Two-dimensional Partitioning of Yield Variation

G.W. Eaton, P.A. Bowen1, and P.A. Jolliffe
Department of Plant Science, University of British Columbia, Vancouver, B.C., Canada, V6T 2A2

Abstract. Total variation in yield per plant of long English cucumbers is partitioned by regression procedures into increments of variation due to successive morphological components. Variation in these increments is partitioned further by conventional analysis of variance. The results of this 2-way partitioning describe how treatments produce differences in yield through their effects on yield component variation. The method (2-dimensional partitioning, TDP) offers a simple way to account for growth and yield variation in crops.

Two continuing goals of crop research are to measure contributions to yield by yield components and to assess effects of treatments on those contributions. Suitable, but separate, analytical techniques exist for each of these goals. No convenient method, however, thus far has been available for the direct, combined analysis of yield components and treatment contributions to yield variation.

One statistical technique that has been used to examine variation in yield components is sequential yield component analysis (SYCA) (1, 2, 4, 5, 11). SYCA partitions total yield variation among orthogonal yield components and is well-suited to crops where yield is the product of a sequence of developmental processes. A related technique, sequential plant growth analysis (SPGA), includes indices of conventional plant growth analysis as yield components (8).

While SYCA and SPGA measure yield component contributions to yield variation, they do not directly assess treatment effects. On the other hand, the analysis of variance (ANOVA) can be used to detect treatment effects on individual yield components or yield (2, 5). This report describes the joint application of SYCA and ANOVA to partition total yield variation in 2 dimensions. In one dimension, yield variation is partitioned among orthogonalized yield components by linear regression analysis. In the 2nd dimension, variation is partitioned further among treatments and other factors of the experimental design. These 2 dimensions become the columns and rows, respectively, of the table in which the results are summarized.

The analysis of results from a simple experiment on hydroponically grown greenhouse cucumbers (1) is used to illustrate the 2-dimensional partitioning procedure.

Data used to exemplify the procedure were obtained during a study of hydroponically grown greenhouse cucumbers (1). These data were chosen because the experiment already had been described (1) and because both the ANOVA model and TDP results were relatively simple. Two treatments, a control and oxygen enrichment of the sawdust medium, were applied to individual plants in a completely random design with 13 replicates of each treatment. Oxygen enrichment was applied once per day for the first 17 days and was increased by one time per day in subsequent 10-day periods, reaching a maximum of 4 times per day. Plants were allowed to grow for 96 days. Fruit were harvested as they matured, starting on the 69th day.

The original data matrix consisted of the following 5 response variables expressed on a per-plant basis: \( U_1 \) = stem length (cm), \( U_2 \) = number of nodes, \( U_3 \) = leaf area (m²), \( U_4 \) = number of fruits, and \( U \) = total weight of cucumbers (g). \( U_1 \), \( U_3 \), and \( U_5 \) were measured at the final harvest.

Yield components were calculated as ratios of the response variables, then transformed to natural logarithms (3) so that correlations with yield may be spurious. With this assumption, the yield components were orthogonalized (10) in the order presented by the model. The first component remained the same, and each of the others was calculated as the residual from a multiple regression on developmentally earlier ones. Yield then could be expressed without residual error from the orthogonal independent variables in a multiple regression. The \( r^2 \) for each component corresponds to its incremental contribution to yield variation as it developed. These values are presented in the bottom line of Table 1. Note that the sum of these \( r^2 \)s is exactly 100%.

The regression procedure just described has been used previously in other studies (1, 4, 5, 7, 11) and is known as sequential yield component analysis (SYCA). Since all variation in yield was attributed to variation in yield components, both treatments and experimental error must have affected yield indirectly by affecting its components. To determine these effects, the orthogonal components were first scaled to units of \( \ln Y \). Scaling was accomplished by multiplying each orthogonal variate by its regression coefficient obtained from the multiple regression. Each scaled component then was analyzed by ANOVA, partitioning its variation into treatment and error sources. At this stage, the sum of the treatment and error sum of squares (SS) equaled the total SS for each component. Moreover the sum of these total SS for components equaled the total SS for \( \ln Y \). This partitioning of the total SS for \( \ln Y \) in 2 dimensions is shown in Table 2. Each \( X \) column contains the treatment and error contributions to yield component variation, and the bottom row contains yield component contributions to variation in \( \ln Y \). In Table 1, these SS are expressed as percentages of the total SS for \( \ln Y \).

To complete Tables 1 and 2 so that the sources of treatment and error variation in \( \ln Y \) are all accounted for, interactions between treatments and component pairs were added to the rows. Such interactions arise when treatments have opposing effects upon components of a pair, but both contribute in the

---

Table 1. Partitioning of yield variation as a percentage of the total SS for \( \ln Y \).

<table>
<thead>
<tr>
<th>Source</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
<th>Products</th>
<th>( \ln Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.19</td>
<td>0.00</td>
<td>0.10</td>
<td>19.08*</td>
<td>0.00</td>
<td>0.19</td>
<td>19.18*</td>
</tr>
<tr>
<td>Error</td>
<td>5.74</td>
<td>0.19</td>
<td>6.04</td>
<td>50.34</td>
<td>18.31</td>
<td>0.19</td>
<td>80.82</td>
</tr>
<tr>
<td>Total</td>
<td>5.94</td>
<td>0.19</td>
<td>6.13</td>
<td>69.43**</td>
<td>18.31*</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

*The \( X \) are the log-transformed, orthogonalized, and scaled yield components listed: \( X_1 \) = stem length/plant, \( X_2 \) = leaves/stem length, \( X_3 \) = leaf area/leaf, \( X_4 \) = fruits/leaf area, \( X_5 \) = weight/fruit, and \( \ln Y \) = natural logarithm of yield.

**Significant at \( P = 0.05 \) and 0.01, respectively. Significance in the treatment row refers to analysis of variance and in the total row refers to regression analysis.
same direction to yield. For example, a treat-
ment may increase one yield component but
decrease another so that there is a small net
effect on yield. In this case, both com-
ponents would have positive SS for treatments,
and their interaction would be negative. These
interactions are calculated as twice the treat-
ment sum of products for each component
pair. As the error value for this interaction
is opposite in sign to the treatment value,
their total is zero. To save space, we have
listed only the sum of these interactions in
Tables 1 and 2, because none was found to
be significant in this study. The 2-dimen-
sional partitioning of variation in ln Y is
complete in Tables 1 and 2 in the sense that
both the row sums and the column sums also
sum to the total SS for ln Y.

The main computational steps in TDP may
be summarized as follows: a) select and
measure primary variates on a common ba-
sis, such as a per plant basis; b) construct
ratios of those variates according to a chron-
ological sequence; c) transform ratios to log-
arithms (3); d) construct orthogonalized variates
(10); e) measure incremental contributions of
successive orthogonal variates; and f) partiti-
on sums of squares and cross products for
each variate according to the experimental
design.

In the present example of cucumber yield
components, yield was affected significantly
by the oxygen treatment (Tables 1 and 2).
Two-dimensional partitioning of SYCA re-
sults revealed that 2 components, fruits/leaf
area ($X_4$) and weight/fruit ($X_5$), were im-
portant contributors to yield (bottom lines of
Tables 1 and 2). Therefore, it can be inferred
that oxygen affected yield by influencing
either or both of these components. Since
oxygen affected fruits/leaf area but not weight/
fruit (columns $X_4$ and $X_5$ of Tables 1 and
2), the component source of the treatment
effect was fruits/leaf area. Variation in weight/
fruit contributed to variation in yield but was
not attributable to treatments. Possible sources
of this variation were either external factors
not considered in the ANOVA or earlier
component effects.

We choose to call the analysis presented in
this paper ‘2-dimensional partitioning’
(TDP) of yield variation. TDP was obtained
by combining ANOVA and SYCA, and it
produces a structural tabulation of sources of
yield variability. The ANOVA accounts for
sources of variation according to the exper-
imental design, while SYCA jointly assesses
variation in yield and yield components. TDP
is not just an alternative method of comput-
ing well-known statistics, but is an extended
framework to study yield components and
their relationships in response to treatment.
It should be evident that the TDP procedure
accommodates other experimental designs,
such as randomized complete blocks, Latin
squares, and split-plots.

The tables produced by the TDP analysis
are condensed summaries of the extent to
which treatment effects on yield components
are responsible for treatment effects on yield.
The chief advantage of the procedure lies in
this condensation. Ordinary SYCA requires a
separate table of results for each treatment
(1, 11), is cumbersome in complex experi-
ments, and the interpretation of treatment ef-
cfects requires subjective comparisons to be
made among the separate tables. In the TDP
procedure, treatment comparisons are done
objectively in one table.

### Literature Cited

fumigation of sawdust medium on the yield
and yield components of greenhouse cucum-
component analysis of winter damage and
flower buds in highbush blueberry. Scientia
Hort. 19:279–286.
rect logarithmic transformation for standard-
izing multiplicative trait variables. Crop Sci.
16:301–303.
component analysis in the cranberry. J. Amer.
5. Fraser, J. and G.W. Eaton. 1983. Applica-
tions of yield component analysis to crop
drology. The Iowa State Univ. Press, Ames.
Plant growth analysis: Additive and multi-
Doust. 1982. Sequential analysis of plant
9. Puri, Y.P., C.O. Qualset, and W.A. Wil-
liams. 1982. Evaluation of yield compo-
nents as selection criteria in barley breeding.
Crop Sci. 22:927–931.
10. Winer, B.J. 1971. Statistical principles in
experimental design, 2nd edition. McGraw-
Hill, New York.
cotton-grass upon yield components of

---

### Table 2. Partitioning of sum of squares for ln Y.

<table>
<thead>
<tr>
<th>Source</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>Products</th>
<th>ln Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>0.002</td>
<td>0.000</td>
<td>0.001</td>
<td>0.196**</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.197*</td>
</tr>
<tr>
<td>Error</td>
<td>0.059</td>
<td>0.002</td>
<td>0.062</td>
<td>0.517</td>
<td>0.188</td>
<td>0.002</td>
<td>0.830</td>
</tr>
<tr>
<td>Total</td>
<td>0.061</td>
<td>0.002</td>
<td>0.153</td>
<td>0.713**</td>
<td>0.188*</td>
<td>0.002</td>
<td>1.027</td>
</tr>
</tbody>
</table>

*The $X_i$ are the log-transformed, orthogonalized, and scaled yield components listed: $X_1 = \text{stem length/plant}$, $X_2 = \text{leaves/stem length}$, $X_3 = \text{leaf area/leaf}$, $X_4 = \text{fruits/leaf area}$, $X_5 = \text{weight/fruit}$, and ln $Y = \text{natural logarithm of yield}$

**Significant at $P = 0.05$ and 0.01, respectively. Significance in the treatment row refers to analysis of variance and in the total row refers to regression analysis.