

Lumpkin Foundation - Illinois Bundleflower Final Report

Spencer Barriball, Bo Meyering, Brandon Schlautman

2025-03-12

Introduction and Summary of Work

Materials and Methods

Accession Germination Test

Field Trial Design

Harvesting

Sample Processing

NIR Spectra Sampling

Plant tissues from harvet were ground and filtered using a 1mm sieve and scanned using the NeoSpectra NIR handheld scanner . . . FILL THIS OUT

Wet Chemistry

A total of 75 samples out of the entire population were selected for targeted wet chemistry analysis. Samples were selected out of the entire population using the Kennard Stone algorithm to select samples with the greatest Euclidean distance from each other, ensuring that the samples reflected the available NIR variance in the population. Samples were processed by xyz method, dried and sent to ABC labs for targeted wet chemistry analysis.. FILL THIS OUT

NIR Modeling

A total of ten wet chemistry metrics were selected for modeling. These included forage moisture content (%), forage dry matter content (%), ADF (units), NDF (units), relative feed value (units), total digestible nutrients (units), net energy gain (Mcal/cwt), net energy maintenance (Mcal/cwt), net energy lactation (Mcal/cwt), and forage yield (kg/Ha) all calculated on a dry matter basis. Spectra were presented as the percent reflectance per wavelength measured. Low quality spectra, i.e. low variance among the response values for all wavelengths, were filtered out of the dataset programatically. All spectra were then linearly resampled by wavelength to ensure equal wavelength intervals between each response value. The spectra were smoothed and detrended using a Savitsky-Golay filter by fitting a 5th order polynomial over a window length of 15 frames and taking the second derivative. Standard mean subtraction scaling was applied per wavelength, but were not scaled by the column variance. Training and testing populations (70% and 30%,

respectively) were selected from the calibration samples by using the Kennard-Stone selection algorithm. A partial least squares regression (PLSR) model was fit to the data for each of the target variables, allowing the number of model components to vary from 1 to 16 total components. Root Mean Squared Error (RMSE) defined below was used to evaluate each models performance on the training and testing populations.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

where y_i represents the ground truth target value of the i^{th} sample and \hat{y}_i represents the model prediction for that sample. The model with the lowest testing population RMSE was selected as the final model.

```
# Example dataset
set.seed(123)
data <- data.frame(
  Genotype = rep(paste0("G", 1:10), each=10),
  Location = rep(paste0("Loc", 1:5), times=20),
  Year = rep(2019:2020, each=5, times=10),
  Yield = rnorm(100, mean=500, sd=50)
)

# BLUES: Treat Genotype as a fixed effect
model_blue <- lm(Yield ~ Genotype + Location + Year, data=data)
blues <- emmeans(model_blue, ~ Genotype)
blues_df <- as.data.frame(blues)
print(blues_df)
```

```
##   Genotype   emmean      SE df lower.CL upper.CL
##   G1        503.7313 14.7434 85 474.4175 533.0451
##   G10       521.8547 14.7434 85 492.5409 551.1685
##   G2        510.4311 14.7434 85 481.1173 539.7449
##   G3        478.7721 14.7434 85 449.4582 508.0859
##   G4        516.1022 14.7434 85 486.7884 545.4161
##   G5        499.5642 14.7434 85 470.2504 528.8780
##   G6        511.0843 14.7434 85 481.7705 540.3981
##   G7        506.1542 14.7434 85 476.8404 535.4680
##   G8        481.8541 14.7434 85 452.5403 511.1679
##   G9        515.6548 14.7434 85 486.3409 544.9686
##
## Results are averaged over the levels of: Location, Year
## Confidence level used: 0.95
```

```
# BLUPs: Treat Genotype as a random effect
model_blup <- lmer(Yield ~ Location + Year + (1 | Genotype), data=data)
```

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

```
blups <- ranef(model_blup)$Genotype
blups_df <- data.frame(Genotype=rownames(blups), BLUP=blups[,1])
print(blups_df)
```

```
##   Genotype      BLUP
## 1      G1 -3.157477e-11
```

```
## 2      G10  6.936893e-10
## 3      G2   2.365388e-10
## 4      G3  -1.030394e-09
## 5      G4   4.634863e-10
## 6      G5  -1.983323e-10
## 7      G6   2.626786e-10
## 8      G7   6.538503e-11
## 9      G8  -9.070564e-10
## 10     G9   4.455797e-10
```

BLUE/BLUP

Selection Index

Results

Germination Results

```
germination_data <- fread('../data/lumpkin_ibf_germination_test.csv')
head(germination_data)
```

```
##      sample_id germination_test germination_pct hard_pct total_viable_pct
##      <char>          <char>          <num>      <num>          <num>
## 1:  IBF-KS-1      200 seed           16.0      15.5           31.5
## 2:  IBF-KS-2      200 seed           18.0      16.5           34.5
## 3:  IBF-KS-3      200 seed           21.0      24.0           45.0
## 4:  IBF-KS-4      200 seed           16.5      19.0           35.5
## 5:  IBF-KS-5      200 seed           18.5      18.0           36.5
## 6:  IBF-KS-6      200 seed           28.0      22.5           50.5
```

Yield Data

```
trial_layout <- fread('../data/trial_layout_metadata.csv') %>%
  mutate(sample_name = paste('LMP', plots, sep='-')) %>%
  dplyr::select(-V1)
yield_data <- fread('../data/lumpkin_ibf_forage_yield.csv') %>%
  left_join(trial_layout, by='sample_name') %>%
  dplyr::select(sample_name, trts, block, rep, row, column, forage_yield_kg_ha)

yield_mm <- lmer(forage_yield_kg_ha ~ trts + (1|block), data=yield_data)
anova(yield_mm)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##      Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## trts 224162941 2837506     79 227.06  1.4554 0.01718 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(yield_mm)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: forage_yield_kg_ha ~ trts + (1 | block)
## Data: yield_data
##
## REML criterion at convergence: 4097
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7993 -0.6403 -0.0022  0.5280  3.2584
##
## Random effects:
## Groups Name Variance Std.Dev.
## block (Intercept) 121474 348.5
## Residual 1949644 1396.3
## Number of obs: 310, groups: block, 4
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)    3362.2     719.6  182.9  4.673 5.75e-06 ***
## trtsID 1006 Year 88    1797.0     987.3  227.0  1.820 0.070066 .
## trtsID 1007 Year 90    2754.6     987.3  227.0  2.790 0.005720 **
## trtsID 1009 Year 90    1025.0     987.3  227.0  1.038 0.300320
## trtsID 1010 Year 88     958.4     987.3  227.0  0.971 0.332748
## trtsID 1014 Year 88    2571.1     987.3  227.0  2.604 0.009820 **
## trtsID 1015 Year 88    2716.2     987.3  227.0  2.751 0.006419 **
## trtsID 1017 Year 88    3191.9     987.3  227.0  3.233 0.001408 **
## trtsID 103 Year 88     2867.7     987.3  227.0  2.904 0.004042 **
## trtsID 1031 Year 90    1668.0    1067.3  227.2  1.563 0.119482
## trtsID 1032 Year 89    1721.7     987.3  227.0  1.744 0.082548 .
## trtsID 1037 Year 89    1623.1     987.3  227.0  1.644 0.101569
## trtsID 1040 Year 90    1841.6    1067.3  227.2  1.726 0.085791 .
## trtsID 1042 Year 89    2413.1     987.3  227.0  2.444 0.015286 *
## trtsID 1043 Year 90    1095.7     987.3  227.0  1.110 0.268256
## trtsID 1045 Year 88    1215.5     987.3  227.0  1.231 0.219561
## trtsID 1046 Year 88    3352.9     987.3  227.0  3.396 0.000807 ***
## trtsID 1047 Year 89    1596.3     987.3  227.0  1.617 0.107304
## trtsID 1056 Year 89    3949.4     987.3  227.0  4.000 8.57e-05 ***
## trtsID 1062 Year 88    1733.7     987.3  227.0  1.756 0.080449 .
## trtsID 1063 Year 88    1674.8     987.3  227.0  1.696 0.091198 .
## trtsID 1064 Year 89    1971.2     987.3  227.0  1.996 0.047076 *
## trtsID 1065 Year 90     996.8     987.3  227.0  1.010 0.313742
## trtsID 1067 Year 88    2510.6     987.3  227.0  2.543 0.011661 *
## trtsID 1070 Year 88    1303.5    1067.3  227.2  1.221 0.223204
## trtsID 1071 Year 89    1583.8     987.3  227.0  1.604 0.110086
## trtsID 1075 Year 88    2233.0     987.3  227.0  2.262 0.024668 *
## trtsID 1076 Year 88    2188.4     987.3  227.0  2.216 0.027652 *
## trtsID 1080 Year 90    1317.9    1067.3  227.2  1.235 0.218149
## trtsID 1081 Year 90    1806.1    1067.3  227.2  1.692 0.091968 .
## trtsID 1095 Year 90    1072.0     987.3  227.0  1.086 0.278735
## trtsID 1096 Year 90     865.2     987.3  227.0  0.876 0.381765
```

```

## trtsID 1097 Year 90 2768.0 987.3 227.0 2.804 0.005492 **
## trtsID 1099 Year 90 2376.3 987.3 227.0 2.407 0.016895 *
## trtsID 1100 Year 90 1917.5 987.3 227.0 1.942 0.053360 .
## trtsID 1101 Year 90 1484.0 987.3 227.0 1.503 0.134205
## trtsID 1104 Year 90 841.5 987.3 227.0 0.852 0.394972
## trtsID 1107 Year 90 2014.4 987.3 227.0 2.040 0.042482 *
## trtsID 1108 Year 90 1863.0 987.3 227.0 1.887 0.060447 .
## trtsID 1109 Year 90 2908.0 987.3 227.0 2.945 0.003562 **
## trtsID 1110 Year 90 1243.6 987.3 227.0 1.260 0.209111
## trtsID 1111 Year 90 1434.2 1067.3 227.2 1.344 0.180361
## trtsID 1112 Year 90 794.6 987.3 227.0 0.805 0.421760
## trtsID 1114 Year 90 3767.7 987.3 227.0 3.816 0.000175 ***
## trtsID 1115 Year 90 2419.3 987.3 227.0 2.450 0.015027 *
## trtsID 1117 Year 90 1970.4 987.3 227.0 1.996 0.047159 *
## trtsID 1118 Year 90 2406.3 987.3 227.0 2.437 0.015572 *
## trtsID 1119 Year 90 2246.3 987.3 227.0 2.275 0.023833 *
## trtsID 1120 Year 90 3507.6 987.3 227.0 3.553 0.000464 ***
## trtsID 1123 Year 90 2175.1 987.3 227.0 2.203 0.028601 *
## trtsID 1124 Year 90 2723.3 987.3 227.0 2.758 0.006286 **
## trtsID 1125 Year 90 1315.4 987.3 227.0 1.332 0.184111
## trtsID 1126 Year 90 1961.5 987.3 227.0 1.987 0.048162 *
## trtsID 1127 Year 90 3127.8 987.3 227.0 3.168 0.001747 **
## trtsID 1129 Year 90 2555.3 987.3 227.0 2.588 0.010275 *
## trtsID 1130 Year 90 2510.1 987.3 227.0 2.542 0.011677 *
## trtsID 1132 Year 90 1221.5 987.3 227.0 1.237 0.217305
## trtsID 1136 Year 90 2720.3 987.3 227.0 2.755 0.006341 **
## trtsID 1137 Year 90 1272.0 987.3 227.0 1.288 0.198941
## trtsID 1139 Year 90 2906.3 987.3 227.0 2.944 0.003582 **
## trtsID 1140 Year 90 1348.8 987.3 227.0 1.366 0.173250
## trtsID 1141 Year 90 1343.3 987.3 227.0 1.361 0.175015
## trtsID 1143 Year 90 1304.2 987.3 227.0 1.321 0.187842
## trtsID 1144 Year 90 4493.5 987.3 227.0 4.551 8.70e-06 ***
## trtsID 275 Year 88 2161.3 987.3 227.0 2.189 0.029614 *
## trtsID 282 Year 90 370.3 1067.3 227.2 0.347 0.728906
## trtsID 289 Year 88 1731.1 1067.3 227.2 1.622 0.106198
## trtsID 291 Year 90 2819.8 987.3 227.0 2.856 0.004689 **
## trtsID 294 Year 90 2450.9 987.3 227.0 2.482 0.013775 *
## trtsID 360 Year 90 1962.2 987.3 227.0 1.987 0.048080 *
## trtsID 371 Year 88 1300.5 987.3 227.0 1.317 0.189094
## trtsID 377 Year 90 1552.1 987.3 227.0 1.572 0.117339
## trtsID 379 Year 89 2630.8 987.3 227.0 2.665 0.008262 **
## trtsID 382 Year 88 1321.9 987.3 227.0 1.339 0.181954
## trtsID 391 Year 90 3226.8 987.3 227.0 3.268 0.001250 **
## trtsID 393 Year 88 1495.7 987.3 227.0 1.515 0.131203
## trtsID 413 Year 90 1845.8 987.3 227.0 1.870 0.062837 .
## trtsID 415 Year 89 2478.9 987.3 227.0 2.511 0.012744 *
## trtsKura Clover 254.4 987.3 227.0 0.258 0.796916
## trtsSainfoin 1123.7 1211.4 227.4 0.928 0.354581
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

##
## Correlation matrix not shown by default, as p = 80 > 12.
## Use print(value, correlation=TRUE) or

```

```
##      vcov(value)      if you need it
```

NIR Modeling

Model	Num. Components	Test RMSE	Train RMSE
Forage Yield	6	1331.441	1238.371
NEL	4	3.937	3.050
NEG	2	2.896	2.246
NEM	2	3.142	2.439
TDN	4	1.848	1.559
RFV	5	11.293	8.491
NDF	5	2.889	2.174
ADF	2	4.744	4.029
Crude Protein	2	1.159	1.053
Dry Matter Content	1	2.997	2.557
Moisture Content	1	2.997	2.557