Multi-Environment Genomic Prediction Models for Hybrid Maize Performance

Hongpeng Wang, Minghua Li ✉

Biotechnology Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, China

✉ Corresponding author: minghua.li@jicat.org

Abstract Hybrid maize breeding relies on accurate prediction of hybrid performance across diverse environments to enhance yield stability and adaptability. This study focused on developing and evaluating Multi-Environment Genomic Prediction (MEGP) models to improve the predictive accuracy of hybrid maize performance under variable environmental conditions. We first examined the principles of genomic prediction and the limitations of single-environment models before implementing MEGP frameworks that integrate genotype-by-environment (G × E) interactions through reaction norm and factor analytic approaches. Environmental variation was quantified using spatial and temporal covariates, while envirotyping provided additional insights into environmental effects on hybrid performance. A multi-year hybrid maize trial was conducted to assess the MEGP models, integrating genotypic, phenotypic, and environmental data. Results demonstrated that MEGP models significantly outperformed single-environment models in predictive accuracy and heritability estimates, highlighting their potential for more robust selection decisions. The study also explored the integration of high-throughput phenotyping, remote sensing, and machine learning techniques to further enhance model performance. Overall, MEGP models present a promising framework for accelerating hybrid maize breeding, improving climate resilience, and supporting global breeding networks through data-driven decision-making.

Keywords Genomic prediction; Hybrid maize; Multi-environment models; G×E interactions; Breeding optimization

1 Introduction

When it comes to hybrid corn breeding, what often comes to mind first are high yield and stable yield. But this outcome is actually not accidental. Since the 20th century, hybrid breeding has been regarded as one of the most core approaches to crop improvement. Based on the principle of "heterosis", researchers have continuously improved the parent combination and eventually cultivated a batch of corn varieties that have performed outstandingly in commercial production-they can not only achieve high yields but also maintain stability under different adverse conditions (Zhao et al., 2025).

Of course, the process is not as simple as it sounds. Breeding work often begins with the selection of suitable inbred lines, which should have ideal "compatibility", taking into account both general and special aspects. Afterwards, these combinations will be placed in multi-environment trials (METs) for verification. There are significant differences in location, climate and soil. A variety may perform extremely well in one place but be unremarkable in another environment. It is precisely this complex relationship between genotypes and the environment (GEI) that makes breeding work full of challenges. To evaluate these interaction effects more accurately, researchers have to rely on more sophisticated statistical tools and computational models (Yu et al., 2020; Supriadi et al., 2024; Popa et al., 2025).

The emergence of genomic prediction (GP) has changed all of this. Results that used to take years of experimentation to obtain can now be "rehearsed" in advance through data simulation. GP uses whole-genome molecular markers to estimate an individual's breeding value. In other words, it uses data to replace field trials and run through them first (Wang, 2025). Compared with traditional marker-assisted selection, it is better at handling complex traits influenced by multiple genes, such as yield and stress resistance. Even for genotypes that have never been tested, the model can provide relatively accurate predictions. This means that the breeding cycle has shortened, costs have decreased, and the pace of genetic improvement has also accelerated. Nowadays, GP has

almost become the "standard configuration" of modern breeding systems, supporting key links such as rapid cycling, optimized training population and multi-omics data integration (Beyene et al., 2019; Alemu et al., 2024; Zhang et al., 2025).

However, even the most advanced algorithms have blind spots. If environmental differences are ignored, model predictions are very likely to "go off track". Introducing multi-environment data into genomic prediction is the key to improving the accuracy of prediction. Practical experience has shown that models that can take into account environmental covariates, spatial variations or label-environment interactions often perform better than single-environment models. It is reported that the prediction accuracy of the main traits of corn can be increased by an average of 12% to 20%. For breeders, this is not merely a numerical advancement, but rather an opportunity to truly screen out more stable and reliable varieties. Especially in today's increasingly intense climate change, the importance of stability has been infinitely magnified. Meanwhile, the introduction of deep learning and multimodal data enables complex GEI patterns to be presented more precisely. The goal of breeding is also quietly changing, from a single "high yield" to "high yield and wide adaptability", laying the foundation for cultivating corn hybrids with greater environmental adaptability (Xu et al., 2022; Gao et al., 2025; Zou et al., 2025).

2 Fundamentals of Genomic Prediction in Maize

2.1 Basic principles of genomic selection and prediction accuracy

In corn breeding, genomic selection (GS) is no longer a novelty. Its core idea is actually quite straightforward-through whole-genome molecular markers, the genetic potential of individuals can be judged in advance, thus enabling the identification of "good signs" at an early stage of breeding. However, the accuracy of this process is not constant. Marker density, the size and composition of the training population, trait heritability, genetic structure, and the type of model used all have an impact on the prediction results (Zhang et al., 2019). Generally speaking, the denser the markers, the larger the population, and the more diverse the genetic differences, the more reliable the prediction results will be, especially for traits with strong heritability.

Interestingly, the information captured by different models also varies. Models like GBLUP, BayesB, and RKHS not only look at the additive effect but also take the non-additive part into account, which makes them perform more stability in the prediction of hybrid maize traits (Kaler et al., 2022). Furthermore, if environmental factors or the interaction between genotype and environment (G×E) are added to the model, the prediction accuracy often reaches a new level, especially for traits that are greatly affected by the environment, such as grain yield and drought resistance, the improvement is particularly significant.

2.2 Role of marker-based models in predicting hybrid performance

In corn breeding, marker-based prediction models are almost the mainstay. They estimate the genomic breeding values (GEBV) of hybrid combinations that have not yet undergone field trials through SNP markers or other genotyping platforms. However, merely having markers is not enough. Incorporating dominant effects, superposition effects and population structure into the model together will make the prediction results closer to reality, especially for traits that have a significant impact on non-additive inheritance (Luo et al., 2024).

Interestingly, some studies have mentioned that when the model incorporates both additive and dominant effects, the prediction accuracy of corn grain yield can increase by approximately 30% (Ferrão et al., 2020). However, if one goes further and combines the functional markers screened out in the GWAS analysis with the trait specific markers, the prediction results tend to be more stable and closer to the true performance (Yu et al., 2022).

For breeders, the significance of such improvements does not lie in the technology itself, but in saving experimental costs. In the past, it took a lot of time and resources to conduct hundreds of field validations. Now, with the help of these models, more promising combinations can be selected before the experiment begins. In other words, it has transformed breeding from "trying it out" to "calculating it out", accelerating the pace of genetic improvement.

2.3 Historical development and limitations of single-environment models

The initial maize genome prediction models were all rather "simple"-targeting a single environment, a single trait, and most of them only analyzed additive genetic effects (Rice and Lipka, 2021). These models do open up new ideas for breeders and the prediction accuracy is relatively moderate. However, the problem is that they hardly take into account environmental differences or the connections among multiple traits (Fritsche-Neto et al., 2021). The result is that once the region is changed or abnormal climate is encountered, the performance of the model is greatly discounted.

As research deepened, people gradually realized that breeding was not carried out in a vacuum. The multi-environment and multi-trait model was thus proposed to be closer to the actual breeding conditions. They can simultaneously simulate genotype × environment interaction (G×E), and also utilize the correlations between traits to enhance prediction accuracy. However, although such models are smarter, they also consume more computing power and have a huge amount of computation. Further optimization is still needed in practical applications (De Oliveira et al., 2020).

3 Multi-Environment Genomic Prediction (MEGP) Models

3.1 Definition and key components of MEGP models

Evaluating the performance of corn hybrids in different environments has never been an easy task. Climate, soil and management methods-each one could cause trouble. The emergence of multi-environment genomic prediction (MEGP) models can be regarded as providing a more systematic solution to this difficult problem. Its principle is not complicated. It can be understood as a method of "mixing genomic information and multi-environmental test data together for calculation", used to predict in advance the possible performance of hybrids under different conditions.

The basic components of the model are actually not complicated: first, genomic marker data (commonly used ones include SNPS); second, phenotypic data obtained from multi-environment experiments; third, explicitly modeled genotype-environment (G×E) interactions; and third, statistical or machine learning algorithms that can comprehensively utilize genetic and environmental covariations. This type of model typically views genotype and environment as random effects, and introduces kernel functions (such as GBLUP or Gaussian kernels) when necessary. Sometimes, deep learning is even employed to capture nonlinear relationships that are difficult to reveal through traditional statistics (Costa-Neto et al., 2020).

3.2 Comparison between single-environment and multi-environment approaches

If the single-environment model is more like "fixed-point observation", then the multi-environment model is closer to "panoramic monitoring". The former only predicts the performance of hybrids in specific environments and does not take into account the interaction between genotypes and the environment (G×E), thus it is easy to "fail to see the whole picture". The MEGP model is different. It incorporates data from various locations and under different conditions and explicitly models the G×E interaction, making the prediction more stable and accurate.

Many studies have actually mentioned similar results: Whether it is key traits such as grain yield or moisture content, the predictive performance of the MEGP model is generally significantly higher than that of the single-environment model, approximately improving by 10% to 20% (Barreto et al., 2024). What's more interesting is that in those new environments without measured data, it can still provide relatively reliable predictions, while single-environment models are often not very effective.

Of course, this does not mean that the environmental model alone is of no value. It can still be useful when there are too few data samples or the calculation conditions are limited, at least providing a general direction for subsequent analysis.

3.3 Model types: G×E interaction, reaction norm, and factor analytic models

Not all multi-environment models follow the same approach. Take the G×E interaction model as an example. It

directly incorporates the interaction term between genotype and environment into the model and uses random effects or kernel methods to depict the mutual influence between genetics and environment. Such a design can take into account both the main effect and environmental specific deviations simultaneously, and is particularly useful for those traits that are strongly influenced by the environment.

The idea of the reaction norm model is somewhat different. It pays more attention to the "expression curve" of genotypes under environmental gradients, that is, how traits respond when different environmental conditions change continuously. This model has unique advantages in describing phenotypic plasticity and has been proven in practice to effectively improve the accuracy of corn grain yield prediction. There is also the factor analysis (FA) model-it does not directly model each environmental effect, but rather decomposes genotype-environment interactions through latent factors to extract the main sources of variation. Due to its high computational efficiency, the FA model is particularly suitable for large-scale multi-environment experiments, and it can simultaneously simulate additive effects, dominant effects and their interactions with the environment (Krause et al., 2020).

4 Sources and Structure of Environmental Variation

4.1 Characterization of spatial and temporal environmental variability

In multi-environment tests (METs), the environment is never constant. The changes can be quite drastic in different locations and different years. The performance differences of hybrid corn often stem from these spatial and temporal heterogeneity. Geographical conditions, soil types and local climates, these factors jointly shape spatial variations. Weather fluctuations and differences in agricultural management constitute unstable factors at the time level.

Some long-term large-scale experimental data indicate that variables such as temperature, rainfall, vapor pressure deficit, and relative humidity often vary significantly between different locations and years (Yue et al., 2022). These fluctuations can trigger complex genotype-environment (G×E) interactions, making it difficult for the same hybrid to show consistent performance in different regions (Figure 1).
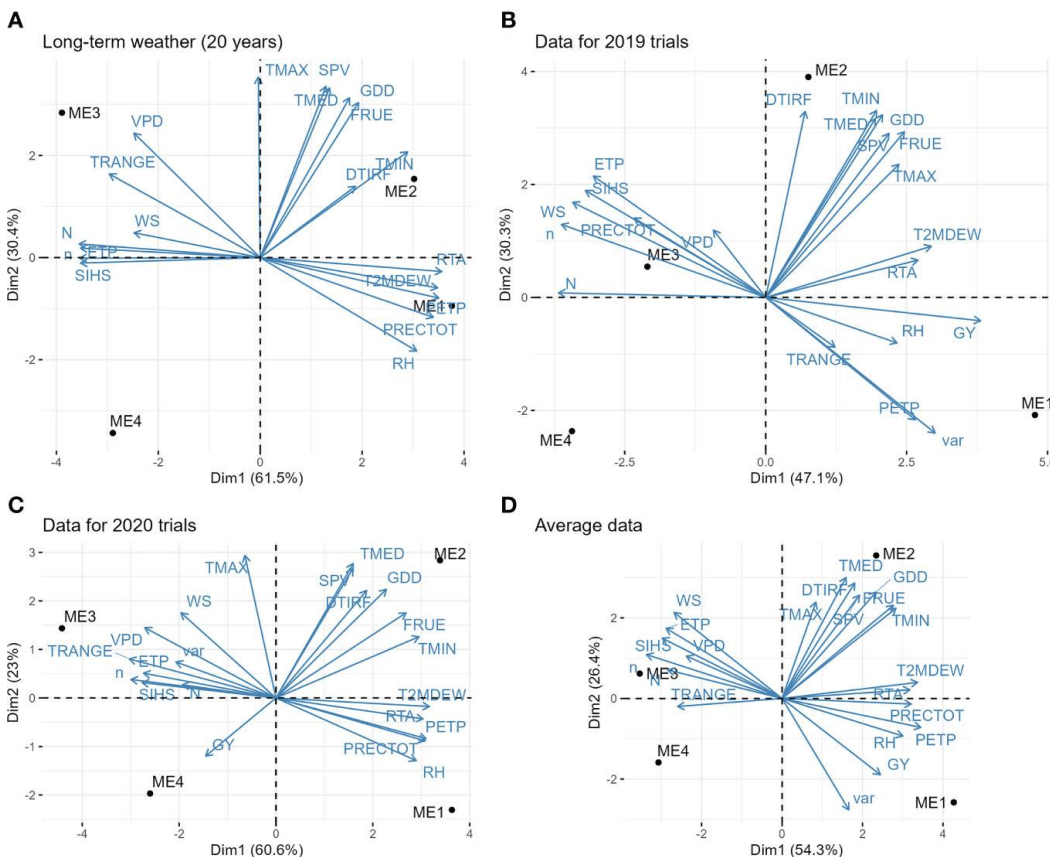

Figure 1 Biplot for the principal component analysis between environmental variables

Later, researchers proposed the concept of "giant environment"-that is, grouping regions with similar climate and soil patterns into one category. Such a classification is not merely a categorization game, but rather helps breeders more effectively locate the adaptation zones of varieties and explain the response patterns of different genotypes under environmental changes.

4.2 Role of environmental covariates and envirotyping in MEGP
In genomic prediction, environmental factors are often regarded as "background noise", but this is not the case. Environmental covariates such as climate and soil, once systematically integrated into the model, often significantly enhance the predictive performance. Environmental typing is precisely for this purpose-it collects long-term meteorological and soil data to identify the similarities and differences in structures among different environments.

In the MEGP model, the usage of environmental covariates is not fixed. They can be directly incorporated into the model or first transformed through some dimensionality reduction methods, such as principal component analysis, before being used. The truly valuable aspect lies in the fact that these variables enable the model to better "understand" the impact of environmental changes on trait performance.

For instance, when both environmental covariates and their interactions with genetic markers are taken into account in the model, the prediction accuracy of traits such as grain yield and drought resistance often improves significantly, especially in regions with large differences in climatic conditions. Furthermore, the addition of environmental typing enables the model to handle the selection tasks of multiple traits and multiple environments simultaneously, making it easier for breeders to identify those hybrids that are both stable and adaptable (Yue et al., 2025).

4.3 Challenges in quantifying genotype-by-environment interactions
Theoretically, the interaction between genotype and environment (G×E) can be characterized by models; But in reality, this matter is far more complicated than imagined. The problem often lies in the data-there are too many genetic and environmental variables, the structure is too complex, and the quality of data from different sources varies greatly. For a model to be effective, it is first necessary to ensure that the training environment is sufficiently similar to the target environment; otherwise, the prediction results will "deviate" (Rogers and Holland, 2021).

Some methods do offer assistance, such as dimensionality reduction techniques like the response specification model or kernel methods, which can simplify the problem to a certain extent. However, not all environmental factors contribute equally to G×E, and the differences in sensitivity among different traits can also interfere with the results. What is more troublesome is that incomplete weather records or the absence of soil data often reduce the accuracy of quantitative analysis. Despite this, with the development of environmental typing, high-throughput phenotypic analysis and machine learning, these obstacles are being gradually weakened. Nowadays, researchers have been able to identify G×E interaction patterns more precisely and are more likely to find stable corn hybrids in complex environments.

5 Case Study: Application of MEGP in a Multi-Year Hybrid Maize Trial
5.1 Study background: experimental design, germplasm, and environments
In corn breeding research, to truly understand the potential of a hybrid, one or two years of field trials alone are often insufficient. Continuous assessment in multiple environments and over multiple years can better reveal the complex nature of the interaction between genotype and environment (G×E). A large-scale study in recent years did exactly this-it tested thousands of hybrid varieties at different locations and in different years, hoping to find those materials that remained stable in a variable environment.

One of the experiments is particularly typical: researchers constructed 2,126 hybrid combinations based on 475 inbred lines and used over 9,000 SNP markers for genotyping. These hybrid varieties were arranged to be planted continuously for two years in 34 environments in the two major corn-producing areas of China (Wang et al., 2025),

with a wide range of climatic conditions and different soil types. Similar multi-environment trials are not uncommon-some cover dozens of locations and multiple years, mainly focusing on core agronomic traits such as grain yield and grain moisture content. The huge volume of data provides a solid foundation for model validation.

5.2 Implementation of MEGP model: data integration and model performance

In these studies, the process of building the MEGP model is not simply "throwing data into the model". It integrates high-density genomic information with detailed environmental descriptions-including 19 climate variables such as temperature, radiation, sunshine duration, etc., as well as principal components extracted based on these variables (Figure 2) (He et al., 2025). To present the interaction between genetics and the environment more realistically, the research team did not use only one method. They tried several different modeling strategies: the traditional GBLUP framework, the response specification model, and some machine learning algorithms were all included.
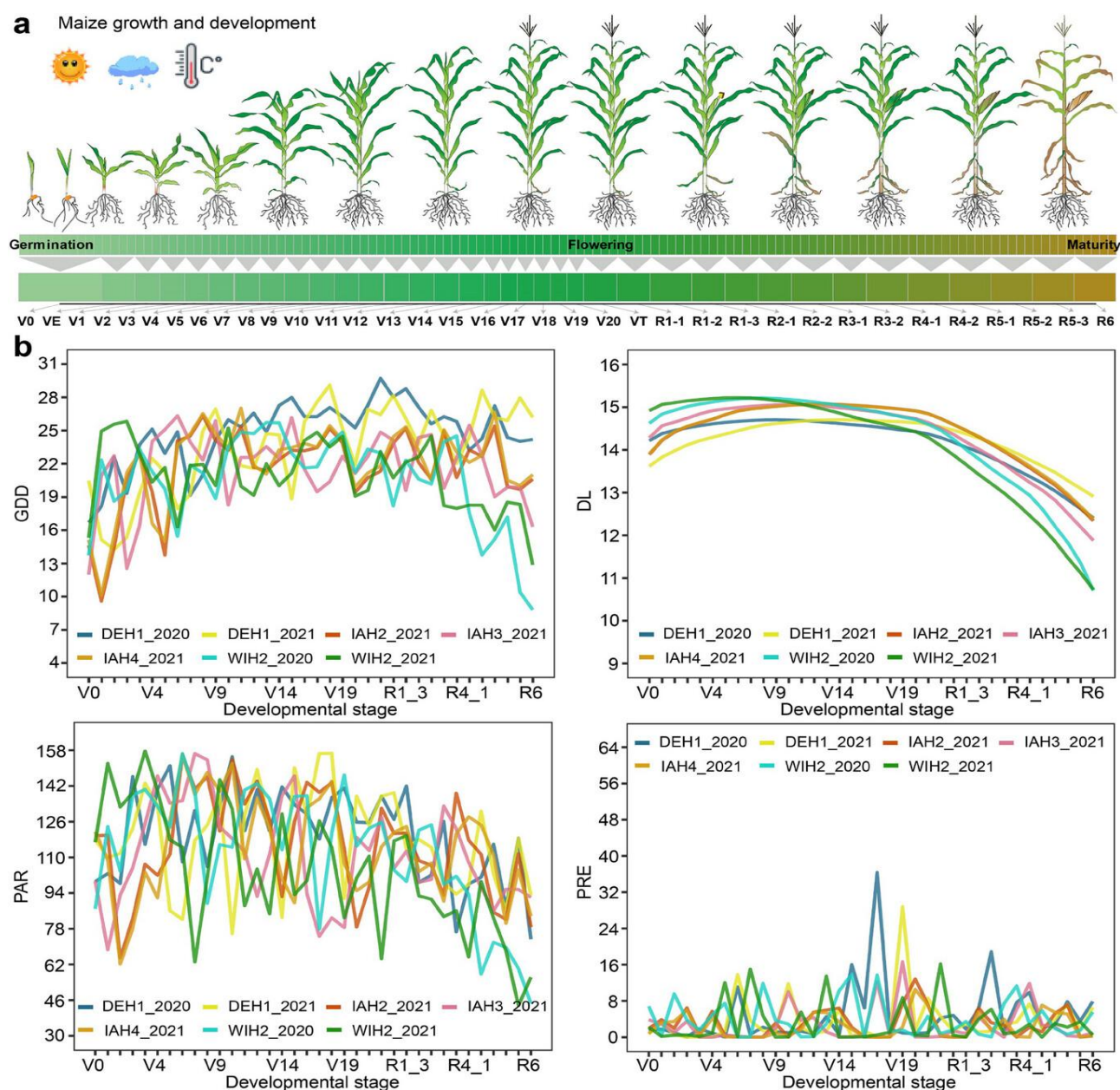

Figure 2 Dimensionality reduction of environmental parameters according to the development period of maize hybrids. a) The 36 development stage-environment windows (V0 to R6) for maize hybrids were defined based on the relationship between the developmental stages and growing degree days (GDD). b) Trends in GDD, day length (DL), photosynthetically active radiation (PAR), and precipitation (PRE) across 36 development stage-environment windows (Adopted from He et al., 2025)

During the modeling process, an environmental similarity matrix was also introduced, and dimensionality reduction methods such as principal component analysis were adopted to alleviate the computational burden-this way, the main information could be retained without making the model run too "heavily". Ultimately, the team adopted cross-validation to test the model's performance and compared the results with those of the single-environment model and the main effect model to ensure that the overall results were not only stable but also understandable.

5.3 Key findings: predictive accuracy, trait heritability, and implications for breeding

The results show that the MEGP model, which takes into account the genotype-environment interaction and environmental covariates, is significantly superior to traditional methods in both prediction accuracy and result stability. Taking grain yield and grain moisture content as examples, the prediction accuracy rates reached 0.33 and 0.73 respectively, and the coincidence degree between the predicted values and the measured superior hybrid varieties exceeded 50%.

When additive effects, dominant effects and environmental variables are simultaneously incorporated into the model, the prediction effect is further enhanced-the prediction accuracy of certain traits even increases by 22%. Interestingly, in an environment with relatively ideal conditions, heritability estimates are generally higher, indicating that genetic differences are more prominent under high-quality ecological conditions (Tolley et al., 2023). These results, from one aspect, confirm the potential of the MEGP model in actual breeding: it can not only help identify excellent combinations that are stable across environments, but also guide resource allocation and selection strategies, thereby making corn breeding more efficient and targeted.

6 Enhancing MEGP Model Accuracy and Utility

6.1 Integration of high-throughput phenotyping and remote sensing data

In the past, researchers mainly relied on field observations and genomic data to assess the performance of hybrids, but such information was often too "static". With the popularization of high-throughput phenotypic analysis (HTP) and remote sensing technology, the situation has changed. Continuous phenotypic data obtained through multispectral imaging, unmanned aerial vehicle (UAV) measurement and other methods enable the model to more truly reflect the dynamic changes of crops at different times and Spaces.

These data often appear in the form of time series phenomics indicators, such as NDVI and other vegetation indices, which can more accurately characterize field heterogeneity. The results show that compared with relying solely on genomic data, the prediction accuracy of key traits such as flowering period and plant height can be improved by approximately 30% after incorporating phenomics information (Adak et al., 2023). More interestingly, this fusion can also reveal the dynamic relationship between genotypes and abiotic stress-for instance, which genes respond most significantly during droughts or high temperatures. Such methods not only enhance predictive performance but also help identify candidate genes and trait markers related to stress resistance, providing more intuitive biological clues for breeding.

6.2 Use of deep learning and bayesian frameworks for improved prediction

In terms of modeling tools, the limitations of traditional linear models are gradually emerging. The introduction of deep learning has brought about a new breakthrough in predictive capabilities. It can automatically identify the complex nonlinear relationships among genotypes, environments, and traits without the need for manual explicit setting of G×E interaction terms. Studies have shown that deep learning models with multiple traits and multiple environments have prediction accuracy approximately 6% to 14% higher than traditional Bayesian or linear models in complex traits such as flowering time and grain yield (Mora-Poblete et al., 2023).

However, the Bayesian method has not been "replaced" by deep learning. Models like BayesB and BMTME still perform quite robustably when considering the interaction between group structure or markers and the environment. Their strength does not lie in speed, but in enabling researchers to clearly understand the internal logic of the model-parameters are adjustable and explanations are intuitive. This makes them still frequently used in multitrait analysis or complex population studies (Yu et al., 2024).

In simple terms, deep learning is more like an automatic and efficient "black box machine", while Bayesian models are more like a controllable "tool system". One emphasizes speed in calculation, while the other focuses on clarity. If they can cooperate with each other, the prediction effect can often be pushed to a more ideal level.

6.3 Strategies for optimizing training population design and marker density

No matter how advanced a model is, it still needs reliable data to support it. How to build the training population and determine the label density, these seemingly detailed issues often determine the overall performance of the MEGP model. Rather than blindly pursuing sample size, it is better to make the data more "representative". Typically, researchers will first select samples based on the genetic background and environmental characteristics of the target population, and then use cluster analysis or principal component analysis to capture the population structure (Gevartosky et al., 2021).

Interestingly, the experimental results show that the training set does not necessarily have to be astonishingly large. As long as it is properly designed, an effective training set accounting for 2% to 13% of the total population is sufficient, and the prediction accuracy can still be maintained at a relatively high level. Especially when the training set contains related individuals of the validation population, the performance of cross-population prediction tends to be better (Guo et al., 2019).

In addition, increasing the density of markers and introducing functional or character-specific markers are also common optimization methods. This can make the model more "sensitive" to complex genetic structures and perform more detailed and closer to the real situation in the prediction results (Roth et al., 2022).

7 Challenges and Limitations in Multi-Environment Prediction

7.1 Computational complexity and data dimensionality issues

In multi-environment genomic prediction (MEGP), the real difficulty is often not the algorithm itself, but rather the excessive amount of data and the overly complex relationships. With tens of thousands of SNP markers, along with dozens of environmental covariates and their interactions, it is very easy for the model to get stuck in a "dimensional quagmire". The more parameters there are, the more difficult the estimation becomes, and the operation time naturally doubles as well.

Some researchers have attempted to solve this problem using machine learning or kernel-based methods, such as CatBoost, XGBoost or deep kernel models. These methods can indeed improve efficiency to a certain extent, but they are not omnipotent. Once a model contains too many variables, even if the dimensionality reduction is handled properly, it may still face the risk of overfitting. Some people have proposed that by reducing the dimension of genetic data and retaining the main environmental variables, it is possible to maintain the predictive ability while also reducing the computational burden. The problem is that such "trade-offs" sometimes sacrifice precision, especially in complex situations where the interaction between genes and the environment is significant. In other words, there is always an unbalanced line between computational feasibility and predictive reliability.

7.2 Limited transferability across populations and environments

Even if the model performs well in the original data, once the environment is changed, the results are often not so ideal. The portability issue of the MEGP model has been repeatedly mentioned-the prediction accuracy is the highest when the environments of the training set and the test set are similar. However, once it moves to a new area or under untested conditions, the accuracy will decrease significantly (Li et al., 2021).

What's more interesting is that the key factor influencing the prediction performance seems not to be genetic similarity, but rather the degree of proximity of the environment. That is to say, the model relies more on "environmental memory" rather than "genetic relationships". In some cases, incorporating the kinship materials of the validation population into the training set can indeed improve cross-environmental prediction, but this operation is not always achievable in actual breeding. What is even more difficult is that the genotype × environment (G×E) interaction itself is extremely complex. Even the most advanced nonlinear models may not be able to fully capture all variations in the new environment (Alves et al., 2021). Therefore, the generalization of the MEGP model remains an unsolved problem.

7.3 Gaps in environmental data standardization and model validation

There are also quite a few problems in the data section. The ways different experiments collect environmental data vary greatly-some focus on climatic factors, while others emphasize soil properties. Some have high temporal resolution, while others only have annual averages. Data deficiency, inconsistency and inconsistent accuracy all make the integration of the model "extremely difficult" (Lopez-Cruz et al., 2023). The result is that cross-study comparisons become difficult and the reproducibility of the model is also limited.

Model verification has actually always been a "hidden" issue. Many studies are still using the same dataset for cross-validation up to now-it is convenient, but the results are often too idealized and a bit far from the real situation. What truly tests the model's capabilities are the data from new environments, different years, or independent breeding cycles. Only under such circumstances can the stability of the model stand up to scrutiny.

To reduce this deviation, researchers are also seeking solutions: unifying data collection standards, developing open-source environmental typing processes, and establishing shareable public databases are gradually becoming a consensus. These measures may not make the problem disappear immediately, but at least they make the model verification more transparent and well-grounded. In other words, they lay the foundation for a more reliable evaluation system in the future.

8 Future Directions and Perspectives

In recent years, discussions on multi-environment genomic prediction (MEGP) models have been increasing. When people talk about it, they no longer merely regard it as a "predictive tool", but rather as a new approach that enables breeders to have a more intuitive understanding of how corn performs in different climates. The instability of the climate has become the norm-high temperatures, droughts and heavy rains take turns to occur. Finding those still stable varieties in such an environment has become the most difficult task at present. The uniqueness of MEGP lies in its ability to simultaneously consider the interaction between genotypes and the environment (G×E), and integrate high-resolution climate data into the model.

Some studies have found that if climate covariates or environmental typing information are incorporated, the prediction accuracy of the model for key traits can be significantly improved, especially in regions with large climate fluctuations. In other words, MEGP can not only predict but also help breeders identify risks in advance, acting like a "climate magnifying glass" that can clearly show the true performance of different materials under environmental stress. Nowadays, with the frequent occurrence of extreme climate events, enhancing the environmental adaptability of corn is no longer an optional question but a core challenge that every breeding program must address.

Another aspect that attracts attention is the speed-up effect brought by MEGP. In the past, screening for superior genotypes relied on several years or even longer field trials, but now, the prediction results of the model can enable this step to be completed ahead of schedule. Even in a new environment that has never been tested, the system can still provide relatively reliable results. The combination of automated machine learning and advanced statistical algorithms makes this prediction method operable and verifiable. Research shows that after introducing environmental information into the model, the average prediction accuracy can be increased by 14% to 28%.

For breeders, this means that potential materials can be identified more quickly among thousands of hybrid combinations. For project managers, this is another means of saving resources. By rationally planning the test environment and sample size, the reliability of the prediction can be maintained even when funds are tight. Compared with traditional methods, this resource-efficient model not only shortens the cycle but also enables genomic selection to truly integrate into large-scale and realistic breeding processes. Ultimately, it helps the team achieve genetic improvement more quickly and also makes the promotion of climate-resilient corn more precise and targeted.

Looking ahead, the development of MEGP is unlikely to remain at the level of a single laboratory or project. Its direction is towards openness and collaboration. With the rise of global breeding networks and digital

decision-making systems, open-source environment typing processes, shared databases, and joint field trials are gradually breaking down regional and institutional barriers. High-quality phenotypic, genotypic and environmental data from different sources were collected, continuously enhancing the robustness and universality of the model. When the achievements of MEGP are further embedded in decision support systems, breeders and even policymakers can make judgments based on real data: which regions are more suitable for growing which varieties, how resources should be allocated, and how strategies for addressing climate risks should be adjusted. It can be foreseen that future corn breeding will no longer be a "single-point breakthrough", but a global collaborative networked system. MEGP may not change everything immediately, but it is becoming an indispensable supporting tool for modern breeding-a key path towards higher yields, more stable yields and more sustainable agriculture.

Acknowledgments

Conflict of Interest Disclosure

References

Adak A., Kang M., Anderson S., Murray S., Jarquín D., Wong R., and Katzfuss M., 2023, Phenomic data-driven biological prediction of maize through field-based high throughput phenotyping integration with genomic data, Journal of Experimental Botany, 74(17): 5307-5326.
https://doi.org/10.1093/jxb/erad216

Alemu A., Åstrand J., Montesinos-López O., Sánchez J., Fernández-González J., Tadesse W., Vetukuri R., Carlsson A., Ceplitis A., Crossa J., Ortiz R., and Chawade A., 2024, Genomic selection in plant breeding: key factors shaping two decades of progress, Molecular Plant, 17(4): 552-578.
https://doi.org/10.1016/j.molp.2024.03.007

Alves F., Galli G., Matias F., Vidotti M., Morosini J., and Fritsche-Neto R., 2021, Impact of the complexity of genotype by environment and dominance modeling on the predictive accuracy of maize hybrids in multi-environment prediction models, Euphytica, 217: 37.
https://doi.org/10.1007/s10681-021-02779-y

Barreto C., Dias K., De Sousa I., Azevedo C., Nascimento A., Guimarães L., Guimarães C., Pastina M., and Nascimento M., 2024, Genomic prediction in multi-environment trials in maize using statistical and machine learning methods, Scientific Reports, 14: 1062.
https://doi.org/10.1038/s41598-024-51792-3

Beyene Y., Gowda M., Olsen M., Robbins K., Pérez-Rodríguez P., Alvarado G., Dreher K., Gao S., Mugo S., Prasanna B., and Crossa J., 2019, Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections, Frontiers in Plant Science, 10: 1502.
https://doi.org/10.3389/fpls.2019.01502

Costa-Neto G., Fritsche-Neto R., and Crossa J., 2020, Nonlinear kernels, dominance, and envirotyping data increase the accuracy of genome-based prediction in multi-environment trials, Heredity, 126: 92-106.
https://doi.org/10.1038/s41437-020-00353-1

De Oliveira A., Resende M., Ferrão L., Amadeu R., Guimarães L., Guimarães C., Pastina M., and Margarido G., 2020, Genomic prediction applied to multiple traits and environments in second season maize hybrids, Heredity, 125: 60-72.
https://doi.org/10.1038/s41437-020-0321-0

Ferrão L., Marinho C., Muñoz P., and Resende M., 2020, Improvement of predictive ability in maize hybrids by including dominance effects and marker × environment models, Crop Science, 60: 666-677.
https://doi.org/10.1002/csc2.20096

Fritsche-Neto R., Galli G., Borges K., Costa-Neto G., Alves F., Sabadin F., Lyra D., Morais P., De Andrade L., Granato Í., and Crossa J., 2021, Optimizing genomic-enabled prediction in small-scale maize hybrid breeding programs: a roadmap review, Frontiers in Plant Science, 12: 658267.
https://doi.org/10.3389/fpls.2021.658267

Gao S., Yu T., Rasheed A., Wang J., Crossa J., Hearne S., and Li H., 2025, Fast-forwarding plant breeding with deep learning-based genomic prediction, Journal of Integrative Plant Biology, 67: 1700-1705.
https://doi.org/10.1111/jipb.13914

Gevartosky R., Carvalho H., Costa-Neto G., Montesinos-López O., Crossa J., and Fritsche-Neto R., 2021, Enviromic-based kernels may optimize resource allocation with multi-trait multi-environment genomic prediction for tropical maize, BMC Plant Biology, 23: 10.
https://doi.org/10.1186/s12870-022-03975-1

Guo T., Yu X., Li X., Zhang H., Zhu C., Flint-Garcia S., McMullen M., Holland J., Szalma S., Wisser R., and Yu J., 2019, Optimal designs for genomic selection in hybrid crops, Molecular Plant, 12(3): 390-401.
https://doi.org/10.1016/j.molp.2018.12.022

He K., Yu T., Gao S., Chen S., Li L., Zhang X., Huang C., Xu Y., Wang J., Prasanna B., Hearne S., Li X., and Li H., 2025, Leveraging automated machine learning for environmental data-driven genetic analysis and genomic prediction in maize hybrids, Advanced Science, 12(17): 2412423.
https://doi.org/10.1002/advs.202412423

Kaler A., Purcell L., Beissinger T., and Gillman J., 2022, Genomic prediction models for traits differing in heritability for soybean, rice, and maize, BMC Plant Biology, 22: 87.
https://doi.org/10.1186/s12870-022-03479-y

Krause M., Dias K., Santos J., Oliveira A., Guimarães L., Pastina M., Margarido G., and Garcia A., 2020, Boosting predictive ability of tropical maize hybrids via genotype-by-environment interaction under multivariate GBLUP models, Crop Science, 60: 3049-3065.
https://doi.org/10.1002/csc2.20253

Li D., Xu Z., Gu R., Wang P., Xu J., Du D., Fu J., Wang J., Zhang H., and Wang G., 2021, Genomic prediction across structured hybrid populations and environments in maize, Plants, 10(6): 1174.
https://doi.org/10.3390/plants10061174

Lopez-Cruz M., Aguate F., Washburn J., De León N., Kaeppler S., Lima D., Tan R., Thompson A., De La Bretonne L., and De Los Campos G., 2023, Leveraging data from the Genomes-to-Fields initiative to investigate genotype-by-environment interactions in maize in North America, Nature Communications, 14: 6904.
https://doi.org/10.1038/s41467-023-42687-4

Luo P., Yang R., Zhang L., Yang J., Wang H., Yong H., Zhang R., Li W., Wang F., Li M., Weng J., Zhang D., Zhou Z., Han J., Gao W., Xu X., Yang K., Zhang X., Fu J., Li X., Hao Z., and Ni Z., 2024, Genomic prediction of kernel water content in a hybrid population for mechanized harvesting in maize in Northern China, Agronomy, 14(12): 2795.
https://doi.org/10.3390/agronomy14122795

Mora-Poblete F., Maldonado C., Henrique L., Uhdre R., Scapim C., and Mangolim C., 2023, Multi-trait and multi-environment genomic prediction for flowering traits in maize: a deep learning approach, Frontiers in Plant Science, 14: 1153040.
https://doi.org/10.3389/fpls.2023.1153040

Popa C., Călugăr R., Varga A., Muntean E., Băcilă I., Vana C., Racz I., Tritean N., Berindean I., Ona A., and Muntean L., 2025, Evaluating maize hybrids for yield, stress tolerance, and carotenoid content: insights into breeding for climate resilience, Plants, 14(1): 138.
https://doi.org/10.3390/plants14010138

Rice B., and Lipka A., 2021, Diversifying maize genomic selection models, Molecular Breeding, 41: 33.
https://doi.org/10.1007/s11032-021-01221-4

Rogers A., and Holland J., 2021, Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data, G3 Genes|Genomes|Genetics,12(2): jkab440.
https://doi.org/10.1093/g3journal/jkab440

Roth M., Beugnot A., Mary-Huard T., Moreau L., Charcosset A., and Fiévet J., 2022, Improving genomic predictions with inbreeding and non-additive effects in two admixed maize hybrid populations in single and multi-environment contexts, Genetics, 220(4): iyac018.
https://doi.org/10.1093/genetics/iyac018

Supriadi D., Bimantara Y., Zendrato Y., Widaryanto E., Kuswanto K., and Waluyo B., 2024, Assessment of genotype by environment and yield performance of tropical maize hybrids using stability statistics and graphical biplots, PeerJ, 12: e18624.
https://doi.org/10.7717/peerj.18624

Tolley S., Brito L., Wang D., and Tuinstra M., 2023, Genomic prediction and association mapping of maize grain yield in multi-environment trials based on reaction norm models, Frontiers in Genetics, 14: 1221751.
https://doi.org/10.3389/fgene.2023.1221751

Wang W., 2025, Review of breeding maize varieties for biofuel production, Journal of Energy Bioscience, 16(3): 151-162
http://dx.doi.org/10.5376/jeb.2025.16.0015

Wang J., Liu L., He K., Gebrewahid T., Gao S., Tian Q., Li Z., Song Y., Guo Y., Li Y., Cui Q., Zhang L., Wang J., Huang C., Li L., Guo T., and Li H., 2025, Accurate genomic prediction for grain yield and grain moisture content of maize hybrids using multi-environment data, Journal of Integrative Plant Biology, 67(5): 1379-1394.
https://doi.org/10.1111/jipb.13857

Xu Y., Zhang X., Li H., Zheng H., Zhang J., Olsen M., Varshney R., Prasanna B., and Qian Q., 2022, Smart breeding driven by big data, artificial intelligence and integrated genomic-enviromic prediction, Molecular Plant, 15(11): 1664-1695.
https://doi.org/10.1016/j.molp.2022.09.001

Yu G., Cui Y., Jiao Y., Zhou K., Wang X., W. W., Xu Y., Yang K., Zhang X., Li P., Yang Z., Xu Y., and Xu C., 2022, Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance, The Crop Journal, 11(2): 490-498.
https://doi.org/10.1016/j.cj.2022.09.004

Yu G., Li F., Wang X., Zhang Y., Zhou K., Yang W., Guan X., Zhang X., Xu C., and Xu Y., 2024, Enhancing across-population genomic prediction for maize hybrids, Plants, 13(21): 3105.
https://doi.org/10.3390/plants13213105

Yu K., Wang H., Liu X., Xu C., Li Z., Xu X., Liu J., Wang Z., and Xu Y., 2020, Large-scale analysis of combining ability and heterosis for development of hybrid maize breeding strategies using diverse germplasm resources, Frontiers in Plant Science, 11: 660
https://doi.org/10.3389/fpls.2020.00660

Yue H., Olivoto T., Bu J., Li J., Wei J., Xie J., Chen S., Peng H., Nardino M., and Jiang X., 2022, Multi-trait selection for mean performance and stability of maize hybrids in mega-environments delineated using envirotyping techniques, Frontiers in Plant Science, 13: 1030521.
https://doi.org/10.3389/fpls.2022.1030521

Yue H., Olivoto T., Bu J., Wei J., Liu P., Wu W., Nardino M., and Jiang X., 2025, Assessing the role of genotype by environment interaction as determinants of maize grain yield and lodging resistance, BMC Plant Biology, 25: 120.
https://doi.org/10.1186/s12870-025-06158-w

Zhang H., Yin L., Wang M., Yuan X., and Liu X., 2019, Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations, Frontiers in Genetics, 10: 189.
https://doi.org/10.3389/fgene.2019.00189

Zhang Z., Wang X., Zhang Y., Zhou K., Yu G., Yang W., Li F., Guan X., Zhang X., Yang Z., Xu C., and Xu Y., 2025, SPDC-HG: an accelerator of genomic hybrid breeding in maize, Plant Biotechnology Journal, 23: 1847-1861.
https://doi.org/10.1111/pbi.70011

Zhao Y.M., Bao Y., Zhou L., Zhang B.H., and Wang W.J., 2025, Gene mapping of mechanization-friendly traits in maize based on SNP markers and breeding for mechanical adaptation, Molecular Plant Breeding, 16(1): 24-34
http://dx.doi.org/10.5376/mpb.2025.16.0003

Zou Q., Tai S., Yuan Q., Nie Y., Gou H., Wang L., Li C., Jing Y., Dong F., Yue Z., Rong Y., Fang X., and Xiong S., 2025, Large-scale crop dataset and deep learning-based multi-modal fusion framework for more accurate G×E genomic prediction, Computers and Electronics in Agriculture, 230: 109833.
https://doi.org/10.1016/j.compag.2024.109833