CellPress
Partner Journal

Molecular Plant
Perspective

# Satellite-enabled enviromics to enhance crop improvement

Rafael T. Resende[1,2,*], Lee Hickey[3], Cibele H. Amaral[4,5], Lucas L. Peixoto[1], Gustavo E. Marcatti[2,6] and Yunbi Xu[7,8,9,*]

[1]Universidade Federal de Goiás (UFG), Agronomy Department, Plant Breeding Sector, Goiânia (GO) 74690-900, Brazil

[2]TheCROP, a Precision-Breeding Startup: Enviromics, Phenomics, and Genomics, No Zip-code, Operating Virtually, Goiânia (GO) and Sete Lagoas (MG), Brazil

[3]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD, Australia

[4]Earth Lab, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80303, USA

[5]Environmental Data Science Innovation & Inclusion Lab, Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80303, USA

[6]Universidade Federal de São João del-Rei, Forest Engineering Department, Campus Sete Lagoas, Sete Lagoas (MG) 35701-970, Brazil

[7]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[8]Peking University Institute of Advanced Agricultural Sciences, Weifang, Shandong 261325, China

[9]BGI Bioverse, Shenzhen 518083, China

***Correspondence: Rafael T. Resende (rafael.tassinari@ufg.br), Yunbi Xu (xuyunbi@caas.cn)**

https://doi.org/10.1016/j.molp.2024.04.005

## ABSTRACT

**Enviromics refers to the characterization of micro- and macroenvironments based on large-scale environmental datasets. By providing genotypic recommendations with predictive extrapolation at a site-specific level, enviromics could inform plant breeding decisions across varying conditions and anticipate productivity in a changing climate. Enviromics-based integration of statistics, envirotyping (i.e., determining environmental factors), and remote sensing could help unravel the complex interplay of genetics, environment, and management. To support this goal, exhaustive envirotyping to generate precise environmental profiles would significantly improve predictions of genotype performance and genetic gain in crops. Already, informatics management platforms aggregate diverse environmental datasets obtained using optical, thermal, radar, and light detection and ranging (LiDAR)sensors that capture detailed information about vegetation, surface structure, and terrain. This wealth of information, coupled with freely available climate data, fuels innovative enviromics research. While enviromics holds immense potential for breeding, a few obstacles remain, such as the need for (1) integrative methodologies to systematically collect field data to scale and expand observations across the landscape with satellite data; (2) state-of-the-art AI models for data integration, simulation, and prediction; (3) cyberinfrastructure for processing big data across scales and providing seamless interfaces to deliver forecasts to stakeholders; and (4) collaboration and data sharing among farmers, breeders, physiologists, geoinformatics experts, and programmers across research institutions. Overcoming these challenges is essential for leveraging the full potential of big data captured by satellites to transform 21st century agriculture and crop improvement through enviromics.**

**Key words:** envirotyping, precision breeding, genotype–environment interactions, remote sensing, predictive models, enviromic information

**Resende R.T., Hickey L., Amaral C.H., Peixoto L.L., Marcatti G.E., and Xu Y.** (2024). Satellite-enabled enviromics to enhance crop improvement. Mol. Plant. **17**, 848–866.

## INTRODUCTION

Alterations in global climate and their subsequent impacts on agricultural landscapes evoke serious concerns about food security. The concept of enviromics, addressing the myriad environmental variables influencing plant growth and development, offers pivotal insights. These insights elucidate how the enhancement of crop improvement can be achieved by

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

 © 2024 The Author.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

deciphering the intricate interactions between genotype, environment, and management (G × E × M) for effective adaptation to changing climates. Terrestrial and atmospheric factors directly influence crop growth, development, and productivity (Xu 2010, 2016; Cooper et al., 2014), and genotypes that excel in one region may perform poorly in another due to specific environmental conditions (Xu 2010). In addition to understanding such G × E interactions, knowledge of the variable features of the environment (Piepho and Blancon, 2023) can inform targeted genomic prediction, genetic improvement, and crop management strategies that boost agricultural productivity.

To assess G × E and G × E × M interactions, replicate genotypes are generally grown in environments distinguished by specific combinations of locations and years. The resulting data can be assessed using approaches such as adaptability and stability analyses (van Eeuwijk et al., 2016; Crossa et al., 2022) or graphical techniques such as genotype plus genotype-by-environment biplot and additive main effects and multiplicative interaction to map the responses of different genotypes to varied environments (Neisse et al., 2018; Olivoto et al., 2019). In addition, multi-environment trial (MET) analysis, which uses statistical models to analyze variance and covariance structures between environments, can detail genotype performance under precise cultivation and management conditions (Malosetti et al., 2013). Factor analytic modeling has also gained popularity recently, as it speeds model convergence when analyzing large, multi-environment datasets (Krause et al., 2020; Smith et al., 2021). Notably, crop growth models generate virtual simulations of genotype behavior under hypothetical scenarios; integrating genetic and environmental information using mechanistic eco-physiological-based models can enhance model accuracy (Bustos-Korts et al., 2019; Rincent et al., 2019).

Enviromics, as an omics, represents a somewhat distinct approach to assess G × E and G × E × M interactions based on environmental data (Resende et al., 2021). The term enviromics first appeared in psychiatric literature in the mid-1990s (Anthony et al., 1995). In plant breeding, it was initially mentioned by Xu (2016), and its deeper exploration began with a 2019 bioRxiv preprint, ultimately leading to a publication by Resende et al. (2021). Since then, enviromics has rapidly gained popularity in the breeding community, highlighted by research from various groups (Costa-Neto et al., 2021a; Cooper and Messina, 2021; Crossa et al., 2021; Resende et al., 2022). An envirome, similar to a genome and phenome, is a set of envirotypes represented by all environmental factors that affect the growth and development of an organism, involving landscape and climatic variables (Xu, 2016; Costa-Neto and Fritsche-Neto, 2021; Resende et al., 2021). The origin of the term envirotype is attributed to Patten (1991) and was revisited by Beckers et al. (2009) in genetic studies of mice. Envirotyping was conceptualized by Xu in 2010 and formally published in 2016 (Xu et al., 2022) to describe the gathering of environmental data to characterize environments, as also discussed by Cooper et al. (2014). For crops, the enviromics approach prioritizes spatial data analysis, integrating both experimental and on-farm data for accurate model validation across various scales. If envirotyping is considered a third typing technology (Xu, 2016), along with genotyping and phenotyping, then enviromics represents a third omics approach, alongside genomics and phenomics (Resende et al., 2021).

The current era has witnessed the rise of digital agriculture, commonly known as precision agriculture (Shaikh et al., 2022). A wealth of information is readily accessible—often free of charge—and many envirotypic inputs can be downloaded using only a few steps. The availability of big data enables enviromics studies, and integrated analyses of genotypes and environments (Xu et al., 2022) can facilitate the identification of genotypes with superior response patterns across diverse conditions. These insights can inform the selection of varieties best suited to withstand abiotic stresses, including heat stress, drought stress, or elevated $CO_2$ conditions caused by climate change. However, imprecise data overlay poses barriers to enviromics analysis, and this vast amount of information must be rigorously validated (Marcatti et al., 2017; Resende et al., 2022).

Satellite sensors are a source of invaluable data for enviromics. For instance, Earth observation satellites such as Landsat, MODIS (Moderate Resolution Imaging Spectroradiomete), and Sentinel monitor surface phenomena including climate, vegetation composition, land use, and air or water pollution (Zhao et al., 2022). Positioned in higher orbits, the meteorological satellites GOES (Geostationary Operational Environmental Satellites) and Meteosat provide data on weather and climate conditions, acquiring parameters such as light, temperature, humidity, and wind speed (Krinitskiy et al., 2023). In addition, reanalysis tools such as Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2) leverage global observations (Reichle et al., 2017) for calibration, uncertainty determination, and data product evaluation, as emphasized by Brönnimann et al. (2018) (see also this 2-min YouTube video for a nice overview of reanalysis data: youtube.com/watch?v=FAGobvUGl24).

The data now derived from satellites offer a detailed view of geographical sites, enabling more complete analysis of the influence of the environment on genotype performance. In this perspective, we examine the relationship among modern envirotypic data and their effects on enviromics and plant breeding, with a particular focus on spaceborne/spatial technologies that facilitate our understanding of crop genotype and envirotype interactions, emphasizing the role of enviromics in providing information on an omics scale. Our aim is to update the knowledge base, addressing the gap in comprehensive reviews amid rapid technological advancements and data analysis innovations to boost precision and efficiency in crop management and improvement in the face of climate change.

## INTEGRATING THE TARGET POPULATION OF ENVIRONMENTS INTO THE ENVIROMICS CONTEXT

The initial step in managing enviromics frameworks for crop studies involves identifying the target population of environments (TPE; Figure 1, step 1). The TPE represents the

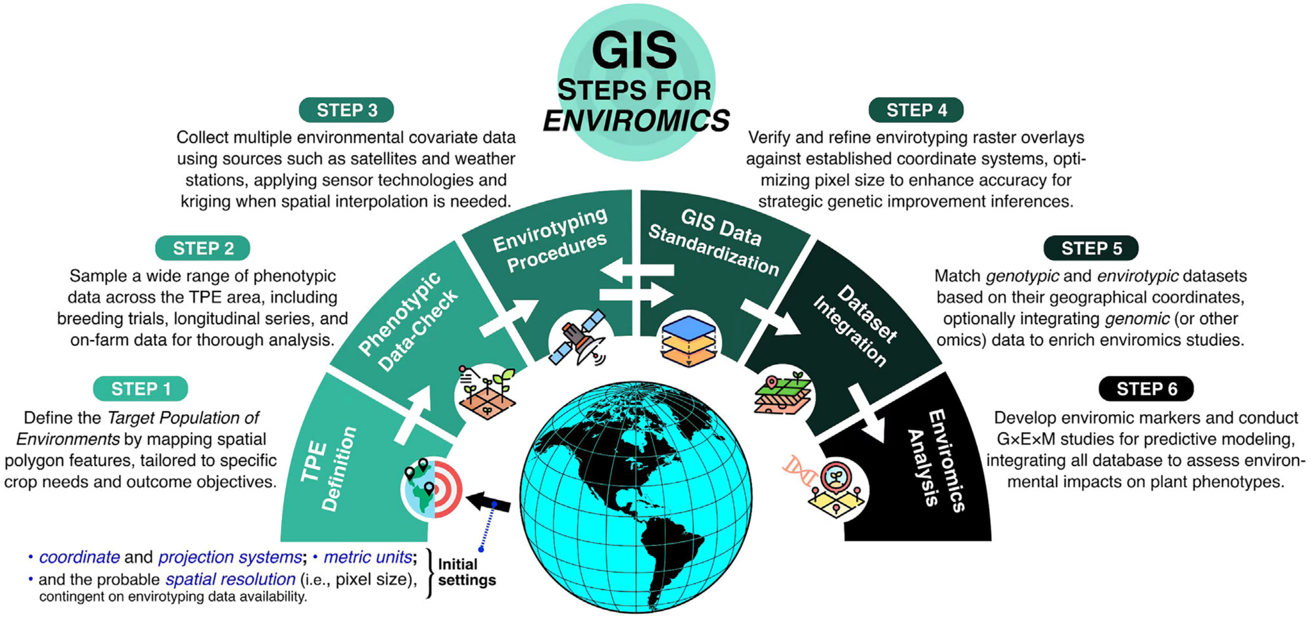


**Figure 1. Necessary GIS steps for enviromics to enhance precision breeding through integrating phenotypic and environmental data.**
This workflow begins with establishing a target population of environments (TPE) and ends with intricate data integration for enviromics analysis. It includes key stages such as the validation of phenotypic data, strategic collection of envirotyping data via remote sensing, meticulous standardization of GIS data, and comprehensive analysis of envirotypic data. This systematic approach is designed to refine predictive modeling and optimize agricultural outcomes by assessing and exploiting the interplay among genotype, the environment, and management practices.

composition and frequency of environmental types within a region targeted by plant breeders and includes the range of environments where candidate genotypes will be evaluated for performance under various growth conditions (Cooper et al., 2014; Chenu, 2015). With the availability of GPS data for civilian use, it has become possible to determine the variability of environmental factors at the regional, farm, and field scales. In enviromics, the focus is not on an individual experimental site, as it would be in MET approaches, but on the entire TPE extent, which is viewed as a virtual geoprocessing shape that directly represents the real environment in the field.

After the TPE is identified, phenotypic data must be examined across diverse settings (Figure 1, step 2). Conducting an enviromics study, or even a comprehensive G × E × M study, with a few representative trials across various environments is akin to conducting a genomics study using a few individual plants or even a few single-nucleotide polymorphism markers (Resende et al., 2021). The environmental range must generate sufficient envirotypic variation to enable enviromics analysis. This involves expanding the data collection to encompass longitudinal studies and on-farm data, providing wide-ranging information on genotype performance in various environmental conditions. Khosla (2023) showcased grid sampling and sensor technology for gathering envirotype data alongside phenotypic performance data for crops. Improved soil sensors can be used for rapid, reliable, cost-effective *in situ* measurements at the plot scale. Through field and laboratory experiments, both crop and envirotypic data layers are collected, providing information about the genotypic conditions within the TPE. By integrating these diverse data sources, the analysis gains depth and reliability, connecting theoretical environmental definitions with tangible, real-world agricultural outcomes.

To date, enviromics has been applied to plant breeding at the cultivar selection/recommendation stage when the TPE is the focus. For instance, enviromics concepts were recently used to quantify the effects of climate on the adaptation of elite common bean (*Phaseolus vulgaris*) germplasm in Brazil, leading to the identification of climate limits and critical developmental phases for each production scenario and guiding efforts in selecting climate-smart varieties (Heinemann et al., 2022). Ultimately, envirotyping needs to be scaled to small experimental plots or even individual plants to achieve the same level of resolution that genotyping and phenotyping can achieve (Xu, 2016). To realize this goal, three types of technical developments and scientific advancements are needed. First, envirotyping for all environmental factors must be possible, likely via coupling of satellite-equipped sensors with ground-based sensors and probes. Second, all sensors and probes must be affordable and have high enough resolution, throughput, and efficiency for envirotyping at the individual-plant level. Third, enviromic information management tools are needed that are equipped with powerful computation and AI-assisted modeling and prediction systems.

## APPLYING ENVIROMICS TO BREEDING: THE NEED FOR EXHAUSTIVE ENVIROTYPING

The term environment is used in different ways. It can refer to (1) the natural conditions that affect human existence, as discussed by environmental scientists; (2) the social and cultural conditions that shape individual or community life, a concept pertinent to social sciences; (3) geographical sites, which exhibit year-to-year variation; and (4) the physical, chemical, biotic, and abiotic factors that influence the growth and development of an organism, which collectively constitute an envirotype. For enviromics, we favor the latter two usages, with a focus on collecting environmental data for georeferenced sites and analyzing their spatial and temporal variation. In particular, viewing the complex layers of environments as envirotypes sets the stage for meticulously examining predictive models in plant breeding, where relationships between environmental factors and genotypic performance must be analyzed with precision and depth (Figure 1, step 3).

From the perspective of quantitative genetics, the environment is one of two terms used to explain phenotypic variation. The envirome is defined as the complete set of external conditions affecting phenotypic performance (Costa-Neto and Fritsche-Neto 2021). Enviromics approaches parse out hidden patterns through the envirotype itself and its interaction with genotype, which is essential for understanding and improving crops. Compared to the genotype, which comprises many genes that determine the phenotype, the envirotype involves numerous environmental factors with different effects on the phenotype (Cooper and Messina, 2021). Some factors may have major effects and be largely predictable, such as photoperiod (day and night length), temperature patterns, annual and seasonal precipitation, soil properties, and specific abiotic stresses.

Much discussion revolves around explanatory versus predictive models (e.g., in a forum led by Leo Breiman, a developer of RandomForest; Breiman, 2001). While explanatory models aim to understand the causal relationships between variables, predictive models aim to forecast future outcomes based on past or current data (Shmueli, 2010). Both prediction and explanation are important for enviromics, each offering unique contributions that together can synergistically enhance crop improvement, making both approaches critical for advancing the field (Costa-Neto et al., 2023). However, it is important to remember that, when two variables appear to be related, the assumption that one variable will accurately predict the other can lead to incorrect population inferences, as what appears to be effective in a set of sample data may not hold true universally. Testing these assumptions is particularly vital for plant breeding, where predictive models are preferred over explanatory models for identifying genotypes with desirable phenotypes, such as yield performance and disease resistance.

The goal of plant breeding is to identify genotypes with desirable phenotypic performance, and predictive models can facilitate this identification. For instance, genomic selection uses genetic markers to build a prediction model for the genetic merits of selected candidates with regard to complex traits (Resende, 2024). Genomic selection offers advantages over traditional quantitative trait locus (QTL)-based marker-assisted selection, particularly for complex traits governed by many genes with small effects (Budhlakoti et al., 2022). High-density genotyping is needed to support genomic prediction, as additional marker data can shed light on the genetic variation contributing to the trait (Sousa et al., 2019). Similarly, more environmental information, when combined with genomic information (and/or other omics inputs), can lead to enhanced prediction outcomes in enviromics (Araújo et al., 2024; Callister et al., 2024). The value of exhaustive data is exemplified by work from Millet et al. (2019), who studied maize (*Zea mays*) across diverse European environments, and Li et al. (2021), who integrated environmental factors in genome-wide association studies of crops such as wheat (*Triticum aestivum*), maize, and oat (*Avena sativa*).

Several approaches can be used to characterize envirotypes. For instance, the agricultural production systems simulator pathway, which categorizes environments for agricultural modeling, employs observed data and model simulations to define distinct types of environments based on factors such as climate, soil, and management practices (Holzworth et al., 2018). Statistical methods for characterizing environment type, such as iclass, rely on analysis of observed datasets, such as crop yield, to identify groups or clusters of environments. These methods aim to minimize crossover G × E to classify environment types for research or breeding purposes (Smith et al., 2021).

Enviromics can harness both agricultural modeling and statistical models, integrating multidimensional information from envirotyping and genotyping through the use of kernel-based or random regression models (Jarquín et al., 2014; Costa-Neto et al., 2021b; Resende et al., 2021). Li et al. (2022) demonstrated that the methods used in genomics and phenomics are also effective in enviromics, applying these methods to predict the impact of climatic conditions on the performance of wheat. Importantly, the envirotype can be regarded as an independent factor that significantly influences phenotype, its prediction, and its selection, rather than merely being included as a cofactor in G × E interactions.

That enviromics integrates various G × E × M methodologies within its framework yet is fundamentally different from models rooted in eco-physiology, such as crop growth models or those derived solely from METs. To acquire sufficient genomic–enviromic–phenomic (G–E–P) data to support the analysis, various unbalanced datasets can be incorporated into the models. Indeed, for predictive models, large, unbalanced datasets are far more advantageous than scarce, balanced data, with all genotypes being represented in all trials (Resende et al., 2021). Statistical strategies can effectively handle genetic predictions and report genetic and residual variance components using unbalanced trial data (Schmidt et al., 2019; Dias et al., 2020). Envirotyping procedures draw inspiration from dissecting G × E interactions through data acquired by exhaustive, or high-throughput, envirotyping (Cooper et al., 2014; Xu, 2016). Figure 2 shows a hypothetical envirotyping data structure that could support envirome-wide selection. The area shown is in Indiana, USA, where diverse layers of envirotyping data can be acquired. Although this hypothetical case study focuses on a single area, the concepts apply to broad regions, from individual countries to entire continents and even intercontinental regions, with the TPE tailored to suit specific interests.

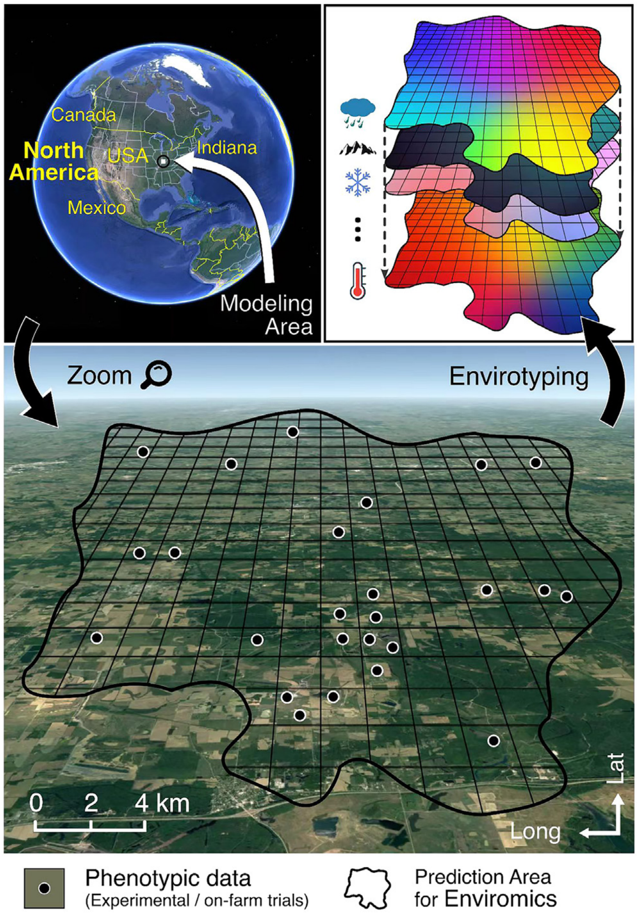


**Figure 2. An example of a prediction area located in North America.**

Twenty-five sampling points containing phenotypic data are shown (experimental or on-farm trials). The prediction area is in Indiana, USA. Various layers of envirotyping data for the area are also depicted.

To integrate envirotyping data from diverse sources, cartography and geodesy use datums, or reference frames, composed of a coordinate and reference system to represent the Earth's surface (an important task for step 4 shown in Figure 1). The ellipsoids and transformation parameters of global datums such as WGS84, NAD83, ED50, SIRGAS, and ITRF, each optimized for specific regions, ensure accurate referencing in data overlays and maps. While WGS84 is a global standard, regional datums continue to be utilized for local applications, especially in high-precision mapping and geodetic studies, with global integration driven by technological advancements. The data cube concept, which utilizes geocoding and image co-registration techniques, handles diverse geospatial data with tools such as gdalcubes and xcube for multivariate analysis. Notably, a gdalcubes library enables on-demand construction and processing of data cubes from satellite image collections (Appel and Pebesma, 2019). In addition, the Python package CGC facilitates co- and tri-clustering of geodata cubes to identify patterns across spatial, temporal, and thematic dimensions (Nattino et al., 2022).

Acquiring myriad data points concerning terrestrial attributes and surface characteristics, such as agricultural crop growth patterns at the farm level, has become feasible, especially when temporal data are considered, segmented according to the accessibility of the environmental monitoring platform and amalgamated with climatological norms (Uthes et al., 2020). Importantly, incorporating climatological norms affords a more robust depiction of average insights pertinent to geographical focal points. This approach facilitates the detection of climatic deviations, such as unusual temperature shifts or precipitation patterns, warranting deeper investigation.

Some environmental factors are predictable, as they are largely determined by longitude, latitude, and altitude, whereas some are unpredictable due to random variable factors such as weather changes. Even for the most seasoned climatologists, predicting aberrant climate events is a formidable undertaking (Brady and Spring, 2021), complicating efforts to identify suitable cultivars for growth in the face of climate variation. Such variations do not invariably result in tangible events. Satellite data offer temporal information, that is, measurements at certain intervals (refer to Supplemental Table 1; discussed in detail below). Some measurements are influenced by atmospheric factors, including (but not limited to) cloud formations, pollutants, lightning, and various forms of radiation such as solar and cosmic rays. Importantly, while enviromics requires less granularity compared to phenomics, it demands attention to detail at the pixel level, even at the levels of individual field plots, blocks, and even individual plants (Xu, 2016; Xu et al., 2022). Furthermore, the concept of "pan-enviromics" encapsulates enviromics across various dimensions, including time, space, multiple locations, and developmental stages. For further discussion on the intricate relationships among envirotype, envirotyping, envirome, enviromics, and pan-enviromics, see Crossa et al. (2021) and Guo and Li (2023).

A variety of sources provide data for enviromics studies. Meteorological stations provide information on climatic parameters such as temperature, humidity, precipitation, and wind speed. Hydrological stations monitor data related to water resources, such as river and lake levels. *In situ* sensors and sensor networks collect data at specific environmental points, sometimes associated with the Internet of Things. Drones and unoccupied aerial vehicles (UAVs) can perform high-resolution data collection in hard-to-reach areas. Crowdsourcing involving public collaboration can be used to collect and provide information. Mobile devices can collect geospatial data in real time. In addition, historical data collections provide information on patterns and trends over time, such as historical climate records, maps, and documents.

## SATELLITE SYSTEMS AND REMOTE SENSING FOR ENVIROTYPING

### Satellites and their classification

Plant scientists are well versed in advanced genotyping technologies, bioinformatics, and increasingly high-throughput phenotyping. Both UAVs and satellites can offer high-resolution imagery, with great potential for precise estimation in breeding plots of various sizes, allowing cost-effective, standardized phenotyping in breeding programs (Pinto et al., 2023). Importantly, enviromics data can sometimes intersect with aerial phenomics and can capture the combined influence of all environmental factors, even when there are small differences. A recent study integrating UAV and Sentinel-2 data significantly improved the prediction of sugarcane (*Saccharum officinarum*) yields, offering a cost-effective solution for yield management (Som-ard et al., 2024). Utilizing multispectral and microwave data from the ALOS (Advanced Land Observing Satellite), specifically AVNIR-2 and PALSAR sensors, Domingues et al. (2023) demonstrated the effectiveness of artificial neural networks for accurately estimating wood volume in a commercial eucalyptus plantation in Brazil. However, few studies have focused on integrating breeding with envirotyping using satellites (Xu et al., 2022).

Satellites are technological marvels tailored for specific missions with unique characteristics. They range from small CubeSats (modular satellites in cube form a few centimeters in size used in low-cost space missions) to large satellites weighing several tons (Levchenko et al., 2018). The satellite choice is determined by payload and launch constraints and by the target applications, from land monitoring to global communications. Earth observation satellites frequently operate in lower orbits to obtain high-resolution images, while communication and navigation satellites operate in higher orbits, such as geostationary orbits (Zhao et al., 2022). The stability and orientation of satellites are ensured by the attitude control system, which can vary; some satellites use gyroscopes and reaction wheels, while others use more sophisticated systems with reaction wheels and magnets.

Satellites can be grouped based on their purpose. Astronomical satellites are used to observe space and celestial bodies (e.g., the celebrated Hubble and James Webb). Communication satellites transmit radio, television, telephone, and Internet signals. Earth observation satellites monitor terrestrial resources, such as vegetation, soil, water, and climate. Meteorological satellites are used for weather forecasting and collecting climatic data. Military satellites are used for defense, espionage, navigation, and communication purposes. Finally, space stations house astronauts and scientific experiments in space (Jakobsen et al., 2022). The most suitable satellites for enviromics are Earth observation and meteorological satellites. These are placed in low orbits to monitor the Earth's surface and collect scientific data (Levchenko et al., 2018) or in geostationary orbits (remaining at high altitudes in a fixed position relative to the Earth) to monitor climate conditions and predict storms (Krinitskiy et al., 2023). Earth observation satellites also vary by sensor type: optical, thermal, radar, and light detection and ranging (LiDAR; Figure 3A). These sensors and their applications in envirotyping are discussed in detail in below.

## Satellite sensors for envirotyping

### *Optical sensors*

Optical sensors capture images of the Earth's surface using the visible and near- and short-wave infrared electromagnetic spectrum, allowing a detailed pixel-based analysis of an agricultural site. By monitoring strategic variables, such as leaf index, soil moisture content, and plant health, breeders gain insights into crop development, as they can detect early signs of issues such as biotic or abiotic stress. By providing continuous, precise aerial views, these optical sensors enable the implementation of more efficient agricultural practices, increasing crop productivity and facilitating informed decision making in managing cultivated areas. These sensors are often employed in drones, and the multispectral approach explores various spectral bands, facilitating their use in phenotyping experiments for genetic improvement.

The spectral bands of optical sensors cover wavelengths useful for agricultural applications, providing powerful envirotypic data for enviromics (Figure 3A and 3B). The blue band (~450–495 nm) is useful for detecting healthy vegetation and measuring nutrients and chlorophyll in plants. The green band (500–575 nm) is used to evaluate plant health, allowing areas with higher vegetation density to be identified. The green 1 band (500–550 nm) is employed to assess leaf health and evaluate plant responses to stress (Yang et al., 2022).

It is important to distinguish between applications in plant phenotyping and environmental envirotyping. Centimeter-level accuracy for detailed plant phenotyping, focusing on individual plants and plots, has generally been achieved using sensors on UAVs. By contrast, envirotyping examines the broader environmental context, making use of satellite imagery with resolutions finer than 4 m or applying kriging (spatial interpolation) at specific resolutions for genetic-improvement projects. Excitingly, advanced satellite imagery technology can now provide centimeter-level detail (Karwowska and Wierzbicki, 2022), which can support applications such as selection in breeding populations at early stages of evaluation (Zhang et al., 2019). This level of detail is also important for allogamous (cross-pollinating) crops such as maize, where a 3–4 $m^2$ plot size is optimal for progeny selection (Chaves and de Miranda Filho, 1992), and for tree crops due to the variable spacing between trees (Marcatti et al., 2017). These advancements highlight the evolving role of remote sensing in agricultural analysis, which seamlessly integrates phenotyping and envirotyping for a thorough genetic and environmental assessment.

In some cases, vegetation behavior acts as a proxy for environmental quality, with healthy, vigorous plants indicating favorable environments for crop development, a notion termed plant-based characterization (Skovsgaard and Vanclay, 2008). Different spectral bands within the electromagnetic spectrum are instrumental for indirectly measuring plant features. The near-infrared (NIR) region (750–1300 nm) is used to assess crop health by reflecting leaf and canopy structures (Kokaly et al., 2003), while red bands (600–750 nm) are used to evaluate chlorophyll levels, underpinning assessment of vegetation cover and productivity (Venancio et al., 2019). Red-edge bands (bridging red and NIR) are used to detect subtle changes in vegetation related to growth and stress. The efficacy of these bands in reporting plant traits depends on crop type, stress levels, and other environmental conditions. Vegetation indices derived from these bands quantify environmental impacts on vegetation: the normalized difference vegetation index (NDVI) gauges plant

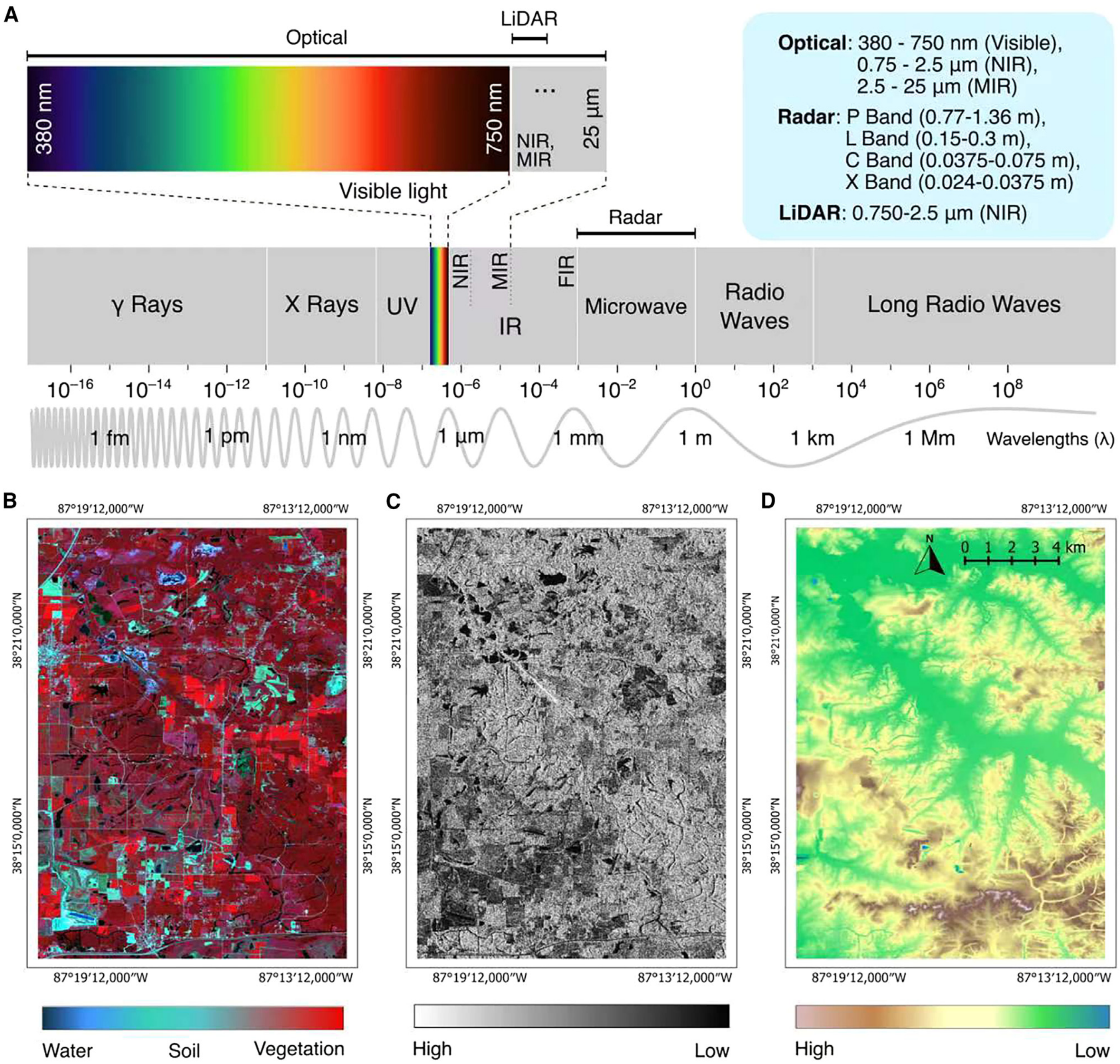


**Figure 3. Imagery captured by three different types of sensors.**
**(A)** Spectrum range covered by different sensor technologies; wavelengths from gamma rays to long radio waves are shown, along with specific bands used by optical, radar, and LiDAR sensors for envirotyping data acquisition; thermal sensors cover NIR, mid-infrared (MIR), and far infrared (FIR) (this categorization is not standardized and can differ across sources).
**(B)** Optical-based sensor, showing a false-color NIR-R-G image from the Sentinel-2 satellite at 10-m resolution.
**(C)** Backscattering signal (microwaves) obtained using Sentinel-1 sensor synthetic aperture radar (SAR) with an approximately 10-m resolution.
**(D)** Topography obtained using airborne LiDAR technology, with an average post spacing of 1.5 m, sourced from IndianaMap (IndianaMap 2011). UV, ultraviolet.

health by contrasting NIR and red light; the soil adjusted vegetation index adjusts NDVI for soil background, enhancing accuracy in regions with sparse vegetation; enhanced vegetation index improves NDVI by correcting for the influences of atmosphere and soil, offering sensitive detection of changes; and normalized difference water index focuses on water content, aiding hydrology studies (Silva et al., 2020). Together, these indices and bands provide comprehensive insights into environmental quality and vegetation, demonstrating the conditional effectiveness of spectral analysis depending on the specific agricultural context.

Optical sensors can also be used in environmental quality assessments incorporating "Earth-based" landscape characteristics (Skovsgaard and Vanclay, 2008). Short-wave infrared (SWIR) bands are useful for identifying soil compositions and pinpointing moisture discrepancies in both the soil and canopy, with SWIR-1 alongside NIR bands being especially valuable for estimating soil moisture content and surveying crop health under varied hydration states (Yue et al., 2019). The less commonly used ultra-blue and yellow bands contribute to research on atmospheric aerosol scattering (Bautista et al., 2022). Hyperspectral sensors have extensive spectral resolution, covering wavelengths from visible to NIR and even into the SWIR spectrum, offering detailed, pixel-based data across numerous narrow bands. Rizzo et al. (2023) constructed a high-definition global soil color map at 30-m precision based on over three decades of Landsat satellite data and ground spectral measurements, laying the foundation for soil resource monitoring and management in the future. These sensors, capped at 30-m spatial resolution, can be used to collect a vast amount of envirotypic data, such as soil types and moisture levels, thereby advancing environmental and agricultural research (Yue et al., 2019).

### *Radar sensors*

Unlike optical sensors, radar sensors emit microwaves and detect the reflected radiation. These sensors have night-vision capabilities, as they can penetrate through clouds, and provide images at lower spatial resolution, making them well suited for adverse weather conditions and three-dimensional (3D) mapping of the Earth's surface (Wang et al., 2018). For example, the image produced by radar sensors shown in Figure 3C demonstrates the phenomenon of backscattering, i.e., the return of microwave energy emitted by the Sentinel-1 synthetic aperture radar (SAR) sensor operating in the C-band to the sensor itself, revealing details about the surface's topography, structure, and moisture based on the signal variation. Radar sensors can thus facilitate envirotyping by providing information about ground conditions, such as vegetation, soil moisture, and water availability. They can also enhance plant health monitoring by detecting changes in crop health related to diseases and water stress (Emmerik et al., 2017). Data from these sensors can be used to analyze phenotype traits associated with plant performance in different environments (Al-Turjman, 2019).

Crop mapping is another important application, as radar sensors can help map the spatial distributions of different crops in large agricultural areas, aiding in the design of plant breeding experiments. By employing data from a dual-polarimetric C-band radar image satellite and a QUEST (Quick, Unbiased, Efficient Statistical Tree) decision tree classifier, Mishra et al. (2017) mapped the spatial distribution of rice (*Oryza sativa*) cultivation areas, with an impressive accuracy of 88.6%. This approach serves as an effective tool for rice crop mapping and could potentially enhance supply chain forecasting.

### *LiDAR sensors*

LiDAR sensors are active remote sensors that emit laser beams in the green and NIR wavelengths. These sensors are capable of accurately modeling the Earth's surface (Figure 3A). LiDAR sensors aboard satellites emit laser pulses toward the Earth's surface and measure the time it takes for the pulse to return to the sensor. The information can be used to calculate the distance between the satellite and the surface point that reflected the pulse. By combining multiple measurements from different shots, LiDAR satellites provide detailed 3D surface information, including the heights of the terrain and vegetation. Even more detailed models can be created using point clouds from both aerial and terrestrial LiDAR systems. The two orbiting LiDAR systems, NASA's IceSat-2 and Global Ecosystem Dynamics Investigation (GEDI), produce data in the form of photon-counting and waveform samples of the Earth's surface that require interpolation to create a raster. Good digital elevation models from orbital platforms can be derived from radar technologies, such as SRTM (Shuttle Radar Topography Mission) and ALOS-PALSAR (Phased Array type L-band Synthetic Aperture Radar).

The Airborne Laser Terrain Mapper is an airborne sensor that performs high-resolution topographic mapping and terrain modeling, providing accurate data. The land, vegetation, and ice sensor mounted on the ICESat satellite measures the height of the Earth's surface, vegetation, and ice, shedding light on climate change and enabling environmental monitoring. The GEDI, coupled with the ICESat-2 satellite, maps the vertical structure of forests, facilitating the study of terrestrial ecosystem dynamics (Alvites et al., 2022). The sophisticated ATLAS LiDAR sensor, which is also integrated into the ICESat-2 satellite, is used to measure the height of polar ice and the Earth's surface, contributing to climate studies and environmental monitoring.

Acquiring LiDAR data is costly, primarily due to the high price of sensors and the expenses involved in integrating them into airborne devices. In addition, the resulting data files are large. Nevertheless, LiDAR sensors serve various purposes in agriculture. Figure 3D illustrates the use of aerial LiDAR data to model the topography of the area shown in Figure 2. The data were obtained through the free IndianaMap Framework LiDAR platform (IndianaMap, 2011) and are displayed for comparative purposes and for exploring LiDAR functionalities. LiDAR sensors can map the topography of agricultural areas, thereby enhancing the precision of land use and irrigation planning (Debnath et al., 2023). Additionally, LiDAR data provide information about crop structure and height, facilitating plant health evaluations and the early detection of issues such as water stress or diseases. When deployed on satellites, LiDAR sensors can be used for large-scale monitoring, swiftly covering large territories—an invaluable asset in modern agriculture, where intelligent resource management and continuous monitoring are needed to enhance productivity. When mounted on UAVs, LiDAR provides geometric measures that are particularly advantageous for capturing height and modeling biomass at the plot level in breeding programs; however, it remains uncertain whether the same level of resolution is achievable with satellite-based LiDAR.

### *Weather-related sensors*

Meteorological satellites equipped with a specific range of sensors monitor atmospheric conditions. High-resolution thermal infrared cameras measure the temperature of the Earth's surface, aiding in climate studies and the detection of thermal variations. Thermal cameras typically exhibit low spatial and spectral resolution. A notable exception is NASA's ECOSTRESS mission, with a spatial capability of 70 m. Lower-level products from the ECOSTRESS mission include information about water-use efficiency and the evaporative stress index (Fisher et al., 2020), which help elucidate the ability of different genotypes to deal with dryer conditions.

Radiometers measure solar and infrared radiation, providing important information about the climate and incident solar energy on Earth. Microwave radiometers measure sea surface temperatures and other oceanic features, contributing to oceanographic studies and the prediction of ocean-related phenomena. Atmospheric sounders are employed to measure temperature, humidity, and pressure at different altitudes, offering information about atmospheric processes and facilitating weather forecasting. Meteorological satellites also include high-temporal-resolution imagers, allowing for detailed images of cloud cover and weather patterns, providing essential data for predicting and monitoring extreme weather.

The surface and weather sensors complement each other, facilitating the study and monitoring of terrestrial and atmospheric environments. The combined use of advanced technologies, such as optical cameras, multispectral and hyperspectral sensors, SAR, radiometers, and atmospheric sensors, provides a broad, accurate view of the Earth and its climatic phenomena. Continued collaborations in these fields will contribute to basic research and the application of satellite data in agriculture and plant breeding, providing a deeper understanding of interactions between genotypes and agro-climatic settings and supporting decision making in enviromics-related sectors such as agriculture, climate science, meteorology, and environmental policy.

## Remote-sensing products for enviromics

The current panorama of remote sensing offers a diverse array of products, each designed for specific applications (Khanal et al., 2020; Lechner et al., 2020; Weiss et al., 2020). These products (detailed in Supplemental Table 1) originate from terrestrial monitoring missions and often comprise multiple satellites. Their various sensors capture "scenes"—images or datasets representing Earth's surface areas at specific times—that are essential for envirotyping. The scenes are presented in a raster format that encapsulates geographic or metric coordinates. The coverage and capture intervals of a scene are usually set during mission planning, with some flexibility in customization for certain products. Scenes can include multiple bands to capture temporal and spectral variations, enriching the data's dimensionality. Selecting a remote-sensing product for enviromics requires an understanding of its resolution types, which determine its suitability for specific applications (Jensen, 2009; Khanal et al., 2020).

Spatial resolution determines how well the smallest identifiable object in a remote-sensing scene can be distinguished. Resolution is closely linked to pixel size, with smaller pixels offering greater detail. High spatial resolution allows specific features to be detected, such as water deficiency in certain crops, making it ideal for analyzing small areas such as experimental sampling units (Jensen, 2009; Khanal et al., 2020). However, products with high spatial resolution (pixel < 5 m) often come with costs, both financially and in terms of the computational resources required for data processing, and they usually cover smaller areas in a single scene. The selection of spatial resolution in a genetic-improvement program is influenced by factors including the type of crop, scale of production, data source, and stage of the breeding cycle. For instance, the space occupied by plant populations can vary significantly, with early-generation populations covering less area compared to later stages (e.g., F5 or F6 generations; Chaves and de Miranda Filho, 1992).

Temporal resolution (the frequency of image capture in the same location) affects the ability to monitor changes over time. A higher temporal resolution means shorter intervals between captures, as exemplified by the Landsat mission's 16-day cycle (Jensen, 2009). However, the practical use of images is affected by atmospheric conditions, especially cloud cover, which can obscure data collection, particularly during the rainy season, when plants are at critical stages of growth. This challenge is especially pronounced for optical sensors. The choice of temporal resolution thus depends on specific requirements, such as the crop's life cycle, the desired level of detail throughout its life cycle, and specific phenological stages (Yang et al., 2022). High temporal resolution increases the likelihood of obtaining usable, cloud-free images for critical periods, such as peak vegetative growth, which is essential for accurate analysis and decision making (Lechner et al., 2020).

Spectral resolution is determined by the sensor's ability to discern different wavelengths across the electromagnetic spectrum (Jensen, 2009). NIR and red bands are essential for calculating vegetation indices (such as NDVI), which help assess plant vigor and stress levels. Blue and green bands contribute to true-color (RGB, Red, Green, Blue) imagery and vegetation and soil indices. Mid-infrared bands aid in evaluating plant water content, while thermal bands are pivotal for measuring land surface temperature. An ideal remote-sensing product combines visible (RGB) and NIR bands, facilitating comprehensive vegetation studies and environmental assessments (Silva et al., 2020; Voitik et al., 2023). Mid-infrared and thermal bands are beneficial for drought-resistance studies. Technological advancements have led to the development of multispectral sensors, offering broad spectral coverage using limited bands, and hyperspectral sensors, providing high spectral resolution using over 100 bands. Although hyperspectral sensors are not yet widespread, their potential for identifying specific plant traits or environmental conditions is significant, promising future advancements in precise plant and environmental monitoring (Terentev et al., 2022).

While spatial, temporal, and spectral resolutions are primary factors in choosing remote-sensing products for enviromics, additional factors are also important, such as the cost of the products, particularly free versus paid options. High-spatial-resolution (<5 m) and high-temporal-resolution (<5 days) products are rarely free, and high-temporal-resolution products typically feature moderate spatial resolution (>30 m). However, for many enviromics applications, ultra-high-resolution data may not be necessary, and data obtained for other purposes, such as infrastructure planning or crop prediction modeling, can be leveraged. The evolution of computational power and data-management tools promises more accessible high-resolution products in the future. Additionally, the operational status of a satellite mission must ensure the long-term viability of enviromics methods. With the continued development of sensor technologies, discontinued sensors are often replaced, ensuring continuous data availability. The use of a diverse array of remote-sensing sources enhances the precision of environmental analysis and reduces reliance on single data sources. Finally, the sensor type—optical very high resolution, optical, radar, LiDAR, or weather—is a key decision for remote sensing, as each offers unique advantages for specific applications, as summarized in Supplemental Table 1.

Choosing the best remote-sensing tool for enviromics requires an understanding of how environmental factors influence a genotype's productive capacity. This can be assessed using Earth-based and plant-based strategies, as outlined by Skovsgaard and Vanclay (2008). Earth-based assessments focus on physical characteristics such as climate, topography, and soil. Weather sensors gather climate data (e.g., precipitation, temperature), while radar and LiDAR sensors are invaluable for topographic modeling, offering information about terrain attributes and soil types. Plant-based assessments utilize optical sensors to evaluate crop-related characteristics, correlating specific optical spectrum bands and derived indices with crop production and stress factors (such as disease and water deficit; Tomar et al., 2014; Khanal et al., 2020; Voitik et al., 2023).

## ENVIROMIC INFORMATION MANAGEMENT PLATFORMS AND CYBERINFRASTRUCTURE

The extract, transform, and load process can also be used to handle the diversity and complexity of environmental data (Aydinoglu, 2016). The process of managing such data can be divided into extraction (represented in Figure 1 by steps 1–2), transformation (Figure 1, steps 3–5), and loading (Figure 1, step 6). Extraction involves gathering information from various sources, such as satellites and sensors (e.g., MODIS and Sentinel-2), weather stations, and climatic repositories (e.g., MERRA-2 and ERA5), encompassing data on climate, radiation, soil, vegetation, and topography. During transformation, the data are processed and prepared for advanced analysis via cleaning, normalization, integration, feature extraction, and relevant index calculations. Finally, the processed data are loaded into an appropriate analytical environment, such as a geographic information system (GIS) or environmental data platform, providing a solid foundation for more detailed analyses. In this section, we address the extraction stage.

Environmental data for enviromics can be extracted from platforms (or repositories) that aggregate datasets and analysis systems, gathering comprehensive and retrospective information about the environment and the Earth's surface. Informatics management platforms integrate a variety of data sources, such as satellite observations, surface measurements, and climate model data, to provide detailed envirotypic information, such as meteorological variables and information about the global climate and climate change on a global or regional scale. Much of the data, such as from EOS, LANDSAT, Sentinel, ALOS, GEDI, and RADARSAT, can be easily acquired using the Google Earth Engine platform (Velastegui-Montoya et al., 2023). In addition, GIS codes facilitate collaboration among software projects. The Open Geospatial Solutions organization on GitHub hosts open-source projects developed and maintained by a community of geospatial software experts, which are free for use and modification and are licensed by Massachusetts Institute of Technology (refer to the Open Geospatial Solutions GitHub page by Wu, Aybar, and Brown for further information: https://github.com/opengeos).

MERRA-2 is a long-term climate dataset that offers a comprehensive, retrospective assessment of past atmospheric conditions, providing detailed information on meteorological variables such as temperature, humidity, wind, and atmospheric pressure (Gelaro et al., 2017). Similarly, ERA5, which was developed by the European Centre for Medium-Range Weather Forecasts, provides high-resolution global climate reanalysis data, covering an extensive period and enabling global-scale climate analysis (Hersbach et al., 2020).

NCEP/NCAR Reanalysis, a collaborative effort between the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), provides a consistent and comprehensive dataset with decades of climate information. The Copernicus Climate Change Service (C3S), part of the European Union's Copernicus program, provides fundamental climate data, including satellite observations, surface data, and climate model results, to analyze climate change and its impacts. MODIS, a satellite sensor aboard NASA's Terra and Aqua missions, is a tool for environmental and climate monitoring that acquires spectral data and high-temporal-resolution images of the Earth's surface, offering a comprehensive, detailed view of our planet's conditions. These platforms can support climate research, studies on environmental changes, weather forecasting, and retrospective analyses, providing data to better comprehend global climate and its complexities over time.

Several R and Python packages can be used to gather data from diverse sources for enviromic modeling. The Python Requests library (Reitz, 2024) makes HTTP requests to web services or application programming interfaces , including various types of environmental data. The pyModis library (Delucchi and Neteler, 2013) extracts MODIS satellite data, and the Sentinelsat package extracts Sentinel satellite data. Geopy enables the extraction of geographical information, and web scraping libraries, such as BeautifulSoup and Scrapy, facilitate data extraction from websites (Kouzis-Loukas, 2016). The Pyproj package assists in extracting geospatial coordinates and transformations. For details on these and other Python packages and procedures for GIS analysis, see Westra (2016).

The R package nasapower (Sparks, 2018) provides access to NASAPOWER data for extracting climatic and meteorological information. The R packages raster, sf, and terra are robust tools for extracting and managing raster geospatial data (Hijmans et al., 2022). The stars package can be used to extract spatiotemporal data (Pebesma and Bivand, 2023), while sen2r can be used to extract data from Sentinel-2 satellites (Ranghetti et al., 2020). Some packages also assist in downloading data, making it easily accessible, such as EnvRtype (Costa-Neto et al., 2021a) and SoilType (Fritsche-Neto, 2023).

Some data-management platforms offer resources and datasets that can be strategically used to acquire envirotyping data. Using detailed climate classifications described by Köppen–Geiger (Cui et al., 2021) and Ecoregions (Dinerstein et al., 2017), it is possible to categorize planting environments based on specific climatic conditions. SoilGrids soil property maps (Poggio et al., 2021) provide information about the physical and chemical properties of multiple soil layers. Solar energy data from nasapower (Sparks, 2018), as well as information on wind, precipitation, seasonal humidity, climate seasonality, and even extreme events via WorldClim1 (Hijmans et al., 2005) and WorldClim2 (Fick and Hijmans, 2017) can be used to model the complex relationships that affect plant growth and development. The Environmental Data Initiative (Gries et al., 2023) provides access to a wide variety of envirotypic data, enriching enviromics studies with information about past and present environmental conditions. NASA GeneLab (Berrios et al., 2021) also contributes to the collection and organization of omics data and provides access to these data from space missions and analogous experiments, fostering scientific discoveries and shedding light on the effects of space environments on biology. All these resources empower researchers to consider myriad environmental factors, enhancing our understanding of G × E interactions and facilitating the selection of plants for sustainable agriculture and the conservation of biodiversity.

Despite the availability of numerous satellite data repositories, processing and analyzing large amounts of data still requires significant computational power and specialized knowledge about Earth data science. The cyberinfrastructure required for such endeavors includes high-performance computing and cloud services such as Amazon Web Services, Planetary Computer/Azure, and Google Earth Engine. The expected rise in the use of web tools, such as Google Earth Engine Applications and R Shiny Dashboards, in the coming decade highlights the increasing need for advanced tools that make data analysis more accessible. In this era of big data analytics, the ability to leverage purpose-driven envirotyping products will empower researchers to engage with the open satellite data revolution, facilitating informed decision making in the environmental and agricultural domains (Khanal et al., 2020; Vance et al., 2024).

## AI-ASSISTED ENVIROMICS

As discussed above, data from various sources are now being seamlessly combined thanks to platforms that provide uniform data access, storage solutions, application-based interfaces, and middleware, facilitating the merger of genotypic–enviromic–phenotypic (G–E–P) data into comprehensive knowledge networks (Lund, 2020). These tools deploy data-mining algorithms to weave together diverse data streams (Marsh et al., 2021). Techniques such as concatenation, transformation, and model-based integration facilitate the effective merging of datasets (Picard et al., 2021). Multiomics datasets make this process challenging due to their varying formats, scales, and dimensions, as they are often noisy, sparse, and collected under different conditions. To navigate these complexities, international standards and deep-learning algorithms are needed to manage nonlinear patterns and facilitate data integration (Selby et al., 2019; Montesinos-López et al., 2022). AI has emerged as an effective strategy to overcome the intricacies of enviromics datasets, enhancing predictions for plant breeding through integrating G–E–P data. This involves considering the structural nuances of breeding data and leveraging statistical methods to optimize predictions and decision-making processes in plant breeding (Xu et al., 2022).

In addition to traditional predictive methods (e.g., mixed models and/or Bayesian), AI and machine learning (ML) methods can assist in enviromics to enhance genotypic predictions and recommendations (Resende et al., 2021; Costa-Neto et al., 2023). Both AI and ML are triggering a paradigm shift in geoprocessing (GIS) and plant breeding. For geoprocessing, AI leverages advanced algorithms and ML models to extract knowledge from spatial data, enabling more precise analysis and decision making in fields such as environmental monitoring and disaster management. AI can enhance the accuracy of mapping, spatial pattern recognition, and predictive modeling, thereby revolutionizing how we understand and interact with geographic information (Khan et al., 2022; Montesinos-López et al., 2022). Emerging AI techniques are useful for identifying and selecting desirable traits in genetic datasets. Their potential to enhance genotype selection, crop yield optimization, and climate adaptability is an active area of research (Hayes et al., 2023; Negus et al., 2024).

Transformative synergy between Geoprocessing + AI (GeoAI) (Song et al., 2023) and plant breeding has the potential to address pressing global challenges related to food security and sustainable land management. Artificial neural networks and other AI techniques have been successfully applied for various purposes in geosciences and geotechnical engineering (Noack et al., 2014; Kim et al., 2019; Samui, 2020). In addition, several studies have demonstrated the superior performance of XGBoost and random forest algorithms in predicting geological properties (Naghibi et al., 2020; Zhang et al., 2021). Notable reviews by Negus et al. (2024) and Khan et al. (2022) highlight the opportunities to similarly exploit AI in plant breeding, which could transform crop improvement and lead to major advancements in agriculture.

## ENVIROMICS FOR CROP IMPROVEMENT

After identifying environmental targets, collecting phenotypic data, and acquiring environmental data, the next step in enviromics involves merging the datasets (Figure 1, step 5). This step combines environmental information with phenotypic observations and (perhaps) genomic or other omics datasets. The core issue is to ensure that the geographical coordinates of the sites containing phenotypic data use the same coordinate system (i.e., Datum) as the envirotyping data. This task ensures a nested evaluation that minimizes analysis noise and clarifies factor interactions, laying the groundwork for comprehensive enviromics analyses. Integrating multidimensional data, big data technology, and AI enables the development of an

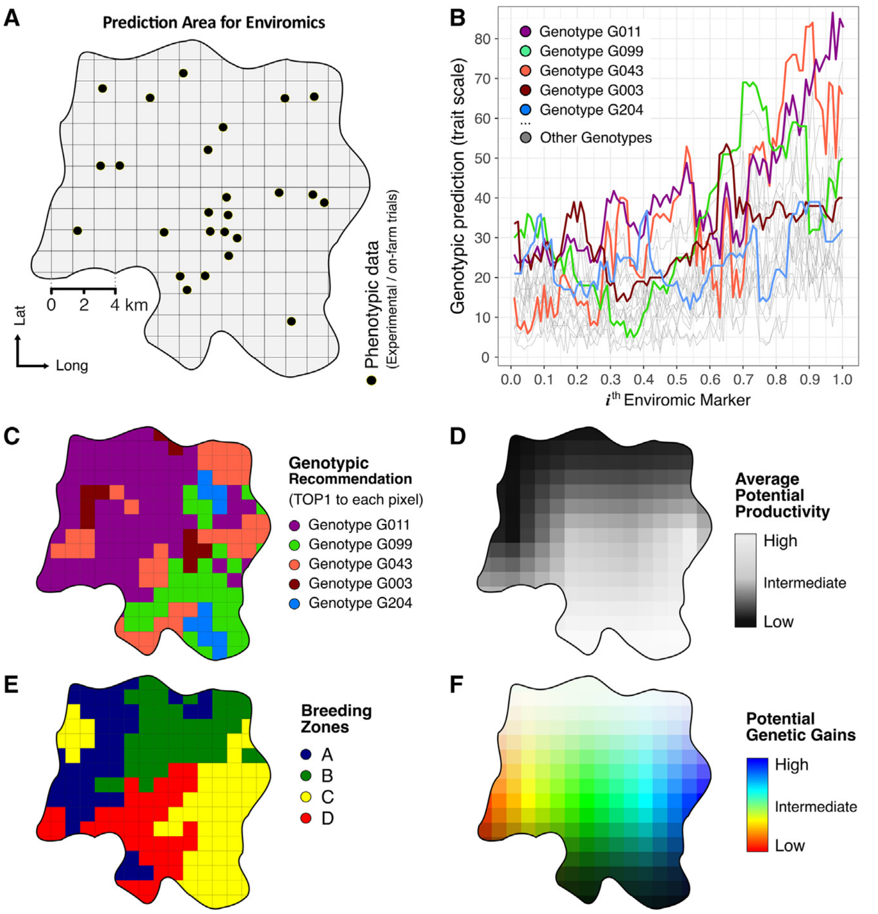


**Figure 4. A hypothetical enviromics analysis.**
This figure serves as a resource for informed local breeding decisions.
**(A)** The prediction area for enviromics or the TPE.
**(B)** Output from an infinitesimal random regression enviromics-based model.
**(C)** Genotypic recommendations: complete ranking of all evaluated genetic materials with predictive extrapolation for all pixels in the area.
**(D)** Potential genetic gains: optimal environments for implementing breeding experiments.
**(E)** Breeding zones: geographical polygons minimizing genotype-by-environment (G × E) interactions.
**(F)** Potential productivity: projection of productivity for the entire area or any other desired phenotypic traits. Results apply to both the population average and selected/recommended genetic materials.

intelligent and integrated G–E–P breeding scheme, leading to more precise phenotype predictions and greater genetic gains through integrative breeding platforms and open-source initiatives (Xu et al., 2022; Vance et al., 2024). Nevertheless, G × E × M interactions pose hurdles to optimizing genetic gains and crop productivity. Driven by advancements in genetic, genomic, and remote-sensing technologies, this could lead to the emergence of enviromics for complex trait prediction (Cooper and Messina, 2021).

Enviromics can also benefit from ecophysiological data integration, offering solutions for climate-smart agriculture, cost-effective field practices, and future plant breeding scenarios (Costa-Neto et al., 2021b). Incorporating probabilistic concepts from Bayesian models further improves cultivar recommendation processes in MET, enhancing our understanding of G × E interactions (Dias et al., 2020). Notably, countries such as Australia and other nations in the International Wheat Improvement Network have obtained substantial genetic gains using remote sensing combined with genomic selection through parent crossing and progeny selection, as highlighted in the comprehensive review by Chen et al. (2022), providing valuable platforms for ongoing research and refining breeding methodologies.

## Enviromics for predictive breeding

Finally, our discussion turns to step 6 in the enviromics process, as depicted in the enviromics analysis (Figure 1), where hypothetical results from Resende et al. (2021) are illustrated in Figure 4A. One model used in enviromics is random regression, where horizontal pulses along the enviromic marker gradient (x axis) represent a new predicted trial (Figure 4B) (i.e., a 100% virtual experiment). This model predicts the behavior of each genetic material and its rank in order, with predictive extrapolation at the site-specific level for all pixels in the area (Figure 4C). The average potential productivity of each selected/recommended genetic material can be predicted, as shown in Figure 4D. For example, integrating climate and geographic data allowed optimal eucalyptus genotypes to be selected across a wide area, tailoring clonal cultivar choices to maximize wood volume for different planting ages; this analysis showcased the power of innovative environmental stratification to optimize productivity (Marcatti et al., 2017).

Compared with genotypes, where inbred or hybrid varieties can generally be replicated, or single-locus genotypes or multi-locus haplotypes can be replicated by groups of individuals, envirotypes for certain environmental factors are considered to be generally replicable for any specific envirotyping location, as determined by longitude, latitude, and altitude. Major environmental factors (envirotypes), such as seasonal day/night length, temperature variations, and managed environments, are generally consistent and largely predictable, while minor environmental factors are largely unpredictable. The prediction accuracy for phenotypic performance is determined by complex combinations of genotypes, envirotypes, and their interactions (Araújo et al., 2024). Therefore, only major environmental factors can be used for classification and prediction. However, similar to the selection indices used to evaluate quantitative phenotypes, enviromic indices can be constructed for each specific envirotypic location/site using information extracted from all environmental factors based on their individual (infinitesimal) contributions to the total envirotypic variation and their envirotypic relationships

(Costa-Neto et al., 2021b; Li et al., 2021; Resende et al., 2021; Piepho, 2022; Xu et al., 2022).

Going further, if a G × E interaction is significantly associated with groups of environments within the TPE, greater genetic gains can be achieved by reorganizing experiments into "mega-environments" (Crespo-Herrera et al., 2021; Krause et al., 2022). This leads to a discussion about "breeding zones" versus mega-environments (Gauch and Zobel, 1997), where breeding zones refer to the re-aggregation of pixels minimizing G × E and not to groups of experiments. In other words, breeding zones are geographical polygons that minimize G × E (Figure 4E and refer Callister et al. (2024)). The partitioning of environments (geographical regions) into homogeneous subgroups has been performed for decades (DeLacy and Cooper, 1990; Ouyang et al., 1995). In this era of enviromics, we have much more diverse and complete envirotypic data available than ever before for establishing homogeneous subgroups by classifying or clustering the environmental trial sites using all enviromic information. Based on enviromic similarity levels or indices, breeding zones, experimental stations, and MET locations can be established and optimized. The more enviromic information used, the better the strategies that can be developed to optimize breeding pipelines and programs.

It is possible to achieve potential selection gains at the pixel level in an area. Optimal environments for implementing breeding experiments are shown in Figure 4F. This information can be especially useful when deploying new experiments in the next rotations. After all, sites with higher potential gains will likely provide more accurate genetic selections. According to Fernandes-Filho et al. (2023), including environmental information in the form of enviromics in genomic prediction models for assessing genotype performance in multi-harvest alfalfa (*Medicago sativa*) breeding experiments resulted in increased genetic variance, reduced error variance, and enhanced predictive capacity, especially for the adaptability and persistence of the evaluated genetic families.

## Field-scale enviromics and optimizing breeding programs

It is currently difficult to predict the effects of genetics and management practices on crop performance in a specific environment at regional-to-global scales. Multiscale crop modeling will allow gene-to-farm systems to be designed for resilient and sustainable crop production in a changing climate. Such modeling could be advanced by representing crop traits, interfacing crop models with large-scale models, improving the representation of physiological responses to climate change and management practices, closing data gaps, and harnessing multi-source data (Peng et al., 2020). Crop growth models provide a way to predict crop productivity in G × E × M scenarios, enabling the rapid design and testing of innovative crop breeding strategies based on an integrated understanding of G × E × M interactions. This will create opportunities to identify and implement pathways to increase productivity through integrating genetic gains from breeding and crop management strategies (Cooper et al., 2020).

At the field level, Guan et al. (2023) proposed a scalable framework to quantify carbon outcomes for farmlands, which measured the emissions of greenhouse gases, including $N_2O$ and $CH_4$, and changes in soil carbon stock. A "system-of-systems" solution was proposed based on integrating various approaches (e.g., diverse observations, sensor/*in situ* data, and modeling). This approach consists of five components: (1) scalable collection of ground-truth data and cross-scale sensing of E, M, and crop conditions at the local field level; (2) advanced modeling to support the quantification; (3) systematic model–data integration at the local farmland level; (4) high computation efficiency and AI to scale to millions of individual fields at low cost; and (5) robust and multi-tier validation systems and infrastructures to ensure solution fidelity and true scalability. This proposed solution should be generalized and used for quantifying other enviromic measurements in plant breeding.

Employing platforms such as The Climate Corporation and the Earth Observing System Data Analytics (EOSDA) Crop Monitoring, the agriculture and plant breeding sectors access extensive datasets, including meteorological data from over 2.5 million sites and 150 billion soil observations, to generate 10 trillion weather simulation data points. This collection of data enables detailed soil moisture monitoring, zoning for efficient resource use, and assessments of vegetation health, thereby driving genetic advancements and promoting sustainable agricultural practices. Specifically, EOSDA excels in providing critical soil and vegetation data through satellite-powered insights, integrating satellite imagery with high-resolution data from UAVs and ground vehicles to enhance the precision of enviromic data for specific locations. This approach, detailed at https://eos.com/blog/how-precision-farming-fights-climate-change/, supports the selection of optimal trial sites based on environmental similarities, accelerates breeding through engineered environments, and enhances breeding strategies. Notably, the adoption of genomic-enviromic prediction methods, as discussed by Xu et al. (2022), could offer improved accuracy over traditional genomic predictions, significantly boosting breeding efficiency and genetic gains.

## Temporal data and adaptation to climate change

Satellite data offer temporal information, providing dynamic measurements at various time points, which increase our understanding of the growth period of a crop and the similarities among different locations. However, some of these measurements can be affected by atmospheric events such as clouds, pollution, lightning, and solar/cosmic radiation. Rustowicz (2017) explored the use of time-series satellite imagery and ML techniques for crop classification. Going further, Pazúr et al. (2021) emphasized the value of fine-temporal-resolution satellite sensors for studying landscape ecology, showing that including temporal information improves the accuracy of landscape mapping and the identification of important landscape elements.

Exploring the impact of climate change on agriculture, the study by Rezaei et al. (2023) delves into the effects of warmer temperatures, elevated $CO_2$ levels, and changing water availability on crop yields. It uncovers varied responses from C3 and C4 crops to drought and high $CO_2$ levels, noting potential plant yield variability increases. Particularly, crops in lower latitudes could experience severe yield reductions, with losses anticipated to be between 7% and 23% without adaptive measures. The research underscores the importance of a multidisciplinary approach, recommending the combination of biophysical yield assessments with economic and environmental analyses to navigate the complex interactions of factors such as nitrogen loss, changes in soil organic matter, and crop nutritional quality (Ciscar et al., 2019). By leveraging envirotyping and accurate climate predictions, it suggests strategies for addressing the broad spectrum of climate-change challenges, advocating for the development of innovative agricultural methods to maintain productivity amid environmental changes. Integrating trial, breeding, and on-farm data with genotype annotations through big data structures clarifies G × E × M interactions, helping to manage the challenges imposed by climate change. Such integration, along with the use of diverse genetic materials in diverse environments, represents a path to improved cultivar performance. Crop environments must be characterized for enhanced breeding and germplasm selection tailored to the TPE (Chenu, 2015). Saltz et al. (2018) delved into the variability of G × E interactions, advocating for studies on their biological bases by examining traits linked to performance in current environments as indicators for future conditions. This approach to temporal continuity is instrumental for forecasting changes in genotypes for plant adaptation to the changing climate. Crafting crop strategies that accommodate these interactions and the realities of climate change is necessary to sustain productivity (Cooper et al., 2021). Notably, a G × M technology framework to adapt to climate change and secure food stability has been proposed (Messina and Cooper, 2022).

## CONCLUDING REMARKS

Enviromics has emerged as a powerful approach to enhance plant breeding, enabling the integration of multidimensional information from satellite-based remote-sensing data. The use of traditional statistics, big data, and AI in conjunction with multiple environmental datasets, many derived from satellite sensors, can provide a precise view of G × E × M interactions. This integration leads to the development of intelligent and integrated G–E–P breeding schemes, enabling more precise phenotypic predictions and greater genetic gains. This is made possible by predicting entire "virtual trials" that closely replicate reality, eliminating the need for physical trials, thereby reducing operational expenses. Enviromics offers a complementary dimension to genomics and phenomics, representing a promising, innovative path toward sustainable advances in crop science to bridge the gap between scientific knowledge and reality in the field and contribute to resilient agricultural crop production in the face of climate change.

With enviromics, the possibilities are vast. Genotypic recommendations with predictive extrapolation for all pixels (i.e., bins) in an area empower breeders to make local decisions based on detailed information about the behavior of genetic materials under different cultivation conditions. Furthermore, the concept of breeding zones, i.e., geographical polygons that minimize G × E (and in some cases G × E × M) interactions, represents an innovative approach for optimizing crop productivity on a large scale. The projection of potential productivity for the entire area and any other desired phenotypic traits highlights the usefulness of enviromics for obtaining predictive insights into crop productivity under different climatic scenarios. These findings can inform decision making regarding the selection of varieties better adapted to specific local environments and help reduce costs in the field under current and future scenarios.

There is a notable gap in the capabilities of many breeding programs due to a lack of proficiency in translating satellite-derived information into practical knowledge to inform decision making. To address this issue, a focused effort is needed to empower professionals working in the plant breeding industry with the necessary knowledge to fully leverage the wealth of satellite data. This would include knowledge about data processing, analytical methods, and the utilization of advanced technologies such as AI and statistical modeling. By nurturing these fundamental skills, plant breeding programs could unlock the tremendous potential of enviromics and effectively utilize satellite data as a potent instrument in advancing sustainable agriculture.

Although challenges remain, such as the need for more freely available, high-resolution satellite data and the widespread use of methods to integrate the data with genomics and phenomics data, it is clear that enviromics offers a new opportunity to enhance agricultural productivity and sustainability. There is a need to integrate low-resolution satellite data with high-resolution UAV data through imputation or AI to reduce the cost associated with obtaining high-resolution imagery at scale. Moreover, advanced AI models can improve several steps in enviromics, from data augmentation to data fusion, the use of complex models, and forecasting. Collaboration among agronomists, physiologists, breeders, geoinformatics experts, and programmers across research institutions is essential for advancing this field and harnessing the full potential of envirotyping data for the transformation of 21st century agriculture.

### SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

### FUNDING

R.T.R., L.L.P., and G.E.M. thank the Brazilian agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support, which was instrumental in the successful execution of this project. L.H. was supported through an ARC Future Fellowship (FT220100350) from the Australian Research Council. C.H.A. was supported by The University of Colorado Boulder Grand Challenge, CIRES Earth Lab. Y.X. was supported by the Agricultural Science and Technology Innovation Program (ASTIP) of the Chinese Academy of Agricultural Sciences, Shenzhen Science and Technology Program (KQTD20230301092839007O), Hebei Science and Technology Program (215A7612D), and the Provincial Technology Innovation Program of Shandong, China.

### AUTHOR CONTRIBUTIONS

R.T.R. led the project, drafted the initial manuscript, revised the manuscript, and created the figures. L.H. wrote the article and reviewed the sections on breeding and phenomics. C.H.A. wrote the article, reviewed the sections on GISs and remote sensing, and reviewed the figures and table. L.L.P. obtained satellite data, prepared the table, and created figures. G.E.M. helped write the sections on GIS and remote sensing and reviewed the article. Y.X. wrote the article and reviewed the sections on breeding and envirotyping.

## ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: November 15, 2023
Revised: April 4, 2024
Accepted: April 11, 2024
Published: April 17, 2024

## REFERENCES

**Al-Turjman, F.** (2019). The road towards plant phenotyping via WSNs: An overview. Comput. Electron. Agric. **161**:4–13. https://doi.org/10.1016/J.COMPAG.2018.09.018.

**Alvites, C., Marchetti, M., Lasserre, B., and Santopuoli, G.** (2022). LiDAR as a tool for assessing timber assortments: A systematic literature review. Remote Sens **14**:4466.

**Anthony, J.C., Eaton, W.W., and Henderson, A.S.** (1995). Looking to the future in psychiatric epidemiology. Epidemiol. Rev. **17**:240–242.

**Appel, M., and Pebesma, E.** (2019). On-demand processing of data cubes from satellite image collections with the gdalcubes library. Data **4**:92.

**Araújo, M.S., Chaves, S.F.S., Dias, L.A.S., Ferreira, F.M., Pereira, G.R., Bezerra, A.R.G., Alves, R.S., Heinemann, A.B., Breseghello, F., Carneiro, P.C.S., et al.** (2024). GIS-FA: an approach to integrating thematic maps, factor-analytic, and envirotyping for cultivar targeting. Theor. Appl. Genet. **137**:80.

**Aydinoglu, A.C.** (2016). Modelling, encoding and transforming of open geographic data to examine interoperability between GIS applications. Geocarto Int. **31**:446–461.

**San Bautista, A., Fita, D., Franch, B., Castiñeira-Ibáñez, S., Arizo, P., Sánchez-Torres, M.J., Becker-Reshef, I., Uris, A., and Rubio, C.** (2022). Crop monitoring strategy based on remote sensing data (Sentinel-2 and Planet), Study case in a rice field after applying Glycinebetaine. Agronomy **12**:708.

**Beckers, J., Wurst, W., and De Angelis, M.H.** (2009). Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. Nat. Rev. Genet. **10**:371–380.

**Berrios, D.C., Galazka, J., Grigorev, K., Gebre, S., and Costes, S.V.** (2021). NASA GeneLab: interfaces for the exploration of space omics data. Nucleic Acids Res. **49**:D1515–D1522.

**Brady, R., and Spring, A.** (2021). Climpred: Verification of weather and climate forecasts. J. Open Source Softw. **6**:2781.

**Breiman, L.** (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Stat. Sci. **16**:199–231.

**Brönnimann, S., Allan, R., Atkinson, C., Buizza, R., Bulygina, O., Dahlgren, P., Dee, D., Dunn, R., Gomes, P., John, V.O., et al.** (2018). Observations for reanalyses. Bull. Am. Meteorol. Soc. **99**:1851–1866.

**Budhlakoti, N., Kushwaha, A.K., Rai, A., Chaturvedi, K.K., Kumar, A., Pradhan, A.K., Kumar, U., Kumar, R.R., Juliana, P., Mishra, D.C., and Kumar, S.** (2022). Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. Front. Genet. **13**:832153.

**Bustos-Korts, D., Malosetti, M., Chenu, K., Chapman, S., Boer, M.P., Zheng, B., and Van Eeuwijk, F.A.** (2019). From QTLs to adaptation landscapes: using genotype-to-phenotype models to characterize G× E over time. Front. Plant Sci. **10**:540.

**Callister, A.N., Costa-Neto, G., Bradshaw, B.P., Elms, S., Crossa, J., and Brawner, J.T.** (2024). Enviromic prediction enables the characterization and mapping of *Eucalyptus globulus* Labill breeding zones. Tree Genet. Genomes **20**:3.

**Chaves, L.J., and de Miranda Filho, J.B.** (1992). Plot size for progeny selection in maize (*Zea mays* L.). Theor. Appl. Genet. **84**:963–970.

**Chen, C.J., Rutkoski, J., Schnable, J.C., Murray, S.C., Wang, L., Jin, L.W.X., Stich, B., Crossa, J., Hayes, B.J., and Zhang, Z.** (2022). Role of the genomics–phenomics–agronomy paradigm in plant breeding. Plant Breed. Rev. **46**:627–673.

**Chenu, K.** (2015). Characterizing the crop environment–nature, significance and applications. Crop Physiol, 321–348.

**Cooper, M., Messina, C.D., Podlich, D., Totir, L.R., Baumgarten, A., Hausmann, N.J., Wright, D., and Graham, G.** (2014). Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci. **65**:311–336.

**Ciscar, J.-C., Rising, J., Kopp, R.E., and Feyen, L.** (2019). Assessing future climate change impacts in the EU and the USA: insights and lessons from two continental-scale projects. Environ. Res. Lett. **14**:84010.

**Cooper, M., and Messina, C.D.** (2021). Can we harness "*Enviromics*" to accelerate crop improvement by integrating breeding and agronomy? Front. Plant Sci. **12**:735143.

**Cooper, M., Tang, T., Gho, C., Hart, T., Hammer, G., and Messina, C.** (2020). Integrating genetic gain and gap analysis to predict improvements in crop productivity. Crop Sci. **60**:582–604. https://doi.org/10.1002/csc2.20109.

**Cooper, M., Voss-Fels, K.P., Messina, C.D., Tang, T., and Hammer, G.L.** (2021). Tackling G× E× M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. Theor. Appl. Genet. **134**:1625–1644.

**Costa-Neto, G., Crespo-Herrera, L., Fradgley, N., Gardner, K., Bentley, A.R., Dreisigacker, S., Fritsche-Neto, R., Montesinos-López, O.A., and Crossa, J.** (2023). Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. G3 **13**:jkac313.

**Costa-Neto, G., Crossa, J., and Fritsche-Neto, R.** (2021b). Enviromic assembly increases accuracy and reduces costs of the genomic prediction for yield plasticity in maize. Front. Plant Sci. **12**:717552.

**Costa-Neto, G., and Fritsche-Neto, R.** (2021). Enviromics: bridging different sources of data, building one framework. Crop Breed. Appl. Technol. **21**:e393521S12.

**Costa-Neto, G., Galli, G., Carvalho, H.F., Crossa, J., and Fritsche-Neto, R.** (2021a). EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3 **11**, jkab040.

**Crespo-Herrera, L.A., Crossa, J., Huerta-Espino, J., Mondal, S., Velu, G., Juliana, P., Vargas, M., Pérez-Rodríguez, P., Joshi, A.K., Braun, H.J., and Singh, R.P.** (2021). Target population of environments for wheat breeding in India: definition, prediction and genetic gains. Front. Plant Sci. **12**:638520.

**Crossa, J., Fritsche-Neto, R., Montesinos-Lopez, O.A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., and Bentley, A.R.** (2021). The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. Front. Plant Sci. **12**:651480.

**Crossa, J., Montesinos-López, O.A., Pérez-Rodríguez, P., Costa-Neto, G., Fritsche-Neto, R., Ortiz, R., Martini, J.W.R., Lillemo, M., Montesinos-López, A., Jarquin, D., et al.** (2022). Genome and environment based prediction models and methods of complex traits incorporating genotype× environment interaction. Methods Mol. Biol. **2467**:245–283.

**Cui, D., Liang, S., Wang, D., and Liu, Z.** (2021). A 1 km global dataset of historical (1979–2013) and future (2020–2100) Köppen–Geiger climate classification and bioclimatic variables. Earth Syst. Sci. Data **13**:5087–5114.

**Debnath, S., Paul, M., and Debnath, T.** (2023). Applications of LiDAR in agriculture and future research directions. J. Imaging **9**:57.

**DeLacy, I.H., and Cooper, M.** (1990). Pattern analysis for the analysis of regional variety trials. In Genotype-by-environment Interaction and Plant Breeding, M.S. Kang, ed. (Louisiana State University), pp. 301–334.

**Delucchi, L., and Neteler, M.G.** (2013). pyModis: the Python library for MODIS data. In FOSS4G 2013, pp. 17–21.

**Dias, K.O.G., Piepho, H.P., Guimarães, L.J.M., Guimarães, P.E.O., Parentoni, S.N., Pinto, M.O., Noda, R.W., Magalhães, J.V., Guimarães, C.T., Garcia, A.A.F., and Pastina, M.M.** (2020). Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theor. Appl. Genet. **133**:443–455.

**Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., et al.** (2017). An ecoregion-based approach to protecting half the terrestrial realm. Bioscience **67**:534–545.

**Domingues, G.F., Souza, G.S.A. de, Barbosa, R.A., Marcatti, G.E., Lorenzon, A.S., and Leite, H.G.** (2023). Estimated volume of eucalyptus plantations through ALOS satellite images. TreeDimensional **11**, e2023025.

**van Emmerik, T., Steele-Dunne, S., Paget, A., Oliveira, R.S., Bittencourt, P.R.L., Barros, F.d.V., and van de Giesen, N.** (2017). Water stress detection in the Amazon using radar. Geophys. Res. Lett. **44**:6841–6849.

**Filho, C.C.F., Andrade, M.H.M.L., Nunes, J.A.R., Jarquin, D.H., and Rios, E.F.** (2023). Genomic prediction for complex traits across multiples harvests in alfalfa (*Medicago sativa* L.) is enhanced by enviromics. Plant Genome **16**:20306.

**Fick, S.E., and Hijmans, R.J.** (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. **37**:4302–4315.

**Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., et al.** (2020). ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the international space station. Water Resour. Res. **56**:e2019WR026058.

**Fritsche-Neto, R.** (2023). SoilType: An R package to interplay soil characterization in plant science. Agron. J. **116**:848–854. https://doi.org/10.1002/agj2.21383.

**Gauch, H.G., and Zobel, R.W.** (1997). Identifying mega-environments and targeting genotypes. Crop Sci. **37**:311–326.

**Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M.G., Reichle, R., et al.** (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. **30**:5419–5454.

**Gries, C., Hanson, P.C., O'Brien, M., Servilla, M., Vanderbilt, K., and Waide, R.** (2023). The Environmental Data Initiative: Connecting the past to the future through data reuse. Ecol. Evol. **13**:e9592.

**Guan, K., Jin, Z., Peng, B., Tang, J., DeLucia, E.H., West, P.C., Jiang, C., Wang, S., Kim, T., Zhou, W., et al.** (2023). A scalable framework for quantifying field-level agricultural carbon outcomes. Earth Sci. Rev. **243**:104462.

**Guo, T., and Li, X.** (2023). Machine learning for predicting phenotype from genotype and environment. Curr. Opin. Biotechnol. **79**:102853.

**Hayes, B.J., Chen, C., Powell, O., Dinglasan, E., Villiers, K., Kemper, K.E., and Hickey, L.T.** (2023). Advancing artificial intelligence to help feed the world. Nat. Biotechnol. **41**:1188–1189.

**Heinemann, A.B., Costa-Neto, G., Fritsche-Neto, R., da Matta, D.H., and Fernandes, I.K.** (2022). Enviromic prediction is useful to define the limits of climate adaptation: a case study of common bean in Brazil. Field Crops Res. **286**:108628.

**Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.** (2020). The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. **146**:1999–2049.

**Hijmans, R.J., Bivand, R., Forner, K., Ooms, J., Pebesma, E., and Sumner, M.D.** (2022). Package 'terra'. Maintainer.

**Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A.** (2005). Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**:1965–1978.

**Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., and Snow, V.** (2018). APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environ. Model. Software **103**:43–51.

**IndianaMap.** (2011). IndianaMap Framework Lidar. Indiana Statewide Imagery and LiDAR Program. Distributed by OpenTopography. https://doi.org/10.5069/G9959FHZ.

**Jakobsen, P., Ferruit, P., de Oliveira, C.A., Arribas, S., Bagnasco, G., Barho, R., Beck, T.L., Birkmann, S., Boker, T., Bunker, A.J., et al.** (2022). The near-infrared spectrograph (nirspec) on the james webb space telescope-i. overview of the instrument and its capabilities. Astron. Astrophys. **661**:A80.

**Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., et al.** (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. **127**:595–607.

**Jensen, J.R.** (2009). Remote sensing of the environment: An earth resource perspective 2/e (Pearson Education India).

**Karwowska, K., and Wierzbicki, D.** (2022). Using super-resolution algorithms for small satellite imagery: A systematic review. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. **15**:3292–3312.

**Khan, M.H.U., Wang, S., Wang, J., Ahmar, S., Saeed, S., Khan, S.U., Xu, X., Chen, H., Bhat, J.A., and Feng, X.** (2022). Applications of artificial intelligence in climate-resilient smart-crop breeding. Int. J. Mol. Sci. **23**:11156.

**Khanal, S., Kc, K., Fulton, J.P., Shearer, S., and Ozkan, E.** (2020). Remote sensing in agriculture - accomplishments, limitations, and opportunities. Rem. Sens. **12**:3783.

**Khosla, R.** (2023). Big data in agriculture: agronomic considerations. Soil and Crop Sciences Section Seminar Series (Cornell University).

**Kim, H.J., Dinoy, P.R.T., Choi, H.S., Lee, K.B., and Mission, J.L.C.** (2019). Spatial interpolation of SPT data and prediction of consolidation of clay by ANN method. Coupled Syst. Mech. **8**:523–535.

**Kokaly, R.F., Despain, D.G., Clark, R.N., and Livo, K.** (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. Remote Sens. Environ. **84**:437–456.

**Kouzis-Loukas, D.** (2016). Learning Scrapy. Livery Place (Packt Publishing).

**Krause, M.D., Dias, K.O.G., Singh, A.K., and Beavis, W.D.** (2022). Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.11.487885.

**Dalsente Krause, M., Dias, K.O.D.G., Pedroso Rigal dos Santos, J., de Oliveira, A.A., Guimarães, L.J.M., Pastina, M.M., Margarido, G.R.A., and Garcia, A.A.F.** (2020). Boosting predictive ability of tropical maize hybrids via genotype-by-environment interaction under multivariate GBLUP models. Crop Sci. **60**:3049–3065.

**Krinitskiy, M., Sprygin, A., Elizarov, S., Narizhnaya, A., Shikhov, A., and Chernokulsky, A.** (2023). Towards the accurate automatic detection of mesoscale convective systems in remote sensing data: from data mining to deep learning models and their applications. Remote Sens **15**:3493.

**Lechner, A.M., Foody, G.M., and Boyd, D.S.** (2020). Applications in remote sensing to forest ecology and management. One Earth **2**:405–412.

**Levchenko, I., Bazaka, K., Ding, Y., Raitses, Y., Mazouffre, S., Henning, T., Klar, P.J., Shinohara, S., Schein, J., Garrigues, L., et al.** (2018). Space micropropulsion systems for Cubesats and small satellites: From proximate targets to furthermost frontiers. Appl. Phys. Rev. **5**:011104.

**Li, X., Guo, T., Bai, G., Zhang, Z., See, D., Marshall, J., Garland-Campbell, K.A., and Yu, J.** (2022). Genetics-inspired data-driven approaches explain and predict crop performance fluctuations attributed to changing climatic conditions. Mol. Plant **15**:203–206.

**Li, X., Guo, T., Wang, J., Bekele, W.A., Sukumaran, S., Vanous, A.E., McNellie, J.P., Tibbs-Cortes, L.E., Lopes, M.S., Lamkey, K.R., et al.** (2021). An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. Mol. Plant **14**:874–887.

**Lund, H.** (2020). Most Common Types of Data Integration Methods. https://www.rapidionline.com/blog/most-common-types-of-data-integration-methods.

**Malosetti, M., Ribaut, J.M., and van Eeuwijk, F.A.** (2013). The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Front. Physiol. **4**:37433.

**Marcatti, G.E., Resende, R.T., Resende, M.D.V., Ribeiro, C.A.A., dos Santos, A.R., da Cruz, J.P., and Leite, H.G.** (2017). GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. For. Ecol. Manage. **392**:144–153.

**Marsh, J.I., Hu, H., Gill, M., Batley, J., and Edwards, D.** (2021). Crop breeding for a changing climate: integrating phenomics and genomics with bioinformatics. Theor. Appl. Genet. **134**:1677–1690.

**Messina, C., and Cooper, M.** (2022). Plant Breeding in the face of climate change. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.07.511293.

**Millet, E.J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., van Eeuwijk, F., and Tardieu, F.** (2019). Genomic prediction of maize yield across European environmental conditions. Nat. Genet. **51**:952–956.

**Mishra, V.N., Prasad, R., Kumar, P., Srivastava, P.K., and Rai, P.K.** (2017). Knowledge-based decision tree approach for mapping spatial distribution of rice crop using C-band synthetic aperture radar-derived information. J. Appl. Remote Sens. **11**:1.

**Montesinos-López, O.A., Montesinos-López, A., and Crossa, J.** (2022). Multivariate Statistical Machine Learning Methods for Genomic Prediction (Springer Nature), p. 691.

**Naghibi, S.A., Hashemi, H., Berndtsson, R., and Lee, S.** (2020). Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors. J. Hydrol. X. **589**:125197.

**Nattino, F., Ku, O., Grootes, M.W., Izquierdo-Verdiguier, E., Girgin, S., and Zurita-Milla, R.** (2022). CGC: a scalable Python package for co-and tri-clustering of geodata cubes. J. Open Source Softw. **7**:4032.

**Negus, K.L., Li, X., Welch, S.M., and Yu, J.** (2024). The role of artificial intelligence in crop improvement. Adv. Agron. **184**:1–66.

**Neisse, A.C., Kirch, J.L., and Hongyu, K.** (2018). AMMI and GGE Biplot for genotype× environment interaction: a medoid–based hierarchical cluster analysis approach for high–dimensional data. Biom. Lett. **55**:97–121.

**Noack, S., Knobloch, A., Etzold, S.H., Barth, A., and Kallmeier, E.** (2014). Spatial predictive mapping using artificial neural networks. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. **40**:79–86.

**Olivoto, T., Lúcio, A.D., da Silva, J.A., Marchioro, V.S., de Souza, V.Q., and Jost, E.** (2019). Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. Agron. J. **111**:2949–2960.

**Ouyang, Z., Mowers, R.P., Jensen, A., Wang, S., and Zeng, S.** (1995). Cluster analysis for genotype × environment interaction with unbalanced data. Crop Sci. **33**:1300–1305.

**Patten, B.C.** (1991). Holoevolution: the genotype–phenotype–envirotype complex in genetic and extragenetic inheritance. Holoecology: The Unification of Nature by Network Indirect Effects, Unpublished Manuscript, Complexity in Ecological Science Series (Columbia University Press).

**Pazúr, R., Price, B., and Atkinson, P.M.** (2021). Fine temporal resolution satellite sensors with global coverage: an opportunity for landscape ecologists. Landsc. Ecol. **36**:2199–2213.

**Pebesma, E., and Bivand, R.** (2023). Spatial Data Science: With Applications in R (CRC Press).

**Peng, B., Guan, K., Tang, J., Ainsworth, E.A., Asseng, S., Bernacchi, C.J., Cooper, M., Delucia, E.H., Elliott, J.W., Ewert, F., et al.** (2020). Towards a multiscale crop modelling framework for climate change adaptation assessment. Nat. Plants **6**:338–348.

**Picard, M., Scott-Boyer, M.P., Bodein, A., Périn, O., and Droit, A.** (2021). Integration strategies of multi-omics data for machine learning analysis. Comput. Struct. Biotechnol. J. **19**:3735–3746.

**Piepho, H.P.** (2022). Prediction of and for new environments: What's your model? Mol. Plant **15**:581–582.

**Piepho, H.P., and Blancon, J.** (2023). Extending Finlay–Wilkinson regression with environmental covariates. Plant Breed. **142**:621–631.

**Pinto, F., Zaman-Allah, M., Reynolds, M., and Schulthess, U.** (2023). Satellite imagery for high-throughput phenotyping in breeding plots. Front. Plant Sci. **14**:1114670.

**Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., and Rossiter, D.** (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil **7**:217–240.

**Ranghetti, L., Boschetti, M., Nutini, F., and Busetto, L.** (2020). "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. Comput. Geosci. **139**:104473.

**Reichle, R.H., Draper, C.S., Liu, Q., Girotto, M., Mahanama, S.P.P., Koster, R.D., and De Lannoy, G.J.M.** (2017). Assessment of MERRA-2 land surface hydrology estimates. J. Clim. **30**:2937–2960.

**Reitz, K.** (2024). Requests Documentation Release 2.31.0. https://requests.readthedocs.io/en/latest/.

**Resende, R.T., Chenu, K., Rasmussen, S.K., Heinemann, A.B., and Fritsche-Neto, R.** (2022). Editorial: Enviromics in plant breeding. Front. Plant Sci. **13**:935380.

**Resende, R.T., Piepho, H.P., Rosa, G.J.M., Silva-Junior, O.B., e Silva, F.F., de Resende, M.D.V., and Grattapaglia, D.** (2021). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. Theor. Appl. Genet. **134**:95–112.

**Resende, R.T.** (2024). Balancing genomic selection efforts for allogamous plant breeding programs. J. Crop Sci. Biotechnol. **27**:349–358. https://doi.org/10.1007/s12892-024-00234-7.

**Rezaei, E.E., Webber, H., Asseng, S., Boote, K., Durand, J.L., Ewert, F., Martre, P., and MacCarthy, D.S.** (2023). Climate change impacts on crop yields. Nat. Rev. Earth Environ. **4**:831–846.

**Rincent, R., Malosetti, M., Ababaei, B., Touzy, G., Mini, A., Bogard, M., Martre, P., Le Gouis, J., and Van Eeuwijk, F.** (2019). Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theor. Appl. Genet. **132**:3399–3411.

**Rizzo, R., Wadoux, A.M.C., Demattê, J.A., Minasny, B., Barrón, V., Ben-Dor, E., Francos, N., Savin, I., Poppiel, R., Silvero, N.E., et al.** (2023). Remote sensing of the Earth's soil color in space and time. Remote Sens. Environ. **299**:113845.

**Rustowicz, R.M.** (2017). Crop classification with multi-temporal satellite imagery. Final Reports (Univ. Stanford).

**Saltz, J.B., Bell, A.M., Flint, J., Gomulkiewicz, R., Hughes, K.A., and Keagy, J.** (2018). Why does the magnitude of genotype-by-environment interaction vary? Ecol. Evol. **8**:6342–6353.

**Samui, P.** (2020). Application of artificial intelligence in geo-engineering. In Information Technology in Geo-Engineering: Proceedings of the 3rd International Conference (ICITG), Guimarães, Portugal, *3* (Springer International Publishing), pp. 30–44.

**Schmidt, P., Hartung, J., Rath, J., and Piepho, H.P.** (2019). Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci. **59**:525–536.

**Selby, P., Abbeloos, R., Backlund, J.E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O.E., Birkett, C., Calaminos, V.C., Carceller, P., Cornut, G., et al.** (2019). BrAPI – an application programming interface for plant breeding applications. Bioinformatics **35**:4147–4155.

**Ayoub Shaikh, T., Rasool, T., and Rasheed Lone, F.** (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Comput. Electron. Agric. **198**:107119.

**Shmueli, G.** (2010). To Explain or to Predict? Stat. Sci. **25**:289–310. https://doi.org/10.1214/10-STS330.

**da Silva, V.S., Salami, G., da Silva, M.I.O., Silva, E.A., Monteiro Junior, J.J., and Alba, E.** (2020). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. Geol. Ecol. Landsc. **4**:159–169.

**Skovsgaard, J.P., and Vanclay, J.K.** (2008). Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry **81**:13–31.

**Som-ard, J., Immitzer, M., Vuolo, F., and Atzberger, C.** (2024). Sugarcane yield estimation in Thailand at multiple scales using the integration of UAV and Sentinel-2 imagery. Precis. Agric. **25**:1581–1608. https://doi.org/10.1007/s11119-024-10124-1.

**Smith, A., Norman, A., Kuchel, H., and Cullis, B.** (2021). Plant variety selection using interaction classes derived from factor analytic linear mixed models: models with independent variety effects. Front. Plant Sci. **12**:737462.

**Song, Y., Kalacska, M., Gašparović, M., Yao, J., and Najibi, N.** (2023). Advances in geocomputation and geospatial artificial intelligence (GeoAI) for mapping. Int. J. Appl. Earth Obs. Geoinf. **120**, 103300.

**Sousa, M., Galli, G., Lyra, D., Granato, Í., Matias, F., Alves, F., and Fritsche-Neto, R.** (2019). Increasing accuracy and reducing costs of genomic prediction by marker selection. Euphytica **215**:1–14. https://doi.org/10.1007/s10681-019-2339-z.

**Sparks, A.** (2018). nasapower: a NASA POWER global meteorology, surface solar energy and climatology data client for R. J. Open Source Softw. **3**:1035.

**Terentev, A., Dolzhenko, V., Fedotov, A., and Eremenko, D.** (2022). Current state of hyperspectral remote sensing for early plant disease detection: a review. Sensors **22**:757. https://doi.org/10.3390/s22030757.

**Tomar, V., Mandal, V.P., Srivastava, P., Patairiya, S., Singh, K., Ravisankar, N., Subash, N., and Kumar, P.** (2014). Rice equivalent crop yield assessment using MODIS sensors' based MOD13A1-NDVI data. IEEE Sensor. J. **14**:3599–3605. https://doi.org/10.1109/JSEN.2014.2329185.

**Uthes, S., Kelly, E., and König, H.J.** (2020). Farm-level indicators for crop and landscape diversity derived from agricultural beneficiaries data. Ecol. Indicat. **108**:105725. https://doi.org/10.1016/j.ecolind.2019.105725.

**van Eeuwijk, F.A., Bustos-Korts, D.V., and Malosetti, M.** (2016). What should students in plant breeding know about the statistical aspects of genotype × environment interactions? Crop Sci. **56**:2119–2140.

**Velastegui-Montoya, A., Montalván-Burbano, N., Carrión-Mero, P., Rivera-Torres, H., Sadeck, L., and Adami, M.** (2023). Google Earth Engine: a global analysis and future trends. Remote Sens **15**:3675.

**Venancio, L.P., Mantovani, E.C., do Amaral, C.H., Usher Neale, C.M., Gonçalves, I.Z., Filgueiras, R., and Campos, I.** (2019). Forecasting corn yield at the farm level in Brazil based on the FAO-66 approach and soil-adjusted vegetation index (SAVI). Agric. Water Manag. **225**:105779.

**Vance, T.C., Huang, T., and Butler, K.A.** (2024). Big data in Earth science: Emerging practice and promise. Science **383**: eadh9607.

**Voitik, A., Kravchenko, V., Pushka, O., Kutkovetska, T., Shchur, T., and Kocira, S.** (2023). Comparison of NDVI, NDRE, MSAVI and NDSI indices for Early Diagnosis of crop problems. Agric. Eng. **27**:47–57. https://doi.org/10.2478/agriceng-2023-0004.

**Wang, A., Wo, J., Luo, X., Wang, Y., Cong, W., Du, P., Zhang, J., Zhao, B., Zhang, J., Zhu, Y., et al.** (2018). Ka-band microwave photonic ultra-wideband imaging radar for capturing quantitative target information. Opt Express **26**:20708–20717. https://doi.org/10.1364/OE.26.020708.

**Weiss, M., Jacob, F., and Duveiller, G.** (2020). Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. **236**, 111402. https://doi.org/10.1016/j.rse.2019.111402.

**Westra, E.** (2016). Python Geospatial Development, 3rd ed. (Packt Publishing).

**Xu, Y.** (2010). Molecular Plant Breeding (CABI Publisher), p. 734pp.

**Xu, Y.** (2016). Envirotyping for deciphering environmental impacts on crop plants. Theor. Appl. Genet. **129**:653–673.

**Xu, Y., Zhang, X., Li, H., Zheng, H., Zhang, J., Olsen, M.S., Varshney, R.K., Prasanna, B.M., and Qian, Q.** (2022). Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Mol. Plant **15**:1664–1695.

**Yang, B., Zhu, W., Rezaei, E.E., Li, J., Sun, Z., and Zhang, J.** (2022). The optimal phenological phase of maize for yield prediction with high-frequency UAV remote sensing. Rem. Sens. **14**:1559.

**Yue, J., Tian, J., Tian, Q., Xu, K., and Xu, N.** (2019). Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. ISPRS J. Photogrammetry Remote Sens. **154**:216–230.

**Zhang, W., Wu, C., Zhong, H., Li, Y., and Wang, L.** (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. **12**:469–477.

**Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., Zhao, X., and Gai, J.** (2019). Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. Remote Sens **11**:2752.

**Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., and Gong, P.** (2022). An overview of the applications of earth observation satellite data: impacts and future trends. Rem. Sens. **14**:1863.