machine learning has become increasingly attractive in sorghum yield prediction because sorghum systems are shaped by non-linear interactions among climate, soils, management, and canopy signals. Algorithms such as random forests, gradient boosting, support vector machines, artificial neural networks, and stacking ensembles can absorb high-dimensional predictor sets and model interactions that are difficult to specify mechanically. Recent sorghum applications illustrate both the promise and the limits of this approach. In South Sudan, machine-learning models combining yield, climate, remote sensing, and conflict-probability data produced useful end-of-season yield predictions, with XGBoost, decision tree, and random forest performing especially well. In tropical and arid experiments, neural networks and ensemble approaches also produced strong fits. But these models can become opaque, and their success depends greatly on training data coverage and quality. When extrapolation is required, or when the user needs biological explanation rather than prediction alone, machine learning is strongest when paired with domain knowledge rather than treated as a black box (Ferraz et al., 2024; Javed and Murad, 2024; Deng et al., 2025; Karongo et al., 2025).