PREDICTING TOMATO YIELD IN CENTRAL EUROPE THROUGH INTEGRATION OF FIELD EXPERIMENTS, DSSAT-CROPGRO-TOMATO SIMULATIONS AND MACHINE LEARNING

Vera POTOPOVÁ[1,2]*, Tudor TRIFAN[1], Miroslav TRNKA[2], Pavel ZAHRADNÍČEK[2], Petr ŠTĚPÁNEK[2], Josef SOUKUP[1], Anxhela HAMETI[1]

[1] Department of Agroecology and Crop Production, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague (ČZU), Praha-Suchdol, Czech Republic

[2] Global Change Research Institute of the Czech Academy of Sciences, Brno, Czech Republic

*Corresponding author: potop@af.czu.cz

https://doi.org/10.52757/bsd26.55

Background: Tomato (*Solanum lycopersicum* L.) is a major vegetable crop worldwide, valued for its nutritional quality, rich in antioxidants such as lycopene, and its economic significance.

The aim of this study was to test three high-yielding tomato varieties (drought-tolerant Cocktail Crush F1, heat-sensitive Nagina F1, and the stable line Momini Salzi) at the Ostra site (2023–2025) to develop models for climate adaptation, disease risk management, and production efficiency.

Materials and methods: The Ostra site features an automated weather station operated by the Department of Agro ecology and Crop Production (ČZU), which provides continuous, canopy-level monitoring of daily meteorological conditions directly within the crop environment. We integrated field observations with the Cropping System Model CSM-CROPGRO-Tomato, part of the Decision Support System for Agrotechnology Transfer (DSSAT) software, and calibrated the model using Generalized Likelihood Uncertainty Estimation with 50,000 evaluations per cultivar, then validated against independent 2025 data.

Results: The calibrated model generated synthetic datasets that captured cultivar-specific physiological responses and were used to train 10 machine learning algorithms. Long Short-Term Memory networks achieved the highest prediction accuracy (R^2 = 0.94, RMSE = 0.31 t ha^{-1}), providing reliable yield forecasts 30 days before the initial harvest. The drought-tolerant hybrid Cocktail Crush F1 demonstrated superior stability, maintaining 85% of yield potential under severe stress compared to 76–78% for other cultivars. Under optimized management (double-row planting, late-May transplanting), yield gains reached 2.8 t ha^{-1} under favourable scenarios and 1.6 t ha^{-1} under high-emission scenarios. Thirty-day forecasts enable early-market entry with 35–45% price premiums and reduce last-minute labour costs by 20–25%.

Conclusions: The key innovation is a hybrid framework that leverages the CSM-CROPGRO-Tomato within DSSAT–a sophisticated Fortran-based system–as a knowledge engine to generate physiologically robust synthetic data. These datasets train accessible machine learning algorithms for practical 30-day yield forecasting. The study presents a hybrid framework combining DSSAT-based modelling with machine learning to deliver practical, accurate farm-level yield predictions.

Keywords: DSSAT-CROPGRO-Tomato; Long Short-Term Memory; machine learning; ClimRisk; Central Europe

Acknowledgments: This work was supported by the PERUN project (Prediction, Evaluation and Research for Understanding National Sensitivity and Impacts of Drought and Climate Change for Czechia), co-financed with state support from the Technology Agency of the Czech Republic under the Environment for Life Programme.