

management alters both climate sensitivity and yield stability, so predictive models and management strategies must jointly consider soil fertility and climate interactions rather than treating them in isolation.

3 Data Sources and Overview of the Study Area

3.1 Natural and agricultural conditions of the study area

The major maize-producing regions of northern and northeastern China are characterized by temperate monsoon climates with distinct growing seasons, where temperature, precipitation, and sunshine jointly determine maize climate suitability at different phenological stages (Wang et al., 2024). In the Northeast, relatively cooler temperatures and variable rainfall make precipitation a key limiting factor, while temperature plays a stronger role in the suitability index than in more southerly zones. In contrast, the Huang-Huai-Hai (3H) region has warmer average temperatures and generally higher comprehensive climate suitability, although spatial differences in precipitation and sunshine still create heterogeneous yield potentials. Across China's broader maize belt, temperature variability and climate perturbations can cause substantial yield losses, especially under warming, but these impacts are spatially heterogeneous (Chen et al., 2024).

Soil conditions in the maize belt range from high-soil organic carbon (SOC) soils in parts of the Northeast to more degraded or compacted soils in other regions, and these differences strongly affect yield responses to climate. High SOC, favorable texture, and adequate field capacity enhance buffering capacity against adverse temperature and moisture perturbations, stabilizing yields under climate variability (Feng et al., 2022). In contrast, soils with higher bulk density, coarser texture, or lower water-holding capacity tend to amplify yield losses under warming, underscoring the importance of soil improvement for resilient production. Regional tillage practices, such as deep ploughing or conservation tillage, also interact with local climate: in cooler sites, practices that improve early-season soil temperature and water availability promote maize emergence and growth, whereas in warmer, windier areas, systems that enhance water retention and aeration can be more beneficial (Qian et al., 2025).

3.2 Data sources and acquisition methods

Maize yield data and associated environmental variables can be obtained from long-term field trials, experimental stations, and statistical records, often at plot or county scales. Multi-year experiments in Northeast China and the North China Plain provide detailed measurements of yield, phenology, and management, suitable for evaluating soil-climate interactions and model performance. In some studies, plot-scale experiments under different fertilization or tillage systems supply yield and soil measurements across contrasting climate conditions, enabling analysis of management impacts on yield and soil properties (Meng et al., 2021; Qian et al., 2025). For broader regional coverage, station networks combining agronomic records with local weather observations support large-scale assessments of yield responses to climate variability and soil attributes.

Climate data are typically derived from ground-based automatic weather stations and gridded meteorological datasets, providing variables such as temperature, precipitation, radiation, humidity, and derived indices (e.g., heat degree days, consecutive dry days) during key growth stages (Dandridge et al., 2024; Wang et al., 2025). Remote sensing products supply complementary environmental information, including vegetation indices, land surface temperature, and solar-induced fluorescence that capture canopy status over time. Soil data come from field sampling, regional soil surveys, and derived soil property databases, covering SOC, texture, bulk density, water-holding capacity, and nutrient indicators. In advanced yield-prediction frameworks, these multi-source datasets—yield, weather, soil, and remote sensing—are integrated into unified databases for machine learning or crop model applications.

3.3 Data preprocessing and quality control

Prior to model construction, environmental and yield data require systematic preprocessing to ensure completeness and consistency. Weather station data are screened for missing values, range violations, and temporal or spatial inconsistencies, often using automated quality-control algorithms tailored to agricultural decision needs. Such systems flag implausible measurements—e.g., unrealistic temperature sequences, saturated relative humidity at too low values, or anomalous rainfall series—enabling early detection and correction or removal of erroneous records. For gridded or satellite-based climate products, temporal aggregation (e.g., daily to