

# Modelling landrace structure in under-utilised crops: the bambara groundnut example.

L.R. Band<sup>1,2</sup>, S.P. Preston<sup>2</sup>, S. White<sup>2</sup>, M. Tindall<sup>3</sup>.

Problem presented by Sean Mayes<sup>4</sup> and Asha Karunaratne<sup>4</sup>.

1. Centre for Plant Integrative Biology, University of Nottingham,
2. School of Mathematical Sciences, University of Nottingham,
3. School of Biological Sciences, University of Reading,
4. School of Biosciences, University of Nottingham.

June 30, 2009

Bambara groundnut is an under-utilised legume crop species which is grown primarily by subsistence farmers in sub-Saharan Africa. Bambara groundnut is tolerant to drought, reasonably free of diseases and pests, and adapted to poor soils; however, despite these advantageous traits, bambara groundnut is not widely grown, and its limited use is thought to be due to low and unpredictable yields, lengthy cooking and processing times, and the cultural perception that bambara groundnut is a ‘woman’s crop’. Improving the yields and perception of under-utilised crops, such as bambara groundnut, could greatly improve future food security in semi-arid environments.

Bambara groundnut exists as landraces, which are dynamic plant populations that have historical origin, distinct identity and lack formal crop improvement; landraces are often locally adapted and associated with traditional farming systems. Research at the Tropical Crop Research Unit at the University of Nottingham focuses on two bambara groundnut landraces: S19-3 which originates in Namibia and is adapted to hot and dry environments, and Uniswa red which originates in Swaziland and is adapted to cold and wet environments. Within a landrace, there can be a range of genotypes, and the genetic diversity can be determined using molecular genetics; initial analysis suggests that the Uniswa red landrace is genetically more diverse than S19-3, and larger-scale investigations are underway to confirm these findings and determine a quantitative measure of the diversity. Despite this genetic diversity, adaptive traits (such as days to harvest) are likely to be reasonably fixed within a given landrace, due to adaptation to the environment in which the landrace is regularly grown (and selected). In contrast, genes for agronomic traits may be highly variable within a landrace.

It is currently not well understood how genetic diversity within a landrace affects the total yield on a field scale, and this report discusses how the relationship can be investigated using statistical analysis and mathematical modelling. The study is motivated by the BAMLINK integrated research project (funded by the European Union Framework 6) that is investigating the barriers to further uptake of bambara groundnut. As part of the BAMLINK project, a field-scale model of bambara groundnut growth (BAMGRO) has been

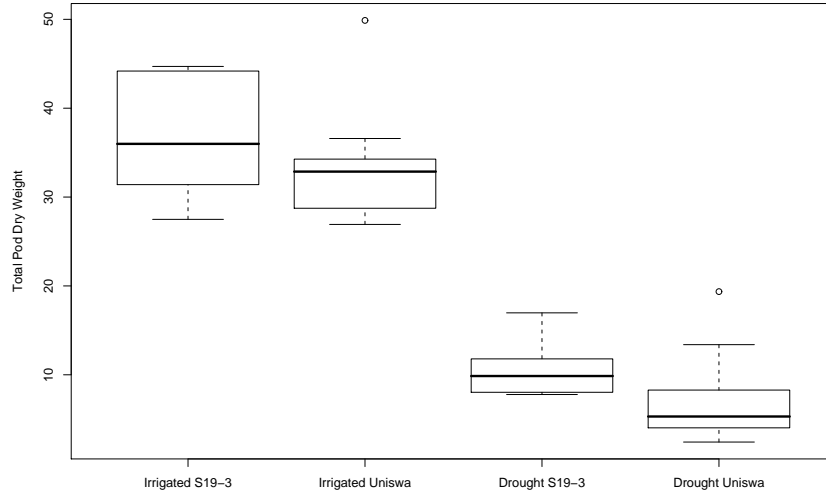


Figure 1: Box and whisker plots comparing the effects on total pod dry weight (grams per plant) of the landrace and irrigation factors.

developed. This model has been calibrated for the Uniswa red landrace using physiological data collected at the Tropical Crop Research Unit. The BAMGRO model is deterministic and on the field scale; however, it would be hugely beneficial to extend the model to incorporate stochastic effects due to both genetic diversity and environmental influences. The report is structured as follows. In section 1, we statistically analyse the physiological data provided by the Tropical Crops Research Unit to assess whether the different genetic diversities exhibited in the S19-3 and Uniswa red landraces lead to differences in the variations of physiological traits. We then consider how to incorporate genetic diversity into a modelling framework. In section 2, we mathematically model the growth of a single plant. We focus on two developmental stages for simplicity; we take the start of the second phase to be determined by the genetically prescribed ‘pod initiation’, and show how pod initiation affects the plant’s total yield. Finally in section 3, we use the single plant model (from section 2) to model a field of  $N$  independent plants, and assess how variations in ‘podding initiation’ affect the yield on the field scale.

## 1 Statistical Analysis of Physiological data

We wish to investigate the relationship between physiological traits of different landraces under different growing conditions. Our dataset consists of measurements for 40 plants of the total pod dry weight (i.e. the response variable) for four combinations of two different factors (i.e. the input variables): the landrace, either i) Uniswa red or ii) S19-3, and the irrigation conditions, either i) well irrigated or ii) drought conditions. The data are summarised by the boxplots shown in figure 1.

## 1.1 Linear models

Initially, consider the simple linear model

$$y_i = \beta_0 + \beta_L x_L + \beta_I x_I + \epsilon_i, \quad (1.1)$$

where  $i$  indexes each plant,  $y_i$  is the total pod dry weight,  $x_L$  and  $x_I$  are indicator variable for the two factors (such that  $x_L$  equals one for Uniswa red and zero for S19-3, and  $x_I$  equals one for plants that are well irrigated and zero for those grown under drought conditions), and constants  $\beta_0, \beta_L, \beta_I$  are to be chosen to minimise the sum of the squares of the residuals  $\epsilon_i$ .

The common assumption in linear modelling is that the residuals are normally distributed with common variance, i.e.  $\epsilon_i \sim N(0, \sigma^2)$ . Under this assumption, fitting the model to the data gives the following results:

	Estimate	Std. Error	t-value	p-value
Intercept, $\beta_0$	10.538	1.480	7.123	1.95e-08
Uniswa, $\beta_L$	-3.312	1.708	-1.939	0.0602
Irrigated, $\beta_I$	26.004	1.708	15.221	<2e-16

The interpretation of the first column is that the mean total pod dry weight of an S19-3 plant under drought conditions is 10.538; the mean is lower by  $-3.312$  with Uniswa red and increases by 26.004 with irrigation. The standard errors in the second column are estimates of the standard deviation of the sampling distribution of the parameter estimates. The final two columns correspond to a t-test of the hypothesis that the true value of each of the parameters estimated in the first column equals zero; small p-values are evidence against this hypothesis. Hence there is strong evidence that  $\beta_0$  and  $\beta_L$  are non-zero, and weak evidence that  $\beta_I$  is also non-zero. Thus irrigation leads to a significant increase in the mean total pod dry weight, and using S19-3 over Uniswa red also leads to a (weakly) significant increase.

There are two disadvantages, however, of using a linear model of this type: (i) it is concerned only about the effects of the two factors on the response (i.e mean total pod dry weight) so we can say nothing about the effect of landrace on the variance of the response, and (ii) the tests require the residuals to be normally distributed; see figure 2 for a qq plot of the residuals of (1.1).

One way to address (i) is to extend the model to incorporate a ‘random effect’ from the landraces, leading to a so-called ‘mixed effects’ model,

$$y_i = (\beta_{L1} + \alpha_{L1})x_{L1} + (\beta_{L2} + \alpha_{L2})x_{L2} + \beta_I x_I + \epsilon_i, \quad (1.2)$$

where  $\{x_{L1}, x_{L2}\}$  equals  $\{1, 0\}$  for Uniswa red and  $\{0, 1\}$  for S19-3, and  $\alpha_{L1} \sim N(0, \sigma_{L1}^2)$  and  $\alpha_{L2} \sim N(0, \sigma_{L2}^2)$ . However, this model is difficult to analyse, and tests based on it still require that the residuals are normally distributed.

## 1.2 Two-sample tests

A simpler approach is consider irrigated and drought conditions separately and perform two-sample tests on the effects of landrace: we can use a t-test to test whether the responses have equal means ( $\mu_U$  and  $\mu_S$ ),

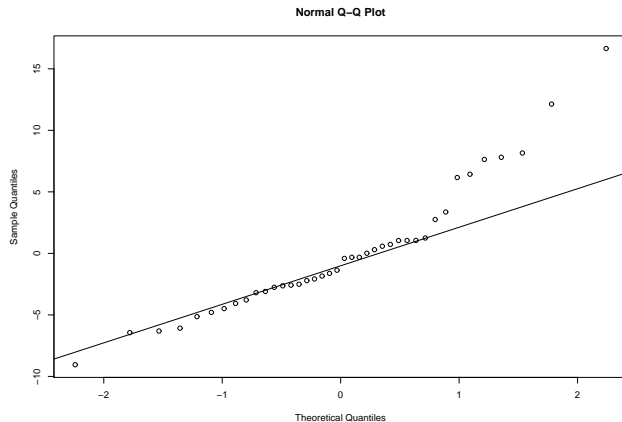


Figure 2: qq plot for the residuals of the simple linear model (1.1) showing deviation from normality.

and an F-test to test whether the variances ( $\sigma_U$  and  $\sigma_S$ ) are equal. Standard tests that rely on asymptotic null distributions still assume that residuals are normally distributed, but this problem can be avoided by using a ‘non-parametric’ procedure called bootstrapping. Bootstrapping involves repeatedly resampling from the samples in a way that mimics sampling under the null hypothesis; in this way, calculating the test statistic for each resample provides an estimate of the distribution of the test statistic under the null hypothesis.

The results are:

Condition	Null & alternative hypotheses	p-val (standard)	p-val (bootstrap)
Irrigated	$H_0 : \mu_U = \mu_S, H_1 : \mu_U < \mu_S$	0.129	0.131
Irrigated	$H_0 : \sigma_U = \sigma_S, H_1 : \sigma_U > \sigma_S$	0.414	0.425
Drought	$H_0 : \mu_U = \mu_S, H_1 : \mu_U < \mu_S$	0.053	0.075
Drought	$H_0 : \sigma_U = \sigma_S, H_1 : \sigma_U > \sigma_S$	0.048	0.142

Note that the standard and bootstrap p-values are similar in the irrigated conditions, but are notably different in drought conditions (in which the qq plots show that the data deviate more from the normal distribution). Under irrigated conditions there is little evidence against different means and variances for Uniswa red and S19-3. Under drought conditions there is weak evidence that the Uniswa red has a lesser mean, and there is evidence, weaker still, that its variance is greater.

In summary, we used linear models (section 1.1) to show that irrigation increases the average total pod weight, and that the Uniswa red landrace appears to have a smaller average pod weight than S19-3 (although this evidence is weak). We then considered the irrigated and drought conditions separately using a two-sample test (section 2.2) and found that under irrigated conditions, the average pod weight is not significantly different between the two landraces; however, under drought conditions, the Uniswa red landrace appears to have a smaller average pod weight than S19-3, which is in agreement with the linear analysis in section 1.1. This analysis confirms that the landraces have become adapted to their conditions

— Uniswa red is normally grown in a wet environment, whereas S19-3 is grown in a dry one, therefore we would expect S19-3 to perform better under drought conditions.

## 2 Mathematical model of the growth of a single plant

### 2.1 Model assumptions

We now use assumptions from the BAMGRO field-scale model together with qualitative understanding of the physiological data to model the growth of a single plant and predict the temporal variations in the plant's leaf mass,  $L(t)$ , and pod mass,  $P(t)$ , where  $t$  denotes time in days.

The plant gains energy from sunlight through photosynthesis (as in the BAMGRO model), with the energy input dependent on the mass of leaves. As the canopy forms, leaves overlap and therefore we prescribe a saturating relationship between the mass of leaves and the rate of energy input and use a Michaelis-Menten form, denoting the maximum rate of energy input by  $k_i$  and the mass of leaves at which the energy input is half this maximum by  $K_l$ . We assume that the plant uses energy for growth of leaves and pods, with negligible energy used for plant maintenance or root growth. We consider two distinct developmental phases: i) pre 'pod initiation', when the plant uses its energy to make leaves; and ii) post 'pod initiation', when the plant uses its energy to make pods (and no leaf growth occurs). We denote the rates at which the plant uses energy to make leaves and pods by  $k_l$  and  $k_p$  respectively, and suppose that the mass of leaves and pods produced per unit energy are given by  $m_l$  and  $m_p$  respectively. We also suppose that leaves decay, with a constant rate of decay,  $\lambda$ . We note that the model describes a fully-irrigated plant that is not at a temperature extreme, as it does not include the impact of water and heat stress.

The time at which a plant switches between its developmental phases depends on the temperature. It is generally supposed that plants are genetically programmed to move to a new developmental phase at a prescribed thermal time. Thermal time,  $T_h$ , is a cumulative measure of how much heat the plant has been exposed to, and is related to time,  $t$ , via

$$\frac{dT_h}{dt} = T_e(t) - B, \quad \text{for } T_e > B, \quad (2.1a)$$

$$\frac{dT_h}{dt} = 0 \quad \text{for } B > T_e, \quad (2.1b)$$

where  $T_e(t)$  is the temperature and  $B$  is the base temperature above which thermal time increases (with the rate of increase equal to the difference between the temperature and the base temperature). In the model, we consider only two developmental stages and suppose that pods initiate at the prescribed thermal time,  $T_h = T_h^*$ .

Thus, we can model the plant's total energy by

$$\frac{dE}{dt} = k_i \frac{L}{L + K_l} - k_l EH(T_h^* - T_h) - k_p EH(T_h - T_h^*), \quad (2.2)$$

where  $H$  denotes the Heaviside function. The first term on the right-hand side represents the energy input from water and sunlight, the second term is the energy loss due to leaf growth when  $T_h < T_h^*$  and the third term is the energy loss due to pod growth when  $T_h > T_h^*$ . The plant's energy increases the leaf mass for

$T_h < T_h^*$ , and leaf mass reduces due to leaf decay, therefore

$$\frac{dL}{dt} = k_l m_l E H(T_h^* - T_h) - \lambda L. \quad (2.3)$$

Finally, for  $T_h > T_h^*$ , the plant's energy increases the pod mass

$$\frac{dP}{dt} = k_p m_p E H(T_h - T_h^*). \quad (2.4)$$

To summarise, the plant growth is described by four coupled ordinary differential equations, (2.1–2.4). We suppose that the plant has initially no energy and no pods, that thermal time is zero and that the leaf weight is  $0.1K_l$  (note that due to the form of the equations, we must start with a nonzero leaf weight to enable the plant to gain energy). Thus, the initial conditions are

$$T_h = E = P = 0, \quad L = 0.1K_l \quad \text{at} \quad t = 0. \quad (2.5)$$

### 2.1.1 Nondimensionalisation

We nondimensionalise the model as follows

$$t = \frac{1}{k_p} \hat{t}, \quad T_h = T_h^* \hat{T}_h, \quad T_e = B \hat{T}_e, \quad (L, P) = K_l (\hat{L}, \hat{P}), \quad E = \frac{k_i}{k_p} \hat{E}, \quad (2.6)$$

where hats denote dimensionless variables. We consider the time scale based on the rate at which energy is used to make pods during the second developmental stage. The pod-initiation thermal time,  $T_h^*$ , is a typical thermal time, and the base temperature,  $B$ , is a typical temperature. The mass of leaves and pods are nondimensionalised using the leaf mass for which the energy adsorption is half of the maximum energy adsorption, and finally, energy is nondimensionalised using the maximum possible energy during the second developmental stage  $k_i/k_p$ . The model then depends on five dimensionless parameters, namely

$$\bar{B} = \frac{B}{k_p T_h^*}, \quad \bar{k}_l = \frac{k_l}{k_p}, \quad \bar{m}_l = \frac{m_l k_i}{k_p K_l}, \quad \bar{m}_p = \frac{m_p k_i}{k_p K_l}, \quad \bar{\lambda} = \frac{\lambda}{k_p}. \quad (2.7)$$

Nondimensionalising (and dropping the hats), the governing equations, (2.1–2.4), become

$$\frac{dT_h}{dt} = \bar{B}(T_e - 1)H(T_e - 1), \quad (2.8a)$$

$$\frac{dE}{dt} = \frac{L}{L+1} - \bar{k}_l E H(1 - T_h) - E H(T_h - 1), \quad (2.8b)$$

$$\frac{dL}{dt} = \bar{k}_l \bar{m}_l E H(1 - T_h) - \bar{\lambda} L, \quad (2.8c)$$

$$\frac{dP}{dt} = \bar{m}_p E H(T_h - 1), \quad (2.8d)$$

and the initial conditions, (2.5), become

$$T_h = E = P = 0, \quad L = 0.1 \quad \text{at} \quad t = 0. \quad (2.9)$$

		S19-3	Uniswa red
Physiological data	Mean $T_h^*$	994.5	1184.1
	S.d. of $T_h^*$	87.38	107.55
Model results	Mean final leaf mass	23.5	30.8
	S.d. of final leaf mass	3.25	4.52
	Mean final pod mass	62.5	52.8
	S.d. of final pod mass	4.49	5.81

Table 1: Physiological data on the podding initiation thermal time, and the corresponding mean and standard deviation of the leaf and pod mass distributions predicted by the model for a crop of  $N = 2000$  plants.

### 2.1.2 Parameter estimates

Key parameters in the model are the pod initiation thermal time,  $T_h^*$ , and the base temperature  $B$ . It is thought that  $T_h^*$  is genetically prescribed, and therefore can vary between landraces, and can vary between plants in a given landrace (if there is a range of genotypes present). For both S19-3 and Uniswa red, the physiological data (obtained at the Tropical Crops Research Unit) provides estimates for the mean and standard deviation of  $T_h^*$  (as shown in Table 1). The data suggest that pods initiate sooner for S19-3 than for Uniswa red, which is possibly because the S19-3 landrace is adapted for the short growing season in Namibia. In addition, the standard deviation is smaller for S19-3, which is likely to be due to the small genetic variation in the S19-3 landrace. The base temperature,  $B$ , may also be genetically prescribed; however, experimental data is not currently available, and therefore we estimate  $B = 10^\circ C$ . Due to time constraints during the study group, we were unable to estimate the rate or Michealis-Menten constants; however, we can still consider the qualitative behaviour of the model results to gain understanding of the mechanisms included. We set the dimensionless rate constants to unity ( $\bar{k}_l = \bar{m}_l = \bar{m}_p = \bar{\lambda} = 1$ ), and dimensional rate constant,  $k_p = 1$ ; these parameter estimates could be investigated in future work (see section 5 for further details).

## 2.2 Results

Figures 3a,b show the temporal variations in energy, leaf mass and pod mass predicted by the model for a plant from each landrace that is at a uniform temperature,  $T_e(t) = 28^\circ C$ , and with  $T_h^*$  given by the mean of the physiological data (see Table 1). For illustrative purposes, figure 3c shows the model results for a plant with a much larger pod initiation time,  $T_h^* = 500$ .

For  $T_h < T_h^*$ , the plant's energy produces leaves; with more leaves the plant takes in more energy and therefore the rate of leaf growth increases with time. Once  $T_h > T_h^*$ , leaves are no longer produced and therefore leaf mass decreases due to leaf decay. During this time,  $T_h > T_h^*$ , the pod mass increases. However, as the leaf mass reduces, the plant takes in less energy; we would expect the rate of increase of podding mass reduces, although this effect does not appear to be significant before the harvesting time of  $t = 120$  days. The pod mass would eventually plateau when the leaf mass reaches zero. Comparing figures 3a-c, we find that with a smaller  $T_h^*$ , the leaf mass is smaller when podding begins, and therefore the plant

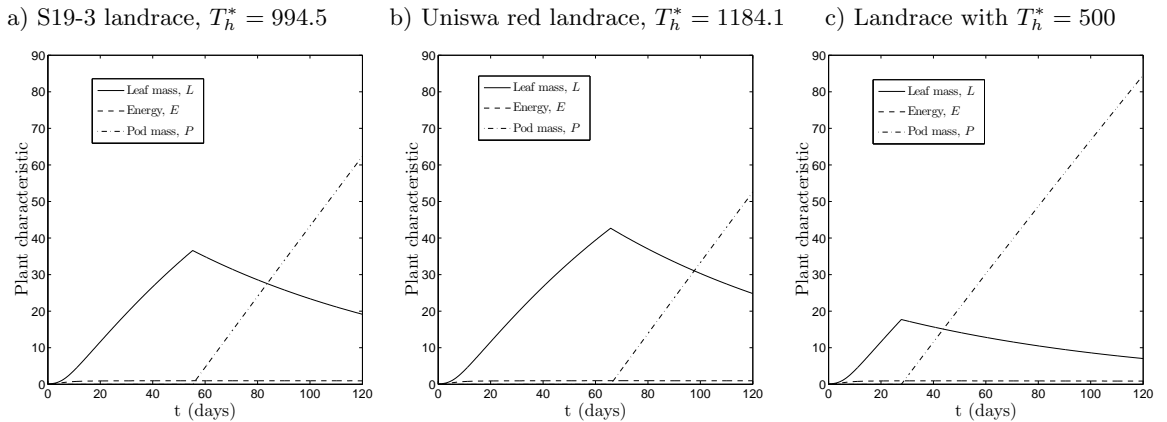


Figure 3: Numerical solution of the energy, leaf weight and pod weight from the single-plant model, for plants at a uniform temperature  $T_e(t) = 28^\circ C$ ; a) landrace S19-3 with pod initiation thermal time  $T_h^* = 994.5$ ; and b) landrace Uniswa red with pod initiation thermal time  $T_h^* = 1184.1$ .

has less energy which leads to a smaller rate of increase of pod mass. Thus, as S19-3 has a smaller  $T_h^*$  than Uniswa red, the pod mass increases more slowly; however, because pod initiation occurs earlier, the final pod mass (at 120 days) is larger for S19-3 than that for Uniswa red.

### 3 From the plant to the field scale

We now use the single-plant model developed in section 2 to predict the yield from a field of  $N$  plants of a given landrace. We suppose that the key difference between the plants is the genetically prescribed podding-initiation thermal time, and that within each landrace this parameter is normally distributed with the mean and standard deviation determined from the physiological data (see Table 1). Although genetic or environmental variations of other parameters are also of interest, a more comprehensive analysis is beyond the scope of this report.

For  $N = 2000$  plants, figure 4 shows the frequency density of the leaf and pod masses predicted by the model at the final ‘harvesting’ time of  $t = 120$  days. Table 1 displays the mean and standard deviation of these results. The final leaf mass and pod mass appear to be normally distributed. The Uniswa red landrace has a larger pod-initiation-thermal-time standard deviation, and the model predicts that it will have larger final leaf-mass and pod-mass standard deviations. This result is likely to be due to the linearity of the model.

## 4 Summary

We have determined appropriate statistical techniques to elucidate the effects of genetic variation versus environmental factors on physiological trait variation. The analysis showed that irrigation has a significant effect on the mean yield. In addition, there is also weak evidence that the landrace affects the variance



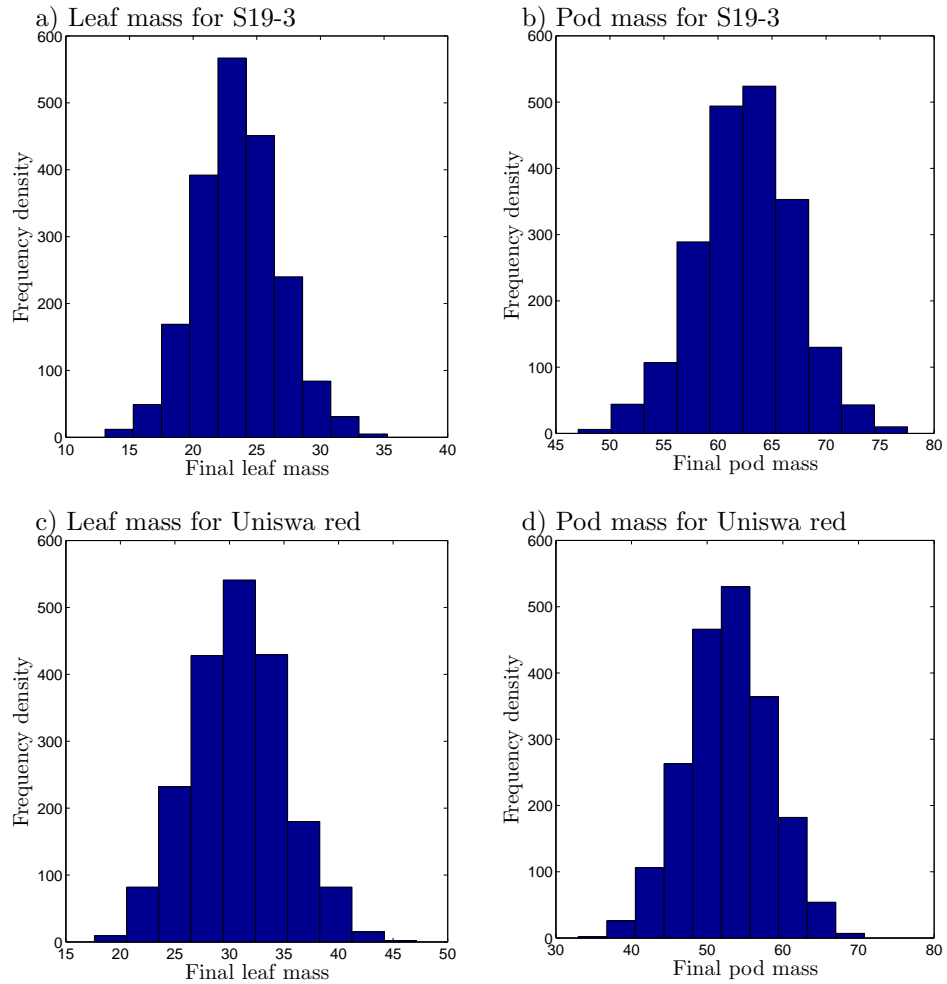


Figure 4: Histograms of the final leaf mass and pod mass at the ‘harvesting time’ of 120 days for  $N = 2000$  plants from each landrace where, for each plant, the pod initiation thermal time is randomly chosen from a normal distribution (with mean and standard deviation detailed in Table 1). The plants are at a uniform temperature  $T_e(t) = 28^\circ C$ . a) the leaf mass for landrace S19-3; b) the pod mass for landrace S19-3; c) the leaf mass for landrace Uniswa red; and d) the pod mass for landrace Uniswa red.

of the yield under drought conditions. However, the analysis was based on only ten observations in each group; we would be able to make stronger conclusions with a larger sample size.

A deterministic model has been developed that describes the growth of a single plant, assuming that the plant's energy is used to produce leaves and pods. A key parameter in the model is the pod initiation thermal time, which determines when the plant switches from producing leaves to producing pods. The model solutions show similar behaviour to the experimental results from the Tropical Crops Research Unit. The model predicts that because Uniswa red has a later pod initiation than S19-3, the rate of pod growth is faster for Uniswa red however, the final pod mass (at 120 days) is still smaller.

The single plant model has been extended to the crop scale by simulating  $N$  independent plants. We have linked this to the available experimental data by simulating the two different landraces via the thermal time to pod initiation. By considering normally distributed pod initiation thermal times, we have included the possible effects of genetic variation amongst the plants within a landrace. The model is linear, so as expected, the model solutions show that genetic variation leads to variations in the final leaf and pod mass of the crop.

## 5 Future work

In this report, we focussed our statistical analysis on a small subset of the bambara groundnut data to illustrate the potential for using statistics to understand the relationships between the input factors (*i.e.* landrace and irrigation conditions) and the output factors (*i.e.* crop yield). It would clearly be beneficial to perform more comprehensive analysis of the full data set to understand the influence of other input factors, such as thermal time.

The mathematical models predict qualitatively correct behaviour; however, many of the parameters used are not based on experimental data. A key challenge is to parameterise model using experimental data from the Tropical Crops Research Unit and from the literature. To determine variation on the field-scale, we would also need to understand which of the parameters may be genetically determined (and therefore vary between plants), or are affected by the environment (and may therefore vary temporally). As pod initiation depends on thermal time, it would be particularly interesting to obtain experimental estimates of the base temperature,  $B$ , and data on whether this parameter varies between landraces, and plants within a given landrace.

The single-plant model could be improved and extended by incorporating more biological phenomena. We currently only consider two developmental stages; however, in practise the plant may have as many as seven developmental stages, and the growth in each stage would be described by different dominant behaviour. It would be beneficial to assess the importance of including additional stages in the model. In addition, we focus here on development in response to the plants' accumulated thermal time; however, in an extended model we could also consider the role of the photoperiod (*i.e.* the amount of time that a plant is exposed to light) on the pod-formation dynamics, as demonstrated in the BAMGRO model. In another model extension, we could also include variation in the water supply and extreme temperatures, and incorporate terms to represent water and heat stresses. In this report, we have simplified the biology by only considering the plants' energy to be used for pod and leaf growth. In practise, some energy will be used for maintenance and root growth, and it would be interesting to see how this would affect the

model results, especially if we were to consider variations in the water supply, as the amount of water taken up by the plant would depend on the root architecture. When considering the crop scale (section 3), we suppose that each plant grows independently; it would be interesting to extend the model to incorporate competition for resources such as light and water.

## **Acknowledgements**

This report is based on the work carried out at the Mathematics in the Plant Sciences study group, January 2009. We would like to thank Kim Kenobi and Theo Kypraios for helpful discussions.