ABSTRACT

ALLIE, NORMAN. Sequential Split Plot Designs. (Under the direction of Jonathan Stallrich).

Split-Plot experiments are an important segment of statistical methods within the field of designed experiments. They are typically analyzed at the end of the experiment, after Whole-Plot and Split-Plot treatments have been applied. This paper presents a method for analyzing Split-Plot experiments with data after Whole-Plot treatments are applied and after Split-Plot treatments are applied. This method has advantages over established methods in certain situations. We review the basic mathematical basis for Split-Plot experiments, and the various established methods used to analyze these experiments. The new method is then introduced and compared to established methods. An omnibus test to verify assumptions associated with the method is described. The paper concludes with simulations demonstrating the capability of the method under various conditions.

Sequential Split Plot Designs

by
Norman Allie

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Statistics

Raleigh, North Carolina
2025

APPROVED BY:

Jason Osborne	Emily Griffith

Jonathan Stallrich
Chair of Advisory Committee

BIOGRAPHY

Norman is a product design and development leader with expertise in Design for Six Sigma (DFSS), Design for Reliability, Systems Engineering, and Industrial Statistics. He holds a BS in Mechanical Engineering from Kettering University, an MS in Electrical Engineering from Rensselaer Polytechnic Institute and an MS in Statistics from NC State University in December 2025. Norman has led cross-functional teams in system design and development across the medical device and commercial industries.

His career includes leadership roles at Pfizer, Medtronic, Becton Dickinson, Baxter Healthcare and Pitney Bowes. He enjoys spending time with his family, weight training, tennis, reading and the financial markets.

ACKNOWLEDGEMENTS

I wanted to thank all of my professors at NC State University for their time during my time as a student, especially Dr. Jon Stallrich for his guidance and coaching through the process of writing this document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

SEQUENTIAL SPLIT PLOT DESIGNS

1.1 Introduction

Split plot experiments are characterized by having two levels of experimental units with separate factors and randomization schemes for each level. The first level is comprised of the whole plot units, or simply whole plots (WPs), which are then further partitioned into smaller split plot units, or split plots (SPs). Therefore, split plots are nested within whole plots and a split plot design can be viewed as one experiment embedded within another experiment. The first experiment involves assigning and applying levels of one or more factors to the WPs. We call such factors the whole plot factors. The second experiment assigns and applies another set of one or more factors to the SPs; these are called the split plot factors. Given that SPs are nested within WPs, the randomization for this experiment follows that of a block design in which the whole plots act as the blocks.

Although initially motivated for agricultural experiments, there have been many successful applications of split plot experiments across multiple disciplines. Box and Jones (1992) describe an experiment where a packaged food manufacturer wanted to develop an optimal formulation of a cake mix. Bisgaard (2000) discusses the use of split-plot experiments in optimizing parameter settings of a manufacturing process. For example, he describes an experiment involving four prototype combustion engine designs and two grades of gasoline. Bingham and Sitter (2001) discuss the effect of various factors on the swelling of a wood product after it has been saturated with water and allowed to dry.

The common practice of implementing a split plot experiment is to apply all WP and SP factors before collecting the response data from the SPs. However, if the WP factors can be applied prior to application of the SP factors, like in some sequential process, it is feasible to collect responses between applying the WP and SP factors. This thesis, to my knowledge, is the first to consider such a data structure and recommend an analysis method to leverage the additional data. Given the implied sequential nature of the experiment and

data collection, I will refer to the proposed analysis as the Sequential Split Plot Design Method (SSPDM).

1.1.1 The PLA Experiment

The motivating example for this thesis comes from Gajjar et al. (2021), who performed a study on process-property relationships for melt-spun poly(lactic acid), or PLA, yarn. This study focused on the effect of the melting process parameters on the mechanical and physio chemical properties of the resulting PLA yarn and the derived property relationships of the process. PLA manufacturing consists of a two-stage process: (a) Stage 1: Melt Extrusion and (b) Stage 2: Drawing. The experiment consisted of five two-level factors, three WP factors applied in Stage 1 (Maximum Extruder Temperature (x_1), Pump Throughput (x_2), and Take-up Speed (x_3)), and two SP factors applied in Stage 2 (Draw Temperature (z_1) and Draw Ratio (z_2). The Stage 2 process involved heating yarn to the desired drawing temperature and then stretching according to some draw ratio. The lowest value of draw ratio is 1, which means the yarn is not stretched. Higher values mean the yarn is stretched more. A draw ratio of 2 means the machine attempts to double the length of the yarn.

It is possible for the yarn to break if the draw ratio is set too high and the maximum draw ratio one could use depended on the settings of the other four factors. To standardize the analysis, the experiments identified a high value of draw ratio, denoted $R_{\max}$, that minimized the chance of breakage. For example, $R_{\max}$ for WP label 3 with $z_1 = -1$ was 2.5 and with $z_1 = +1$ it was 3.7. The low value was then set to $R_{\min} = (R_{\max} + 1)/2$, the midpoint between $R_{\max}$ and 1. The experimenters then recoded these draw ratio values as -1 for the low value and $+1$ for the high value, which is common for two-level factorial experiments. The experimenters collected data at the end of Stage 1 and the end of Stage 2, but the results in

Gajjar et al. (2021) only used the Stage 2 data. Both data sets are shown in Table 1.1 with the response variable Denier which is essentially a measure of the durability of the thread.

There were some interesting challenges with the PLA data that were not originally discussed in Gajjar et al. (2021), which are due to the complicated process of finding $R_{\max}$ for all 2^4 settings. There are four cases (WP labels 1, 5, 7, and 8) in which the WP Response exactly equals the SP response where $z_1, z_2 = -1$. For all these cases, the experimenters set $R_{\min} = 1$, meaning the yarn was not drawn at all. This is because the $R_{\max}$ values were either also 1 (which happened for WP label 5) or were just barely above 1. Apparently, the experimenters decided to not actually perform the SP runs for these settings and instead simply imputed the WP responses, since one could view the WP responses as not drawing the yarn at all. In the case of WP label 5 with $z_1 = -1$, the experimenters determined $R_{\max} = 1$ and so they could not vary z_2 anymore. This produced one less SP response than all the other WP labels. This made the experiment unbalanced because not all 32 possible combinations of the five factors were performed. For most of the paper, we ignore these challenges as they are particular to this specific application, but they will become important when we apply the SSPDM to the data.

1.1.2 Contributions and Overview

The first contribution of this thesis is a simple two-stage analysis that uses both WP and SP responses. The basic concept of the method is to break up a linear mixed effects model for a Split-Plot experiment (S Model) into two linear models without random effects. The first model is a linear model of the response after only the WP treatments have been applied (W Model). The second model is a linear model of the response after SP treatments have been applied. In this second model, the WP effects are removed (S-W Model). When the

Table 1.1: Data from the PLA experiment. The WP factors, Maximum Extruder Temperature, Pump Throughput, and Take-up Speed, are denoted by x_1, x_2, and x_3, respectively. The SP factors, Draw Temperature and Draw Ratio, are denoted by z_1 and z_2, respectively. The runs are arranged in groups of the WP labels. The symbols $-$ and $+$ denote the low and high levels of each factors, respectively. The responses are the Denier values.

WP Label	x_1	x_2	x_3	z_1	z_2	WP Response	SP response
1	$-$	$-$	$-$	$-$	$-$	445.5	445.5
1	$-$	$-$	$-$	$-$	$+$		437.1
1	$-$	$-$	$-$	$+$	$-$		198.0
1	$-$	$-$	$-$	$+$	$+$		151.8
2	$-$	$-$	$+$	$-$	$-$	145.8	147.0
2	$-$	$-$	$+$	$-$	$+$		88.5
2	$-$	$-$	$+$	$+$	$-$		90.0
2	$-$	$-$	$+$	$+$	$+$		74.2
3	$-$	$+$	$-$	$-$	$-$	914.7	890.0
3	$-$	$+$	$-$	$-$	$+$		864.9
3	$-$	$+$	$-$	$+$	$-$		317.2
3	$-$	$+$	$-$	$+$	$+$		249.3
4	$-$	$+$	$+$	$-$	$-$	302.1	299.2
4	$-$	$+$	$+$	$-$	$+$		292.9
4	$-$	$+$	$+$	$+$	$-$		164.2
4	$-$	$+$	$+$	$+$	$+$		125.5
5	$+$	$-$	$-$	$-$	$-$	434.2	434.2
5	$+$	$-$	$-$	$+$	$-$		231.0
5	$+$	$-$	$-$	$+$	$+$		113.4
6	$+$	$-$	$+$	$-$	$-$	145.8	144.9
6	$+$	$-$	$+$	$-$	$+$		67.5
6	$+$	$-$	$+$	$+$	$-$		73.3
6	$+$	$-$	$+$	$+$	$+$		55.8
7	$+$	$+$	$-$	$-$	$-$	882.6	882.6
7	$+$	$+$	$-$	$-$	$+$		891.3
7	$+$	$+$	$-$	$+$	$-$		334.5
7	$+$	$+$	$-$	$+$	$+$		207.9
8	$+$	$+$	$+$	$-$	$-$	277.2	277.2
8	$+$	$+$	$+$	$-$	$+$		274.8
8	$+$	$+$	$+$	$+$	$-$		168.6
8	$+$	$+$	$+$	$+$	$+$		120.3

model assumptions hold, I show that estimators under the SSPDM have better statistical properties than the traditional approach which analyzes the SP responses only.

The assumptions associated with SSPDM are somewhat strict, so the second contribution of the thesis is a test of model adequacy that we recommend before reporting the results from the SSPDM. The test is a comparison of the second linear model (S-W Model) to the full linear mixed effects model (S Model) using ANOVA, where the S-W Model is the Null Model. If the test fails, we recommend only analyzing the SP responses and to investigate why the two models are different.

This thesis is structured as follows. Section 1.2.1 provides the necessary background on existing methods for analyzing balanced and unbalanced split plot designs. In particular, I focus on completely randomized split plot designs (CRSPDs) in this section and througout the thesis to simplify its presentation. The approach can technically be applied to many different split plot experiment structures. Section 1.3 describes the SSPDM and Model Adequacy Test. Section 1.4 contains the specification and results from simulation studies conducted under various conditions. Section 1.5 applies SSPDM and the Model Adequacy test to the PLA experiment. I conclude the thesis in Section 1.6 with a short summary of the thesis and possible areas of future work.

1.2 Background

1.2.1 Completely Randomized Split Plot Designs

In a CRSPD, n WPs are randomly assigned to levels of w WP factors. Each WP is then divided into m SPs. The SPs are then randomly assigned to levels of the s SP factors. The randomization is done separately for each WP, meaning the WPs should be treated as random blocks for the SP level of the experiment. To demonstrate, consider an experiment

to study the performance (e.g., hydrophobicity) of aftermarket automotive paint coatings based on three brands of ceramic coatings applied using two different application methods (hand versus spray). There are $3 \times 2 = 6$ possible treatment combinations but suppose we only have six cars available for the experiment. If we applied one coating and one application to each car, we would only replicate each treatment combination one time, which would make the statistical analysis challenging if not impossible as there would be no degrees of freedom for estimating variance. The only option would be to assume two factor interaction effects were negligible, which is a strong assumption.

To create more experimental units, we could break up each car into three sections: hood, roof, and trunk. While changing the coating type for different sections is easy, changing application method requires additional setup time, equipment changes, and possibly technician retraining. A CRSPD would then first assign the two application methods to the cars completely at random, ensuring equal replication of each application level. This would make the cars the $n = 6$ WPs and the $w = 1$ WP factor is application method. Next, the three brands are assigned at random to the sections within each car. This experiment resembles a randomized complete block design with the blocks comprised of the cars. The sections are the $m = 3$ SPs per WP and the $s = 1$ SP factor is coating brand.

A potential random assignment is shown in Table 1.2. The randomization ensures that each Brand appears once for every car. This equal replication is not guaranteed for each section of the car. Note that Hood has Brands 1, 2, and 3 replicated 3, 2, and 1 times, respectively. If we were concerned about variation between the sections of the car, we would need to randomize the experiment differently. However, this would no longer be considered a CRSPD which falls outside the scope of the thesis.

We say a CRSPD is balanced when

1. Every combination of the WP factors is assigned to the same number of WPs;

Table 1.2: Application Method and Sections for Cars

Car	Application	Hood	Roof	Trunk
1	Hand	Brand 1	Brand 3	Brand 2
2	Spray	Brand 1	Brand 3	Brand 2
3	Spray	Brand 3	Brand 1	Brand 2
4	Hand	Brand 1	Brand 2	Brand 3
5	Spray	Brand 3	Brand 2	Brand 1
6	Hand	Brand 2	Brand 3	Brand 1

2. Every combination of the SP factors is assigned to the same number of SPs for every WP.

A balanced CRSPD ensures that combinations of all factors (both WP and SP) appear the same number of times. Otherwise, the design is said to be unbalanced. The PLA experiment mentioned earlier is an example of an unbalanced design. While it has equal replication of the WP factors to the WPs, WP label 5 only has three of the possible four combinations of the SP factors. Balanced designs are often taught in textbooks such as Montgomery (2020) because their analysis is relatively straightforward. The analysis of unbalanced designs can be much more complicated.

1.2.2 Analysis of CRSPD

In this section and the remainder of the paper, we focus on the analysis of CRSPDs involving only two-level WP and SP factors. Such designs are the most commonly used in practice and in the design literature. Traditionally, CRSPDs only collect and analyze the SP responses. Let Y_{ij}^S denote the SP response collected from the j-th SP contained within the i-th WP, so $i = 1,\ldots,n$ and $j = 1,\ldots,m$. Next, let $x_{ip} = \pm 1$ denote the level of the p-th WP factor assigned to the i-th WP (and so is also assigned to all SPs within that WP). Similarly, let $z_{ijr} = \pm 1$ denote the level of the r-th SP factor assigned to the j-th SP in the i-th WP. If

three-way and higher-order interactions are negligible, the statistical model for a CRSPD is then

$$Y_{ij} = \beta_0 + \sum_{p=1}^{w} \beta_p x_{ip} + \sum_{p=1}^{w-1} \sum_{p'=p+1}^{w} \beta_{pp'} x_{ip} x_{ip'} + e_i^W + \sum_{r=1}^{s} \gamma_r z_{ijr} + \sum_{r=1}^{s-1} \sum_{r'=r+1}^{s} \gamma_{rr'} z_{ijr} z_{ijr'} + \sum_{p=1}^{w} \sum_{r=1}^{s} \gamma'_{pr} x_{ip} z_{ijr} + e_{ij}^S, \tag{1.1}$$

where $e_i^W \overset{\text{iid}}{\sim} N(0, \sigma_W^2)$ represents the WP experimental error and $e_i^S \overset{\text{iid}}{\sim} N(0, \sigma_S^2)$ represents the SP experimental error. The model also assumes that all e_i^W and e_{ij}^S are mutually independent. Because we take repeated measurements from each WP, the SP responses from model (1.1) that come from the same whole plot will be correlated: $\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_W^2$.

Model (1.1) can be written in matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{e} \ , \tag{1.2}$$

where $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Note that the model matrix $\boldsymbol{Z}$ contains all terms involving at least one split plot factor, including the interaction terms, $x_{ip} z_{ijr}$. Note that $\boldsymbol{Z}$ is not to be confused with the mixed linear effects model that uses Z as the model matrix for random effects. After sorting the responses in $\boldsymbol{y}$ by their whole plot labels, the covariance matrix, $\boldsymbol{\Sigma}$, is block diagonal where the covariance matrix for the m responses coming from WP i is

$$\mathbf{V}_i = \sigma_W^2 \mathbf{J}_m + \sigma_S^2 \mathbf{I}_m \ , \tag{1.3}$$

where $\boldsymbol{J}_m$ is an $m \times m$ matrix of all 1's. The least-squares estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are given

by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{Z} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{y} \end{pmatrix}. \tag{1.4}$$

A balanced CRSPD makes all columns within and between $\boldsymbol{X}$ and $\boldsymbol{Z}$ mutually orthogonal. This simplifies the expression for these estimators to:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \frac{1}{mn}\boldsymbol{X}^T\boldsymbol{y} \tag{1.5}$$

$$\hat{\boldsymbol{\gamma}} = (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{y} = \frac{1}{mn}\boldsymbol{Z}^T\boldsymbol{y}\,. \tag{1.6}$$

There are unbalanced designs that still produce orthogonal estimators, so long as a subset of the factorial effects are to be estimated. Such designs are sometimes referred to as orthogonal designs and include regular/nonregular fractional factorial designs.

The split plot model's covariance structure leads to the following variances for (1.5) and (1.6):

$$\text{Var}(\hat{\beta}_p) = \frac{1}{n}\sigma_W^2 + \frac{1}{mn}\sigma_S^2 = \frac{1}{mn}(m\sigma_W^2 + \sigma_S^2) \tag{1.7}$$

$$\text{Var}(\hat{\gamma}_r) = \frac{1}{mn}\sigma_S^2. \tag{1.8}$$

Given that SPs are nested with WPs, it is often assumed that $\sigma_S^2 < \sigma_W^2$, which causes $\text{Var}(\hat{\beta}_p)$ to sometimes be larger than $\text{Var}(\hat{\gamma}_r)$. However, there are examples where this may not hold. For example, the drawing process of the PLA experiment is a more stressful process than the Stage 1 process because the yarn is heated and stretched. As we will see, this can cause issues when estimating the two variance components and when performing inferences on the model effects.

The analysis of split plot designs is more involved than ordinary least squares. The

difficulty rises for unbalanced data because the inference requires reliable estimation of both σ_W^2 and σ_S^2. We list here the most popular frequentist analysis methods that have been used to analyze linear mixed effects models, including split plot designs, that are covered in Næs et al. (2007).

Ordinary Least Squares (OLS)

The first approach is to ignore the split plot structure and to analyze the experiment as if it were a completely randomized factorial experiment with $w + s$ factors. The Ordinary Least Squares (OLS) estimator (1.12) is used to estimate fixed effects and its estimated covariance matrix is $(\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma}^2$ where $\hat{\sigma}^2 = \text{MSE}$ calculated using the traditional residual sum of squares. Letsinger et al. (1996)argued that if $\frac{\sigma_W^2}{\sigma_S^2} > 1$, the OLS approach should not be used. Their argument is based on comparing the variances of the two classes of estimators as well as the expected values of the two mean square error terms.

Half-Normal Plots

For balanced or orthogonal CRSPDs, it is possible to identify significant effects graphically by plotting the absolute values of the least-squares estimates with a Half Normal Q–Q plot. The idea is that under the null hypothesis that all effects are negligible (ignoring the intercept) the estimates may be considered as independent draws from a Normal distribution with mean 0. Equations (1.7) and (1.8) show that the whole plot factorial effect estimators have equal variance and the split plot factorial effect estimators have a different variance. Therefore, the whole plot factorial effect estimators are independent and identically distributed under the null hypothesis. Taking the absolute value of these effect estimates (the sign of the effects is not important for statistical significance if $H_0 : \beta = 0$) these estimates are independent draws from a half-normal distribution and so their effects can be compared to the quantiles of a theoretical half-normal distribution. The same is

true with the group of the split plot level effects. If the null hypothesis is true, the plotted effects should follow a straight line. Otherwise, we expect any nonzero effects to appear as “outliers" from the straight line. A point that appears as an outlier enables us to justify declaring the observed effect as significant. It is not recommended to use this method for nonorthogonal designs as they would have estimators that violate the independent and identically distributed assumptions.

This method is fairly subjective because one has to identify where in the plot the “outliers" appear. This is most easily seen when at least half of the plotted effects are insignificant and so produce a good reference for what the straight line should look like. Figure 1.1 gives an example of a nearly ambiguous plot in which it may be difficult to tell which effects are important. It appears that effects labeled E4 to E9 follow a straight line, while the remaining three effects in the top right look like potential outliers.

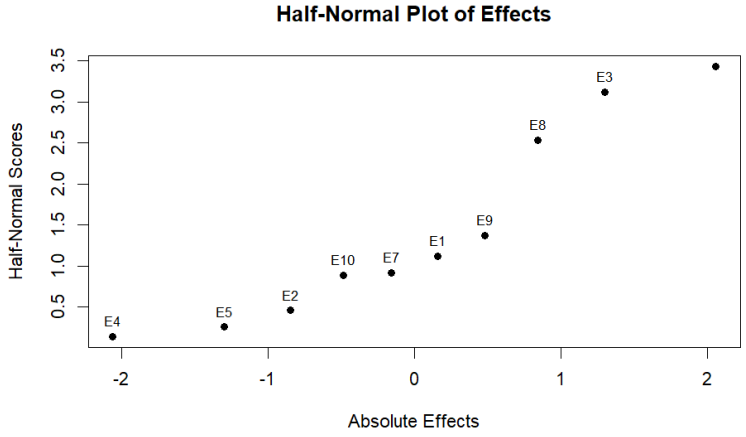

Figure 1.1: Half-Normal QQ plot for detecting active effects from an orthogonal design.

The Lenth Method

The Lenth method is an extension to the Q-Q plot approach. The goal is to generate an estimate of the effect estimator's variance using the principle of effect sparsity. The Lenth plot is similar to the analysis of means plot with additional limits provided. The estimates that fall within the margin of error (ME) limits are unlikely to be significant; the estimates that fall within the simultaneous margin of error (SME) limits but outside the margin of error limits (ME) are possibly significant, and those that fall outside the SME limits are likely significant (Lawson (2015)).

Analysis of Variance (ANOVA) using Expected Mean Squares

To perform traditional inference using T- and F-statistics, we need unbiased estimators for (1.7) and (1.8). For a balanced CRSPD, these follow easily from the whole plot and split plot mean square errors:

$$\mathrm{E}(\mathrm{MSE}_W) = m\sigma_W^2 + \sigma_S^2 \tag{1.9}$$

$$\mathrm{E}(\mathrm{MSE}_S) = \sigma_S^2 \tag{1.10}$$

The degrees of freedom for MSE_W, denoted df_W, equals $n-2^w$. The degrees of freedom from MSE_S will then be $\mathrm{df}_S = \mathrm{df}_W \times (2^s-1)$. By having smaller variances and larger degrees of freedom, the analysis of the split plot level effects is expected to be more powerful than that of the whole plot level effects.

Although unnecessary, an unbiased estimator for σ_W^2 can also be derived using the method of moments:

$$\hat{\sigma}_W^2 = \frac{\mathrm{MSE}_W - \mathrm{MSE}_S}{m}\,. \tag{1.11}$$

However, it is possible for $\text{MSE}_W < \text{MSE}_S$, leading to a negative estimate. This can occur with relatively high probability when $\sigma_W^2 < \sigma_S^2$. Fortunately, this will not cause issues with the inference for the whole plot effects.

For unbalanced data, one may still calculate expressions for the estimator variances and the expected values of the expected mean squares. However, unlike with balanced data where both $\text{Var}(\hat{\beta}_p)$ and $\text{E}(\text{MSE}_W)$ shared the term $m\sigma_W^2 + \sigma_S^2$, there may be a disconnect between these two expressions. Intead, some linear combination of $\hat{\sigma}_W^2$ and $\hat{\sigma}_S^2$ is needed to estimate $\text{Var}(\hat{\beta}_p)$. Practitioners should then be skeptical of the analysis results for the whole plot effects because (1) an approximation such as Satterthwaite is needed for the degrees of freedom and (2) the estimator for σ_W^2 may be negative.

Restricted Maximum Likelihood

REML involves maximizing a likelihood function that is constructed to be independent of the fixed effects. Unlike traditional maximum likelihood estimation, which estimates both the fixed effects and variance components, REML adjusts its variance estimates for the loss of degrees of freedom that come with estimating the fixed effects. This adjustment produces unbiased estimators of the variance components.

REML is an iterative procedure that usually starts with calculating the OLS estimators of the fixed effects (which are unbiased but not the best linear unbiased estimators). The residuals are then calculated and used to estimate the unknown variance components. This produces an estimate $\hat{\boldsymbol{\Sigma}}$ which for a balanced CRSPD will be block diagonal with $\hat{\boldsymbol{V}}_i = \hat{\sigma}_W^2 \boldsymbol{J}_m + \hat{\sigma}_S^2 \boldsymbol{I}_m$. The fixed effect estimators are then updated using the Generalized Least Squares (GLS) estimator

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\tau}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y} \end{pmatrix}. \tag{1.12}$$

Note that the GLS estimator is also unbiased for any positive definite $\hat{\boldsymbol{\Sigma}}^{-1}$ and will be the best linear unbiased estimator when $\hat{\boldsymbol{\Sigma}}^{-1}$ employs the true variance components. REML then calculates new residuals based on this GLS estimator to update the variance components. The process continues until convergence.

For balanced CRSPDs, REML and ANOVA will give the same results for both the fixed effects and variance components. The two methods will differ for unbalanced data. In these situations, the REML method is often preferred because it provides unbiased estimates of variance components by accounting for the structure and imbalance in the data. Similar to ANOVA, it is possible for the REML method to produce negative estimators for the variance components unless a lower bound is set when optimizing for the variance components. This will lead to REML simply setting the negative variance components equal to 0.

Another disadvantage of REML compared to ANOVA is that it is not as straightforward to perform inference because the sampling distribution of the estimators, which are found numerically, is less tractable. The method will produce estimates of fixed effects and their estimated standard errors so the common approach is to take the ratio of the estimates and standard errors and assume it follows a T distribution. The challenge though is with determining the corresponding denominator degrees of freedom.

Assuming that a test statistic approximately follow an F-distribution (recall that a squared T-distribution is an F-distribution), several methods for computing denominator degrees of freedom (or otherwise modifying the test statistics) have been proposed and implemented in commercial statistical software. Fai and Cornelius (1996) proposed a method for multi-degree-of-freedom tests in unbalanced split plot designs. More details of this method can be found in Schaalje et al. (2001). Kenward and Roger (1997) proposed a method for tests in mixed linear effects models based on any covariance structure.

Let $\boldsymbol{\theta}$ denote all fixed effects in a linear mixed effects model and we want to perform

simultaneous inference on the estimable functions $\boldsymbol{C\theta}$. The KR test statistic is given by

$$F^* = \delta F_{KR} = \delta \frac{(\boldsymbol{C}\hat{\boldsymbol{\theta}})'(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}^*\boldsymbol{C})^{-1}(\boldsymbol{C}\hat{\boldsymbol{\theta}})}{q},$$

where

F^***:** Adjusted (Kenward-Roger) F-statistic;

δ**:** Scale factor to adjust the F-statistic;

F_{KR}**:** Adjusted Wald-Type F-statistic;

$\hat{\Sigma}^*$**:** Adjusted covariance matrix of $\hat{\boldsymbol{\theta}}$ (Kenward-Roger adjustment);

q**:** Rank of $\boldsymbol{C}$.

Kenward and Roger use a second order Taylor series expansion of $(\boldsymbol{C}\hat{\boldsymbol{\Sigma}}^*\boldsymbol{C})^{-1}$ about the REML estimates of the variance components, and conditional expectation relationships to yield $\tilde{\mathrm{E}}(F_{KR})$ and $\widetilde{\mathrm{Var}}(F_{KR})$. The moments of F^* are then generated and equated to the moments of an F distribution to solve for δ and u,

$$u = 4 + \frac{q+2}{q\gamma - 1} \qquad \text{and} \qquad \delta = \frac{u}{\tilde{\mathrm{E}}(F_{KR})(u-2)},$$

where u is the denominator degrees of freedom and

$$\gamma = \frac{\widetilde{\mathrm{Var}}(F_{KR})}{2\,\tilde{\mathrm{E}}(F_{KR})^2}.$$

Schaalje et al. (2001) claim that the KR Method produces inflated Type I error rates for cases where the covariance structure is complex and the sample size is small. In addition, Alnosaier (2023) showed that the KR Method was unstable for some data sets with negative

variance estimates. Alnosaier also found that constraining the variance estimates to zero produced stable results. However, for split plot designs, we would never assume that σ^2_W would equal zero since this implies there is no variation between the whole plots. Therefore, for split plot designs with a small number of whole plots, the KR method may produce poor results.

The relative performance of the methods discussed above depend on the underlying structure of the experimental design. For balanced CRSPDs, the OLS and GLS estimators can be shown to be equivalent. In some cases, $\hat{\sigma}^2_W$ may be zero or even negative, but we can estimate the variance of the estimators without estimating the individual variance components. However, for all other cases and for hypothesis testing, knowledge of the error variances is essential (Goos and Vandebroek 2001).

In addition to these areas of concern, the likelihood based methods for fitting linear mixed effects models and computing test statistics for fixed effects do not generate correct denominator degrees of freedom (ddf) values automatically. In addition, for complex mixed effects models involving complicated covariance structures and/or unbalanced data sets, the exact distributions of the test statistics are unknown. The F-test demands correctly specifying the numerator and denominator degrees of freedom if we are to trust the result. These problems have been the subject of research and solutions have been integrated into statistical algorithms in software such as SAS.

In addition to all the issues described above, there are also problems related to replication. The above methods work well with experiments that contain replicates, but do not work for situations like the PLA experiment under consideration in this document, where there are no WP replicates. For unreplicated split plot experiments, $n-2^w=0$ so there are no degrees of freedom available for the estimation of WP variance other than those from higher order interactions that are not included in the model. This is especially problematic

for the PLA experiment because $w = 3$ so there is only one three factor interaction degree of freedom that we might consider removing from the model.

1.2.3 Split-Plot Design Analysis of PLA Experiment

In this section, we demonstrate the different analysis methods discussed in Section 1.2.2 on the PLA data. The model is given by

$$\begin{aligned} Y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} + \\ + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \gamma_{12} z_{1j} z_{2j} + \gamma'_{11} x_1 z_{1j} + \gamma'_{12} x_1 z_{2j} + \gamma'_{21} x_2 z_{1j} \\ + \gamma'_{22} x_2 z_{2j} + \gamma'_{31} x_3 z_{1j} + \gamma'_{32} x_3 z_{2j} + e_i^W + e_{ij}^S, \end{aligned} \tag{1.13}$$

Note that the QQ-plot and Lenth method are not strictly valid here because the effects are not mutually orthogonal, but we use these methods to illustrate the advantages of the sequential method. All analyses were performed in R Studio; for clarity, I have included the code and output produced by R in this document.

In the following code, the data is denoted as `Ys_data` and the WP factors are denoted as `WP`. Factors x_1, x_2, x_3 are denoted by `T`, `P`, and `S`, respectively. The SP factors z_1 and z_2 are denoted by `D` and `R`.

Ordinary Least Squares

```
WP_factors <- colnames(Ys_data)[1:3]
SP_factors <- colnames(Ys_data)[4:5]
linear_terms <- c(WP_factors, SP_factors)
# Model with all main effects and 2-way interactions
formula_1 <- as.formula(paste("Y_S ~ (",
```

```
    paste(linear_terms, collapse = " + "),")^2"))
Method_1 <- summary(lm(formula_1, data = Ys_data))
Method_1
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  299.0906     9.0640  32.998 2.03e-15 ***
T             -3.1156     9.0640  -0.344 0.735815
P             98.4344     9.0640  10.860 1.67e-08 ***
S           -145.0969     9.0640 -16.008 7.72e-11 ***
D           -131.9031     9.0640 -14.552 2.97e-10 ***
R            -19.4969     9.0640  -2.151 0.048181 *
T:P            0.2406     9.0640   0.027 0.979171
T:S           -3.0781     9.0640  -0.340 0.738867
T:D           -0.9719     9.0640  -0.107 0.916033
T:R           -2.8156     9.0640  -0.311 0.760350
P:S          -37.0906     9.0640  -4.092 0.000962 ***
P:D          -54.6844     9.0640  -6.033 2.29e-05 ***
P:R            0.3344     9.0640   0.037 0.971059
S:D           86.8969     9.0640   9.587 8.68e-08 ***
S:R            2.9406     9.0640   0.324 0.750093
D:R          -10.4156     9.0640  -1.149 0.268500
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49.74 on 15 degrees of freedom
```

```
Multiple R-squared:  0.981, Adjusted R-squared:  0.9619
F-statistic: 51.53 on 15 and 15 DF,  p-value: 3.718e-10
```

Restricted Maximum Likelihood using Satterwaithe

```
# Model with all main effects, 2-way interactions, and random WP effect
formula_2A <- as.formula(paste("Y_S ~ (",
    paste(linear_terms, collapse = " + "),")^2 + (1|WP)"))
Method_2A <- summary(lmer(formula_2A, data = Ys_data))
```

```
boundary (singular) fit: see help('isSingular')
```

```
Method_2A
```

```
Fixed effects:
             Estimate Std. Error       df t value Pr(>|t|)
(Intercept)  299.0906     9.0640  15.0000  32.998 2.03e-15 ***
T             -3.1156     9.0640  15.0000  -0.344 0.735815
P             98.4344     9.0640  15.0000  10.860 1.67e-08 ***
S           -145.0969     9.0640  15.0000 -16.008 7.72e-11 ***
D           -131.9031     9.0640  15.0000 -14.552 2.97e-10 ***
R            -19.4969     9.0640  15.0000  -2.151 0.048181 *
T:P            0.2406     9.0640  15.0000   0.027 0.979171
T:S           -3.0781     9.0640  15.0000  -0.340 0.738867
T:D           -0.9719     9.0640  15.0000  -0.107 0.916033
T:R           -2.8156     9.0640  15.0000  -0.311 0.760350
P:S          -37.0906     9.0640  15.0000  -4.092 0.000962 ***
P:D          -54.6844     9.0640  15.0000  -6.033 2.29e-05 ***
```

```
P:R            0.3344     9.0640   15.0000    0.037 0.971059
S:D           86.8969     9.0640   15.0000    9.587 8.68e-08 ***
S:R            2.9406     9.0640   15.0000    0.324 0.750093
D:R          -10.4156     9.0640   15.0000   -1.149 0.268500
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Restricted Maximum Likelihood using Kenward Roger

```
# Model with all main effects, 2-way interactions, and random WP effect
formula_2B <- as.formula(paste("Y_S ~ (",
    paste(linear_terms, collapse = " + "),")^2+ (1|WP)"))
# Fit model with Kenward-Roger method
model_2B <- lmer(formula_2B, data = Ys_data, REML = TRUE)
```

```
boundary (singular) fit: see help('isSingular')
```

```
Method_2B <- summary(model_2B, ddf = "Kenward-Roger")
Method_2B
```

```
Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
(Intercept)  299.0906     9.1398    0.9845  32.724   0.0205 *
T             -3.1156     9.1398    0.9845  -0.341   0.7915
P             98.4344     9.1398    0.9845  10.770   0.0610 .
S           -145.0969     9.1398    0.9845 -15.875   0.0417 *
D           -131.9031     9.1398   14.1074 -14.432 7.66e-10 ***
R            -19.4969     9.1398   14.1074  -2.133   0.0509 .
```

```
T:P            0.2406     9.1398    0.9845   0.026   0.9833
T:S           -3.0781     9.1398    0.9845  -0.337   0.7939
T:D           -0.9719     9.1398   14.1074  -0.106   0.9168
T:R           -2.8156     9.1398   14.1074  -0.308   0.7625
P:S          -37.0906     9.1398    0.9845  -4.058   0.1568
P:D          -54.6844     9.1398   14.1074  -5.983 3.25e-05 ***
P:R            0.3344     9.1398   14.1074   0.037   0.9713
S:D           86.8969     9.1398   14.1074   9.507 1.63e-07 ***
S:R            2.9406     9.1398   14.1074   0.322   0.7524
D:R          -10.4156     9.1398   14.1074  -1.140   0.2734
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Expected Mean Squares (ANOVA)

```r
# Model with all main effects, 2-way interactions,
#and split-plot error structure
formula_4A <- as.formula(paste("Y_S ~ (",
    paste(linear_terms, collapse = " + "),")^2 + Error(WP)"))
Method_4A <- summary(aov(formula_4A, data = Ys_data))
Method_4A
```

```
Error: WP
    Df Sum Sq Mean Sq
T    1   2251    2251
P    1 356982  356982
S    1 613959  613959
```

```
D    1  23273   23273
T:P  1   2808    2808
T:S  1   9591    9591
P:S  1  23020   23020

Error: Within
          Df Sum Sq Mean Sq F value   Pr(>F)
D          1 527529  527529 199.115 1.14e-09 ***
R          1  13710   13710   5.175   0.0392 *
T:D        1     16      16   0.006   0.9388
T:R        1    685     685   0.259   0.6190
P:D        1 107948  107948  40.745 1.70e-05 ***
P:R        1    927     927   0.350   0.5635
S:D        1 226165  226165  85.366 2.47e-07 ***
S:R        1    377     377   0.142   0.7116
D:R        1   3290    3290   1.242   0.2839
Residuals 14  37091    2649
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that no tests of significance were performed for the whole plot factors because there is only one denominator degree of freedom from the three-way interaction. The R function does not produce tests in this instance.

QQ Plot and Lenth Method

Figure 1.2 shows the plots of the QQ plot and Lenth method. The left column shows

Half-Normal Plots and the right column shows Lenth Plots. As discussed earlier, two plots are required for both methods, one for the WP effects and SP effects. Again, these methods are not completely valid as the effects are not mutually orthogonal due to the data being unbalanced.

Based on the QQ plot for the WP effects, it appears that factors P and S, as well as their interaction, are statistically significant, while all effects involving T are not statistically significant. Similar conclusions are reached by the Lenth plot. The QQ plot for the SP effects reveals that factor D is important as well as its interaction with S and P. Factor R is on the border of significance but it seems more likely to be grouped with the insignificant effects according to the QQ plot. For the Lenth method, R is outside the margin of error limit but within the simultaneous margin of error limit. The R effect is not significant under the more conservative criterion.

The results of the SPD Analysis of the PLA experiment by the some of the methods discussed above are summarized in Table 1.3 below. With $\alpha = 0.05$, the OLS and REML using the Satterwaithe approximation (REML(S)) produce nearly the same results. This is because REML estimates σ_w^2 to be zero and OLS implicitly assumes $\sigma_w^2 = 0$. REML using the Kenward-Roger approximation deems P, R and P:S insignificant, although P and R would be significant at $\alpha = 0.10$. The R software for ANOVA was unable to produce p-values on the WP effects because there is only one degree of freedom for estimating the effect variances. Similar to REML(KR), the QQ Plots are also borderline on the R effect and T effect, but includes the P:S effect. If we only included effects beyond the SME value, the Lenth Plot would have the same effects as the QQ plot but with P:S removed. If we included effects beyond the smaller ME value, the Lenth plot would include both P:S and R. While there are many effects for which the analysis methods agree are important, there are some effects that are potentially significant. We are also unsatisfied with the poor estimation of σ_W^2, which we believe is truly a positive value.

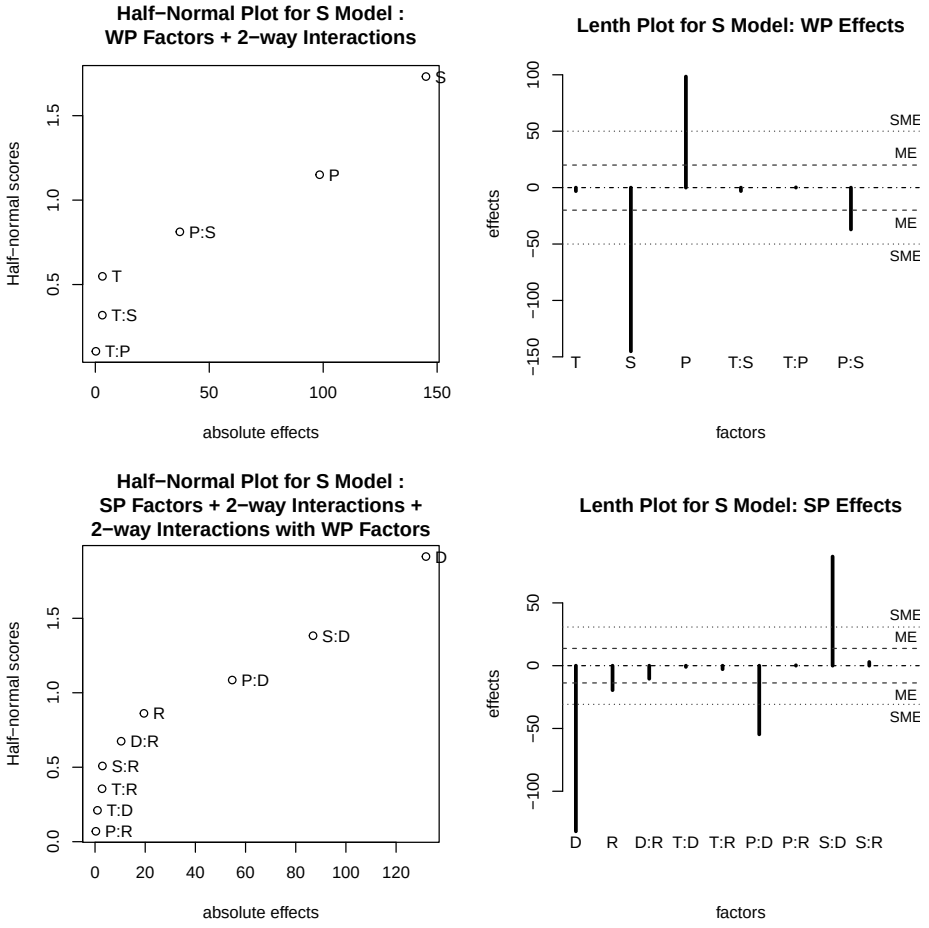

Figure 1.2: QQ Plots and Lenth Plots for S Model

Table 1.3: Significant Factors and Variance Components by Analysis Methods, $\alpha = 0.05$

Fixed Effects and Variance Components																		
Analysis Methods		**T**	**P**	**S**	**D**	**R**	**T:P**	**T:S**	**T:D**	**T:R**	**P:S**	**P:D**	**P:R**	**S:D**	**S:R**	**D:R**	σ^2_w	σ^2_s
	OLS		X	X	X	X					X	X		X			0	2474
	REML(S)		X	X	X	X					X	X		X			0	2474
	REML(KR)			X	X							X		X			0	2474
	ANOVA	NA	NA	NA	X	X	NA	NA		NA		X		X			NA	2649
	QQ		X	X	X						X	X		X			NA	NA
	LENTH		X	X	X	X					X	X		X			NA	NA

X ≡ statistically significant at $\alpha = 0.05$

1.3 Sequential Split Plot Design (SSPD) Method

1.3.1 The Method

In most split plot experiments, data is collected after both WP and SP factors have been applied. However, when the WP factors and SP factors are applied sequentially, it may be possible to take measurements from the WPs which are unaffected by the SP factors. This is what was done in the PLA experiment, which took measurements at the end of the Melt Extrusion process (Stage 1) and then again at the end of the Drawing process (Stage 2). We now describe a straightforward model and analysis for such two-stage data. The most remarkable feature of our proposed method is that it eliminates the need to perform mixed modeling and produces positive estimates of both variance components.

Let the n responses taken from Stage 1, which only involve the WP factors, be denoted by Y_i^W. Recall the Stage 2 responses are denoted by Y_{ij}^S. Again ignoring three-way and higher-order interactions, we model the two sets of responses by:

$$Y_i^W = \beta_0 + \sum_{p=1}^{w} \beta_p x_{ip} + \sum_{p=1}^{w-1} \sum_{p'=p+1}^{w} \beta_{pp'} x_{ip} x_{ip'} + e_i^W \tag{1.14}$$

$$Y_{ij}^S = \beta_0^* + Y_i^W + \sum_{r=1}^{s} \gamma_r z_{ijr} + \sum_{r=1}^{s-1} \sum_{r'=r+1}^{s} \gamma_{rr'} z_{ijr} z_{ijr'} + \sum_{p=1}^{w} \sum_{r=1}^{s} \gamma'_{pr} x_{ip} z_{ijr} + e_{ij}^S \,. \tag{1.15}$$

That is, the Y_i^W responses are embedded in the Y_{ij}^S responses. This model will be referred to as the "W Model"; the second model will be referred to as the "S Model". All of the information about the effects involving only the WP factors is contained within the W Model. Therefore, we propose the first step is to analyze the W Model just as one would for any fixed factorial experiment.

Inference for the WP-only effects based on the W Model has two advantages over trying to perform this inference based on the S Model. First recall that for balanced CRSPD designs,

an analysis on the Y_{ij}^S responses had $\text{Var}(\hat{\beta}_p) = \frac{1}{n}\sigma_W^2 + \frac{1}{mn}\sigma_S^2$. However, the estimators under the W Model would have $\text{Var}(\hat{\beta}_p^W) = \frac{1}{n}\sigma_W^2$. Therefore, our proposed analysis should produce estimators with greater precision. This improvement will be significant when $\sigma_W^2 << \sigma_S^2$, which we anticipate for the PLA experiment.

The other advantage the W Model provides is its ability to guarantee that $\hat{\sigma}_W^2 > 0$, since its $\hat{\sigma}_W^2$ comes from the residual sum of squares of a linear model:

$$\text{MSE}_W = \frac{1}{n - \frac{w(w+1)}{2} - 1} \sum_{i=1}^{n} (Y_i^W - \hat{Y}_i^W)^2 = \frac{\text{SSE}_W}{\text{dfE}}, \tag{1.16}$$

which is unbiased for σ_W^2 and always positive. In contrast, as we saw in Section 1.2.2, the estimator $\hat{\sigma}_W^2$ after fitting the S Model can be either 0 or negative depending on whether REML or the Method of Moments is used. This is especially common when $\sigma_W^2 << \sigma_S^2$.

From the presentation of the W and S Models in (1.14) and (1.15), it is clear that the transformed response $Y_{ij}^{SW} = Y_{ij}^S - Y_i^W$ can be modeled as

$$\begin{aligned} Y_{ij}^{SW} = (\beta_0^* - \beta_0) + \sum_{r=1}^{s} \gamma_r z_{rij} + \sum_{r=1}^{s-1} \sum_{s'=r+1}^{s} \gamma_{rs'} z_{rij} z_{s'ij} \\ + \sum_{p=1}^{w} \sum_{r=1}^{s} \gamma'_{pr} x_{ijp} z_{ijr} + e_{ij}^S, \end{aligned} \tag{1.17}$$

which involves an intercept $(\beta_0^* - \beta_0)$, effects involving only the SP factors, and the interactions between the SP and WP factors. We refer to this model as the "S-W Model." An analysis under this model avoids specifying a mixed model and inference can be carried out for the remaining effects using traditional methods.

The steps for the proposed method are detailed here:

1. Perform inference on the WP-only effects using the Y_i^W responses and the W Model.

2. Calculate the Y_{ij}^{SW} responses and perform inference on the remaining split plot model effects using the S-W Model.

3. Combine estimates from (1) and (2) to arrive at an estimate for the S Model if desired.

What may be most appealing to practitioners is that SSPDM avoids the mixed model analysis, which is prone to issues for smaller data sets. Instead, it performs two simple ordinary least-squares analyses on two models. Moreover, this property holds whether the CRSPD is balanced or unbalanced. Section 1.4 demonstrates this property via a simulation study for both kinds of data structure.

1.3.2 SSPD Method Model Adequacy Test

The simplicity of the SSPD method relies on the modeling assumptions between the W and S Models, particularly that Y_i^W is embedded in the Y_{ij}^S responses. If these assumptions did not hold, then there may be some remaining parts of the WP model leftover in Y_{ij}^{SW}. The most extreme situation would be if there is no overlap of effects, making the true S and S-W Models:

$$\begin{aligned} Y_{ij}^S &= \beta_0^* + \sum_{p=1}^{w} \beta_p^* x_{ip} + \sum_{p=1}^{w-1} \sum_{p'=p+1}^{w} \beta_{pp'}^* x_{ip} x_{ip'} + e_i^{W*} \\ &\quad + \sum_{r=1}^{s} \gamma_r z_{ijr} + \sum_{r=1}^{s-1} \sum_{r'=r+1}^{s} \gamma_{rr'} z_{ijr} z_{ijr'} + \sum_{p=1}^{w} \sum_{r=1}^{s} \gamma'_{pr} x_{ip} z_{ijr} + e_{ij}^S \qquad (1.18) \\ Y_{ij}^{SW} &= (\beta_0^* - \beta_0) + \sum_{p=1}^{w} (\beta_p^* - \beta_p) x_{ip} + \sum_{p=1}^{w-1} \sum_{p'=p+1}^{w} (\beta_{pp'}^* - \beta_{pp'}) x_{ip} x_{ip'} + (e_i^{W*} - e_i^W) \\ &\quad + \sum_{r=1}^{s} \gamma_r z_{ijr} + \sum_{r=1}^{s-1} \sum_{r'=r+1}^{s} \gamma_{rr'} z_{ijr} z_{ijr'} + \sum_{p=1}^{w} \sum_{r=1}^{s} \gamma'_{pr} x_{ip} z_{ijr} + e_{ij}^S \,. \qquad (1.19) \end{aligned}$$

The original S-W Model (1.17) would then be underspecified, leading to biased estimation of the fixed effects and inflated estimation of σ_S^2. Therefore, it is crucial to first perform a test of the hypothesis that assumptions for (1.14) and (1.15) hold.

As model (1.17) is nested within model (1.19), the well-known F-test comparing full and reduced models can be used to determine whether (1.17) is underspecified. The null hypothesis, H_0, is that (1.17) is correct, while the alternative hypothesis, H_1, is that (1.19) is correct. The corresponding F-statistic is then

$$F = \frac{\left(\frac{\text{SSE}(H_0)-\text{SSE}(H_1)}{\text{dfE}(H_0)-\text{dfE}(H_1)}\right)}{\left(\frac{\text{SSE}(H_1)}{\text{dfE}(H_1)}\right)}, \tag{1.20}$$

where $\text{SSE}(H_0)$ and dfE_{H_0} denote the residual sum-of-squares and error degrees of freedom for the model under H_0, with similar definitions for the two terms involving H_1. These calculations may be done automatically in statistical software such as R. Under H_0, F follows a central F distribution with $\text{dfE}(H_0) - \text{dfE}(H_1)$ and $\text{dfE}(H_1)$ numerator and denominator degrees of freedom, respectively.

If the null hypothesis is rejected, we recommend the experimenters try to understand what about the process may have led to a violation of the assumptions. Rejecting H_0 can happen due to a number of possibilities, all of which suggest that the effect of the WP factors may change by just subjecting the WPs to additional experimentation, regardless of the settings of the SP factors. This seems plausible for many applications, but for the case of the PLA experiment, running yarn through the drawing process at room temperature with a draw ratio of 1 should not change the material properties of the yarn. Therefore, we have a reasonable expectation that the model assumptions hold.

1.4 Simulation Study

In order to demonstrate the principles and advantages of the SSPD Method, an R Shiny App was developed based on the SSPD Method and the model adequacy test discussed above. Three scenarios will be examined:

1. the SP design is balanced and the SSPD model is adequate.
2. the SP design is unbalanced and the SSPD model is adequate.
3. the SP design is balanced and the SSPD model is inadequate.

For all simulation scenarios, the inputs are shown in the tables in the respective simulation report sections. All simulations share the settings shown in Table 1.4, WP represents the number of WP factors, SP represents the number of SP factors, r represents the number of replicates for each treatment combination and *nsim* represents the number of simulation runs. The remaining tables shown in the respective sections contain values of the parameters specified in the row headers of each table. Two simulations were run for each case. The first simulation has more non-zero parameters and σ^2_W larger than σ^2_S. The second simulation has less non-zero parameters and σ^2_W smaller than σ^2_S.

Table 1.4: Simulation Settings

WP	SP	r	nsim
3	2	2	1000

1.4.1 Scenario 1 - Balanced Design Simulations

The results of the Balanced Design Simulations are shown below. The histograms show comparisons of the fixed effects and variance components between the S, W and SW

Models. The Summary of F-Test Results shows the results of the SSPD Method Adequacy Test discussed in Section 1.3.2

1.4.1.1 Balanced Design Simulation 1

Table 1.5: Intercept and Error Terms for Balanced Design Simulation 1

β_0	σ_W^2	σ_S^2
10	1	5

Table 1.6: Whole Plot Main Effects and Interactions for Balanced Design Simulation 1

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
3	2	-6	1	1	-2

Table 1.7: Split Plot Main Effects and Interactions for Balanced Design Simulation 1

γ_1	γ_2	γ_{12}
1	3	2

Table 1.8: Split Plot, Whole Plot Interactions for Balanced Design Simulation 1

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	1	1	0	0	-3

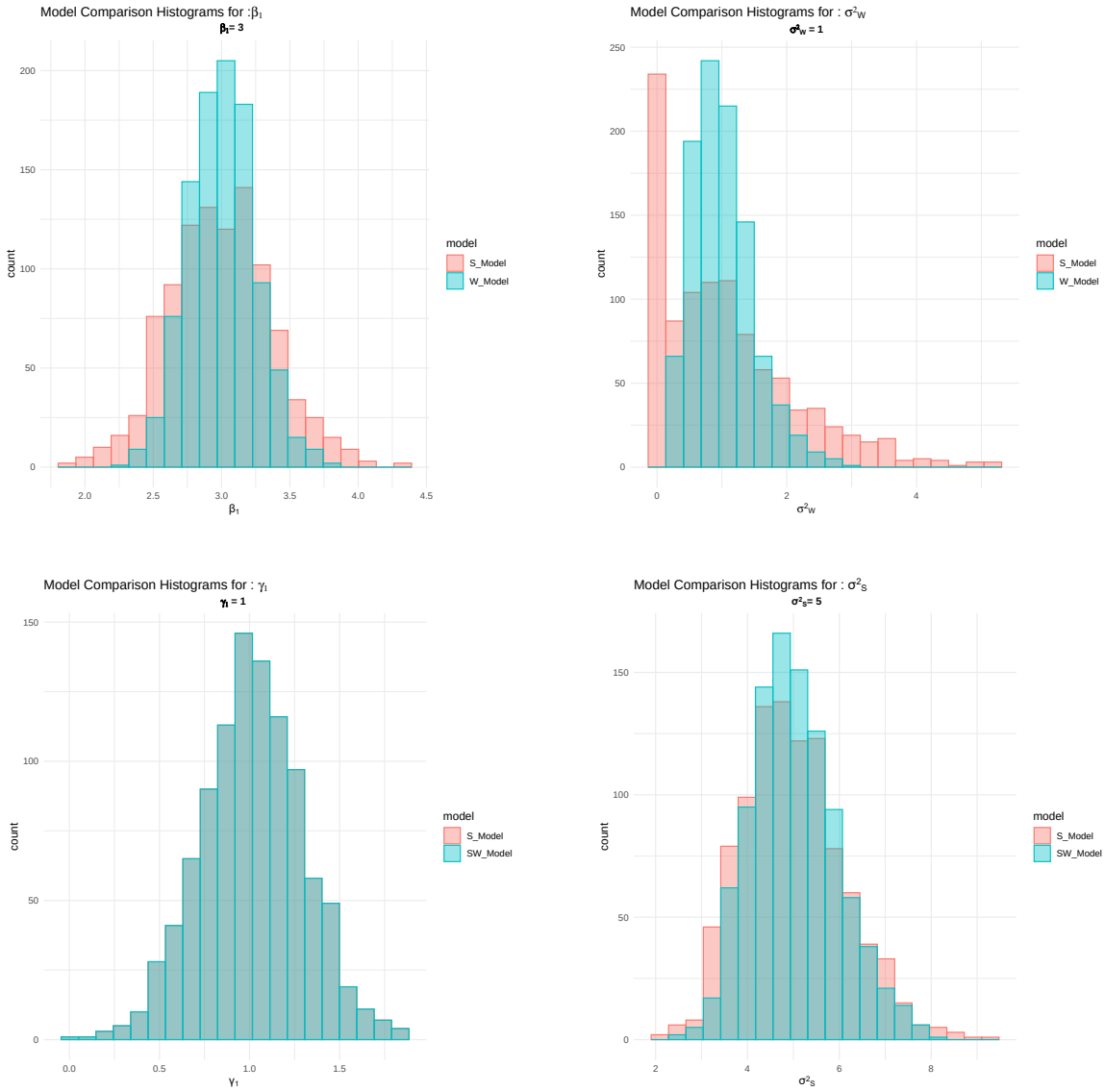

Figure 1.3: Results for Balanced Design Simulation 1

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	1.053
2	Mean F-critical Value	1.9313
3	Proportion Significant	0.06
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.4: F-Test Summary for Balanced Design Simulation 1

1.4.1.2 Balanced Design Simulation 2

Table 1.9: Intercept and Error Terms for Balanced Design Simulation 2

β_0	σ_W^2	σ_S^2
5	2	1

Table 1.10: Whole Plot Main Effects and Interactions for Balanced Design Simulation 2

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
2	1	0	1	0	0

Table 1.11: Split Plot Main Effects and Interactions for Balanced Design Simulation 2

γ_1	γ_2	γ_{12}
1	0	2

Table 1.12: Split Plot, Whole Plot Interactions for Balanced Design Simulation 2

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	0	0	-1	0	0

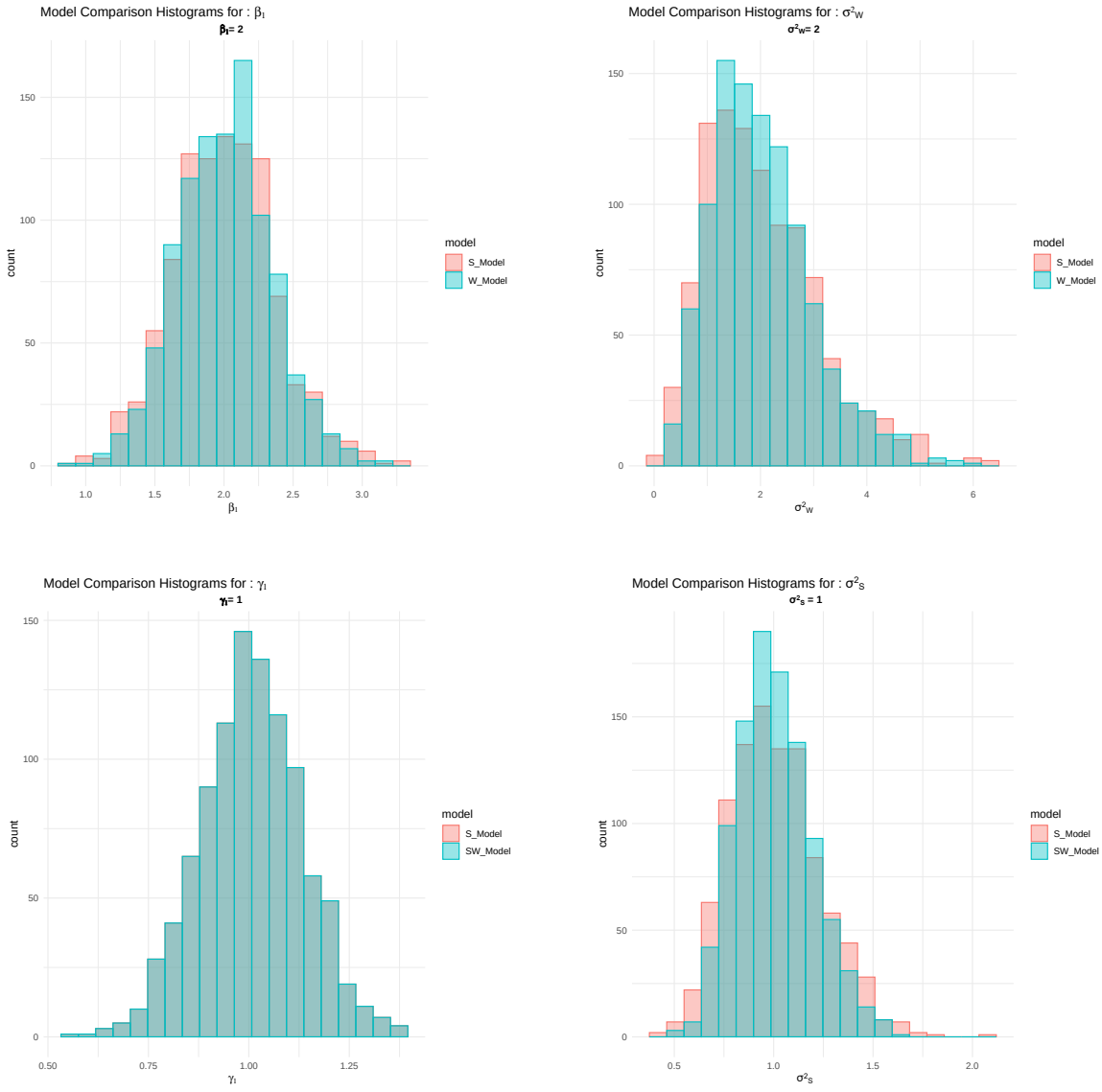

Figure 1.5: Results for Balanced Design Simulation 2

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	1.053
2	Mean F-critical Value	1.9313
3	Proportion Significant	0.06
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.6: F-Test Summary for Balanced Design Simulation 2

1.4.1.3 Analysis of Balanced Design Simulation Studies

From inspection of the histograms of the the balanced design simulation, the width of the histogram for the W Model is narrower than the S Model for the same factor. On the other hand, the histograms for the SW Model and S Model for the SP factor is identical. Histograms for the remaining SP and WP factors show similar results.

The most significant difference can be seen in the histograms for the estimator of σ_w^2. The histogram for Simulation 1 for the S Model has a tall bar at zero, compared to the W Model which has very low frequency count at 0 and is centered at approximately two. This result support the claim that estimates of WP factors and σ_w^2 will be more accurate using the SSPD Method. At the same time, the variation of the S Model is preserved.

The histogram for σ_s^2 between the S Model and SW Model suggests a slightly narrower distribution for the SW Model than the S Model, but they are nearly identical. The histogram for Simulation 2 for the S Model and W Model are nearly identical. This result illustrates the concept that the SSPD method is most useful in scenarios where σ_w^2 is smaller than σ_s^2.

With regard to the Model Adequacy tests, the results for both simulations are shown shown in Figure 1.4 and Figure 1.6 , the Mean F-Statistic Value of 1.053 is far below the

Mean F-critical value of 1.93. In addition, the proportion of significant simulation runs is near the expected Type I Error Rate of 0.05. This table shows that the software doing the F-Test is working properly and that the model is adequate.

1.4.2 Scenario 2 - Unbalanced Design Simulations

The results of the Unbalanced Design Simulations are shown below. For the first simulation, two rows were removed. For the second simulation, four rows of the experimental design were removed. The histograms show comparisons of the fixed effects and variance components between the S, W and SW Models. The Summary of F-Test Results show the results of the SSPD Method Adequacy Test discussed in Section 1.3.2

1.4.2.1 Unbalanced Design Simulation 1

Table 1.13: Intercept and Error Terms for Unbalanced Design Simulation 1

β_0	σ^2_W	σ^2_S
10	1	5

Table 1.14: Whole Plot Main Effects and Interactions for Unbalanced Design Simulation 1

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
3	2	-6	1	1	-2

Table 1.15: Split Plot Main Effects and Interactions for Unbalanced Design Simulation 1

γ_1	γ_2	γ_{12}
1	3	2

Table 1.16: Split Plot, Whole Plot Interactions for Unbalanced Design Simulation 1

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	1	1	0	0	-3

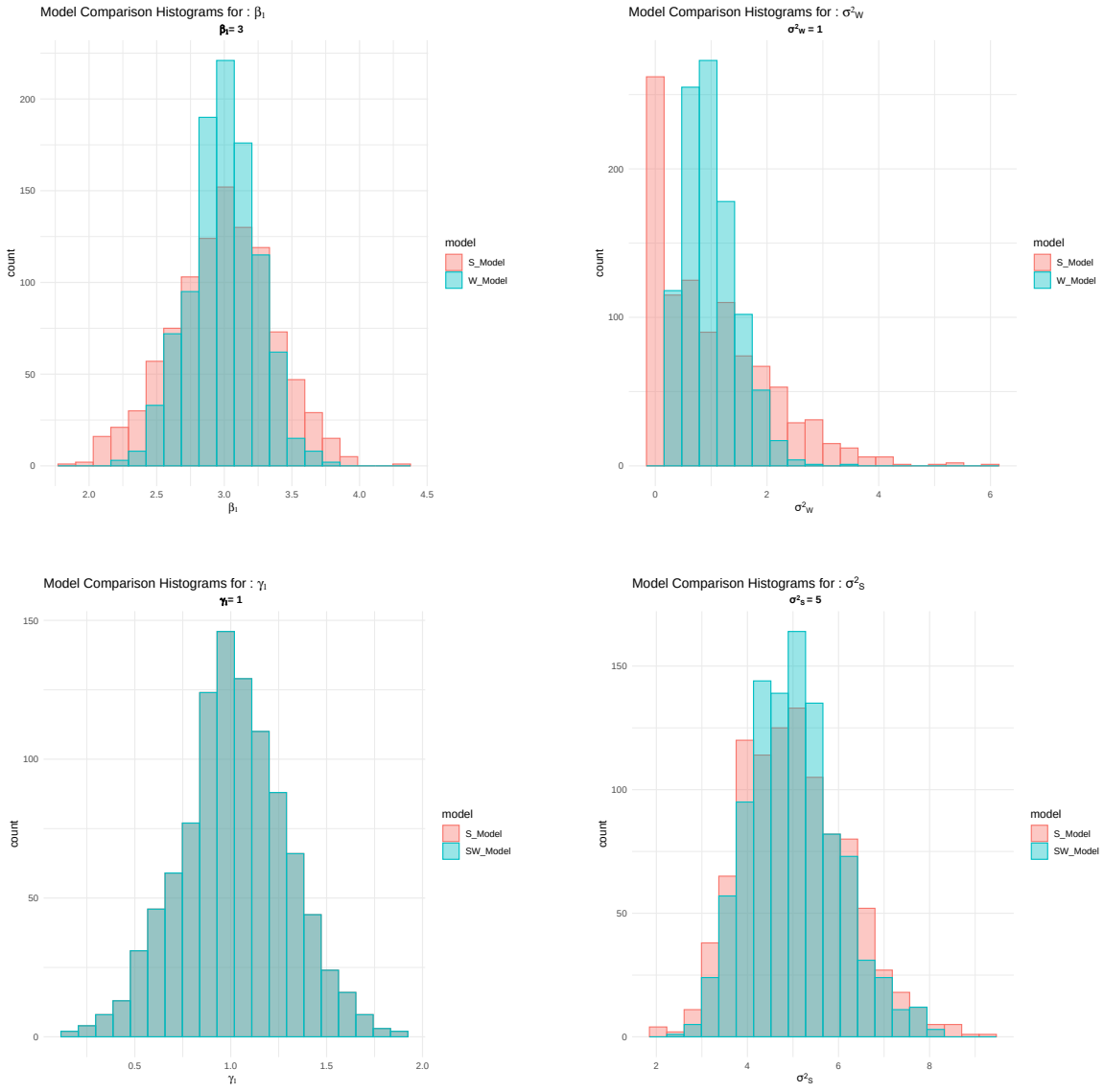

Figure 1.7: Results for Unbalanced Design Simulation 1

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	1.0035
2	Mean F-critical Value	1.9462
3	Proportion Significant	0.042
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.8: F-Test Summary for Unbalanced Design Simulation 1

1.4.2.2 Unbalanced Design Simulation 2

Table 1.17: Intercept and Error Terms for Unbalanced Design Simulation 2

β_0	σ^2_W	σ^2_S
5	2	1

Table 1.18: Whole Plot Main Effects and Interactions for Unbalanced Design Simulation 2

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
2	1	0	1	0	0

Table 1.19: Split Plot Main Effects and Interactions for Unbalanced Design Simulation 2

γ_1	γ_2	γ_{12}
1	0	2

Table 1.20: Split Plot, Whole Plot Interactions for Unbalanced Design Simulation 2

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	0	0	-1	0	0

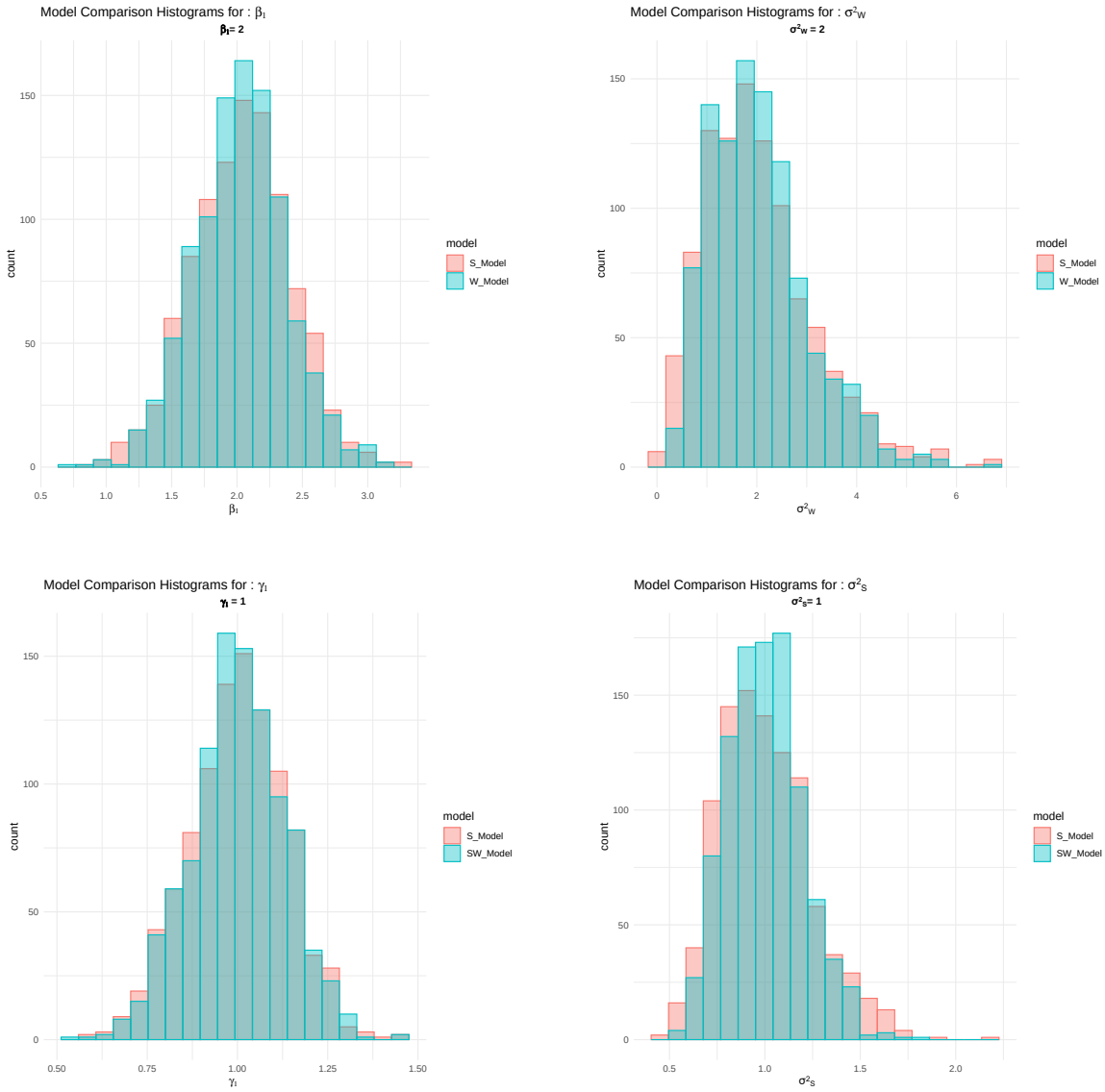

Figure 1.9: Results for Unbalanced Design Simulation 2

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	1.0698
2	Mean F-critical Value	1.9629
3	Proportion Significant	0.062
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.10: F-Test Summary for Unbalanced Design Simulation 2

1.4.2.3 Analysis of Unbalanced Design Simulation Studies

From inspection of the histograms for the model inadequate case, the results for Simulation 1 (2 rows removed) look very similar to the corresponding results for the Balanced Design. The results for Simulation 2 (4 rows removed) also look very similar to the corresponding results for the Balanced Design. These simulations indicate that SSPD is insensitive to moderately unbalanced designs.

1.4.3 Scenario 3 - Inadequate Model Simulations

The results of the Model Inadequate Simulations are shown below. The histograms show comparisons of the fixed effects and variance components between the S, W and SW Models. The Summary of F-Test Results show the results of the SSPD Method Adequacy Test discussed in Section 1.3.2

1.4.3.1 Inadequate Model Simulation 1

Table 1.21: Intercept and Error Terms for Inadequate Model Simulation 1

β_0	σ^2_W	σ^2_S
10	1	5

Table 1.22: Whole Plot Main Effects and Interactions for Inadequate Model Simulation 1

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
3	2	-6	1	1	-2

Table 1.23: Split Plot Main Effects and Interactions for Inadequate Model Simulation 1

γ_1	γ_2	γ_{12}
1	3	2

Table 1.24: Split Plot, Whole Plot Interactions for Inadequate Model Simulation 1

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	1	1	0	0	-3

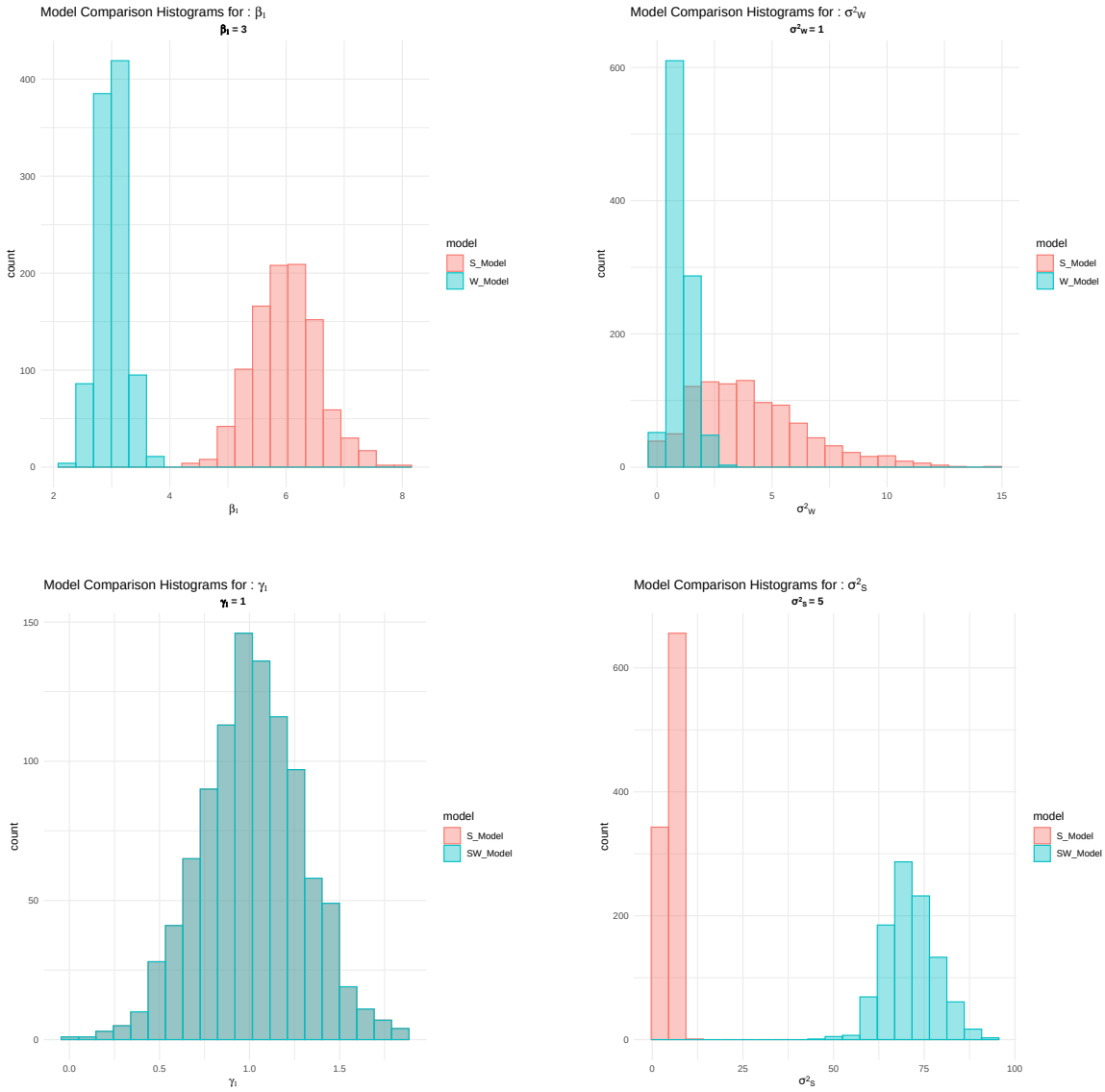

Figure 1.11: Results for Inadequate Model Simulation 1

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	50.8511
2	Mean F-critical Value	1.9313
3	Proportion Significant	1
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.12: F-Test Summary for Inadequate Model Simulation 1

1.4.3.2 Inadequate Model Simulation 2

Table 1.25: Intercept and Error Terms for Inadequate Model Simulation 2

β_0	σ^2_W	σ^2_S
5	2	1

Table 1.26: Whole Plot Main Effects and Interactions for Inadequate Model Simulation 2

β_1	β_2	β_3	β_{12}	β_{13}	β_{23}
2	1	0	1	0	0

Table 1.27: Split Plot Main Effects and Interactions for Inadequate Model Simulation 2

γ_1	γ_2	γ_{12}
1	0	2

Table 1.28: Split Plot, Whole Plot Interactions for Inadequate Model Simulation 2

γ'_{11}	γ'_{21}	γ'_{31}	γ'_{12}	γ'_{22}	γ'_{32}
1	0	0	-1	0	0

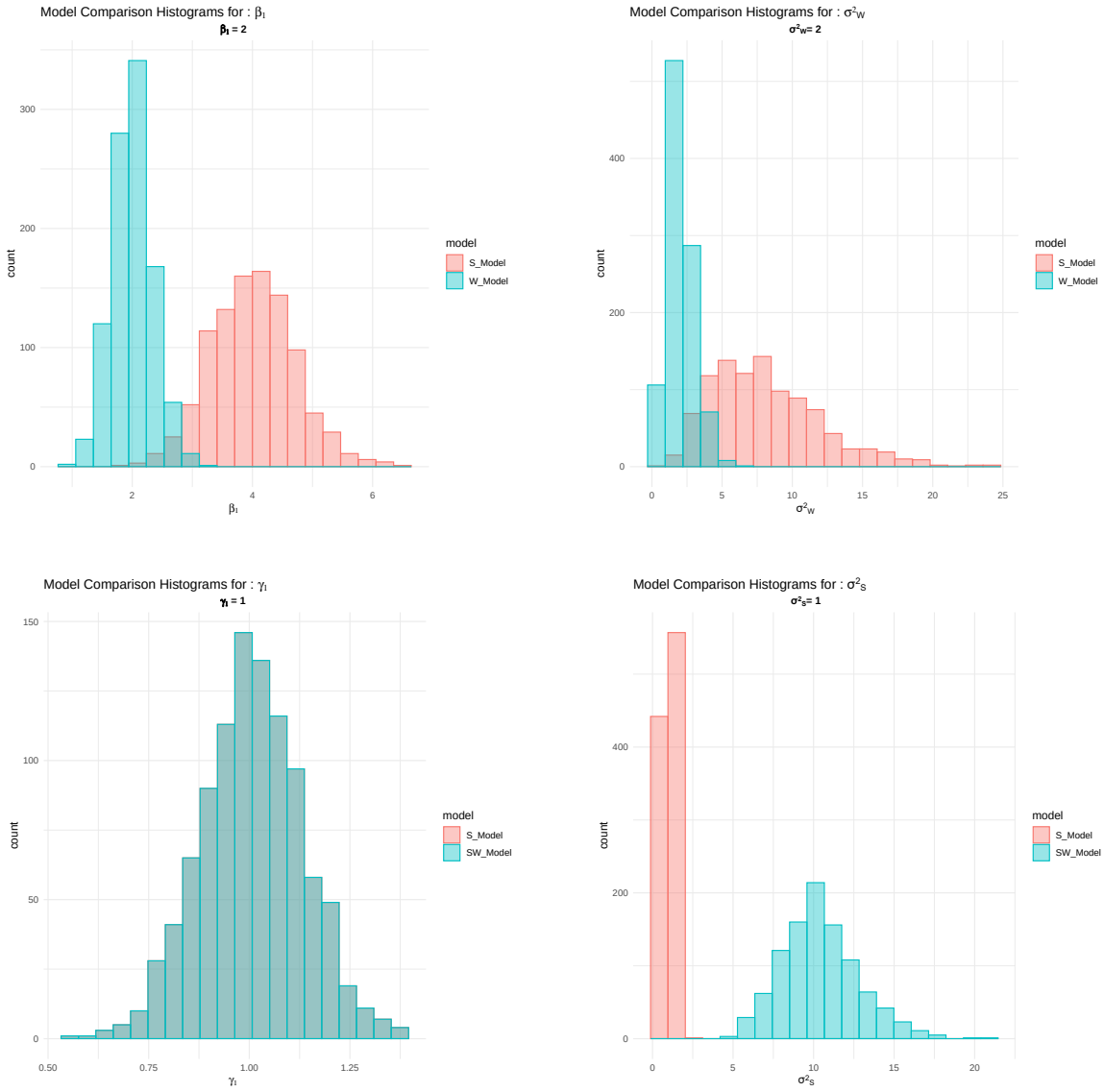

Figure 1.13: Results for Inadequate Model Simulation 2

Summary of F-Test Results: S_Model vs SW_Model

	Statistic	Value
1	Mean F-Statistic Value	36.3599
2	Mean F-critical Value	1.9313
3	Proportion Significant	1
4	Number of Successful Tests	1000
5	Number of Failed Tests	0

Figure 1.14: F-Test Summary for Inadequate Model Simulation 2

1.4.3.3 Analysis of Inadequate Model Simulation Studies

From inspection of the histograms for the model inadequate case, the results are quite different compared to the balanced model adequate case. While the histograms for WP main effect is still much narrower, the effect is now shifted to the right. The SP main effect appears to be unaffected. Here, σ_w^2 for the W Model has a very high peak near zero compared to the S Model. σ_s^2 for the SW Model is shifted to the right by orders of magnitude.

All this demonstrates that the stated assumptions for the SSPD method are important. If they do not hold, the estimates for WP factors and variance components may not be reliable. With regard to the Model Inadequate Test, under this scenario, as expected, the F-Test is significant 100% of the time. The Mean F-Statistics are 36 and 51 compared to the Mean F-Critical Value of 1.93, which while exaggerated shows the importance of an adequate model. The inflation of error variance is likely related to the model inadequacy.

1.5 PLA Experiment Analysis using SSPDM

The W Model and SW Model for the PLA experiment are

$$Y_i^W = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} + e_i^W, \tag{1.21}$$

$$\begin{aligned} Y_{ij}^{(S-W)} = (\beta_0^* - \beta_0) &+ \gamma_1 z_{1j} + \gamma_2 z_{2j} + \gamma_{12} z_{1j} z_{2j} \\ &+ \gamma'_{11} x_{i1} z_{1j} + \gamma'_{12} x_{i1} z_{2j} + \gamma'_{21} x_{i2} z_{1j} \\ &+ \gamma'_{22} x_{i2} z_{2j} + \gamma'_{31} x_{i3} z_{1j} + \gamma'_{32} x_{i3} z_{2j} + e_{ij}^S, \end{aligned} \tag{1.22}$$

The W Model and the S-W Model for the PLA experiment were created and analyzed in R Studio. The code and results of the analyses are shown below.

```
#-------------------------------------------------------------
#S Model - SP Factors (WP main effects, SP main effects,
#2-way interactions within WP and SP and 2-way interactions
# with WP Factors)
formula_S <- as.formula(paste("Y_S ~ (", paste(linear_terms,
                                  collapse = " + "),")^2"))
#Create model matrix for S Model
Y_S <- Ys_data$Y_S
X_S <- model.matrix(formula_S, data = Ys_data)
#-------------------------------------------------------------
#W Model - WP factors (main WP effects, 2-way interactions)
formula_W <- as.formula(paste("Y_W ~ (",
              paste(WP_factors, collapse = " + "),")^2"))
```

```r
#Create model matrix for W Model
X_W <- model.matrix(formula_W, data = Yw_data)
# Replicate each row 4 times and drop row 20.
X_W <- X_W[rep(1:nrow(X_W), each = 4), ][-20, ]

Y_W <- rep(Yw_data$Y_W, each = 4)[-20]
df_W <- data.frame(Y_W, X_W)

cat("Linear Model Fit for W Model (OLS)", "\n")
```

```
Linear Model Fit for W Model (OLS)
```

```r
W_Model <- summary(lm(formula_W, data = Yw_data))
W_Model
```

```

Call:
lm.default(formula = formula_W, data = Yw_data)

Residuals:
      1       2       3       4       5       6       7       8
 0.5125 -0.5125 -0.5125  0.5125 -0.5125  0.5125  0.5125 -0.5125

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept)  443.4875     0.5125  865.341 0.000736 ***
T             -8.5375     0.5125  -16.659 0.038170 *
```

```
P            150.6625     0.5125  293.976 0.002166 **
S           -225.7625     0.5125 -440.512 0.001445 **
T:P           -5.7125     0.5125  -11.146 0.056962 .
T:S            2.3125     0.5125    4.512 0.138845
P:S          -78.7375     0.5125 -153.634 0.004144 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.45 on 1 degrees of freedom
Multiple R-squared:      1, Adjusted R-squared:      1
F-statistic: 5.075e+04 on 6 and 1 DF,  p-value: 0.003398
```

```r
#-------------------------------------------------------------
#S-W Model by Satterwaithe
#Create model matrix for S-W Model
X_SW <- X_S[, SP_effects_names]
Y_SW <- Y_S - Y_W
df_SW <- data.frame(Y_SW, X_SW, check.names = FALSE)
  formula_SW <- as.formula(
  "Y_SW ~ D + R + `D:R` + `T:D` + `T:R` + `P:D` + `P:R` + `S:D` + `S:R`")
SW_Model <-lm(formula_SW, data = df_SW)


cat("\n")
```

```r
cat("Linear Model Fit for SW1 Model (OLS)", "\n")
```

```
Linear Model Fit for SW1 Model (OLS)
```

```r
SW_Model_summary <- summary(SW_Model)
SW_Model_summary
```

```

Call:
lm.default(formula = formula_SW, data = df_SW)

Residuals:
    Min      1Q  Median      3Q     Max 
-220.51  -68.83   37.78   86.55  152.59 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -145.2415    24.6531  -5.891 7.56e-06 ***
D           -131.0585    24.6531  -5.316 2.85e-05 ***
R            -20.3415    24.6531  -0.825  0.41858    
`D:R`         -9.5710    24.6531  -0.388  0.70176    
`T:D`         -0.1273    24.6531  -0.005  0.99593    
`T:R`         -3.6602    24.6531  -0.148  0.88339    
`P:D`        -55.5290    24.6531  -2.252  0.03512 *  
`P:R`          1.1790    24.6531   0.048  0.96231    
`S:D`         86.0523    24.6531   3.491  0.00218 ** 
```

```
‘S:R‘          3.7852    24.6531   0.154  0.87944
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1

Residual standard error: 136.4 on 21 degrees of freedom
Multiple R-squared:  0.6929,    Adjusted R-squared:  0.5613
F-statistic: 5.265 on 9 and 21 DF,  p-value: 0.0008274

#------------------------------------------------------------
```

The analysis finds all WP main effects to be statistically significant at $\alpha = 0.05$, as well as the two-factor interaction between factors P and S. The WP variance is estimated to be $\sigma_W^2 = 1.45^2 = 2.11$. Finding statistical significance might be somewhat surprising, given that there is only one degree of freedom for estimating σ_W^2, which comes entirely from the sum of squares of the three-way interaction between the WP factors. But the effect sizes are large and the standard errors are small, so the results are reasonable. The analysis mostly agrees with previous analyses except for the inclusion of the T main effect.

As a point of comparison, a QQ Plot and Lenth Plot of the W Model are also included in this block of analysis, shown in Figures 1.15 and 1.16. These plots also include the three-way interaction as a reference point. The plots suggest that the T main effect should not be included, but the method is mostly graphical. It was ultimately decided to keep the T main effect to avoid committing a potential Type II error. Table 1.29 compares the results of SSPDM to the other methods from Section 1.2.2.

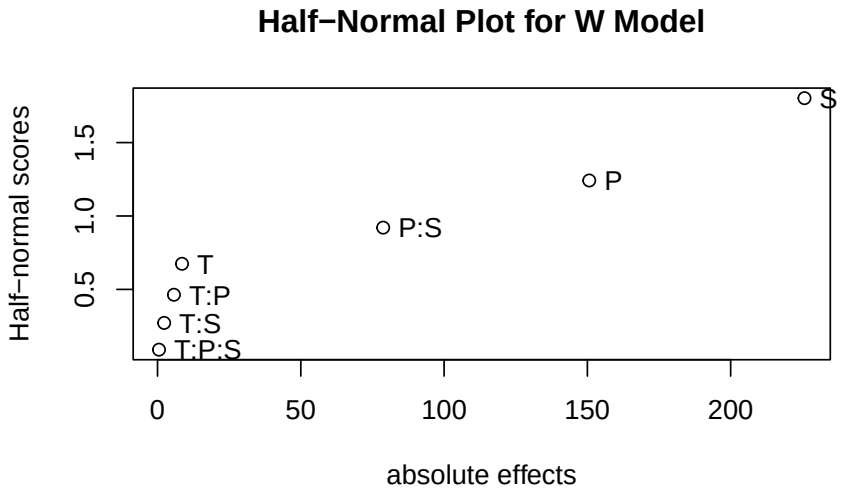

Figure 1.15: QQ Plot for W Model

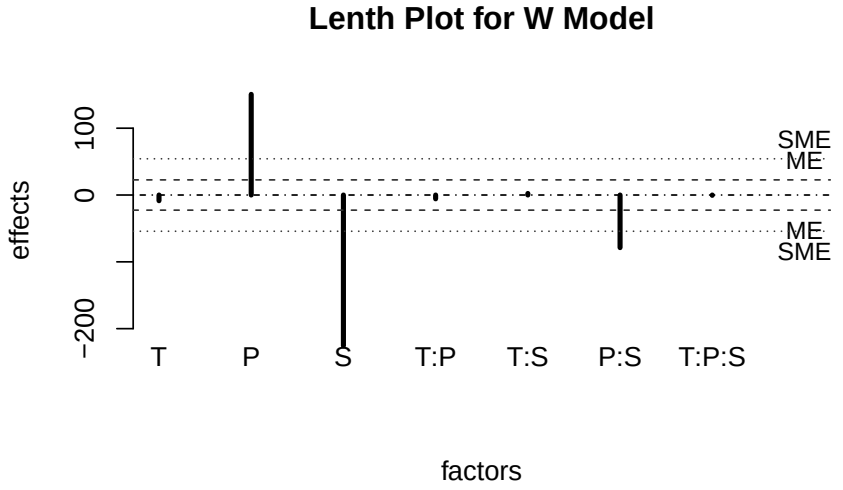

Figure 1.16: Lenth Plot for W Model

The SSPD Method identifies all effects identified by the other methods, except it notably excludes main effect of draw ratio. Notice that SSPDM is the only method that estimates a positive value for σ^2_W, but its estimate for σ^2_S is orders of magnitude larger than the other methods. It is interesting to note that the Lenth Plot of the W Model shows P, S and P:S as significant.

Table 1.29: Significant Factors and Variance Components by Analysis Methods including SSPD Method, $\alpha = 0.05$

Fixed Effects and Variance Components																			
Analysis Methods		**INT**	**T**	**P**	**S**	**D**	**R**	**T:P**	**T:S**	**T:D**	**T:R**	**P:S**	**P:D**	**P:R**	**S:D**	**S:R**	**D:R**	σ^2_w	σ^2_s
	OLS	X		X	X	X	X					X	X		X			NA	2474
	REML(S)	X		X	X	X	X					X	X		X			0	2474
	REML(KR)	X			X	X							X		X			0	2474
	ANOVA	NA	NA	NA	NA	X	X	NA	NA		NA		X		X			NA	2649
	QQ (S)	NA		X	X	X						X	X		X			NA	NA
	LENTH (S)	NA		X	X	X							X		X			NA	NA
	SSPD	X	X	X	X	X						X	X		X			2.10	18604

X ≡ statistically significant at $\alpha = 0.05$

We started this section with the SSPDM results for demonstration purposes. In practice, one should start first with the model adequacy test. The null hypothesis model for the PLA experiment is

$$\begin{aligned} Y^{S-W}_{ij} = (\beta^*_0 - \beta_0) + \gamma_1 z_{1j} + \gamma_2 z_{2j} + \gamma_{12} z_{1j} z_{2j} \\ + \gamma'_{11} x_{i1} z_{1j} + \gamma'_{12} x_{i1} z_{2j} + \gamma'_{21} x_{i2} z_{1j} \\ + \gamma'_{22} x_{i2} z_{2j} + \gamma'_{31} x_{i3} z_{1j} + \gamma'_{32} x_{i3} z_{2j} + e^S_{ij}, \end{aligned} \tag{1.22}$$

and the alternative model is given by

$$Y^{S-W}_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_{12} x_{i1} x_{i2} + \beta_{13} x_{i1} x_{i3} + \beta_{23} x_{i2} x_{i3} +$$

$$+\gamma_1 z_{1j} + \gamma_2 z_{2j} + \gamma_{12} z_{1j} z_{2j} + \gamma'_{11} x_1 z_{1j} + \gamma'_{12} x_1 z_{2j} + \gamma'_{21} x_2 z_{1j}$$
$$+\gamma'_{22} x_2 z_{2j} + \gamma'_{31} x_3 z_{1j} + \gamma'_{32} x_3 z_{2j} + e_i^W + e_{ij}^S \,. \tag{1.23}$$

The Model Adequacy Test results from R Studio are shown in Figure 1.17. The Alternative Model (Model 2) fits the data significantly better than the Null Model (Model 1), with a p-value < 0.0001. Therefore we should reject the null hypothesis in favor of the more complicated alternative model.

```
Analysis of Variance Table

Model 1: Y_SW ~ D + R + `D:R` + `T:D` + `T:R` + `P:D` +
`P:R` + `S:D` +
    `S:R`
Model 2: Y_SW ~ (T + P + S + D + R)^2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     21 390667
2     15  37151  6    353517 23.789 7.349e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’
0.1 ‘ ’ 1
```

Figure 1.17: Model Adequacy Test for PLA Experiment

When introducing the PLA data in Table 1.1, we noted that in some cases the WP and SP responses were the same for four out of the eight WP labels. The reason for this is because the experimenters did not find the maximum draw ratio, $R_{\max}$ to be much larger than the minimum value of 1 when the drawing temperature was set to its lower bound. For these WPs, the experimenters decided to set $R_{\min} = 1$ and decided not to actually perform the second stage experiment, opting instead to impute the WP response values to the SP response values. This was not a problem when they increased the drawing temperature to its high level. This means that the modeling assumptions we made are incorrect, and this more than likely explains the failure of the Model Adequacy Test.

1.6 Discussion

This paper presented a straightforward analysis method for split plot designs in which the WP and SP factors are applied sequentially, and additional responses are collected from the WPs prior to applying the SP factors. I showed that under certain model assumptions, the variance of the WP effect estimators can be dramatically decreased, and the estimator of σ_W^2 is guaranteed to be positive. The analysis method is practical and easy to interpret, involving fitting two traditional linear models. It is also not sensitive to the data structure, having similar performance for both balanced and unbalanced CRSPDs. The only hurdle comes with checking the model assumptions, which we have done using a simple F-test to compare nested models. If the test fails, we recommend only analyzing the SP responses but to investigate why the two models are different. A detailed investigation of how the test fails is the subject of future work, but we demonstrate some insights for the PLA experiment. Adding estimates of Monte-Carlo Standard Errors for alpha and the F-Statistic would be a useful addition to this component of future work.

I have chosen to focus on frequentist methods for this thesis, as they are more commonly used in practice, although Bayesian methods have been considered for the analysis of split plot designs. Gilmour and Goos (2009) discuss the analysis of split plot designs in industrial experiments, motivated by a study done at the University of Reading, UK, to study the effects of several process factors on the retention of volatile compounds in the freeze-drying of coffee. They used Bayesian methods and various priors to demonstrate their claim that a Bayesian approach is superior to REML analysis. It would be interesting to see how the sequential data could be put into a Bayesian framework.

REFERENCES

Alnosaier, W. (2023). Impact of negative variance component estimates on the kenward-roger test for fixed effects in linear mixed models. *The Egyptian Statistical Journal*, 67(1):1–18.

Bingham, D. R. and Sitter, R. R. (2001). Design issues in fractional factorial split-plot experiments. *Journal of Quality Technology*, 33(1):2–15.

Bisgaard, S. (2000). The design and analysis of blocked fractional factorial experiments. *Quality Engineering*, 12(4):637–650.

Box, G. E. P. and Jones, B. (1992). Split–plot design for robust production experiment. *Journal of Applied Statistics*, 19:3–26.

Fai, A. H. T. and Cornelius, P. L. (1996). Approximate f-tests of multiple degree of freedom hypotheses in the general linear model. *Journal of Statistical Computation and Simulation*, 54(2):101–119.

Gajjar, C. R., Stallrich, J. W., Pasquinelli, M. A., and King, M. W. (2021). Process–property relationships for melt-spun poly(lactic acid) yarn. *ACS Omega*, 6(25):16292–16301.

Gilmour, S. G. and Goos, P. (2009). Analysis of data from non-orthogonal multistratum designs in industrial experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(4):467–484.

Goos, P. and Vandebroek, M. (2001). Optimal split-plot designs. *Journal of Quality Technology*, 33(4):436–450.

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53(3):983–997.

Lawson, J. (2015). *Design and Analysis of Experiments with R*. Chapman and Hall/CRC.

Letsinger, L. S., Myers, R. H., and Lentner, M. (1996). Split-plot designs—experiences with applications to industrial problems. *Journal of Quality Technology*, 28(1):31–46.

Montgomery, D. C. (2020). *Design and Analysis of Experiments*. John Wiley & Sons, 10 edition.

Næs, T., Aastveit, A. H., and Sahni, N. S. (2007). Analysis of split-plot designs: An overview and comparison of methods. *Quality and Reliability Engineering International*, 23(8):801–820.

Schaalje, G. B., McBride, J. B., and Fellingham, G. W. (2001). Approximations to distributions of test statistics in complex mixed linear models using sas proc mixed. *Proceedings of the SAS Users Group International (SUGI 26)*. Paper 262-26.