PREDICTING THE CROP CYCLE FOR SUB-SAHARAN AFRICA

Abstract

The purpose of this study is to address drought risk to cowpea farmers by identifying possible mixed lots of seed that spread risk. The method employed was a Poisson regression model using multi-year data for flowering time for 160 cowpea accessions. Two possible pairs of varieties for mixed seed release were identified, 1393-1-2-3(-) paired with Cameroon 12-58 and Sasaque paired with Tvu-9557. This study provides a new tool that can be used to predict the crop cycle for the lines and environments in the data set. Further studies should consider consumer preferences, yield, disease resistance, and seed type.

Introduction

Cowpea (*Vigna unguiculata)* is a warm-season legume typically farmed as a staple by small-holder families in sub-Saharan Africa. Cowpea is a self-pollinated diploid species (Hall 1997). Cowpea cultivation began about 4,000 years before present (Hancock 2012).

Flowering time is an important trait for cowpea domestication and is often considered to be a part of domestication syndrome. The suite of traits could include, early flowering, self-pollination, non-shattering, and large seed size (Ladizinsky, 1998).

Cowpea is adapted to marginal environments and is more tolerant to abiotic stress than soybean (*Glycine max)*, for example (University of Florida Extension, 2013). An aspect of cowpea reproduction that is essential to the problem is that heat is a confounding factor to flowering (Mutters, 1998). Floral abortion occurs at 34°C in susceptible genotypes. Therefore, lines that flower normally under heat and drought stress must be identified to produce food for Africa in the context of climate change.

Discussion of drought risk mitigation has maintained its relevance due to the expected effects of climate change (Hassan 2019). Drought can occur during cowpea cultivation at the early seedling stage, mid-season vegetative phase, or during the late-season pod filling stage. One way to help ensure a harvest would be to plant both early and late season varieties. If farmers can purchase a single bag of seed that has both early and late season plants, they can possibly escape a late or midseason drought.

Purpose

The objectives of this paper include to:

- Address the risk of drought to smallholder cowpea farmers by considering mixed lots of seeds that spread risk,
- Identify late and early flowering cultivars that can be paired together and marketed as a single variety in West Africa,
- Predict flowering time in different growing environments in Burkina Faso,
- Compare observed mean values for flowering time with the predicted response mean to assess the reliability of the model, and to
- Create a framework for obtaining the probability of a cowpea accession flowering at the mean flowering time.

One option to spread drought risk would be to stagger plantings of the same cultivar. However, this creates more work for farmers and requires more resources. Additionally, the seeds could be delivered as near-isogenic lines. The NILs could be made by introgression of an early flowering segment of the chromosome into a late flowering line. After a few rounds of backcrossing, both of the NILs could be sold together in the same bag as a single cultivar.

However, since UC Riverside has the third-largest germplasm collection in the world with over 5,500 accessions, it is possible that there are compatible pairs of seed types in existence. The lines could be readily increased by self-pollination. That would allow quicker delivery of the seeds to farmers than going through the marker-assisted breeding cycle.

One of the most important traits for adapting to climate change is drought tolerance. Drought tolerance is linked to early flowering time (Muchero, 2013). Cowpea is attacked by the opportunistic plant fungal pathogen *Macrophomina phaseolina* under drought conditions. *Macrophomina* resistance is linked to late flowering time (Muchero, 2011). Given this, it would be difficult to pyramid traits for drought tolerance and *Macrophomina* using traditional breeding methods and current sources of resistance. However, it may be possible to circumvent the necessity to breed both of these resistances into a single cultivar if a late and early variety could be marketed together in the same seed lot.

Although there is an intuitive aspect of farming the land, farmers will want to know what to expect from their purchased seed varieties. It is important to use statistical inference to aid farmers to make crop decisions. Planting these lines does not produce the same mean value for days to flowering in any given season. Mathematical models can be used to predict the mean values for days to flowering.

Hypothesis Testing

The null hypotheses to be tested in this model are the following:

- H01: T1=T2=T3….Ti=0, in i=1 to *a* cowpea accessions

The cowpea accession grown is not a significant factor in determining flowering time.

- H02: B1=B2=B3….Bk=0, in k=1 to *b* environments

The environment in which cowpea is grown is not a significant factor in determining flowering time.

Tukey's Test

The null hypotheses tested by pairwise comparisons are the following:

H01: $\mu1=\mu2$

H02: $\mu1=\mu3$

H03: $\mu1=\mu4$

H04: $\mu2=\mu3$

H05: $\mu2=\mu4$

H06: $\mu3=\mu4$

Where $\mu1$= mean of environment 1, $\mu2$=mean of environment 2, $\mu3$=mean of environment 3, and μ4 is equal to environment 4.

Methods

Data Collection

50% flowering time is defined as a count of the number of days until 50% of the plants in the plot have open flowers.

Data for 50% flowering time was collected in Burkina Faso from 2008-2009 in two locations.

Table 1

Summary of climatic conditions for cowpea field experiments conducted in Burkina Faso (Muchero, 2013)

Experiment	Months	Max. Air Temp (°C)	Min. Air Temp (°C)	Evapo-transpiration (mm)
BF_2008	Aug, Sept, Oct	32.1, 33.1, 36.8	22.8, 23.0, 22.5	40.0, 50.0, 58.1
BF_2009A&B	Aug, Sept, Oct	31.8, 34.0, 35.1	24.2, 23.5, 28.1	58.1, 58.0, 76.1
BF_2009C	Aug, Sept, Oct,	31.1, 33.6, 37.0	22.5, 23.0, 24.2	62.0, 50.0, 59.3

In Table 1 above, environment 1 is BF_2008 from Kamboinse, environments 2 and 3 are BF_2009A&B in Pobe, and environment 4 is BF_2009C from Kamboinse, Burkina Faso. Using the model is dependent on knowing the climatic conditions of the area.

Cowpea was grown in a completely randomized design with three replications in each environment. The average flowering time values for all replications in each environment were used as the flowering time values for each accession and environment in the model. Data from 160 cowpea accessions was used for the analysis.

Model Selection

Statistical analysis was carried out in R. The first model that was considered for this analysis was the two-way analysis of variance (ANOVA). The proposed model was

$$Y = Ti + Bj + (TB)ij + Eijk$$

Where T represented the effects of the cowpea accession used, B represented the environmental effects, TB was the genotype by environ interaction, and E was the error.

The summary of the ANOVA indicated that the interaction between the two factors was significant for some lines. A closer look at the output showed that the factorial ANOVA model was not appropriate. The plots below were generated to assess the homogeneity of variances assumption and normality of the residuals.

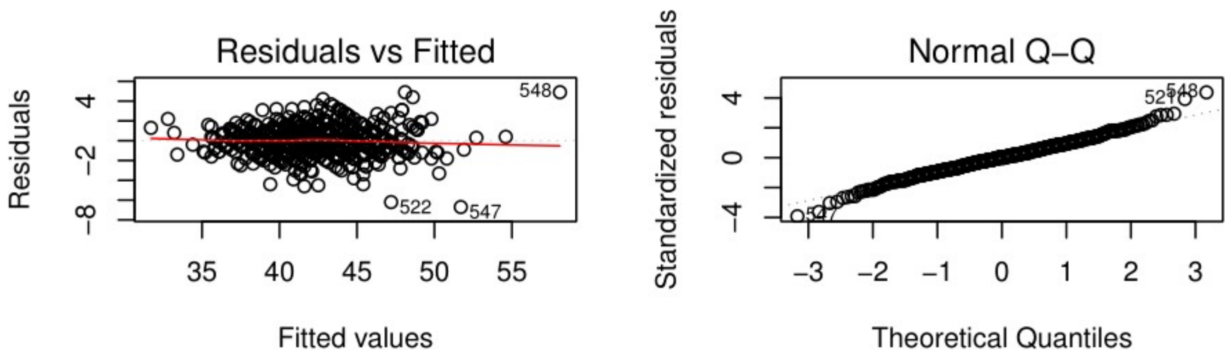

Figure 1. The res*fitted plot for cowpea flowering time shows a megaphone shape and the QQ plot suggests that the data may be nonnormal.

In addition to the evidence in Figures 1 and 2, the Jarque-Bera test for normality was used and the output is shown below.

Table 2

Results of Jarque-Bera Normality Test for cowpea flowering time data

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 65.6155
  P VALUE:
    Asymptotic p Value: 5.662e-15
```

Jarque-Bera's test provided strong evidence of nonnormality (p=5.7x10^-15) and it was clear that another model would have to be considered (see Table 2 and Figure 2).

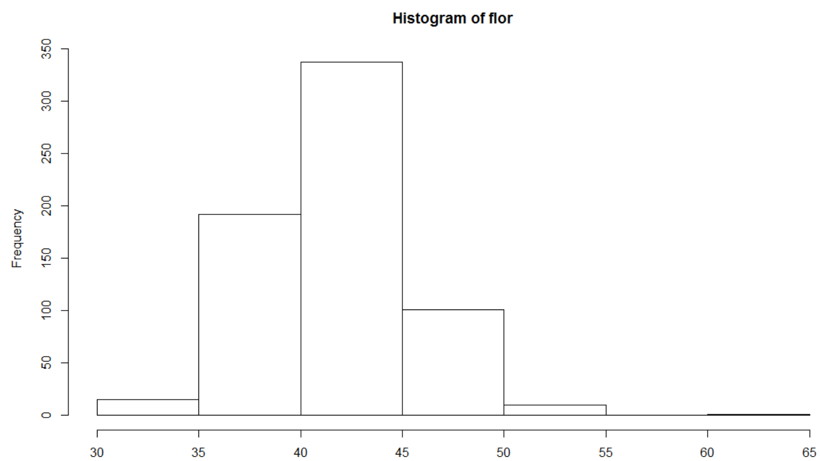

Figure 2. The histogram of cowpea flowering time data allows us to visualize the skewed distribution.

In light of these preliminary findings, a two-factor Poisson regression model was considered. The Poisson model is a good candidate when working with count data such as flowering time (Fávero 2019). The model to be tested is

$$\ln(\mu i) = \mu + \beta + T$$

where μ is the global mean and variance, β is the effect of the environment, and T is the effect of the cowpea accession grown.

Findings

Table 3

Predicted and observed values for the cowpea days to 50% flowering trait generated by Poisson regression

Accession	Predicted Days to 50% Flower	Observed Days to 50% flowering	Difference	Significant p-value
1393-1-2-3(-)	37 days	39 days	2	yes
Cameroon12-58	52 days	49 days	3	yes
CRSPNiebe	47 days	46 days	1	yes
Sasaque	34 days	34 days	0	no
Tvu-9557	49 days	51 days	2	no
Montiero	48 days	46 days	2	yes
Tvu-16594	47 days	47 days	0	yes
UCR2567	49 days	49 days	0	yes

Using Table 3, we can identify some potential pairs of lines that could be combined into a cultivar release with a ratio of x/y days to 50% flowering. A potential pair could be 1393-1-2-3(-) with Cameroon 12-58. The ratio for 1393-1-2-3(-) and Cameroon 12-58 is 0.72.

Another potential pair is Sasaque with Tvu-9557. The ratio for Sasaque and Tvu-9557 is 0.71. A third match is Sasaque with Cameroon12-58. The ratio for Sasaque with Cameroon12-58 is 0.66.

The Poisson model has been used as a quantitative tool in risk evaluation for credit decisions, for example (Rongda 2014). In this study, a lower ratio means that there is a larger spread of the days to flowering. That could translate to a better spread of the drought risk for farmers. However, the predictors did not have a significant p-value for Sasaque or Tvu-9557. This reduces the confidence level in the estimations for these lines. Nevertheless, it is worth investigating if these seed types could be compatible with consumer preferences. Increasing confidence in these predictions requires the evaluation of the overall fitness of the model.

Table 4

Evaluation of the Poisson regression model for cowpea days to 50% flowering with environment and accession included as factors

```
Analysis of Deviance Table (Type II tests)

Response: flor
        LR Chisq  Df Pr(>Chisq)
acc      138.647 163    0.91699
environ    3.167   1    0.07515 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A unexpected result in Table 4 is that accession is not a significant factor, whereas some lines are known to be early or late, such as Early Scarlet. Environment is almost significant (p=0.075), which suggests that the model can be improved.

Table 5

Test for interaction in the cowpea 50% flowering time Poisson regression model

```
Analysis of Deviance Table

Model 1: flor ~ acc
Model 2: flor ~ acc + environ
  Resid. Df Resid. Dev Df Deviance
1       492     60.770
2       491     57.603  1   3.1668
```

There is no significant evidence of interaction for accession and environment according to Table 5. However, some of the lines had significant genotype by environment interactions (see supplemental materials). These included both of the late flowering accessions, Cameroon12-58 and Tvu-9557. The model was run taking only environment into consideration to analyze the result of excluding the accession effects.

Table 6

Evaluation of the Poisson regression model for cowpea 50% days to flowering with the accession variable dropped

```
Analysis of Deviance Table (Type II tests)

Response: flor
        LR Chisq Df Pr(>Chisq)
environ   17.768  3  0.0004912 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Indeed, with the model

$$\ln(\mu i) = \mu + \beta$$

there is a highly significant effect for environment, as seen in Table 6. Therefore the model is improved by dropping the variable accession. In addition, this model allows us to perform Tukey’s test since we had a positive Chi square result for environment ($p<0.0005$). The results of Tukey’s test are shown below in Table 7.

Table 7

Results of Tukey's multiple comparison test for the environments where cowpea was grown

```
	 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: glm(formula = flor ~ environ)

Linear Hypotheses:
           Estimate Std. Error z value Pr(>|z|)
2 - 1 == 0 -0.40854    0.39326  -1.039  0.72658
3 - 1 == 0  0.97561    0.39326   2.481  0.06305 .
4 - 1 == 0  0.88415    0.39326   2.248  0.11026
3 - 2 == 0  1.38415    0.39326   3.520  0.00262 **
4 - 2 == 0  1.29268    0.39326   3.287  0.00564 **
4 - 3 == 0 -0.09146    0.39326  -0.233  0.99557
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

The results of Tukey's test show that the growing conditions in environment 3 and 2 were significantly different in terms of their effect on days to 50% flowering ($p<0.003$). In addition, the p-value of 0.006 for the comparison of environment 4 to environment 2 is an indication that the conditions were significantly different.

Next, it is useful to look at our four chosen accessions in more detail. It is important to evaluate if there is a big difference in predicted values across environments. For this analysis, we are forced to rely on the first model which had accession included in order to make the prediction. The result of the analysis is shown in Table 8.

Table 8

Predicted days to 50% flowering detailed by environment for the four candidate accessions for variety release

Accession	Environment	Predicted 50% flowering
1393-1-2-3(-)	Environment 2	37 days
1393-1-2-3(-)	Environment 3	38 days
1393-1-2-3(-)	Environment 4	38 days
Cameroon12-58	Environment 2	52 days
Cameroon12-58	Environment 3	52 days
Cameroon12-58	Environment 4	53 days
Sasaque	Environment 2	34 days
Sasaque	Environment 3	34 days
Sasaque	Environment 4	35 days
Tvu-9557	Environment 2	48 days
Tvu-9557	Environment 3	49 days
Tvu-9557	Environment 4	49 days

According to Table 8, there is not an obvious difference in the number of days to 50% flowering across environments. However, there is a general trend that in environment 4 the plants take longer to flower. This also can give farmers the understanding that they may experience a small difference in the flowering time we predict versus the days to flower in their field.

Discussion and Conclusion

Two possible pairs of varieties for mixed seed release were identified, 1393-1-2-3(-) paired with Cameroon 12-58 and Sasaque paired with Tvu-9557. To further determine whether the traits for these accessions will conform to consumer preferences, it will be necessary to look at yield, disease resistance, and seed type. A possible improvement to the study would be to group the accessions into early and late binomial categories to analyze the effects for the accession grown. It would also be helpful to augment the data with more environments in order to expand the applicability of the predictions. However, this study provides a new tool that we can use to predict the crop cycle for the lines and environments in the data set.

Acknowledgements

This research was conducted with support from USAID Feed the Future Innovation Lab for Climate Resilient Cowpea (Cooperative Agreement AID-OAA-A-13-00070) at UC Riverside and NIH BUILD PODER at CSU Northridge. Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number RL5GM118975. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Thank you to the TJ Close Lab at UCR for providing data, and Ndeye Diop for data collection. Thank you to Elias Tarver, Aerospace Engineer for technical writing assistance. Thank you to Yoshie Hanzawa, PhD for reviewing and editing the manuscript.

Conflict of Interest Statement

The author declares that there is no conflict of interest regarding the publication of this paper.

Data Availability Statement

Cowpea Flowering time data is available in supplemental materials.

REFERENCES

Luiz Paulo Fávero, Patrícia Belfiore, Chapter 15 - Regression Models for Count Data: Poisson and Negative Binomial, Editor(s): Luiz Paulo Fávero, Patrícia Belfiore, Data Science for Business and Decision Making, Academic Press, 2019, Pages 617-703, ISBN 9780128112168, https://doi.org/10.1016/B978-0-12-811216-8.00015-X.

Hall, A., Singh, B., & Ehlers, J. (1997). Cowpea Breeding. *Plant Breeding Reviews, 15*, 215-174.

Hancock, J. (2012). *Plant Evolution and the Origin of Crop Species*. Wallingford, Oxfordshire, UK ; Cambridge, MA: Cabi.

A.G. Hassan, M.A. Fullen, D. Oloke, Problems of drought and its management in Yobe State, Nigeria, Weather and Climate Extremes, Volume 23, 2019, 100192, ISSN 2212-0947, https://doi.org/10.1016/j.wace.2019.100192.

Ladizinsky, G. (1998). *Plant Evolution under Domestication*. Dordrecht; Boston: Kluwer Academic Publishers.

Muchero, W., Ehlers, J.D., Close, T.J., Roberts, P.A. (2011). Genic SNP markers and legume synteny reveal candidate genes underlying QTL for *Macrophomina phaseolina* resistance and maturity in cowpea [*Vigna unguiculata* (L) Walp.]. BioMed Central Genomics 12:8.

Muchero, W., Roberts, P.A., Diop, N.N., Drabo, I., Cases, N., Close, T.J., Muranaka, S., Boukar, O., Ehlers, J.D. (2013). Genetic architecture of delayed senescence, biomass, and grain yield under drought stress in cowpea. PLOS ONE 8: e70041.

Mutters, R., Hall, A., & Patel, P. (1989). Photoperiod and Light Quality Effects on Cowpea Floral Development at High Temperatures. *Crop Science,* 1501-1501.

Rongda Chen and Huanhuan Yu, "Risk Measurement for Portfolio Credit Risk Based on a Mixed Poisson Model," Discrete Dynamics in Nature and Society, vol. 2014, Article ID 597814, 9 pages, 2014. https://doi.org/10.1155/2014/597814.

University of Florida Extension. (2013). Annual Warm-Season Legumes for Florida and the US Gulf Coast: Forage Yield, Nutritional Composition, and Feeding Value. Retrieved November 30, 2014, from http://edis.ifas.ufl.edu/.