Genotype-by-Environment Interaction

Model-Based Reconstruction from Test-Field Data

L.L. Molenaar

TU Delft

Genotype-by-Environment Interaction

Model-Based Reconstruction from Test-Field Data

by

L.L. Molenaar

to obtain the degree of Bachelor of Science
at the Delft University of Technology,
to be defended publicly on Thursday 10th of July

Student number:	5406293	
Project duration:	April 24, 2025 – July 10, 2025	
Thesis committee:	Dr. N.V. Budko,	TU Delft, supervisor
	Dr. N. Parolya,	TU Delft
	C. Verburg,	TU Delft, co-supervisor

TU Delft

Abstract

This report evaluates methods for reconstructing the coefficients of a model that describes the relationship between environmental conditions and the performance of potato genotypes, with performance measured by the weight of storage organs. For each genotype, this relationship is modeled as a linear combination of environmental factors weighted by genotype-specific environmental coefficients. The model incorporates the soil moisture field effect, wich varies both over the growing season and spatially within the field, with the latter captured through a linear combination of basis functions weighted by field effect coefficients. Initially, synthetic environmental data is generated, and fixed field effect and genotype-specific environmental coefficients are established. The model then calculates the expected performance based on these coefficients.

Subsequently, two approaches — a two-step method and a one-step method — are tested to reconstruct the original coefficients using the environmental and performance measurements. The robustness of both methods to noise is evaluated, and the minimum required number of environmental measurement locations within the field, as well as the optimal number of intermediate harvests during the growing season, are determined.

This is important because, once refined and expanded, the model and reconstruction methods can be applied to real multi-environment data to provide insights into how environmental factors influence the growth of different genotypes. Such analyses are important for advancing plant breeding efforts.

Contents

1

Introduction

According to the International Potato Center, worldwide, the potato ranks as the third most cultivated crop for human consumption [6] and the Netherlands is the world leader in the production and export of seed potatos [5]. However, potato cultivation is hindered by numerous pathogens and insects [5]. Farmers often rely on substantial amounts of crop protection products to achieve high yields [5]. In addition, the potato is sensitive to the effects of climate change [5]. To ensure food security, it is essential to develop new genotypes that are more resilient to these challenges and contribute to higher yields, better quality, and more sustainable cultivation practices.

Plant breeding is the process of selecting plants based on desirable traits, such as increased resistance to environmental stressors like high temperatures, drought, or nitrogen surplus [2]. To identify such traits, multi-environment trials (MET) are conducted, in which the field is divided into plots — small sections where approximately sixty potato plants of a single genotype are grown. Different genotypes are planted across the field and simultaneously exposed to varying environmental conditions. The performance of each genotype is then evaluated [2].

Agricultural systems are inherently complex and dynamic, shaped by interacting biological, physical, and chemical processes, as well as human management and unpredictable environmental factors such as weather, soil variability, pests, and diseases. These systems are particularly challenging to manage due to their living components, which respond to environmental conditions in nonlinear and time-varying ways [12].

In the context of modern plant breeding, mathematical analysis and modeling play a critical role by enabling the prediction of genotype-by-environment interactions, thereby improving selection and decision-making processes [4]. Chapter 2 explores various modeling approaches and methods commonly used in the analysis of MET data.

The aim of this project is to evaluate a method for reconstructing the coefficients of a model that describes the relationship between genotype and environment. In addition, the project investigates the robustness of this reconstruction method under noise and aims to determine the optimal number of measurement time points and locations.

The genotype-environment relationship is modeled as a linear combination of five environmental variables along with their interaction terms, resulting in a second-order polynomial model. This form is chosen because second-order polynomials provide flexibility and are easy to interpret. Moreover, they enable realistic modeling of both global and local effects of predictors — here, the environmental variables — on the response variable, which in this case is plant performance [7]. In this project, plant performance will be measured using the weight of storage organs (WSO). The extent to which these environmental factors affect the WSO is expressed through corresponding genotype-specific environmental coefficients, which are the values we aim to reconstruct.

The environmental variables that are considered in this project are: potential evaporation for a reference crop leaf, the average daily temperature and soil moisture. We also calculate the integral of the average daily temperature over time, known as thermal time, because it shows the total heat the plant experiences, which is important for understanding how temperature affects growth under changing conditions [8]. We also include the leaf area index (LAI) as an environmental variable. In reality, LAI is not solely influenced by environmental

factors; it interacts with WSO in a more complex way, acting both as an outcome and a driver within the system. For simplicity, this project treats LAI as part of the environment.

These five environmental coefficients are not the only unknowns. This is because four of the five environmental factors in this project vary only over time, while one — soil moisture — also varies across locations within the field. This spatial variation can result from differences in soil structure, elevation, or nearby ditches. To account for this, the spatial effect of soil moisture is modeled as a linear combination of known basis functions. Although the basis functions are predefined, their corresponding coefficients must be reconstructed by fitting the basis functions to the field effect measurements.

Thus, there are two categories of coefficients that must be estimated: the environmental coefficients per genotype ($\boldsymbol{\alpha}$) and the field effect coefficients ($\boldsymbol{\beta}$). The reconstruction can be carried out in two ways. In a *two-step method* discussed in Chapter 4, the $\boldsymbol{\beta}$ values are estimated first, followed by the $\boldsymbol{\alpha}$, each by solving a linear problem. Alternatively, in Chapter 5 a *one-step method* is used, in which both sets of coefficients are simultaneously estimated by solving a single nonlinear problem.

In practical terms, this reconstruction method can be applied to measurements of the weight of storage organs (WSO) in potato plants alongside measurements of the environmental conditions. However, the equation used in this report to describe the weight of the storage organs is simplified: in reality, more than five environmental factors play a role, and the interaction between them may be more complex than simple product terms. Therefore, this model may not yet be suitable for real-world data. Instead, synthetic environmental variables and their coefficients, based on a realistic scenario, are defined and used to calculate WSO. The process of generating this synthetic data is described in detail in Chapter 3.

To collect environmental data such as evaporation, temperature, soil moisture, and leaf area index, at least one measurement location in the field is required. However, the hypothesis is that a single measurement point may be insufficient to accurately capture the spatial variation in soil moisture. Since installing measurement locations is costly, we aim to minimize the number of required measurement locations. One of the questions investigated in this report is how many measurement points are minimally required to accurately capture the relationship between crop performance and environmental conditions.

The weight of storage organs is determined by harvesting a few plants from each plot across the field and weighing their yield. Because the potatos are not yet fully mature, they must be discarded after harvesting, making this intermediate harvest a destructive process. Minimizing the number of intermediate harvests is thus important. Therefore, another research questions addressed in this report is determining the minimum number of intermediate harvests required to accurately reconstruct the relationship between crop performance and environmental conditions.

2

Related work

This chapter reviews key methods and models relevant to the analysis of multi-environment trial (MET) data and crop growth modeling. We explore mixed model approaches that capture genotype-by-environment (GxE) interactions, models addressing spatial variability within field trials, and dynamic system models that simulate crop growth over time.

The work discussed in this chapter relates closely to this project, as we will be using an ordinary differential equation-based dynamic system model to capture genotype-by-environment interactions. Additionally, the model incorporates spatio-temporal field effects. However, unlike mixed models, the model used in this project includes only fixed effects and does not account for random effects.

2.1. Mixed Model Approaches

Mixed model approaches, which incorporate both fixed and random effects, are widely applied in the analysis of MET data [11]. The analysis of crop genotype trials originally relied on Analysis of Variance (ANOVA) methods, an early form of linear mixed modeling [11]. While ANOVA effectively partitions total variation, it does not provide insight into GxE interactions.

Additive Main Effects and Multiplicative Interaction (AMMI) is a mixed model developed to capture and analyze GxE interactions in MET data. AMMI combines traditional ANOVA with principal component analysis (PCA). It begins by extracting the interaction effects from the ANOVA model, forming a matrix that represents GxE interactions. PCA is then applied to this matrix to decompose the complex interaction into a series of simpler components known as principal components. Typically, the first few components capture most of the essential variation in the GxE interaction, making the results easier to interpret and potentially offering insights into how genotypes respond differently across environments [11].

Another approach is the Factor Analytic (FA) mixed model, which offers a more rigorous, random-effects-based framework for analyzing MET data. A major advantage of the FA model is its flexible variance structure for G×E interactions. While the most general option is an unstructured variance-covariance matrix, it can be computationally demanding, especially with many environments. In contrast, the FA model provides a robust approximation using a limited number of multiplicative terms, making it more practical for large datasets [11].

2.2. Spatial Effects Models

Additionally, spatial field effects, which arise from trends within the field, have been addressed through differt kind of models.

Spatial variability and FA models can be integrated within a single linear mixed-model framework improves the analysis of MET data. A comparison of randomized complete block designs, spatial-only models, and models combining spatial and G×E effects across ten Ethiopian grain yield trials demonstrated that the integrated approach not only captures complex spatial patterns and reduces residual variance, but also improves the interpretation of G×E interactions [1].

An alternative approach to incorporating spatial field effects in field trials is through a spatial mixed model based on tensor-product P-splines. This method reformulates the spatial component as an ANOVA-type de-

composition into five smooth additive surfaces, enabling the model to efficiently capture both local and global spatial trends. It simultaneously accounts for genotype effects, block structure (i.e., groups of similar plots), and the influence of repeated treatments across the field to model random variation. Notably, the approach links genetic contribution to trait variation (heritability) with the model's flexibility in attributing variation to genetic factors, as measured by the effective degrees of freedom [9].

Moreover a spatial mixed model analysis of MET data can also be extended by incorporating multiplicative mixed models, which enable a flexible and interpretable modeling of G×E interactions through a FA structure, while also accounting for spatial field trends [10].

2.3. Dynamic System Models

Dynamic system models mathematically represent how components of a system change over time, often using differential or difference equations. In agriculture, they capture interactions among biological, physical, and chemical processes and their response to environmental factors. These models help predict system behavior under varying conditions, supporting better decision-making for yield, profitability, and sustainability [12].

An Example would be The Simple Maize Crop Model (SMCM), which simulates the growth of a maize crop in a homogeneous field area using three state variables: above-ground biomass, physiological age (expressed as thermal time), and leaf area index (LAI). The model assumes ideal growth conditions with no limitations from water, nutrients, pests, or weeds. This simplified model captures key crop growth processes and facilitates applications such as parameter estimation and sensitivity analysis [12].

3

Generating Synthetic Data

In this chapter, we describe the construction of a synthetic dataset designed to simulate the performance of five potato genotypes across a spatially structured field over one growing season. The data are generated using an ordinary differential equation-based model that captures genotype-by-environment interactions. Environmental variables are derived from realistic data, and extended with a spatio-temporal field effect in soil moisture.

3.1. Setting up the Field

We create a field consisting of a grid of 6×10 plots. In each plot, one of five different genotypes is planted. The spatial position of each plot is denoted by $\vec{r} = (x, y)$, where x and y are discretized over the interval $[0, 1]$.

We track development in the field over a single growing season, from early April to early October. Time is divided into 48 time steps across the interval $[0, 1]$, meaning each time step corresponds to approximately half a week.

3.2. Environmental Variables

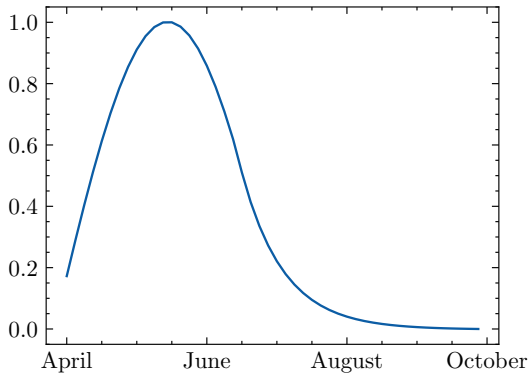
The environmental vector $\phi(\vec{r}, t)$ includes five environmental variables and their interactions, represented as a 20-element vector:

$$\phi(\vec{r}, t) = \left[w_1(t),\, w_2(t),\, w_3(t),\, w_4(\vec{r}, t),\, w_5(t),\, w_1^2(t),\, w_2^2(t), \cdots, w_5^2(t),\, w_1(t)w_2(t),\, w_1(t)w_3(t), \cdots,\, w_4(\vec{r}, t)w_5(t) \right]^\top$$

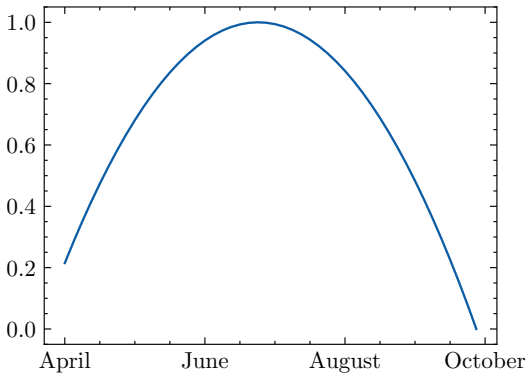
Where:

- $w_1(t)$: Potential evaporation for a reference crop leaf
- $w_2(t)$: Average daily temperature
- $w_3(t)$: Temperature integral of daily average temperature $\int_{t_0}^{t} w_2(\tau)\, d\tau$
- $w_4(\vec{r}, t)$: Soil moisture
- $w_5(t)$: Leaf area index

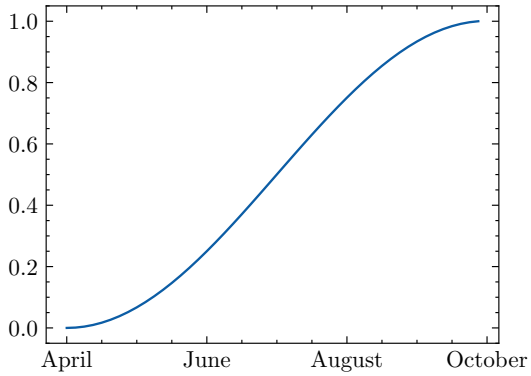
The variables $w_1(t), w_2(t), w_3(t)$ and $w_5(t)$ vary only over time and are uniform throughout the field. Their simulation is based on observational data representing realistic conditions and subsequently normalized to the interval $[0, 1]$. Figure 3.1 shows the temporal evolution of these time-dependent environmental factors over one growing season.



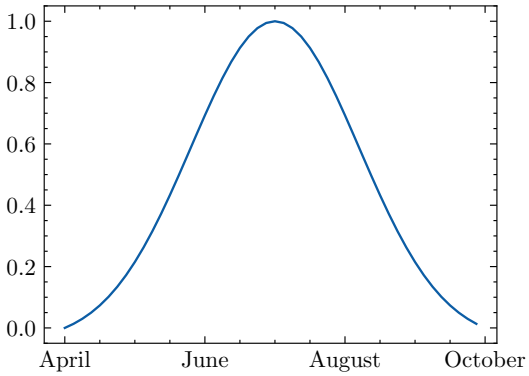
(a) $w_1(t)$: Potential evaporation for a reference crop leaf.



(b) $w_2(t)$: Average daily temperature.



(c) $w_3(t)$: Temperature integral of daily average temperature.



(d) $w_5(t)$: Leaf area index.

Figure 3.1: The development of time-dependent environmental variables throughout the growing season. The variables are normalized to the interval $[0, 1]$.

3.3. Soil Moisture

The soil moisture depends on both time and space and is modeled as:

$$w_4(\vec{r}, t) = w_4^c(t) + v(\vec{r}, t)$$

Where:

- $w_4^c(t)$: The field constant (uniform over space but time-dependent)
- $v(\vec{r}, t)$: The spatio-temporal field effect

3.3.1. Soil Moisture Field Effect

We assume that the spatio-temporal field effect component of soil moisture can be represented as:

$$v(\vec{r}, t) = \sum_{\omega=1}^{3} \sum_{\kappa=1}^{6} \beta_{\omega\kappa} \cdot \psi_\omega(t) \cdot \chi_\kappa(\vec{r}) \tag{3.1}$$

Where:

- $v(\vec{r}, t)$: The spatio-temporal field effect
- $\boldsymbol{\psi}(t)$: The 3-element basis fucntions vector for time t
- $\boldsymbol{\chi}(\vec{r})$: The 6-element basis fucntions vector for space $\vec{r} = (x, y)$
- $\boldsymbol{\beta} \in \mathbb{R}^{18}$: The vector of field effect coefficients

3.3.2. Reference Point

We impose that:

$$v(\vec{r}_c, t) = 0$$

where $\vec{r}_c$ is a reference point, chosen to be the center of the field. By setting a fixed reference point at the center of the field, the field effect becomes relative to this location, such that we can interpret field effect values as deviations from the center. We enforce the field effect to be zero at the reference point by modifying Equation 3.1 to :

$$v(\vec{r}, t) = \sum_{\omega=1}^{3} \sum_{\kappa=1}^{6} \beta_{\omega\kappa} \cdot \psi_\omega(t) \cdot \left(\chi_\kappa(\vec{r}) - \chi_\kappa(\vec{r}_c)\right)$$

3.3.3. Field Effect Basis Functions

We choose the basis functions for time t as:

$$\psi_1(t) = \frac{1}{5}\sin(2\pi t), \quad \psi_2(t) = \frac{1}{5}\cos(2\pi t), \quad \psi_3(t) = \frac{1}{5}\sin(2\pi t)\cos(2\pi t)$$

The multiplication by $\frac{1}{5}$ in the time-dependent basis functions scales the field effect to vary approximately between -0.2 and 0.2.

The basis functions for space $\vec{r} = (x, y)$ are chosen as:

$$\chi_1(x, y) = x, \quad \chi_2(x, y) = y, \quad \chi_3(x, y) = x^2, \quad \chi_4(x, y) = y^2, \quad \chi_5(x, y) = xy, \quad \chi_6(x, y) = x^2 y^2$$

When selecting basis functions, one must balance between complexity and simplicity: overly complex functions risk overfitting the data, while overly simple ones may fail to capture essential patterns, leading to underfitting. It is, however, a strong assumption that the true field effect can be accurately represented using this specific set of basis functions. In this project, we make an even stronger assumption by using the same basis functions both for generating the data and for reconstructing the coefficients from the data — a practice known as inverse crime. This means we are solving an inverse problem under idealized conditions, where the model used to recover parameters is exactly the same as the one used to simulate them.

We do this deliberately in this initial stage to isolate and evaluate the performance of the reconstruction method itself, without the influence of model mismatch.

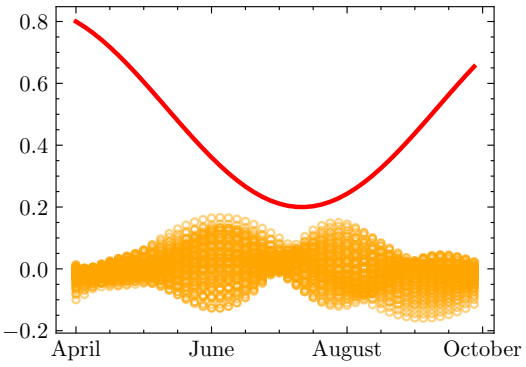
3.3.4. Field Effect Coefficients

The 18 field effect coefficients β are randomly drawn from a uniform distribution over $[-1, 1]$.

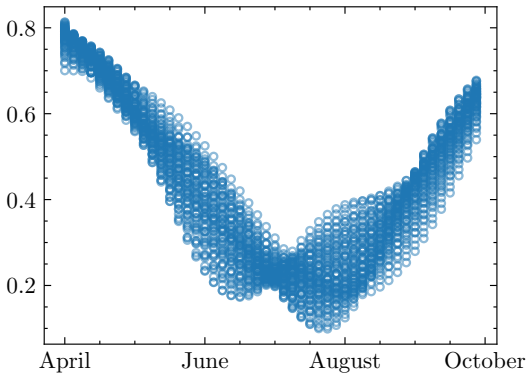
3.3.5. Soil Moisture Field Constant

The simulation of the field constant component of soil moisture is based on observational data representing realistic conditions. These values are then normalized to the interval $[0.2, 0.8]$ to ensure that the soil moisture — calculated as the sum of the field constant and the field effect — remains within the range $[0, 1]$.

Figure 3.2 shows the field constant, the field effect, and their sum, which together determine the soil moisture.



(a) The field constant $w_4^c(t)$ is displayed as a red line, alongside the field effect $v(\vec{r}, t)$ represented by orange scatter points. Each time point includes 60 orange circles, each corresponding to a specific plot within the field.



(b) The scatter plot of the soil moisture $w_4(\vec{r}, t)$ which is the sum of the field constant and the field effect. Each time point includes 60 blue circles, each corresponding to a specific plot within the field.

Figure 3.2: Soil moisture and its components over the course of the growing season, constructed to ensure that soil moisture values remain within the range $[0, 1]$.

The variation and development of soil moisture become clearer when displayed alongside the spatial distribution of genotypes. Figure 3.3 presents the soil moisture across the field at four different moments in the growing season.

3.4. Genotype-Environment Interaction

The relationship between the weight of storage organs and the environmental factors is described by the equation:

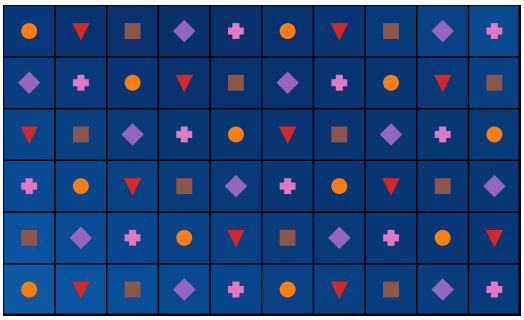
$$\frac{dY_i(\vec{r}, t)}{dt} = \alpha_{21,i} + \sum_{\xi=1}^{20} \alpha_{\xi,i} \phi_\xi(\vec{r}, t) \tag{3.2}$$

where:

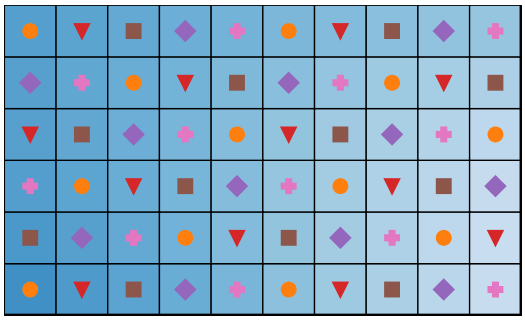
- $Y_i(\vec{r}, t)$: The weight of storage organs of potatos of genotype i at location $\vec{r}$ and time t
- $\phi_\xi(\vec{r}, t)$: The value of environmental factor ξ at location $\vec{r}$ and time t
- $\alpha_{\xi,i}$: The genotype-specific environmental coefficient corresponding to factor ϕ_ξ for genotype i
- $\alpha_{21,i}$: The constant coefficient for genotype i

Since the storage organs have no initial weight at the beginning of the growing season, we set the initial condition for all locations $\vec{r}$ in the field as:

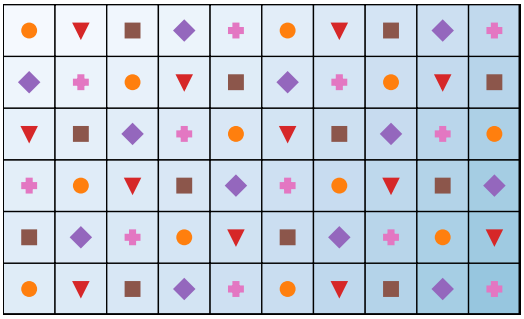
$$Y_i(\vec{r}, t_0) = 0$$



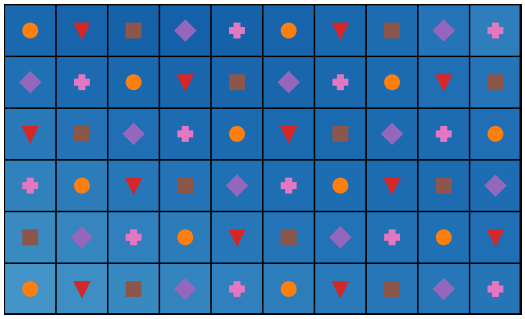
(a) April.



(b) June.



(c) August.



(d) October.

Figure 3.3: Development of soil moisture within the field, which consists of 60 plots. The intensity of blue shading indicates the soil moisture level in each plot, with deeper blue representing higher moisture values. The spatial distribution of genotypes is also depicted, with each genotype represented by a distinct combination of color and marker shape: genotype 0 as an orange circle, genotype 1 as a red triangle, genotype 2 as a brown square, genotype 3 as a purple diamond, and genotype 4 as a pink plus-symbol.

3.5. Genotype-Specific Environmental Coefficients

For simplicity, the constant coefficient α_{21} is included in the set of environmental coefficients. These 21 coefficients α are derived from a table of realistic values for different genotypes. Since the study considers five genotypes, this results in a total of $21 \times 5 = 105$ environmental coefficients.

Since this table of coefficients was generated by fitting the model to environmental data from a different time domain than the one used in this project, the coefficients require slight adjustment. This modification ensures that the potatoes exhibit an overall increase in the weight of their storage organs throughout the growing season. Specifically, the coefficients α_4, α_6, α_9 and α_{11} have been slightly increased — without changing their signs — to produce this behavior. Here, α_4 and α_{11} have negative values and α_6, and α_9 have positive values.

3.6. Performance: Weight of Storage Organs

The plant's performance is quantified by evaluating a key trait—in this case, the weight of storage organs.

To determine the change in this trait, we use Equation 3.2. We solve this differential equation by integrating, yielding the following expression:

$$Y_i(\vec{r}, t) = Y_i(\vec{r}, t_0) + \int_{t_0}^{t} \left(\alpha_{21,i} + \sum_{\xi=1}^{20} \alpha_{\xi,i} \phi_\xi(\vec{r}, \tau) \right) d\tau$$

where $Y_i(\vec{r}, t_0)$ is the initial value of the trait for genotype i. Because this initial value is zero for all genotypes across all locations in the field, the expression simplifies to:

$$Y_i(\vec{r}, t) = \int_{t_0}^{t} \left(\alpha_{21,i} + \sum_{\xi=1}^{20} \alpha_{\xi,i} \phi_\xi(\vec{r}, \tau) \right) d\tau \tag{3.3}$$

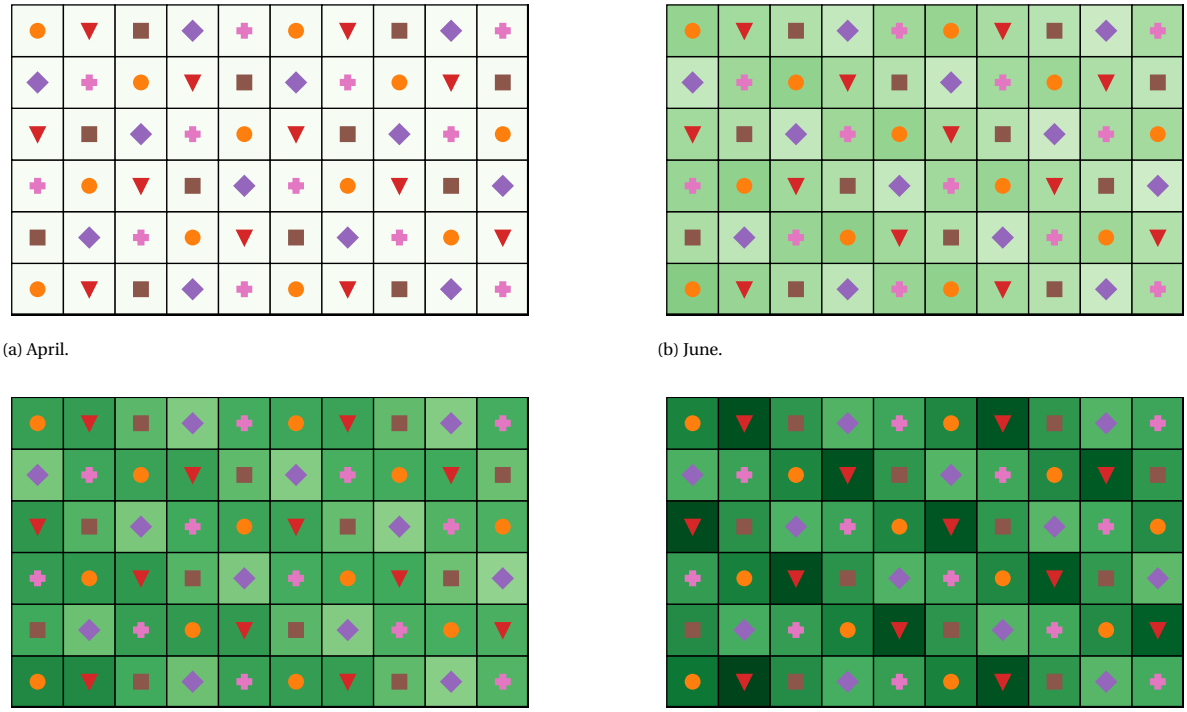
We numerically solve this differential equation using Euler's method, a simple and widely used numerical approach for solving ordinary differential equations by approximating solutions through discrete time steps

[3].

However, since we work with average values over half-week intervals, each step value represents the average environmental variable across that half week. Thus, summing these averaged values exactly computes the cumulative integral over time. Therefore, in this case, the Forward-Euler method is not an approximation, and Equation 3.3 can be computed exactly using the formula:

$$Y_i(\vec{r}, t_k) = Y_i(\vec{r}, t_{k-1}) + \Delta t \left[\alpha_{21,i} + \sum_{\xi=1}^{20} \alpha_{\xi,i} \phi_\xi(\vec{r}, t_k) \right]$$

This approach yields the crop performance over time. Figure 3.4 illustrates the weight of storage organs at several stages of the growing season, displayed alongside the spatial distribution of genotypes. The figures clearly show that genotypes differ in performance. Moreover, the potatoes' performance varies across the field, but the influence of the soil moisture field effect on the performance is relatively subtle and may be difficult to discern in these color-coded images.



(a) April.

(b) June.

(c) August.

(d) October.

Figure 3.4: Development of Performance within the field, which consists of 60 plots. The intensity of the green shading in each plot reflects the WSO, with darker green indicating higher values. Variations between the genotypes are explained by the genotype-specific coefficients α that describe how environmental factors influence genotype performance. Variations in performance within the same genotype are explained by the field effect. The spatial distribution of genotypes is also depicted, with each genotype represented by a distinct combination of color and marker shape: genotype 0 as an orange circle, genotype 1 as a red triangle, genotype 2 as a brown square, genotype 3 as a purple diamond, and genotype 4 as a pink plus-symbol.

To gain further insight into the development of genotypes throughout the growing season and the variation in performance across the field, refer to Figure 3.5. This figure shows the mean weight of storage organs for each genotype over time, along with error bars that represent the spatial variation within the field at each time point.

Note that in this figure, the WSO values are normalized to lie within the range $[0, 1]$.

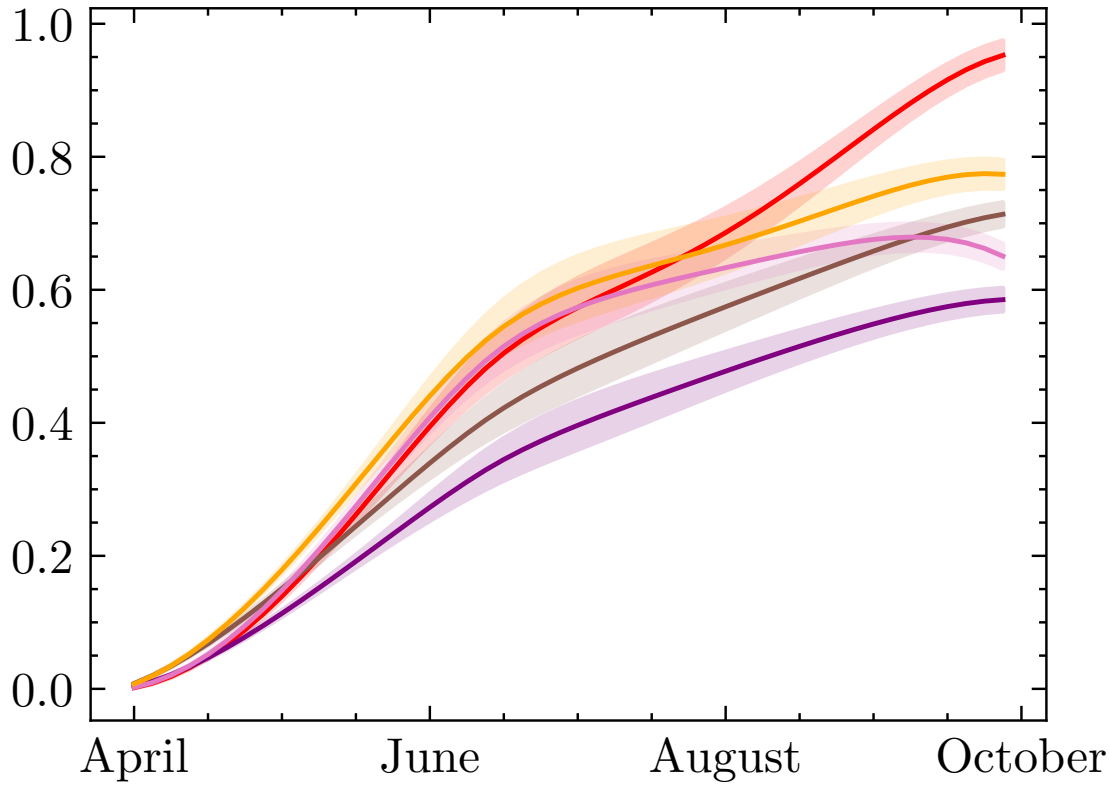


Figure 3.5: Normalized mean performance in WSO values throughout the growing season. Each color corresponds to one of the five genotypes, with error bars indicating the standard deviation to capture spatial variability across the field. Genotype 0 is shown in orange, genotype 1 in red, genotype 2 in brown, genotype 3 in purple, and genotype 4 in pink.

4

Two-Step Method

Now that we have generated the synthetic environmental variables and computed the WSO-values, we can proceed to reconstruct the model coefficients. Although the values for the environmental variables and WSO are synthetically generated, we refer to them as "measured," since the model is ultimately intended for application to real, measured data.

In this chapter the reconstruction process follows a two-step approach. First, we use the field effect measurements to estimate the field effect coefficients (β); then, using these coefficients, we reconstruct the soil moisture. With the reconstructed soil moisture, the measurements of environmental variables, and the WSO-measurements, we can then estimate the genotype-specific environmental coefficients (α).

4.1. Least-Squares Estimation of Field Effect Coefficients

The field effect coefficients β can be reconstructed using the least-squares method, which provides the values that minimize the residuals — the difference between the observed values (field effect $\mathbf{v}$) and the modeled values $B\boldsymbol{\beta}$. Since β is genotype-independent, we describe one linear system of the form: $B\boldsymbol{\beta} = \mathbf{v}$, where:

- $\boldsymbol{\beta} \in \mathbb{R}^{18}$: The vector of field effect coefficients.
- $\mathbf{v} \in \mathbb{R}^{r_{\text{size}} \cdot t_{\text{size}}}$: The observed field effect at different times and locations.
- $B \in \mathbb{R}^{r_{\text{size}} \cdot t_{\text{size}} \times 18}$: The matrix built from the products of basis functions

In this project, we work with 48 time steps and 60 spatial locations, so $t_{\text{size}} = 48$ and $r_{\text{size}} = 60$. The system $B\boldsymbol{\beta} = \mathbf{v}$ therefore becomes:

$$\begin{bmatrix} \zeta_{1,1}(\vec{r}_0, t_1) & \zeta_{1,2}(\vec{r}_0, t_1) & \cdots & \zeta_{3,6}(\vec{r}_0, t_1) \\ \zeta_{1,1}(\vec{r}_1, t_1) & \zeta_{1,2}(\vec{r}_1, t_1) & \cdots & \zeta_{3,6}(\vec{r}_1, t_1) \\ \vdots & \vdots & & \vdots \\ \zeta_{1,1}(\vec{r}_{60}, t_{48}) & \zeta_{1,2}(\vec{r}_{60}, t_{48}) & \cdots & \zeta_{3,6}(\vec{r}_{60}, t_{48}) \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix} = \begin{bmatrix} v(\vec{r}_0, t_1) \\ v(\vec{r}_1, t_1) \\ \vdots \\ v(\vec{r}_{60}, t_{48}) \end{bmatrix} \tag{4.1}$$

where

$$\zeta_{\omega\kappa}(\vec{r}, t) = \psi_\omega(t) \left(\chi_\kappa(\vec{r}) - \chi_\kappa(\vec{r}_c) \right)$$

We aim to find the vector $\boldsymbol{\beta}$ that minimizes the Euclidean norm of the residuals between the observed field effects and the model prediction, solving:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|B\boldsymbol{\beta} - \mathbf{v}\|_2^2$$

The minimizer is given by the analytical solution:

$$\hat{\boldsymbol{\beta}} = (B^\top B)^{-1} B^\top \mathbf{v}$$

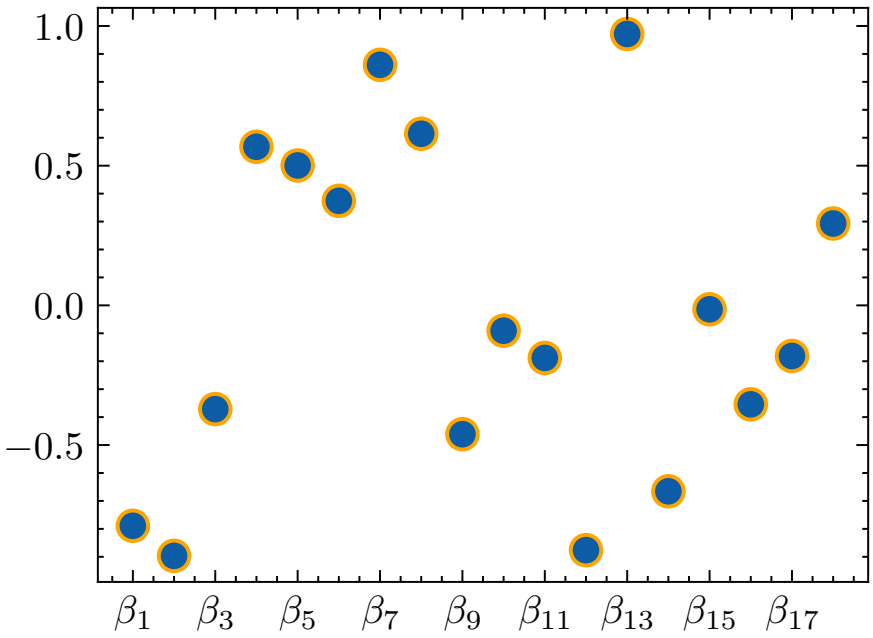


Figure 4.1: Least-squares estimation of the field effect coefficients $\boldsymbol{\beta}$. The filled blue circles represent the original coefficients, while the open orange circles indicate the reconstructed coefficients.

Since the synthetic data is noise-free and the matrix B has full rank, the least-squares solution $\hat{\boldsymbol{\beta}}$ will exactly recover the true coefficient vector $\boldsymbol{\beta}$, up to numerical precision. This is confirmed in Figure 4.1 The relative error (RE) is defined as:

$$\mathrm{RE}(\boldsymbol{\beta}) = \frac{\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2}{\|\boldsymbol{\beta}\|_2^2} \tag{4.2}$$

The relative error (RE) in the computed field effect coefficient, was evaluated using the Euclidean norm (L2 norm) as implemented in NumPy's `linalg.norm` function. We obtain:

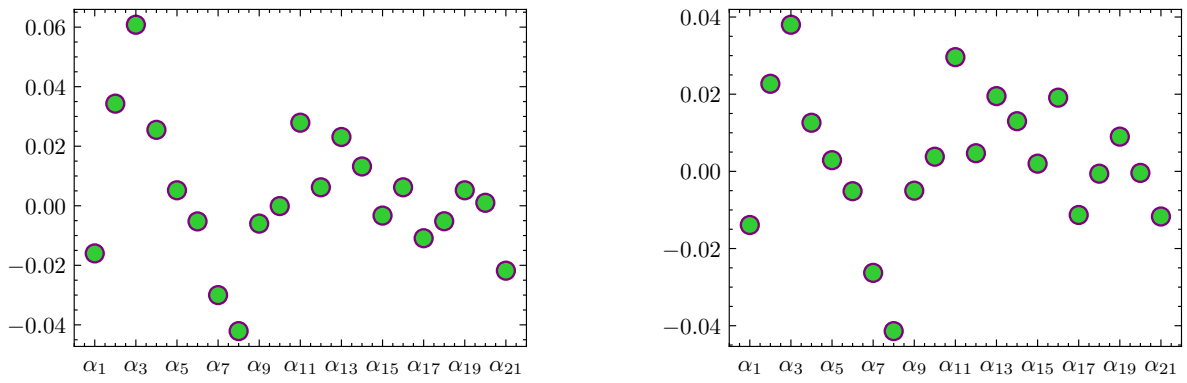
$$\mathrm{RE}(\boldsymbol{\beta}) = 2.57521 \times 10^{-28}$$

This extremely small error arises solely due to numerical round-off and floating-point precision limitations in the computation.

4.2. Least-Squares Estimation of Genotype-Specific Environmnetal Coefficients

In the second step of the two-step method, we first reconstruct the soil moisture using the previously estimated field effect coefficients. Then, using the measurements of the environmental-variables and the weight of storage organs, we estimate the α coefficients by applying the least-squares method once again. This method finds the parameters that minimize the residuals between the WSO measurements $\mathbf{Y}_i$ and the modeled output $A_i\boldsymbol{\alpha}_i$. Since α is genotype-dependent, we solve a system for each genotype of the form: $A_i\boldsymbol{\alpha}_i = \mathbf{Y}_i$, where:

- $\boldsymbol{\alpha}_i \in \mathbb{R}^{21}$: The vector of parameters to be estimated
- $\mathbf{Y}_i \in \mathbb{R}^{m \cdot t_{\text{size}}}$: The observed WSO values across all time points and across all spatial locations where genotype i is present
- $A_i \in \mathbb{R}^{m \cdot t_{\text{size}} \times 21}$: The system-matrix constructed from the integrated environmental variables ϕ_ξ across all time points and across all spatial locations where genotype i is present



(a) Genotype 1

(b) Genotype 4

Figure 4.2: Least-squares estimation of the genotype-specific environmental coefficients $\boldsymbol{\alpha}$. Filled green circles denote the original coefficients, while open purple circles represent the reconstructed estimates.

Here, m represents the number of spatial locations (plots) at which a genotype is present. Since there are 60 spatial locations evenly divided among 5 genotypes, we have $m = \frac{60}{5} = 12$. Therefore, each genotype is observed at 12 locations.

The system $A_i \boldsymbol{\alpha}_i = \mathbf{Y}_i$ can be expressed as:

$$\begin{bmatrix} \int_{t_0}^{t_1} w_1(\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4(\vec{r}_{0,i},\tau)w_5(\tau)\,d\tau & t_1 \\ \int_{t_0}^{t_1} w_1(\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4(\vec{r}_{1,i},\tau)w_5(\tau)\,d\tau & t_1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_{t_0}^{t_{48}} w_1(\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_4(\vec{r}_{12,i},\tau)w_5(\tau)\,d\tau & t_{48} \end{bmatrix} \begin{bmatrix} \alpha_{1,i} \\ \alpha_{2,i} \\ \vdots \\ \alpha_{21,i} \end{bmatrix} = \begin{bmatrix} Y(\vec{r}_{0,i},t_1) \\ Y(\vec{r}_{1,i},t_1) \\ \vdots \\ Y(\vec{r}_{12,i},t_{48}) \end{bmatrix} \tag{4.3}$$

For each data point — i.e., every combination of time t and position $\vec{r}$ — the integral values of the environmental variables ϕ_ξ in matrices A_i are computed using Euler's method.

We aim to find the vector $\boldsymbol{\alpha_i}$ that minimizes the Euclidean norm of the residuals between the observed weight of storage organs and the model prediction, solving:

$$\hat{\boldsymbol{\alpha}}_i = \arg\min_{\boldsymbol{\alpha}_i} \|A_i \boldsymbol{\alpha}_i - \mathbf{Y}_i\|_2^2$$

The least-squares solution can be expressed analytically as:

$$\hat{\boldsymbol{\alpha}}_i = (A_i^\top A_i)^{-1} A_i^\top \mathbf{Y}_i$$

Although each matrix A_i is full rank, the corresponding product $A_i^\top A_i$ has rank 20, indicating near-linear dependence among its columns. This makes the classic normal equation approach for solving linear least-squares problems — based on $(A_i^\top A_i)^{-1} A_i^\top$ — numerically unstable and potentially inaccurate. To address this, we estimate the coefficients $\boldsymbol{\alpha}$ using the *Moore–Penrose pseudoinverse* via `np.linalg.pinv`, which provides a stable and robust solution even in the presence of ill-conditioning or rank deficiency. As a result, the reconstructed $\boldsymbol{\alpha}$ closely approximates the true coefficients up to numerical precision. This high-accuracy reconstruction is illustrated in Figure 4.2 for genotypes 1 and 4. The relative errors in $\boldsymbol{\alpha}$ and the reconstructed WSO are computed as in equation 4.2. We obtain:

$$\mathrm{RE}(\boldsymbol{\alpha}) = 3.81833 \times 10^{-21}, \qquad \mathrm{RE}(\mathbf{Y}) = 8.9105 \times 10^{-22}$$

These tiny discrepancies are again explained by floating-point round-off errors.

4.3. Effect of Measurement noise in the Two-Step Method

In real-world scenarios, observations are affected by noise, and measurements deviate from the true values. These noisy measurements can be modeled as:

$$\tilde{\mathbf{v}} = B\boldsymbol{\beta} + \mathbf{n}_v, \quad \text{with} \quad \mathbf{n}_v \sim \mathcal{N}(0, \sigma_v^2) \quad \text{with} \quad \sigma_v = \epsilon \cdot \max(|\mathbf{v}|)$$

$$\tilde{\mathbf{Y}}_i = A_i\boldsymbol{\alpha}_i + \mathbf{n}_{Y_i}, \quad \text{with} \quad \mathbf{n}_{Y_i} \sim \mathcal{N}(0, \sigma_{Y_i}^2) \quad \text{with} \quad \sigma_{Y_i} = \epsilon \cdot \max(|\mathbf{Y_i}|)$$

Here, $\mathbf{n}$ denotes additive Gaussian noise with zero mean and a variance that scales with the magnitude of the true signal and the noise level (ϵ). Despite this noise, the model coefficients can still be estimated using the standard least-squares formulas:

$$\hat{\boldsymbol{\beta}} = (B^\top B)^{-1} B^\top \tilde{\mathbf{v}}, \qquad \hat{\boldsymbol{\alpha}}_i = (A_i^\top A_i)^{-1} A_i^\top \tilde{\mathbf{Y}}_i$$

We reconstruct the coefficients by adding varying levels of noise to the field effect and WSO measurements, respectively.

In Figure 4.3, we present the reconstruction of both the field effect coefficients and the genotype-specific environmental coefficients using the least-squares method under varying noise levels added to the field effect measurements. Only the results for genotype 1 are displayed, as the reconstruction performance shows little variation across different genotypes. Table 4.1 presents the corresponding relative errors in α, β, and Y.

Figure 4.4 illustrates how the environmental coefficients are reconstructed when noise is added to the WSO measurements. The reconstruction of the field effect coefficients is not shown here, as adding noise to the WSO measurements does not influence their reconstruction — it remains exact. Grey vertical lines are drawn from a coefficient to the x-axis whenever the reconstruction error is particularly small. This highlights that α_4, α_9, α_{13}, α_{16}, α_{18} and α_{20} are reconstructed well. All these coefficients, except for α_{18}, correspond to either the soil moisture term or interaction terms involving soil moisture, suggesting that access to field effect information enhances the model's robustness to noise. This effect can be attributed to the spatial variation of soil moisture across the field, which introduces greater variability in the system matrix A_i. As a result, the rows of A_i become more linearly independent, which improves the conditioning of the least-squares problem and enables more reliable estimation of the associated coefficients. The reason α_{18} is reconstructed well may be that it has the largest absolute value among the coefficients, making it easier for the algorithm to accurately estimate parameters with stronger signals.

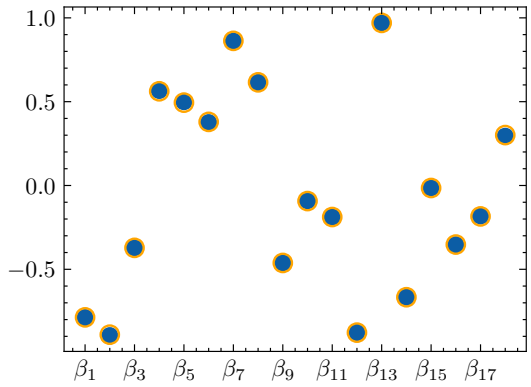
The relative errors in α and Y under different noise levels in the WSO measurements, can be found in Table 4.2. Noise in WSO significantly increases errors in genotype-specific coefficients (α), up to a factor of 10^7. The performance Y, however, is reconstructed with a relative error up to a factor of 10^{-2}. The reason Y can be reconstructed relatively accurately despite the large relative error in α may be that some errors from incorrectly reconstructed α cancel each other out.

Table 4.1: Relative error (RE) between measured and reconstructed values of the field effect coefficients (β), the genotype-specific environmental coefficients (α) and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the field effect (FE). Results are obtained using the two-step method.

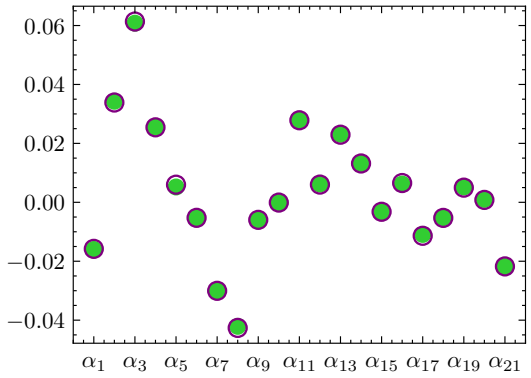
	Noise level in FE = 1%	**Noise level in FE = 5%**	**Noise level in FE = 10%**
$RE(\beta)$	4.01215×10^{-5}	3.98077×10^{-3}	8.69498×10^{-3}
$RE(\alpha)$	8.36684×10^{-4}	2.21691×10^{-1}	8.18840×10^{-1}
$RE(Y)$	5.14281×10^{-9}	1.84097×10^{-7}	1.19770×10^{-6}

Table 4.2: Relative error (RE) between measured and reconstructed values of the genotype-specific environmental coefficients (α) and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the weight of storage organs (WSO). Results are obtained using the two-step method.

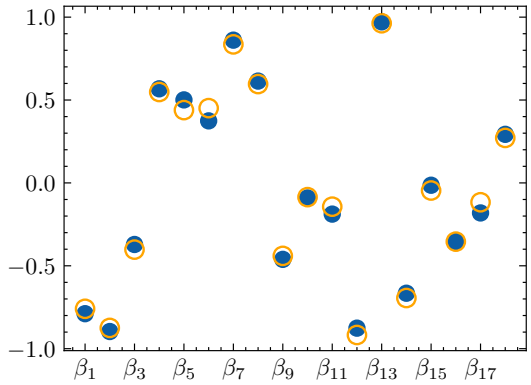
	Noise level in WSO = 1%	**Noise level in WSO = 5%**	**Noise level in WSO = 10%**
$RE(\alpha)$	5.68096×10^{5}	6.27501×10^{7}	4.40664×10^{7}
$RE(Y)$	2.34260×10^{-4}	5.56991×10^{-3}	2.26304×10^{-2}



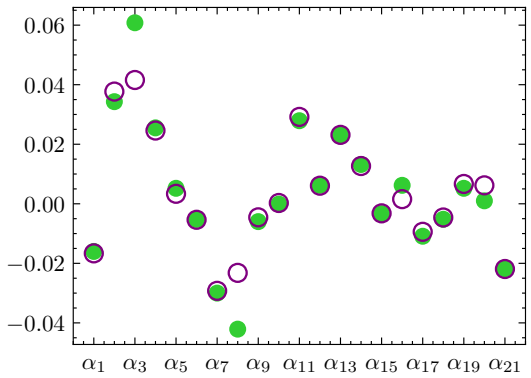
(a) Original and reconstructed field effect coefficients with 1% noise in the field effect measurements.



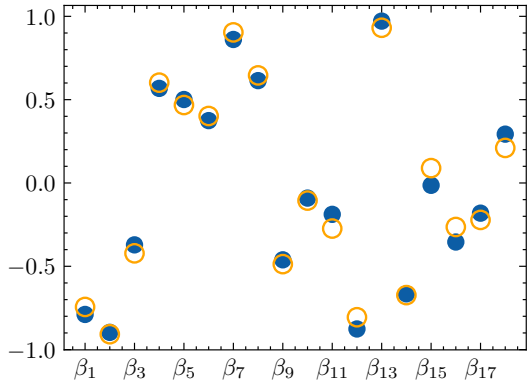
(b) Original and reconstructed environmental coefficients for Genotype 1 with 1% noise in the field effect measurements.



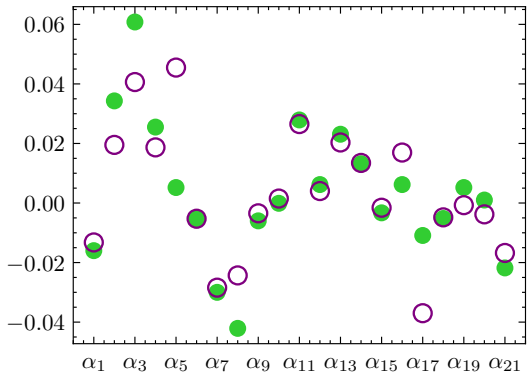
(c) Original and reconstructed field effect coefficients with 5% noise in the field effect measurements.



(d) Original and reconstructed environmental coefficients for Genotype 1 with 5% noise in the field effect measurements.

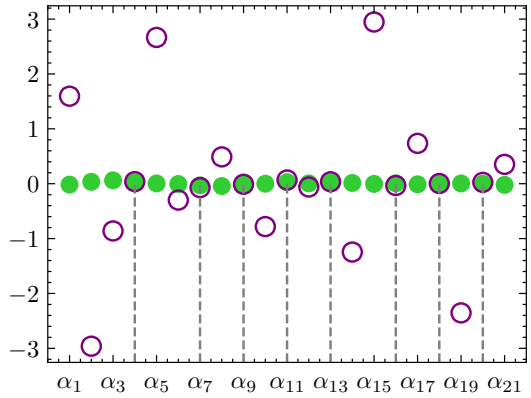


(e) Original and reconstructed field effect coefficients with 10% noise in the field effect measurements.

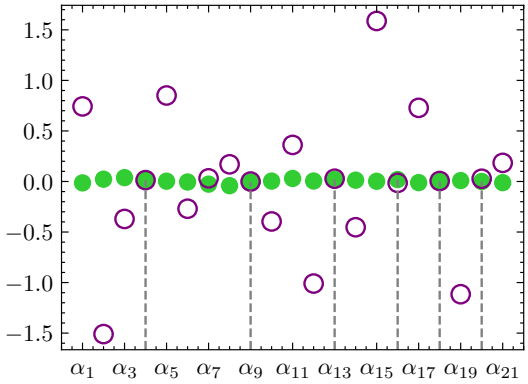


(f) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the field effect measurements.

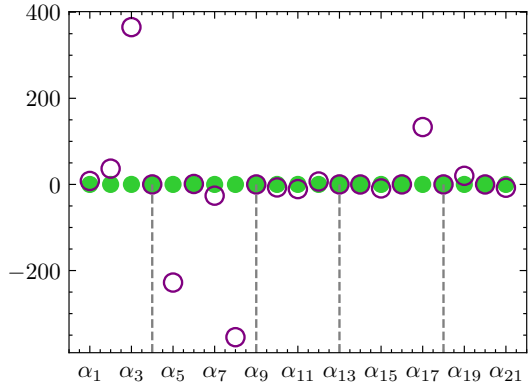
Figure 4.3: Least-squares estimation results for varying noise levels in the field effect measurements, using the two-step method. The left column displays the field effect coefficients $\boldsymbol{\beta}$, with filled blue circles representing the original coefficients and open orange circles showing the reconstructed values. The right column presents the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, where filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates.



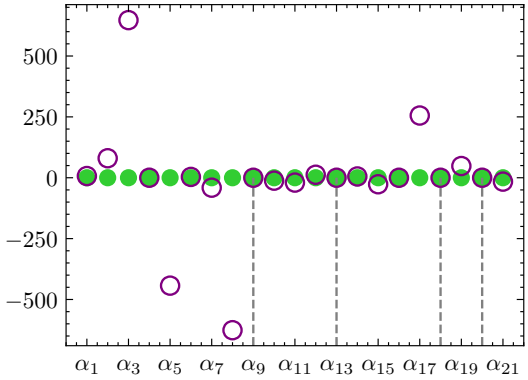
(a) Original and reconstructed environmental coefficients for Genotype 1 with 1% noise in the WSO measurements.



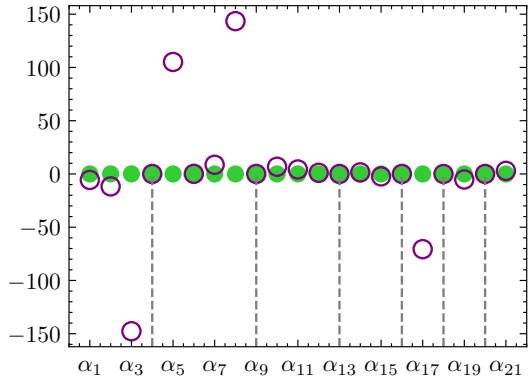
(b) Original and reconstructed environmental coefficients for Genotype 4 with 1% noise in the WSO measurements.



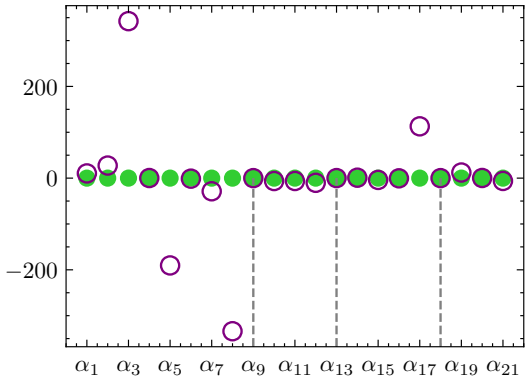
(c) Original and reconstructed environmental coefficients for Genotype 1 with 5% noise in the WSO measurements.



(d) Original and reconstructed environmental coefficients for Genotype 4 with 5% noise in the WSO measurements.



(e) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the WSO measurements.



(f) Original and reconstructed environmental coefficients for Genotype 4 with 10% noise in the WSO measurements.

Figure 4.4: The least-squares estimation results for varying noise levels in the weight of storage organs measurements. The right column shows the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, whereas the left column displays those for genotype 4. The filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates. The grey vertical lines are drawn from a coefficient to the x-axis whenever the reconstruction error is particularly small.

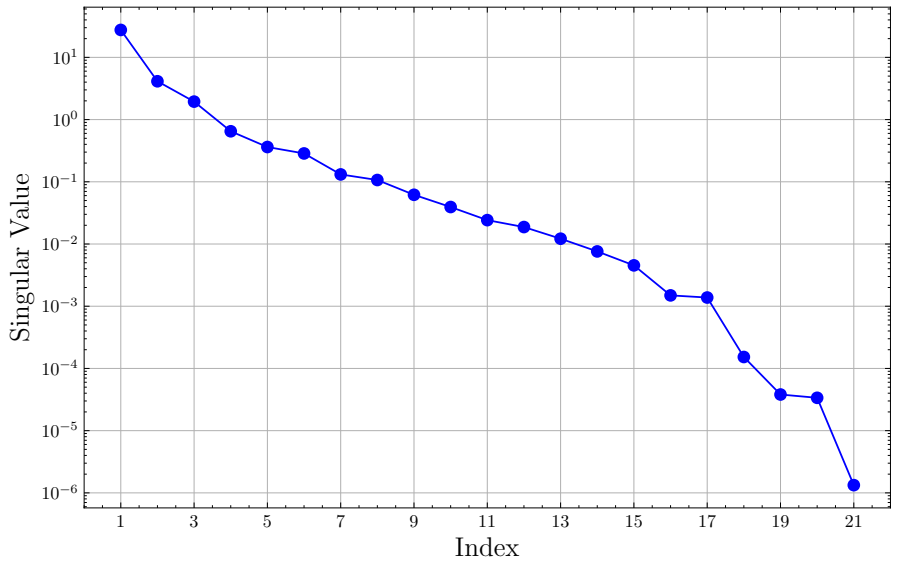


Figure 4.5: Singular values of the system matrix A_1 for genotype 1, plotted on a logarithmic scale.

4.3.1. Regularization with Truncated SVD Method

As can be seen in Table 4.2, the error in the estimated coefficients $\boldsymbol{\alpha}$ increases rapidly as noise is introduced into the WSO measurements, indicating that the underlying problem is ill-conditioned. Small perturbations in the data result in disproportionately large errors in the coefficient estimates due to the amplification of small singular values during the matrix inversion process.

Figure 4.5 displays the singular values of the matrix A_1 on a logarithmic scale for genotype 1, illustrating a steep decay towards zero. Similar decay patterns are observed across the other genotypes.

The severity of this ill-conditioning is quantified by the condition number, defined as:

$$\kappa(A_i) = \frac{\sigma_{\max}(A_i)}{\sigma_{\min}(A_i)}$$

where $\sigma_{\max}(A_i)$ and $\sigma_{\min}(A_i)$ denote the largest and smallest singular values of matrix A_i, respectively. Which gives:

$$\kappa(A_0) = 2.0739\times10^{7}, \quad \kappa(A_1) = 2.0733\times10^{7}, \quad \kappa(A_2) = 2.0737\times10^{7}, \quad \kappa(A_3) = 2.0725\times10^{7}, \quad \kappa(A_4) = 2.0725\times10^{7}$$

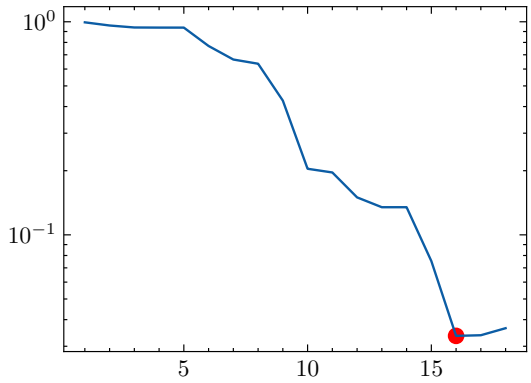
These high condition numbers further confirm the ill-conditioning of the problem.

To stabilize the reconstruction, we use a regularization method called truncated singular value decomposition (TSVD). TSVD keeps only the largest singular values of the system matrix and discards the smaller, noise-sensitive ones. By applying TSVD regularization, we suppress noise amplification and obtain a more stable and robust estimate of α, while reducing the risk of overfitting to noisy data.

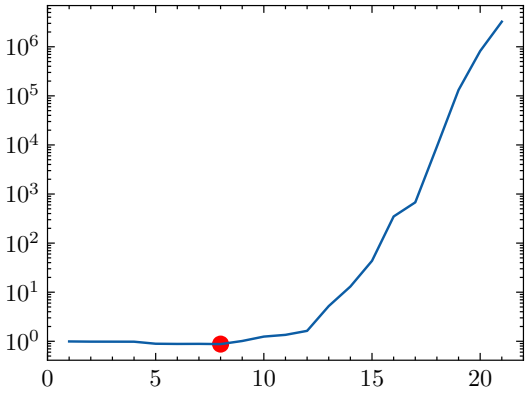
To determine the truncation threshold — that is, the number of singular values to keep — we calculate the relative error in α for each potential threshold and select the one that yields the smallest error. Similarly, for the matrix B, we identify the threshold that minimizes the error in β. Figure 4.6 shows the relative error of the parameter plotted as a function of the truncation threshold values.

In practice, this approach is not feasible, as it requires knowledge of the true α and β values, which are unknown in real-world scenarios. Our goal here is simply to demonstrate that an optimal threshold exists which enables robust reconstruction. In practical applications, one must choose both a regularization method and a regularisation-parameter selection strategy that do not rely on inaccessible ground-truth data.

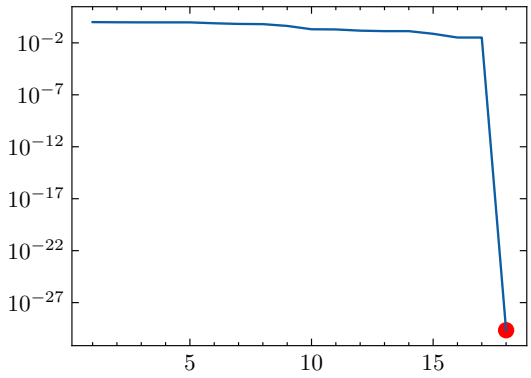
After applying TSVD regularization with the selected threshold, we reconstruct the coefficients and recalculate the relative errors to evaluate the improvement in reconstruction accuracy.



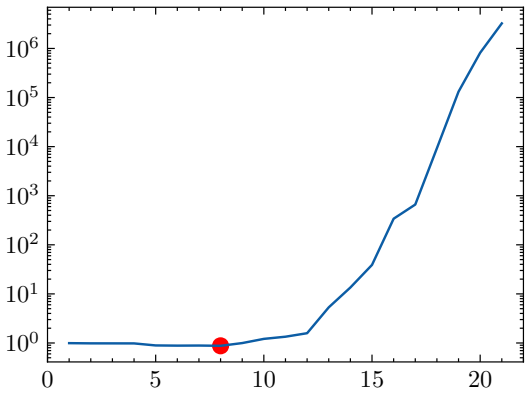
(a) Relative error in β plotted against the truncation threshold values. With 5% noise in the field effect. Chosen threshold is 16.



(b) Relative error in α plotted against the truncation threshold values, for genotype 4. With 5% noise in the field effect. Chosen threshold is 8.



(c) Relative error in β plotted against the truncation threshold values. With 5% noise in the WSO. Chosen threshold is 18.



(d) Relative error in α plotted against the truncation threshold values, for genotype 4. With 5% noise in the WSO. Chosen threshold is 8.

Figure 4.6: Relative error of the parameter (x-axis) plotted against the truncation threshold values (y-axis). The threshold that minimizes the relative error is highlighted with a red dot.

Figure 4.7 shows the reconstruction of both field effect and environmental coefficients under varying noise levels added to the field effect measurements, using the least-squares method with the described regularization applied. The corresponding relative errors in α, β, and Y are shown in Table 4.3.

Figure 4.8 illustrates the reconstruction of environmental coefficients when noise is added to the WSO measurements, with regularization applied. The corresponding relative errors in α and performance (Y) for different noise levels are presented in Table 4.4.

Table 4.3: Relative error (RE) between measured and reconstructed values of field effect coefficients (β), genotype-specific environmental coefficients (α) and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the field effect (FE). Results are obtained using the two step method, where the truncated singular value decomposition is employed for regularization.

	Noise level in FE = 1%	**Noise level in FE = 5%**	**Noise level in FE = 10%**
$RE(\beta)$	1.60481×10^{-4}	5.44900×10^{-3}	2.49362×10^{-2}
$RE(\alpha)$	6.21471×10^{-3}	1.08487×10^{-2}	1.43672×10^{-1}
$RE(Y)$	1.85823×10^{-8}	1.04279×10^{-7}	3.10231×10^{-7}

Table 4.4: Relative error (RE) between measured and reconstructed values of genotype-specific environmental coefficients (α) and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the weight of storage organs (WSO). Results are obtained using the two step method, where the truncated singular value decomposition is employed for regularization.

	Noise level in WSO = 1%	**Noise level in WSO = 5%**	**Noise level in WSO = 10%**
$RE(\alpha)$	7.50615×10^{-1}	8.53175×10^{-1}	8.51549×10^{-1}
$RE(Y)$	2.30128×10^{-4}	5.92710×10^{-3}	2.36792×10^{-2}

4.4. Minimizing Environmental Measurements and Intermediate Harvests

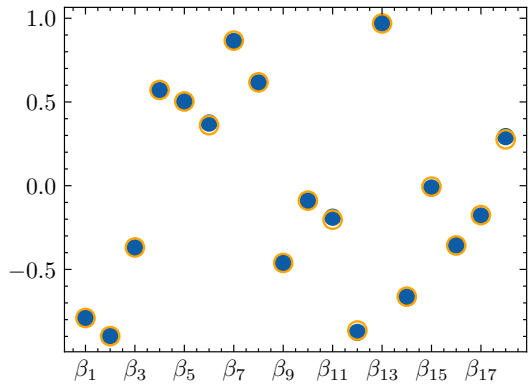
Accurately modeling the relationship between the weight of storage organs (WSO) and environmental variables requires both environmental measurements and crop yield observations. However, collecting these data can be both costly and labor-intensive. Environmental variables such as temperature, evaporation, and leaf area index are typically assumed to be uniform across the field and are often measured at a single location. However, soil moisture exhibits substantial spatial variation due to differences in soil composition, elevation, and drainage. For this reason, multiple measurement locations may be necessary to capture this variability. It is important to note, however, that once a measurement station is installed, it can continuously collect data at little additional cost or effort. Thus, environmental data collection is primarily limited by the number of spatial locations.

In contrast, estimating crop yield involves destructive intermediate harvests, which are undesirable because they remove immature plants from production. Additionally, it necessitates planting significantly more crops at the beginning of the growing season, demanding larger test fields. These intermediate harvests, however, are conducted across the entire field — typically by harvesting a few plants per plot and measuring the WSO for each genotype. As a result, the primary limitations are increased seed requirements, manual labor, test-field and plot size constraints, and the loss of immature plants.

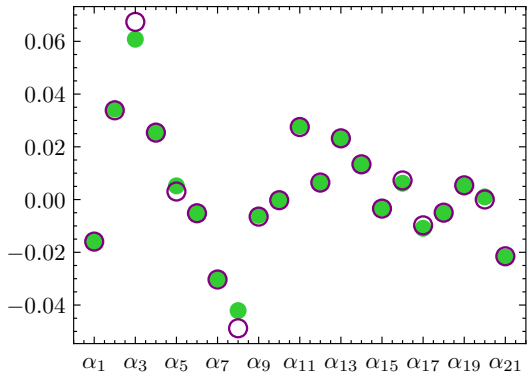
This section investigates how to minimize the number of environmental measurement locations and intermediate harvest time points needed, while still enabling accurate reconstruction of the relationship between genotype performance and environmental variables.

4.4.1. Environmental Measurement Locations

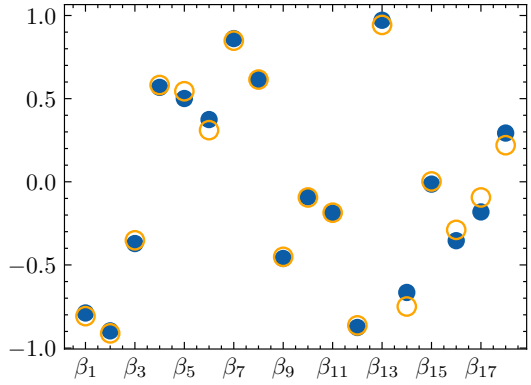
First, we develop an algorithm to select a specified number of measurement locations uniformly distributed across the entire field. The algorithm starts by finding two integers c and r who are as close together as possible, whose product is equal to the desired number of measurements n. With the constraints that $c \leq 10$ (reflecting the 10 plots in the x-direction) and $r \leq 6$ (reflecting the 6 plots in the y-direction). If no exact factor pair exists for n, the algorithm incrementally increases n until it finds a suitable pair (c, r). The field is then divided into an $c \times r$ grid, with a measurement location placed at the center of each grid cell. To ensure the reference point is included among the measurement locations, it is added explicitly, and then existing



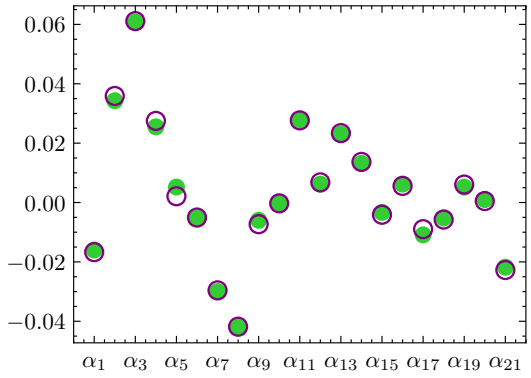
(a) Original and reconstructed field effect coefficients with 1% noise in the field effect measurements.



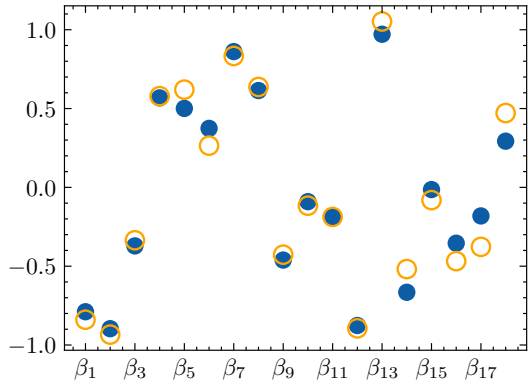
(b) Original and reconstructed environmental coefficients for Genotype 1 with 1% noise in the field effect measurements.



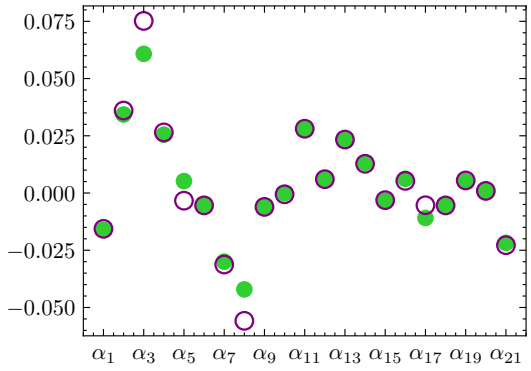
(c) Original and reconstructed field effect coefficients with 5% noise in the field effect measurements.



(d) Original and reconstructed environmental coefficients for Genotype 1 with 5% noise in the field effect measurements.

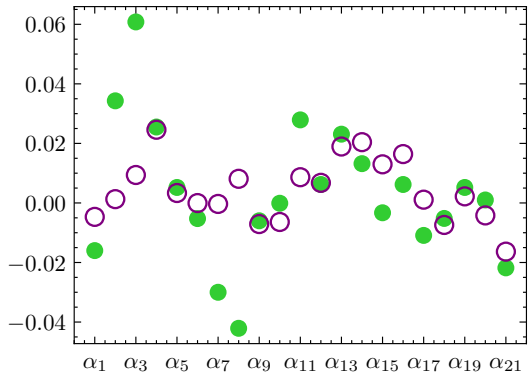


(e) Original and reconstructed field effect coefficients with 10% noise in the field effect measurements.

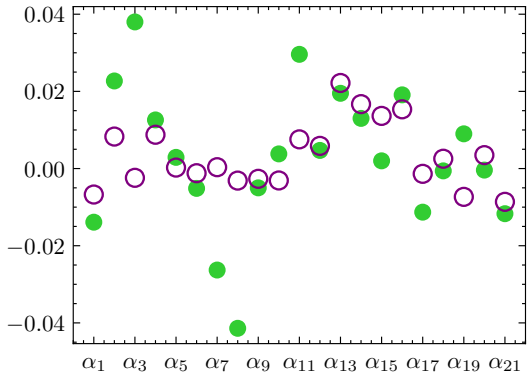


(f) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the field effect measurements.

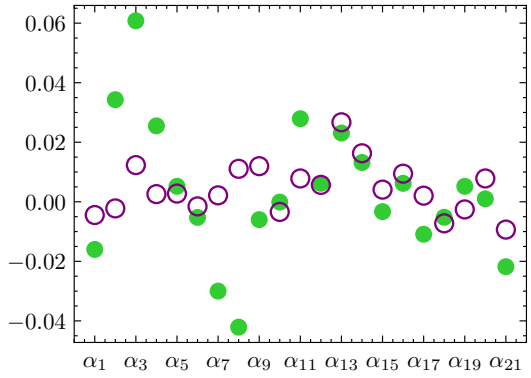
Figure 4.7: Least-squares estimation results for varying noise levels in the field effect measurements, using truncated SVD regularization on the two-step method. The left column displays the field effect coefficients $\boldsymbol{\beta}$, with filled blue circles representing the original coefficients and open orange circles showing the reconstructed values. The right column presents the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, where filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates.



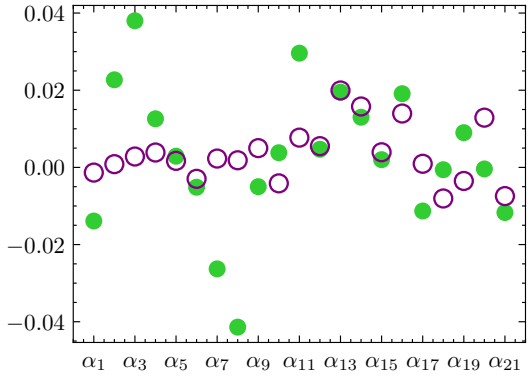
(a) Original and reconstructed environmental coefficients for Genotype 1 with 1% noise in the WSO measurements.



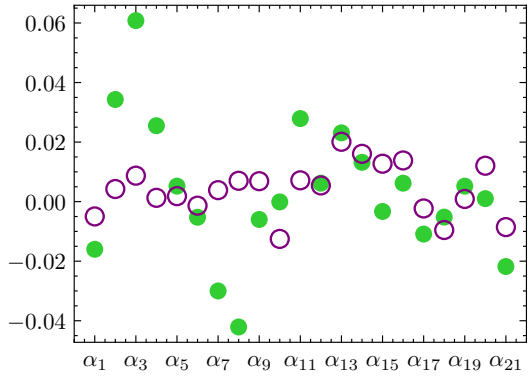
(b) Original and reconstructed environmental coefficients for Genotype 4 with 1% noise in the WSO measurements.



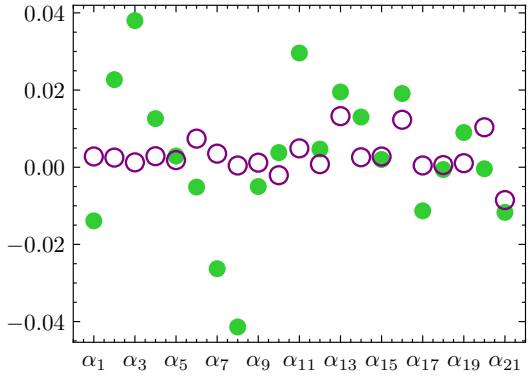
(c) Original and reconstructed environmental coefficients for Genotype 1 with 5% noise in the WSO measurements.



(d) Original and reconstructed environmental coefficients for Genotype 4 with 5% noise in the WSO measurements.



(e) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the WSO measurements.



(f) Original and reconstructed environmental coefficients for Genotype 4 with 10% noise in the WSO measurements.

Figure 4.8: Least-squares estimation results under varying noise levels in the weight of storage organs measurements, using truncated SVD regularization on the two-step method. The right column shows the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, whereas the left column displays those for genotype 4. The filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates.

locations are removed as needed until the total number of measurements matches the original input n.

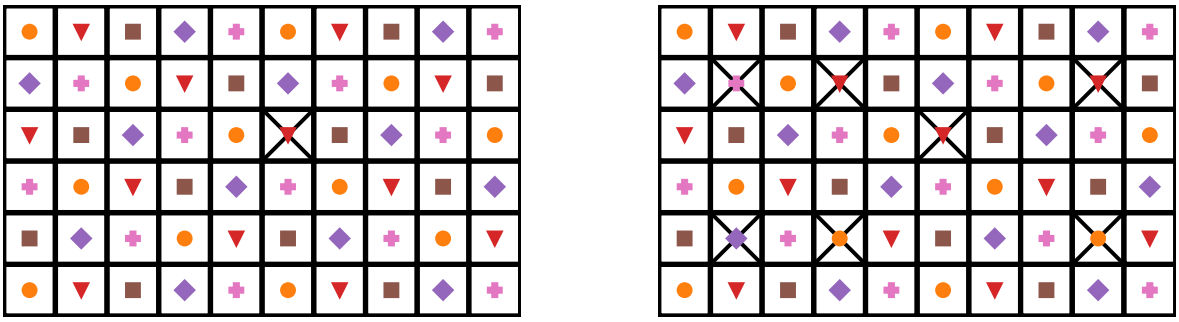
This algorithm ensures that the environmental variability is sampled as evenly as possible, helping to capture spatial patterns. Once the measurement locations are chosen, we retain only the synthetic environmental data corresponding to these selected points. This subset of data is then used to reconstruct the model coefficients, allowing us to assess how well the reduced set of measurements can represent the underlying environmental effects on the crop.

It is important to note that the measurement location at the center of the field, denoted as the reference point $\vec{r}_c$, is fixed and always included in the selection. This is because we impose the condition

$$v(\vec{r}_c, t) = 0,$$

which sets the field effect to zero at the reference point. By always including the reference-point in the selection, we ensure comparability in the coefficient reconstruction across different subsets of measurement locations.

Figure 4.9 demonstrates how the measurement points are spatially arranged to ensure even coverage.



(a) One evironmental measurement location.

(b) Seven environmental measurement locations

Figure 4.9: Environmental measurement locations are uniformly distributed across the field, which consists of 60 plots. The measurement locations are indicated by black crosses on the corresponding plots. The spatial distribution of genotypes is also depicted, with each genotype represented by a distinct combination of color and marker shape: genotype 0 as an orange circle, genotype 1 as a red triangle, genotype 2 as a brown square, genotype 3 as a purple diamond, and genotype 4 as a pink plus-symbol.

4.4.2. Intermediate Harvests

We develop an algorithm to select a fixed number of uniformly spaced measurement time points across the time domain. It divides the time interval into equal segments, selects midpoints as measurement indices, and combines these with spatial locations for each genotype, ensuring consistent, evenly distributed temporal sampling throughout the dataset.

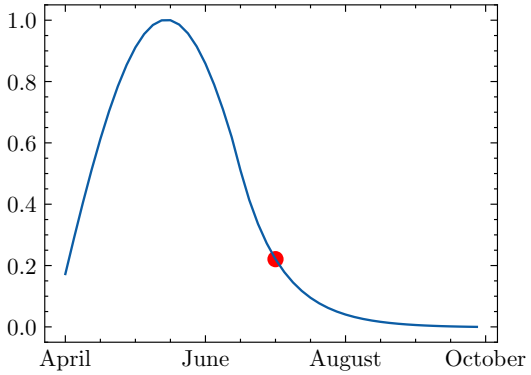
Figure 4.10 illustrates how the algorithm selects timepoints for early harvesting, as well as the corresponding data extracted from the five environmental variables at these selected timepoints.

4.4.3. Finding The Optimal Combination

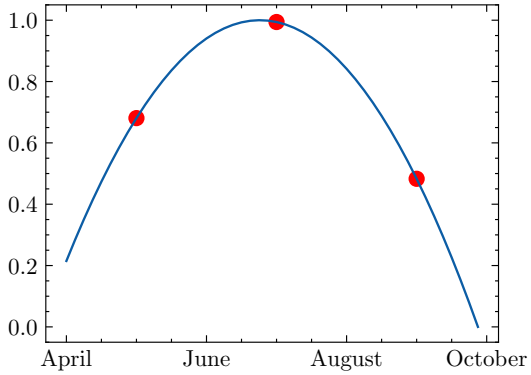
These time-points, along with corresponding spatial locations, are used to reduce the available data. Specifically, the selected time-points determine which rows to retain in the system of Equation 4.1. After reconstructing the field effect coefficients, we use Equation 4.3 and again select the rows corresponding to the chosen time-points to perform the second-stage reconstruction.

We create an algorithm that loops through all combinations of number of intermediate harvests (1 to 48) and number of environmental measurement locations (1 to 60). For each combination, the algorithm reconstructs both the field effect coefficients and the environmental coefficients using the two-step reconstruction method. After each reconstruction, we compute the relative error to assess the accuracy of the estimates. The results are visualized in Figure 4.11 with heat maps, illustrating how reconstruction accuracy varies across different data availability scenarios.

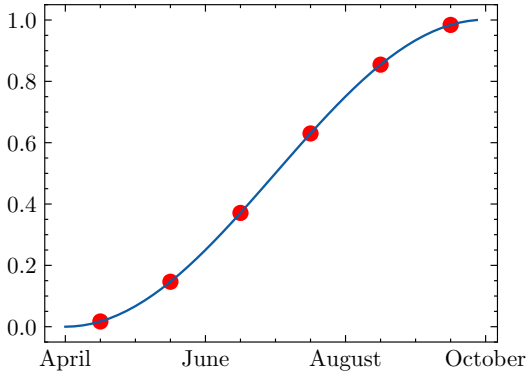
The heatmap for $\boldsymbol{\beta}$ indicates that at least 7 environmental measurement locations are required for reliable reconstruction. Increasing the number of locations from 6 to 7 reduces the relative error by a factor of 10^{24}.



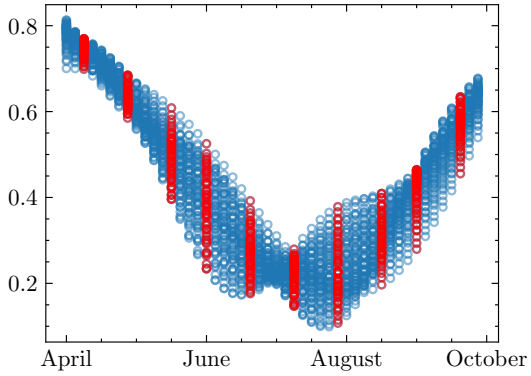
(a) w1: Potential evaporation for a reference crop leaf throughout the growing season. One intermediate harvest.



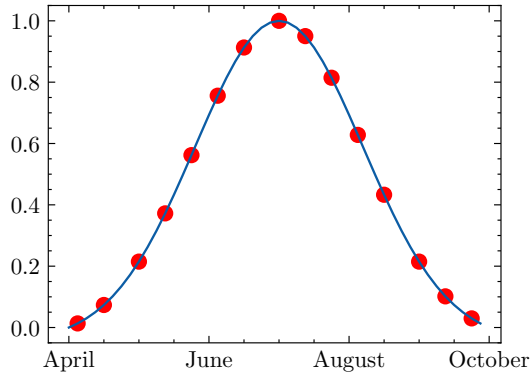
(b) w2: Average daily temperature throughout the growing season. Three intermediate harvests.



(c) w3: Temperature integral of daily average temperature throughout the growing season. Six intermediate harvests.

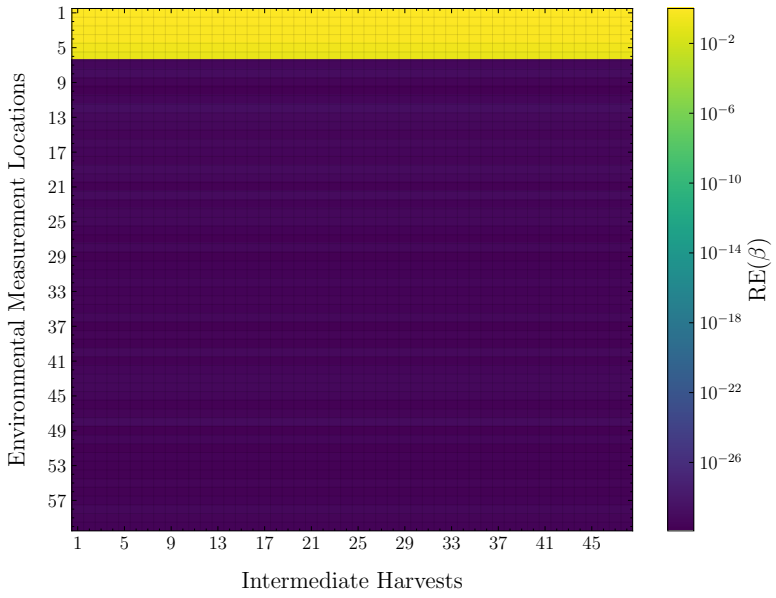


(d) w4: Soil moisture of daily average temperature throughout the growing season. Ten intermediate harvests.

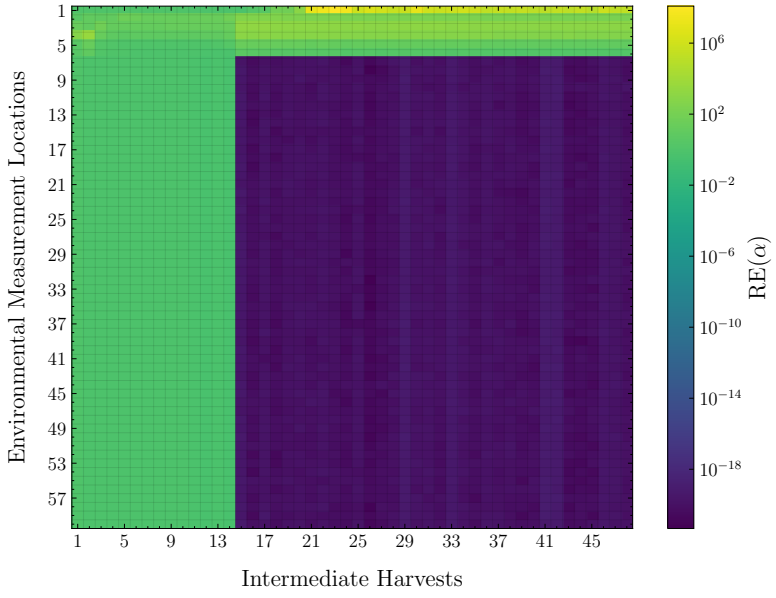


(e) w5: Leaf area index throughout the growing season. Fifteen intermediate harvests.

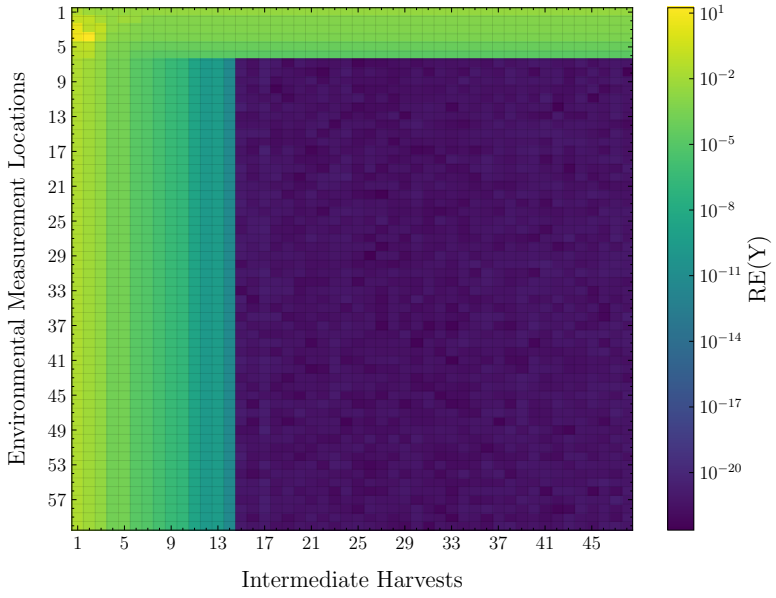
Figure 4.10: Overview of the environmental variables used in the model, shown across the growing season. Each subplot illustrates a different variable, with a different number of selected timepoints. The uniformly distributed red dots indicate the timing of the intermediate harvests.



(a) Heatmap of relative errors in the field effect coefficients $\boldsymbol{\beta}$.



(b) Heatmap of relative errors in the genotype-specific environmental coefficients $\boldsymbol{\alpha}$.



(c) Heatmap of relative error in performance $\boldsymbol{Y}$, i.e., the difference between measured and reconstructed WSO.

Figure 4.11: Heatmaps showing the relative error (on a logarithmic scale) for all combinations of the number of intermediate harvests (x-axis) and environmental measurement locations (y-axis) using the Two-Step Method. Yellow indicates higher relative error; dark blue indicates lower error. The colorbar on the right maps colors to error levels.

Since the reconstruction of the field effect coefficients in the two-step method relies exclusively on field effect measurements — obtained from the environmental measurement locations — it is unaffected by the WSO-measurements obtained from intermediate harvests. Hence the reconstruction of the field effect coefficients is not impacted by changing the number of intermediate harvests.

Accurate estimation of $\boldsymbol{\alpha}$ also requires a minimum of 7 environmental measurement locations. In addition, a minimum of 15 intermediate harvests is necessary to achieve a stable and precise reconstruction of the genotype-specific environmental coefficients. The improvements are again abrupt: increasing from 6 to 7 locations or from 14 to 15 harvests results in error reductions on the order of 10^{16}.

The performance heatmap (Y) — which compares measured and reconstructed WSO values — shows a smoother decline in error as the number of intermediate harvests increases than is observed in the heatmap for $\boldsymbol{\alpha}$. However, similar to the heatmaps for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the performance heatmap also exhibits an abrupt transition with respect to the number of environmental measurement locations. Notably, the relative error drops significantly when increasing from 14 to 15 harvests (by a factor of 10^{10}) or from 6 to 7 measurement locations (by a factor of 10^{16}).

These results suggest that the optimal experimental design includes 7 environmental measurement locations and 15 intermediate harvest timepoints.

5

One-Step Method

In this chapter, we introduce an alternative approach to the reconstruction problem: the *One-Step Method*. Unlike the Two-Step Method discussed previously — which decouples the estimation of field effect coefficients ($\boldsymbol{\beta}$) and genotype-specific environmental coefficients ($\boldsymbol{\alpha}$) — the One-Step Method simultaneously estimates both parameter sets within a unified optimization framework.

This joint estimation approach is motivated by the observation that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are inherently coupled through the underlying model. By solving for both parameter types simultaneously, the One-Step Method has the potential to exploit the full structure of the data, potentially leading to improved accuracy.

This chapter begins with a formulation of the One-Step reconstruction problem, followed by a description and the result of the nonlinear least-squares approach used to solve it. We then analyze the robustness of the method under various noise levels, and select the optimal combination of number of environmental measurement locations and number of early harvests.

5.1. Formulation of the Nonlinear Model

To this end, we begin with the integrated form of the model equation, which describes the relationship between the weight of storage organs and environmental factors:

$$Y_i(\vec{r},t) = Y_i(\vec{r},t_0) + \int_{t_0}^{t} \left(\alpha_{21,i} + \sum_{\xi=1}^{20} \alpha_{\xi,i} \phi_\xi(\vec{r},\tau) \right) d\tau \tag{5.1}$$

In this formulation, the variable w_4 appears in six different terms within the environmental vecor $\phi(\vec{r},\tau)$. Recall that w_4 is defined as the sum of a field constant component and a spatially varying component:

$$w_4(\vec{r},t) = w_4^c(t) + v(\vec{r},t) = w_4^c(t) + \sum_{\omega=1}^{3} \sum_{\kappa=1}^{6} \left[\beta_{\omega\kappa} \cdot \psi_\omega(t) \cdot \left(\chi_\kappa(\vec{r}) - \chi_\kappa(\vec{r}_c) \right) \right]$$

Substituting this expression for w_4 into Equation 5.1 results in the following expanded form:

$$
\begin{aligned}
\mathbf{Y_i}(\vec{\mathbf{r}},\mathbf{t}) = \int_{t_0}^{t} \Bigg(& \alpha_{21,i} + \sum_{\xi=1}^{3} \alpha_{\xi,i}\phi_\xi(\vec{r},\tau) \\
& + \alpha_{4,i}\left[w_4^c(\tau) + \sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right] \\
& + \sum_{\xi=5}^{8} \alpha_{\xi,i}\phi_\xi(\vec{r},\tau) \\
& + \alpha_{9,i}\left[w_4^c(\tau)^2 + 2\,w_4^c(\tau)\sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right. \\
& \left. + \sum_{\omega=1}^{3}\sum_{\omega'=1}^{3}\sum_{\kappa=1}^{6}\sum_{\kappa'=1}^{6}\left[\beta_{\omega\kappa}\,\beta_{\omega'\kappa'}\cdot\psi_\omega(\tau)\,\psi_{\omega'}(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\left(\chi_{\kappa'}(\vec{r})-\chi_{\kappa'}(\vec{r}_c)\right)\right]\right] \\
& + \sum_{\xi=10}^{12} \alpha_{\xi,i}\phi_\xi(\vec{r},\tau) + \alpha_{13,i}\left[w_1(t)\,w_4^c(\tau) + w_1(\tau)\sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right] \\
& + \sum_{\xi=14}^{15} \alpha_{\xi,i}\,\phi_\xi(\vec{r},\tau) + \alpha_{16,i}\left[w_2(\tau)\,w_4^c(\tau) + w_2(\tau)\sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right] \\
& + \alpha_{17,i}\,\phi_{17}(\vec{r},\tau) + \alpha_{18,i}\left[w_3(\tau)\,w_4^c(\tau) + w_3(\tau)\sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right] \\
& + \alpha_{19,i}\,\phi_{19}(\vec{r},\tau) + \alpha_{20,i}\left[w_5(\tau)\,w_4^c(\tau) + w_5(\tau)\sum_{\omega=1}^{3}\sum_{\kappa=1}^{6}\left[\beta_{\omega\kappa}\cdot\psi_\omega(\tau)\cdot\left(\chi_\kappa(\vec{r})-\chi_\kappa(\vec{r}_c)\right)\right]\right] \Bigg) d\tau
\end{aligned}
$$

This reformulated equation expresses the output $Y_i(\vec{r},t)$ as an integral over both α and β, making it suitable for joint reconstruction of the environmental and field effect coefficients. The next step is to rewrite this expression in matrix form to facilitate numerical solution and optimization.

$$
\begin{bmatrix} Y_i(\vec{r}_0,t_1) \\ Y_i(\vec{r}_1,t_1) \\ \vdots \\ Y_i(\vec{r}_m,t_{48}) \end{bmatrix} = \begin{bmatrix} \int_{t_0}^{t_1} w_1(\tau)\,d\tau & \int_{t_0}^{t_1} w_2(\tau)\,d\tau & \int_{t_0}^{t_1} w_3(\tau)\,d\tau & \int_{t_0}^{t_1} w_4^c(\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4^c(\tau)w_5(\tau)\,d\tau & t_1 \\ \int_{t_0}^{t_1} w_1(\tau)\,d\tau & \int_{t_0}^{t_1} w_2(\tau)\,d\tau & \int_{t_0}^{t_1} w_3(\tau)\,d\tau & \int_{t_0}^{t_1} w_4^c(\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4^c(\tau)w_5(\tau)\,d\tau & t_1 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \int_{t_0}^{t_{48}} w_1(\tau)\,d\tau & \int_{t_0}^{t_{48}} w_2(\tau)\,d\tau & \int_{t_0}^{t_{48}} w_3(\tau)\,d\tau & \int_{t_0}^{t_{48}} w_4^c(\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_4^c(\tau)w_5(\tau)\,d\tau & t_{48} \end{bmatrix} \begin{bmatrix} \alpha_{1,i} \\ \alpha_{2,i} \\ \alpha_{3,i} \\ \alpha_{4,i} \\ \vdots \\ \alpha_{20,i} \\ \alpha_{21,i} \end{bmatrix}
$$

$$
+\alpha_{4,i} \begin{bmatrix} \int_{t_0}^{t_1} \zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} \zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} \zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} \zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} \zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} \zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} \zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} \zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} \zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix}
$$

$$
+2\,\alpha_{9,i} \begin{bmatrix} \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_4^c(\tau)\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} w_4^c(\tau)\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} w_4^c(\tau)\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_4^c(\tau)\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix}
$$

$$
+\alpha_{9,i} \begin{bmatrix} \int_{t_0}^{t_1} \zeta_{1,1}\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} \zeta_{1,1}\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} \zeta_{3,6}\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} \zeta_{1,1}\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} \zeta_{1,1}\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} \zeta_{3,6}\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} \zeta_{1,1}\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} \zeta_{1,1}\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} \zeta_{3,6}\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1}\beta_{1,1} \\ \beta_{1,1}\beta_{1,2} \\ \vdots \\ \beta_{3,6}\beta_{3,6} \end{bmatrix}
$$

$$
+\alpha_{13,i} \begin{bmatrix} \int_{t_0}^{t_1} w_1(\tau)\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} w_1(\tau)\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_1(\tau)\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} w_1(\tau)\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} w_1(\tau)\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_1(\tau)\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} w_1(\tau)\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} w_1(\tau)\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_1(\tau)\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix}
$$

$$
+\alpha_{16,i} \begin{bmatrix} \int_{t_0}^{t_1} w_2(\tau)\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} w_2(\tau)\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_2(\tau)\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} w_2(\tau)\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} w_2(\tau)\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_2(\tau)\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} w_2(\tau)\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} w_2(\tau)\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_2(\tau)\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix}
$$

$$
+\alpha_{18,i} \begin{bmatrix} \int_{t_0}^{t_1} w_3(\tau)\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} w_3(\tau)\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_3(\tau)\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} w_3(\tau)\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} w_3(\tau)\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_3(\tau)\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} w_3(\tau)\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} w_3(\tau)\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_3(\tau)\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix}
$$

$$
+\alpha_{20,i} \begin{bmatrix} \int_{t_0}^{t_1} w_5(\tau)\zeta_{1,1}(\vec{r}_0,\tau)\,d\tau & \int_{t_0}^{t_1} w_5(\tau)\zeta_{1,2}(\vec{r}_0,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_5(\tau)\zeta_{3,6}(\vec{r}_0,\tau)\,d\tau \\ \int_{t_0}^{t_1} w_5(\tau)\zeta_{1,1}(\vec{r}_1,\tau)\,d\tau & \int_{t_0}^{t_1} w_5(\tau)\zeta_{1,2}(\vec{r}_1,\tau)\,d\tau & \cdots & \int_{t_0}^{t_1} w_5(\tau)\zeta_{3,6}(\vec{r}_1,\tau)\,d\tau \\ \vdots & \vdots & & \vdots \\ \int_{t_0}^{t_{48}} w_5(\tau)\zeta_{1,1}(\vec{r}_m,\tau)\,d\tau & \int_{t_0}^{t_{48}} w_5(\tau)\zeta_{1,2}(\vec{r}_m,\tau)\,d\tau & \cdots & \int_{t_0}^{t_{48}} w_5(\tau)\zeta_{3,6}(\vec{r}_m,\tau)\,d\tau \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix} \tag{5.2}
$$

Where $\zeta_{\omega\kappa}(t,\vec{r}) = \psi_\omega(t)\left(\chi_\kappa(\vec{r}) - \chi_\kappa(\vec{r}_c)\right)$

So far, we have only used our measurements of the WSO, the environmental variables, the field-constant component of w_4, and our knowledge of the field effect basis functions. However, we also possess direct measurements of the field effect itself, which can be used to reconstruct the beta coefficients. This system can be described in the same manner as in the previous chapter:

$$\begin{bmatrix} v(\vec{r}_0,t_1) \\ v(\vec{r}_1,t_1) \\ \vdots \\ v(\vec{r}_{60},t_{48}) \end{bmatrix} = \begin{bmatrix} \zeta_{1,1}(\vec{r}_0,t_1) & \zeta_{1,2}(\vec{r}_0,t_1) & \cdots & \zeta_{3,6}(\vec{r}_0,t_1) \\ \zeta_{1,1}(\vec{r}_1,t_1) & \zeta_{1,2}(\vec{r}_1,t_1) & \cdots & \zeta_{3,6}(\vec{r}_1,t_1) \\ \vdots & \vdots & & \vdots \\ \zeta_{1,1}(\vec{r}_{60},t_{48}) & \zeta_{1,2}(\vec{r}_{60},t_{48}) & \cdots & \zeta_{3,6}(\vec{r}_{60},t_{48}) \end{bmatrix} \begin{bmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{3,6} \end{bmatrix} \tag{5.3}$$

5.2. Nonlinear Least-Squares Approach

The objective is to find the least-squares solution coefficients that simultaneously fit both the system in Equation 5.3 and all instances of Equation 5.2, corresponding to each genotype.

To achieve this, we use the `least_squares` optimizer from `SciPy`, which minimizes the sum of squared residuals—that is, the differences between model predictions and observed data. Starting from an initial guess, the optimizer iteratively updates the parameter values to reduce the residual error. The algorithm approximates gradients numerically using finite differences. Convergence is achieved when changes in the cost function, parameters, and gradient norm fall below specified tolerances.

To incorporate all systems into a single optimization problem, we construct the combined vector:

$$\begin{bmatrix} \mathbf{Y}_0 \\ \vdots \\ \mathbf{Y}_5 \\ \mathbf{v} \end{bmatrix}$$

This allows us to define a residual vector that merges the normalized differences between model outputs and observations for both $\mathbf{Y}$ and $\mathbf{v}$. Normalizing these differences ensures that both components contribute fairly to the optimization, regardless of their original scales.

By solving the optimization problem usin the `least_squares` optimizer from `SciPy`, we obtain the coefficient estimates and a reconstructed WSO, along with the corresponding relative errors, which demonstrate a higher precision compared to the one-step method.

$$\mathrm{RE}(\boldsymbol{\beta}) = 3.28411 \times 10^{-25}. \qquad \mathrm{RE}(\boldsymbol{\alpha}) = 1.71000 \times 10^{-14}, \qquad \mathrm{RE}(\boldsymbol{Y}) = 4.5548 \times 10^{-24}.$$

5.3. Effect of Measurement Noise in the One-Step Method

As in Section 4.3, we now investigate the robustness of the One-Step method by introducing varying levels of noise. We use additive Gaussian noise with zero mean and a variance that scales with the magnitude of the true signal and with the noise level. We add varying levels of noise to eather the field effect or the WSO measurements, and thereafter solve the optimization problem using the `least_squares` optimizer from `SciPy` in the exact way as we did in Section 5.2.

In Figure 5.1, we present the reconstruction of both the field effect coefficients and the genotype-specific environmental coefficients using the least-squares optimizer under varying noise levels added to the field effect measurements. Only the results for genotype 1 are displayed, as the reconstruction performance shows little variation across different genotypes.

Table 5.1 presents the relative errors in α, β, and performance (Y) when noise is added to the field effect measurements.

Figure 5.2 illustrates how the environmental coefficients are reconstructed when noise is added to the WSO measurements. The reconstruction of the field effect coefficients is not shown here, as adding noise to the WSO measurements turned out to not influence their reconstruction significantly. This is likely because the field effect coefficients appear in System 5.3, which relies exclusively on field effect measurements. As a result, noise introduced only in the WSO data has little impact on the algorithm's ability to accurately reconstruct the field effect coefficients.

Similarly as in Section 4.3 the coefficients that are reconstructed well are: α_4, α_9, α_{13}, α_{16}, α_{18} and α_{20}. All these coefficients, except for α_{18}, correspond to either the soil moisture term or interaction terms involving soil moisture, suggesting that access to field effect information enhances the model's robustness to noise. This effect can be attributed to the spatial variation of soil moisture across the field, which introduces greater variability in the system matrices in System 5.2. As a result, the rows of the system matrices become more linearly independent, which improves the conditioning of the least-squares problem and enables more reliable estimation of the associated coefficients. The reason α_{18} is reconstructed well may be that it has the largest absolute value among the coefficients, making it easier for the algorithm to accurately estimate parameters with stronger signals.

Table 5.2 presents the relative errors α, β, and performance (Y) when noise is added to the WSO measurements.

Table 5.1: Relative error (RE) between measured and reconstructed values of field effect coefficients (β), genotype-specific environmental coefficients (α), and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the field effect (FE). Results are obtained using the one-step method.

	Noise level in FE = 1%	**Noise level in FE = 5%**	**Noise level in FE = 10%**
$RE(\beta)$	$7.94674e^{-2}$	$3.75340e^{-3}$	$1.50907e^{-2}$
$RE(\alpha)$	$2.05676e^{-2}$	$1.06416e^{-2}$	$9.00153e^{-1}$
$RE(Y)$	$1.77601e^{-8}$	$1.55257e^{-7}$	$9.71626e^{-7}$

Table 5.2: Relative error (RE) between measured and reconstructed values of genotype-specific environmental coefficients (α), and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the weight of storage organs (WSO). Results are obtained using the one-step method.

	Noise level in WSO = 1%	**Noise level in WSO = 5%**	**Noise level in WSO = 10%**
$RE(\beta)$	2.79289×10^{-8}	2.88038×10^{-6}	5.01422×10^{-6}
$RE(\alpha)$	5.27384×10^{5}	6.57592×10^{6}	1.19014×10^{8}
$RE(Y)$	1.52550×10^{-5}	2.07317×10^{-4}	9.14598×10^{-4}

In Table 5.1, increasing the noise level in the field effect (FE) measurements results in relative errors $RE(\beta)$ that remain low, ranging roughly between 10^{-2} and 10^{-3}. The relative error in the environmental coefficients α also stays below 10^{-1} even as noise increases to 10%. The performance variable Y is reconstructed with very high accuracy, with relative error remaining below 10^{-6}.

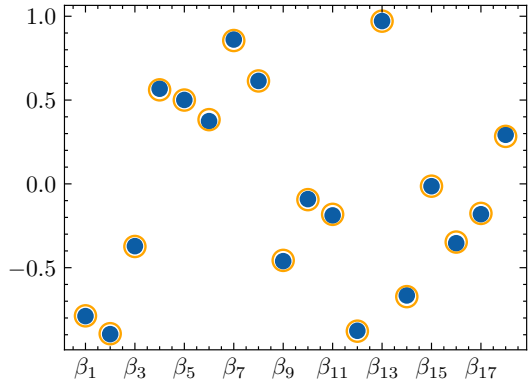
When noise is added to the weight of storage organs (WSO) measurements, as shown in Table 5.2, the relative error in β starts on the order of 10^{-8} at 1% noise, and increases modestly to about 10^{-6} at 10% noise. However, the relative error in α rises sharply with increasing noise, reaching values on the order of 10^8 at the highest noise level. Despite the dramaticly bad reconstruction of α the relative error in Y remains below 10^{-3} for all noise levels considered.

The reason Y can be reconstructed accurately despite α having a very large relative error may be that some errors cancel each other out. This occurs because the model is overparameterized — meaning it has more parameters than necessary—allowing multiple parameter combinations to fit the data well. As shown in Figure 3.5, the WSO curve effectively lies in a low-dimensional space and could likely be reconstructed using only three parameters.

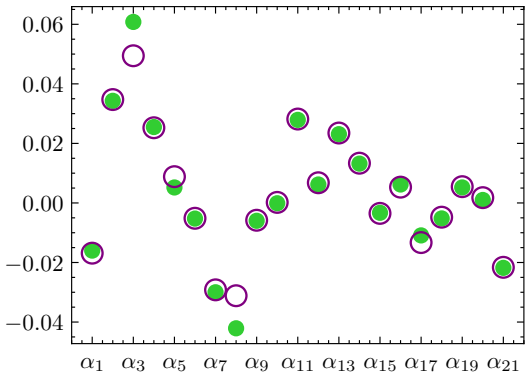
To gain deeper insight into the method's robustness, incorporating a regularization technique would be beneficial. However, due to the complexity of this nonlinear system, developing a suitable regularization approach for noise-affected coefficient reconstruction falls outside the scope of this project.

5.4. Minimizing Environmental Measurement Locations and Intermediate Harvests

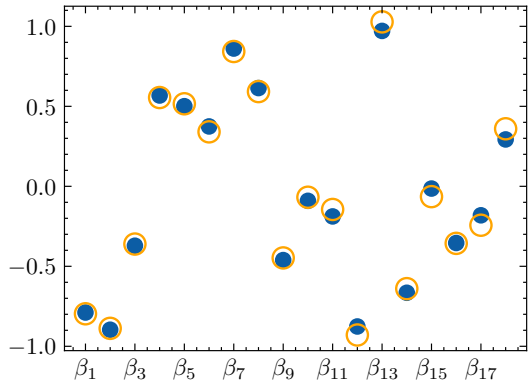
As discussed earlier, it is advantageous to minimize both the number of environmental measurement locations and the number of intermediate harvests. This is because installing measurement locations incurs sig-



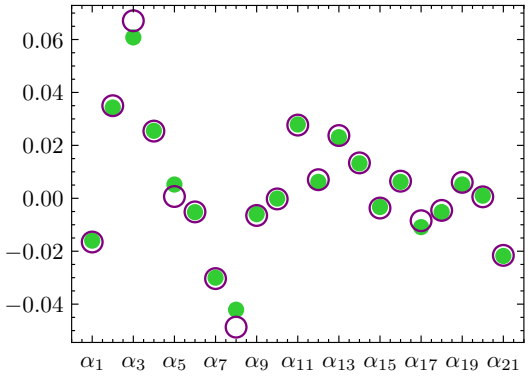
(a) Original and reconstructed field effect coefficients with 1% noise in the field effect measurements.



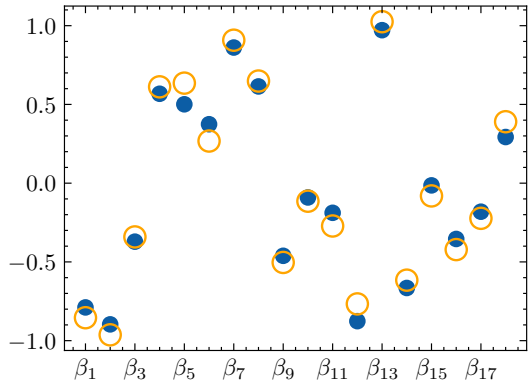
(b) Original and reconstructed environmental coefficients for genotype 1 with 1% noise in the field effect measurements.



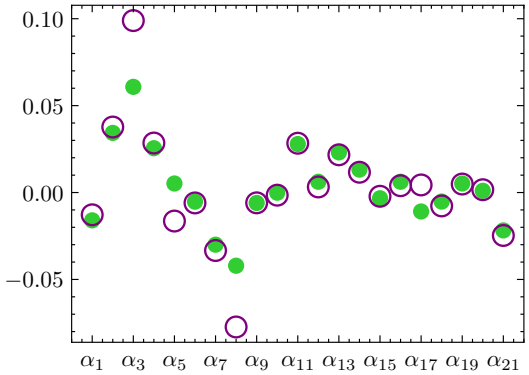
(c) Original and reconstructed field effect coefficients with 5% noise in the field effect measurements.



(d) Original and reconstructed environmental coefficients for Genotype with 5% noise in the field effect measurements.

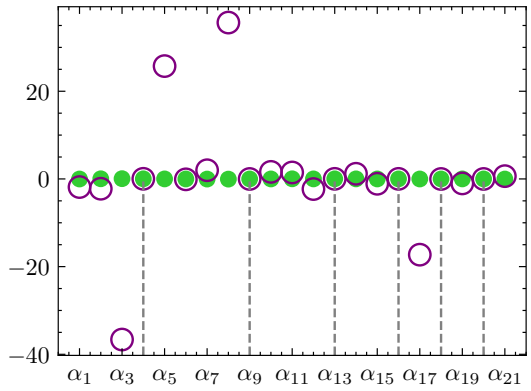


(e) Original and reconstructed field effect coefficients with 10% noise in the field effect measurements.

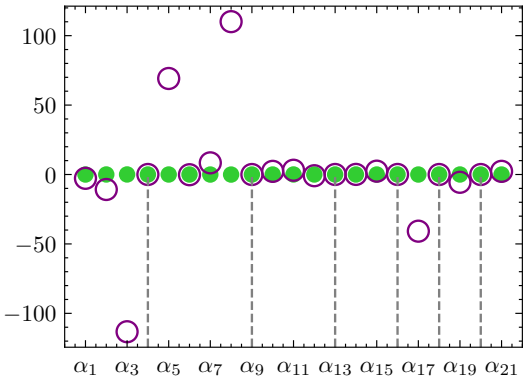


(f) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the field effect measurements.

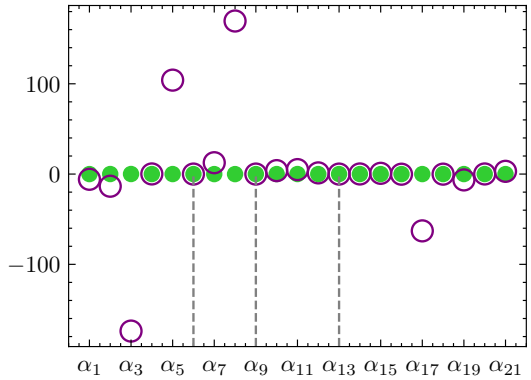
Figure 5.1: Least-squares estimation results for varying noise levels in the field effect measurements, using the one-step method. The left column displays the field effect coefficients $\boldsymbol{\beta}$, with filled blue circles representing the original coefficients and open orange circles showing the reconstructed values. The right column presents the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, where filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates.



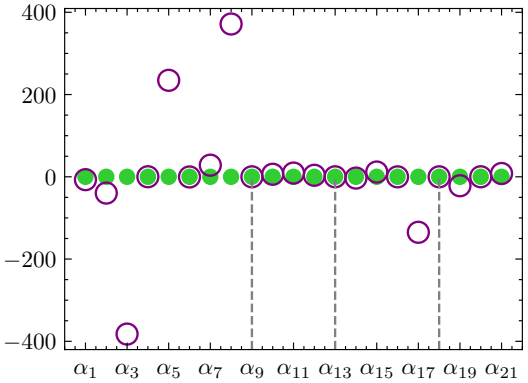
(a) Original and reconstructed environmental coefficients for Genotype 1 with 1% noise in the WSO measurements.



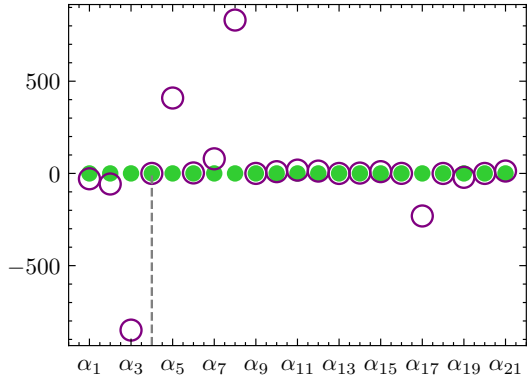
(b) Original and reconstructed environmental coefficients for Genotype 4 with 1% noise in the WSO measurements.



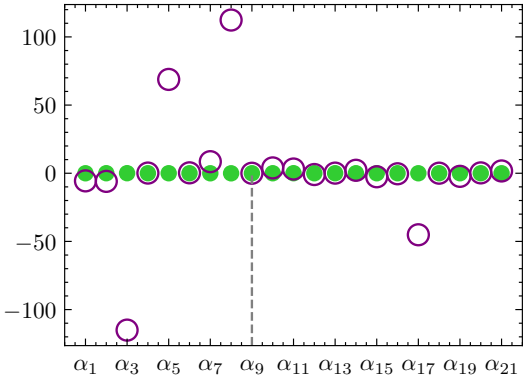
(c) Original and reconstructed environmental coefficients for Genotype 1 with 5% noise in the WSO measurements.



(d) Original and reconstructed environmental coefficients for Genotype 4 with 5% noise in the WSO measurements.



(e) Original and reconstructed environmental coefficients for Genotype 1 with 10% noise in the WSO measurements.



(f) Original and reconstructed environmental coefficients for Genotype 4 with 10% noise in the WSO measurements.

Figure 5.2: Least-squares estimation results under varying noise levels in the weight of storage organs measurements, using the one-step method. The right column shows the genotype-specific environmental coefficients $\boldsymbol{\alpha}$ for genotype 1, whereas the left column displays those for genotype 4. The filled green circles denote the original coefficients and open purple circles indicate the reconstructed estimates. The grey vertical lines are drawn from a coefficient to the x-axis whenever the reconstruction error is particularly small.

nificant costs, and intermediate harvests reduce the final yield. Therefore, our goal is to identify the minimal number of environmental measurement locations and intermediate harvests required.

We use the same algorithms described in Section 4.4 to uniformly select a number of measurement locations and intermediate harvests.

In the optimization problem addressed in Section 5.2, we select the rows corresponding to the chosen spatial points from the system described by Equation 5.2, since the left-hand side represents the WSO data obtained through intermediate harvesting. Similarly, we select the rows corresponding to the chosen time points in the system given by Equation 5.3, where the left-hand side corresponds to the field effect derived from measurements taken at different locations in the field.

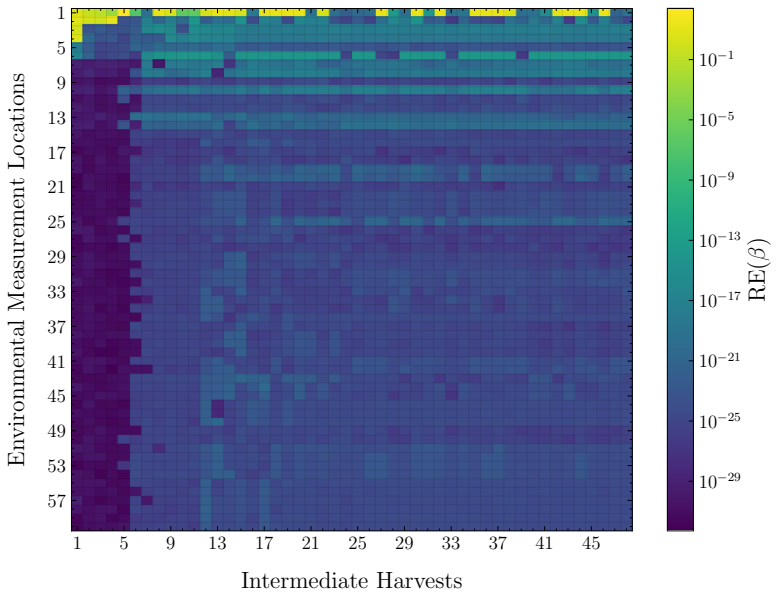
We create an algorithm that loops through all combinations of number of intermediate harvests (1 to 48) and number of environmental measurement locations (1 to 60). For each combination, the algorithm reconstructs both the field effect coefficients and the environmental coefficients using the one-step reconstruction method. After each reconstruction, we compute the relative error to assess the accuracy of the estimates. The results are visualized in Figure 5.3 with heat maps, illustrating how reconstruction accuracy varies across different data availability scenarios.

The heatmap for β exhibits significant fluctuations in relative error along the x-direction when only one environmental measurement location is used. This occurs because System 5.3 contains just a single row of data for reconstructing β, making the estimate heavily dependent on the selected rows in System 5.2. Since these rows vary with the choice of intermediate harvest timepoints, the relative error in β correspondingly fluctuates. However, with at least 2 measurement locations and 5 intermediate harvests, the relative error becomes consistently low. The lowest error occurs with at least 7 measurement locations and 5 or fewer intermediate harvests. In this case, the field effect coefficients are mainly driven by the environmental data, suggesting that the field effect system alone offers more reliable information for estimating the field effect coefficients than when combined with the WSO measurements.

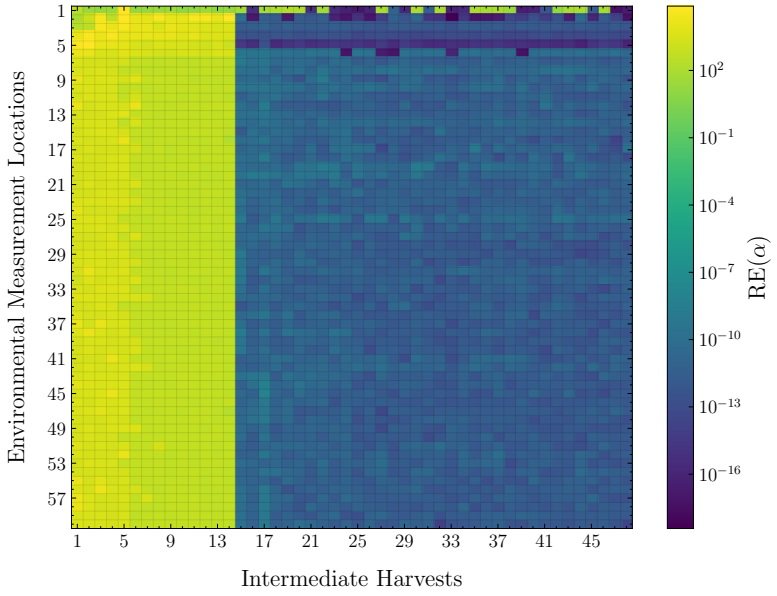
The heatmap for α also shows significant fluctuations in relative error along the x-direction when only one environmental measurement location is used. This can be explained by the fact that System 5.3 contains only a single row of data for reconstructing β, making its estimate heavily dependent on the selected rows in System 5.2. Since α is also determined by System 5.2, some information that would primarily support the reconstruction of α is now shared with the reconstruction of β. Because the rows used in System 5.2 vary with the choice of intermediate harvest timepoints, the relative error in α fluctuates correspondingly. This reflects the trade-off between estimating both α and β using almost exclusively information from a single system.

For accurate estimation of both α and the performance Y, at least 2 environmental measurement locations and 15 intermediate harvests are needed to achieve stable, low relative errors. Notably, the drop in relative error between 14 and 15 harvests is sharp, resulting in an error reduction on the order of 10^{14}.

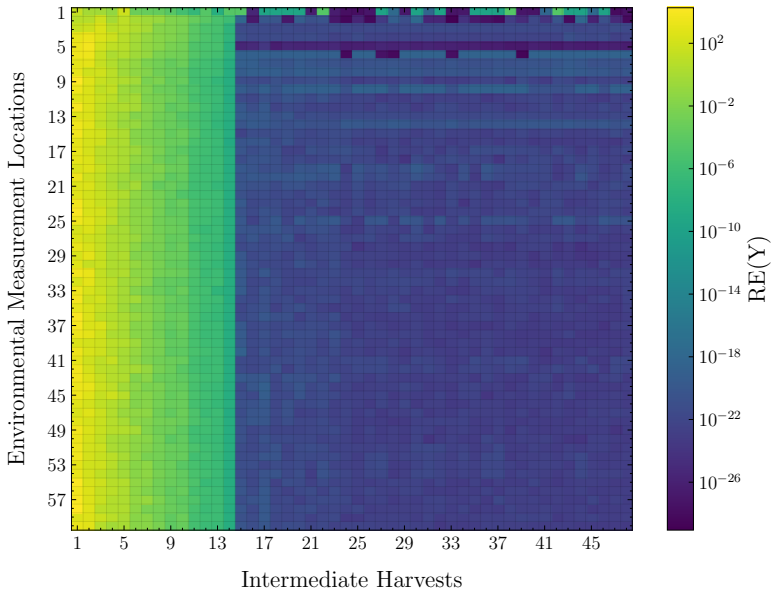
The optimization algorithm converges in just 7 iterations when using data from 2 measurement locations and 15 intermediate harvests. This low number of iterations further justifies for selecting this combination as the optimal.



(a) Heatmap of relative errors in the field effect coefficients $\boldsymbol{\beta}$.



(b) Heatmap of relative errors in the genotype-specific environmental coefficients $\boldsymbol{\alpha}$.



(c) Heatmap of relative error in performance $\boldsymbol{Y}$, i.e., the difference between measured and reconstructed WSO.

Figure 5.3: Heatmaps showing the relative error (on a logarithmic scale) for all combinations of the number of intermediate harvests (x-axis) and environmental measurement locations (y-axis) using the One-Step Method. Yellow indicates higher relative error; dark blue indicates lower error. The colorbar on the right maps colors to error levels.

6

Discussion

6.1. Key Findings

In this project, we presented a model that captures the relationship between two types of data: the weight of storage organs (WSO), obtained from intermediate harvests of selected potato plants, and environmental variables measured at specific locations in the field. These environmental factors include evaporation, temperature, temperature integral, soil moisture, and leaf area index. All environmental variables are time-dependent and uniform across the field, with the exception of soil moisture. Soil moisture is modeled as a combination of a time-dependent component and a spatio-temporal field effect, where the latter is represented as a sum of time and space basis functions, each scaled by corresponding field effect coefficients. The primary objective is to estimate the model coefficients that best describe the relationship between the basis functions and the field effect, as well as the coefficients that characterize the relationship between environmental conditions and WSO for a given genotype.

6.1.1. The Performance of the Reconstruction Methods

We evaluated two reconstruction approaches for estimating the models coefficients.

In the two-step method, we reconstructed the coefficients in two stages: first estimating field effect coefficients, then genotype-specific environmental coefficients. This was done using linear equations and solved analytically using the classic normal equation for least-squares problems.

In the one-step method, we estimated all coefficients simultaneously using a nonlinear formulation. This method applied an iterative optimization procedure to minimize the residuals between observed measurements and model predictions, using the `least_squares` optimizer from `SciPy`.

The relative errors for the genotype-specific environmental coefficients α, the field effect coefficients β and the Performance Y when evaluating these two models, are shown in Tabel 6.1. These results show that the two-step method performs better in reconstructing α and β, while the one-step method performs better in reconstructing Y.

Table 6.1: Relative errors (RE) for field effect coefficients (β), genotype-specific environmental coefficients (α), and weight of storage organs (Y) obtained using the two-step and one-step methods.

	$RE(\beta)$	$RE(\alpha)$	$RE(Y)$
Two-Step Method	2.57521×10^{-28}	3.81833×10^{-21}	8.9105×10^{-22}
One-Step Method	3.28411×10^{-25}	1.71000×10^{-14}	4.5548×10^{-24}

6.1.2. Effect of Measurement Noise

After testing the reconstruction methods in idealized situations, we have added noise to the field effect and WSO measurements respectively and re-evaluated the accuracy of the methods. For the two-step method, we have created a regularization method, namely the truncated singular value decomposition. The relative

errors of the two-step method, the regularized two-step method and the one step method are shown in Table 6.2 where noise is added to the field effect, and in Table 6.3 where noise is added to the WSO.

The relative error for β remains unchanged at 2.57521×10^{-28} when the (regularized) two-step method is used with noise added to the weight of storage organs. This is because, in thet two-step method, β is reconstructed independently of the WSO measurements. As a result, the reconstruction is consistently more accurate than when using the one-step method under the same noise conditions. In contrast, when noise is added to the field effect, both methods yield similar accuracy in reconstructing β.

When noise is added to the WSO, reconstruction of α is poor for both methods due to model overparameterization. Multiple parameter combinations can fit Y well, so noise in WSO weakens constraints on α, leading to large errors even when Y is accurately reconstructed through compensation by other parameters.

However, the accuracy of the two-step method is significantly improved by its regularized version, because truncated singular value decomposition (TSVD) filters out unstable components associated with small singular values, thereby reducing the amplification of noise in the reconstruction of α. When noise is added to the field effect, both methods yield similar accuracy in reconstructing α.

The performance Y is reconstructed with similar errors across all methods, with the One-Step Method performing slightly better than the two-step method when noise is added in the WSO.

Table 6.2: Relative error (RE) between measured and reconstructed values of field effect coefficients (β), genotype-specific environmental coefficients (α), and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the field effect (FE). Results are obtained using the one-step method, the two-step-method and the two-step method where the truncated singular value decomposition is employed for regularization (indicated with 'Reg.').

Method	**RE**(β)	**RE**(α)	**RE**(**Y**)
Noise level in FE = 1%			
Two-Step	4.01215×10^{-5}	8.36684×10^{-4}	5.14281×10^{-9}
Two-Step (Reg.)	1.60481×10^{-4}	6.21471×10^{-3}	1.85823×10^{-8}
One-Step	7.94674×10^{-2}	2.05676×10^{-2}	1.77601×10^{-8}
Noise level in FE = 5%			
Two-Step	3.98077×10^{-3}	2.21691×10^{-1}	1.84097×10^{-7}
Two-Step (Reg.)	5.44900×10^{-3}	1.08487×10^{-2}	1.04279×10^{-7}
One-Step	3.75340×10^{-3}	1.06416×10^{-2}	1.55257×10^{-7}
Noise level in FE = 10%			
Two-Step	8.69498×10^{-3}	8.18840×10^{-1}	1.19770×10^{-6}
Two-Step (Reg.)	2.49362×10^{-2}	1.43672×10^{-1}	3.10231×10^{-7}
One-Step	1.50907×10^{-2}	9.00153×10^{-1}	9.71626×10^{-7}

6.2. Minimizing Environmental Measurement Locations and Early Harvests

Since installing environmental measurement locations and performing intermediate harvests both incur costs, it is important to minimize them. We calculated the relative error for each combination of the number of intermediate harvests and measurement locations.

For the two-step method, the optimal combination was 7 measurement locations and 15 intermediate harvests. This combination results in the following relative errors:

$$\mathrm{RE}(\boldsymbol{\beta}) = 4.8614 \times 10^{-30}. \qquad \mathrm{RE}(\boldsymbol{\alpha}) = 1.6553 \times 10^{-20}, \qquad \mathrm{RE}(\boldsymbol{Y}) = 7.7479 \times 10^{-22}.$$

Table 6.3: Relative error (RE) between measured and reconstructed values of field effect coefficients (β), genotype-specific environmental coefficients (α), and the weight of storage organs (Y). Gaussian noise at varying levels is added to the measurements of the weight of storage organs (WSO). Results are obtained using the one-step method, the two-step-method and the two-step method where the truncated singular value decomposition is employed for regularization (indicated with 'Reg.').

Method	**RE**(β)	**RE**(α)	**RE**(**Y**)
Noise level in WSO = 1%			
Two-Step	2.57521×10^{-28}	5.68096×10^{5}	2.34260×10^{-4}
Two-Step (Reg.)	2.57521×10^{-28}	7.50615×10^{-1}	2.30128×10^{-4}
One-Step	2.79289×10^{-8}	5.27384×10^{5}	1.52550×10^{-5}
Noise level in WSO = 5%			
Two-Step	2.57521×10^{-28}	6.27501×10^{7}	5.56991×10^{-3}
Two-Step (Reg.)	2.57521×10^{-28}	8.53175×10^{-1}	5.92710×10^{-3}
One-Step	2.88038×10^{-6}	6.57592×10^{6}	2.07317×10^{-4}
Noise level in WSO = 10%			
Two-Step	2.57521×10^{-28}	4.40664×10^{7}	2.26304×10^{-2}
Two-Step (Reg.)	2.57521×10^{-28}	8.51549×10^{-1}	2.36792×10^{-2}
One-Step	5.01422×10^{-6}	1.19014×10^{8}	9.14598×10^{-4}

For the one-step method, the optimal combination was 2 measurement locations and 15 intermediate harvests.

$$\text{RE}(\boldsymbol{\beta}) = 2.3553e \times 10^{-15}. \qquad \text{RE}(\boldsymbol{\alpha}) = 1.6207 \times 10^{-10}, \qquad \text{RE}(\boldsymbol{Y}) = 1.2189 \times 10^{-18}.$$

These results show that the two-step method performs better in reconstructing all unknowns. However, the two-step method requires five more environmental measurement locations than the one-step method, and such locations are expensive to install and maintain. Therefore, the choice between the two-step and one-step methods depends on the trade-off between reconstruction accuracy and the cost of environmental measurement locations.

6.3. Limitations and Future Work

Several limitations must be acknowledged, along with clear opportunities for future improvement. Most notably, the current model does not capture the full complexity of factors influencing real-world potato growth. It considers only five environmental variables, while other important factors — such as hours of sunshine and nitrogen and oxygen levels — are excluded. Although interaction terms were incorporated it is possible that additional or alternative combinations of environmental factors could improve the model's realism. Furthermore, all results in this study are based on synthetic data, in which environmental variables are represented by smooth, idealized curves. To make the model applicable to real-world scenarios, it must be extended and refined to handle the variability and complexity of real environmental measurements.

6.3.1. Field Effect Basis Functions

In this report, we assumed that the spatio-temporal field effect could be expressed as a linear combination of a predefined set of basis functions. Importantly, the same basis functions were used both to generate synthetic data and to reconstruct the coefficients from that data. This is a strong assumption that creates a idealized setting for the reconstruction process.

The benefit of this controlled setup is that it allows us to evaluate whether the reconstruction method is capable, in principle, of recovering the true underlying coefficients when there is no model mismatch. However, it does not reflect the challenges encountered in real-world applications, where the true structure of the field effect is unknown and may not align perfectly with the chosen basis.

In future work, it will be essential to extend the approach to real data and explore how to select appropriate basis functions that adequately capture the field effect. Additionally, a key step forward will be testing the

reconstruction method in scenarios where it does not have access to the true generative structure.

6.3.2. The Role of Leaf Area Index

When generating the synthetic data, we treated w_5 (leaf area index, LAI) as an independent environmental factor. However, LAI is more complex as it not only correlates with WSO but is also influenced by genotype and environmental factors like temperature, evaporation, and soil moisture. Since soil moisture varies across the field, LAI indirectly captures spatial effects as well. This makes LAI both an outcome and a driver in the system - an interaction not fully accounted for in the current model.

6.3.3. Jacobian in SciPy's Least-Squares

In the one-step method, we evaluate reconstructions across different combinations of intermediate harvest counts and number of environmental measurements. When only one measurement location is used, the optimization process becomes significantly slower. Supplying the Jacobian to the `least_squares` optimizer from `SciPy`, could improve both speed and accuracy.

6.3.4. Regularization Technique for One-Step Method

As demonstrated in Section 5.3, the one-step method is not robust to noise, making the development of a regularization technique highly desirable. The method involves two types of systems: a linear field effect system and five nonlinear WSO systems. One possible approach is to apply the Truncated Singular Value Decomposition (TSVD) method, selecting a truncation threshold individually for each matrix. However, identifying an optimal strategy for choosing the two regularization parameters is not the primary focus of this thesis. Therefore, this topic could be explored further in future work.

7

Conclusion

This report has presented a modeling approach to quantify the relationship between environmental conditions and the weight of storage organs (WSO) in potato plants, with the aim of supporting genotype evaluation in multi-environment trials. Two estimation strategies were explored: a two-step method, which separately reconstructs field effect and genotype-specific environmental coefficients using linear least-squares, and a one-step method, which jointly estimates all coefficients via nonlinear optimization.

Under ideal, noise-free conditions, both methods accurately reconstructed the coefficients and performance values. However, the two-step method performs slightly better in reconstructing α and β, while the one-step method performs better in reconstructing Y.

The study also evaluated the robustness of both methods in the presence of measurement noise. When noise was added to the spatio-temporal field effect, the two-step method — particularly when regularized — consistently outperformed the one-step method in reconstructing both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The reconstruction of Y, however, was similar across all three methods.

When noise was introduced into the WSO measurements, both methods performed poorly in reconstructing $\boldsymbol{\alpha}$. However, applying truncated singular value decomposition (TSVD) to the two-step method significantly improved its robustness. Additionally, the reconstruction of $\boldsymbol{\beta}$ was exact in the two-step method, as noise in the WSO measurements does not affect the first step in this method. The reconstructions of Y were comparable across all three methods.

Furthermore, we examined how to minimize the number of costly intermediate harvests and environmental measurement locations. The one-step method proved more efficient, requiring only 2 measurement locations and 15 intermediate harvests to achieve very low reconstruction errors. The two-step method required more data — 7 locations and 15 harvests — but yielded higher accuracy. Choosing between the two methods thus involves a trade-off between cost and accuracy.

Despite helpfull results, the model has several limitations. It currently assumes a simplified relationship between environmental variables and crop performance, and synthetic data were used rather than real-world measurements. Moreover, we assumed that the spatio-temporal field effect could be represented as a linear combination of a predefined set of basis functions. This led to an inverse crime, as the same basis functions were used both to generate the synthetic data and to reconstruct the corresponding coefficients. Additionally, the role of the leaf area index (LAI) as both an driver and an outcome in the system is currently not fully reflected in the model. Moreover future work could focus on incorporating regularization into the one-step method.

In conclusion, the study demonstrates that mathematical modeling can offer powerful tools for understanding genotype-environment interactions in crop performance. With further refinement, the approaches developed here may support more efficient and sustainable plant breeding practices.

Bibliography

[1] Tarekegn Argaw, Berhanu Amsalu Fenta, Habtemariam Zegeye, Girum Azmach, and Assefa Funga. Multi-environment trials data analysis: Linear mixed model-based approaches using spatial and factor analytic models. *Euphytica*, 219(7):109, 2023. doi: 10.1007/s10681-023-03299-1.

[2] Abraham Blum. *Plant Breeding for Stress Environments*. CRC Press, 1st edition, 1988. ISBN 9781351075718. doi: 10.1201/9781351075718. eBook published January 18, 2018.

[3] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2nd edition, 2008.

[4] F. Cormier, R. Messmer, A. Pask, J. Enjalbert, M. P. Boer, and L. Moreau. Modeling genotype-by-environment interaction in plant breeding: Methods and applications in wheat and maize. *Crop Science*, 59(4):1152–1163, 2019. doi: 10.2135/cropsci2018.10.0646.

[5] CropXR. Potato, 2024. URL `https://cropxr.org/potato/`. Accessed: 2025-06-30.

[6] International Potato Center. Potato facts and figures, 2023. URL `https://cipotato.org/potato/potato-facts-and-figures/`. Accessed: 2025-06-30.

[7] E. Kroc and O. L. Olvera Astivia. The case for the curve: Parametric regression with second- and third-order polynomial functions of predictors should be routine. *Psychological Methods*, 2023. doi: 10.1037/met0000629. URL `https://doi.org/10.1037/met0000629`. Advance online publication.

[8] B. Parent, F. Tardieu, and C. Welcker. Thermal time and plant development: Integrating temperature effects through mathematical models. *Plant Physiology*, 189(4):1451–1467, 2022. doi: 10.1104/pp.21.01432. URL `https://doi.org/10.1104/pp.21.01432`.

[9] Mara Xosé Rodríguez-Álvarez, Martin P. Boer, Fred A. van Eeuwijk, and Paul H. C. Eilers. Spatial models for field trials. arXiv preprint arXiv:1607.08255, July 2016. URL `https://arxiv.org/abs/1607.08255`.

[10] A. B. Smith, B. R. Cullis, and R. Thompson. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics*, 57(4):1138–1147, 2001. doi: 10.1111/j.0006-341x.2001.01138.x.

[11] A. B. Smith, B. R. Cullis, and R. Thompson. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *The Journal of Agricultural Science*, 143(6):449–462, 2005. doi: 10.1017/S0021859605005587.

[12] Daniel Wallach, David Makowski, James W. Jones, and François Brun. *Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment*. Academic Press, San Diego, 2 edition, 2014. ISBN 978-0-12-397008-4.