Archaeobotanical Analysis of Huizui Seeds

Abstract

Descriptive and regression analyses on archaeobotanical data from the Huizui site, Henan, China reveal millets as the main staple throughout the Neolithic period while exotic cereal (rice, wheat, and barley) and bean consumptions gradually increased. When comparing densities of crop taxa by period and feature types, the Yangshao period samples have a significantly less millet density than Erlitou samples, both alone and within the same feature type. Additionally, samples from houses have a significantly less bean density than ash pit samples, both alone and within the same period. Future research could gain more soil samples and construct a better dataset with balanced data across periods and feature types.

Introduction

The Huizui site is located in the Yiluo valley in Yanshi County, western Henan Province, China (Ford 2004, 71; Lee and Bestel 2007, 49). It was occupied during the Yaoshao period (5000-3000 BC), the Majiayao period (3300-2000 BC), the Longshan period (3000-1900 BC), and the Erlitou period (1900-1500 BC). The Huizui site has been considered a stone tool production locus during the late Longshan and Erlitou times and a secondary center in the hinterland of the Erlitou state (Ford 2004; Liu et al. 2013, 278-279). The Yiluo archaeological team has conducted intensive excavations at the Huizui site since 1998, revealing a total of 665 square meters of human occupation of the site (Lee and Bestel 2007, 49). Previous archaeobotanical studies on Erlitou period Huizui site have confirmed a dominant dryland farming subsistence in this region. At least eighteen archaeobotanical taxa were identified, with foxtail millet as the most common single crop, followed by broomcorn millet (51). However, very few statistical analyses have been done to examine possible contextual factors contributing to the development of dryland farming in this area. This paper will conduct a statistical analysis on charred seeds from the Huizui site. It aims to study the relationship between certain seed densities (foxtail millets, beans, etc.) and some contextual factors such as archaeological feature types (ash pits, ditches, houses, etc.) and time periods (Erlitou, Longshan, etc.). This analysis will help us better understand the farming patterns and possible drivers of dryland farming in the Huizui area.

Data and Methods

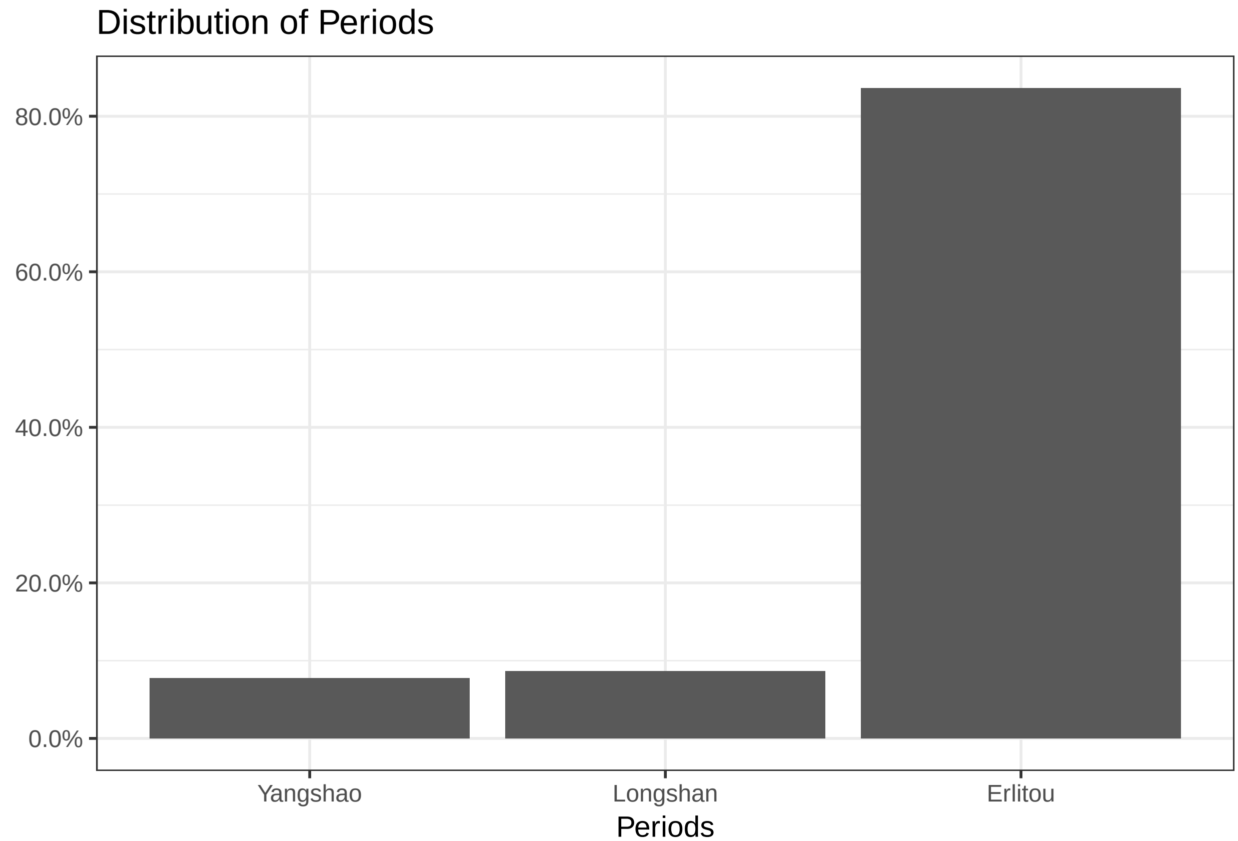
The Huizui site today is separated by a modern gully, and the two separated areas of the site are referred to as Huizui East (10 ha) and Huizui West (4 ha), the latter only being occupied during the Erlitou period (Lee and Bestel 2007, 49; Liu et al. 2013, 291). The Yiluo archaeological team gathered soil samples from the Huizui site from 2000 to 2006 for archaeobotanical analysis. The samples were collected from different excavation grids covering the whole site. The University of Oregon Archaeobotany Lab at the Anthropology Department has been conducting the analyzing work. At the lab, the soil samples were processed by flotation and only the light fractions were kept. The light portions then went through a sieving process to be further classified into 2mm, 1mm, 0.4mm and 0.2mm groups. The samples were then ready to be examined and seeds identified under a Nikon microscope. Every sample was recorded with sample number, feature (the trench and unit #), feature type (ditch/house/ash pits, etc.), historical period, total weight, charcoal weight, volume, number of seeds found for each taxa, and the percentage of work done for different sieving groups. There was a total of 126 samples from the Huizui site with 62 variables available to be used in the following analysis.

During the data cleaning process, certain variables are modified. I'm only interested in three large groups of taxa: millets (foxtail millet + broomcorn millet), luxury cereals (LC) (rice + wheat + barley), and beans (soybean + azuki bean + *Kummerowia* species + *Lespedeza* or *Melilotus* species + other wild beans). Therefore, these single species were combined into the three large groups of taxa, and other species not of interest were deleted from the dataset. I also deleted those columns with no valid data (only NAs), and only kept useful variables that are period (Period), feature type (FeatureType), light fraction weight (LightFractionWt), volume of the soil sample (Vol), number of total seed found in the sample (TotalSeedNo), total seed density (TotalSeedDensity), number of millets found in the sample (Millets), number of luxury cereals found in the sample (LuxuryCereal), number of beans found in the sample (Bean), millet percentage (MilletPct), luxury cereal percentage (LCPct), and bean percentage (BeanPct). Three new variables were added to the dataset—MilletDensity, LCDensity, and BeanDensity. I substituted NAs with 0s for those quantitative variables, rounded certain columns to have 2 decimal places, and standardized periods to only show "Yangshao", "Majiayao", "Longshan", and "Erlitou". However, the sample sizes for each categories of periods and feature types are not balanced. In order to avoid biases, I deleted periods and feature types with only one or two observations. The final version of Period variable contains Yangshao, Longshan, and Erlitou periods, chronologically organized, and the FeatureType variable contains ash pit, ditch, and house. The dataset after cleaning has 116 observations and 18 variables.

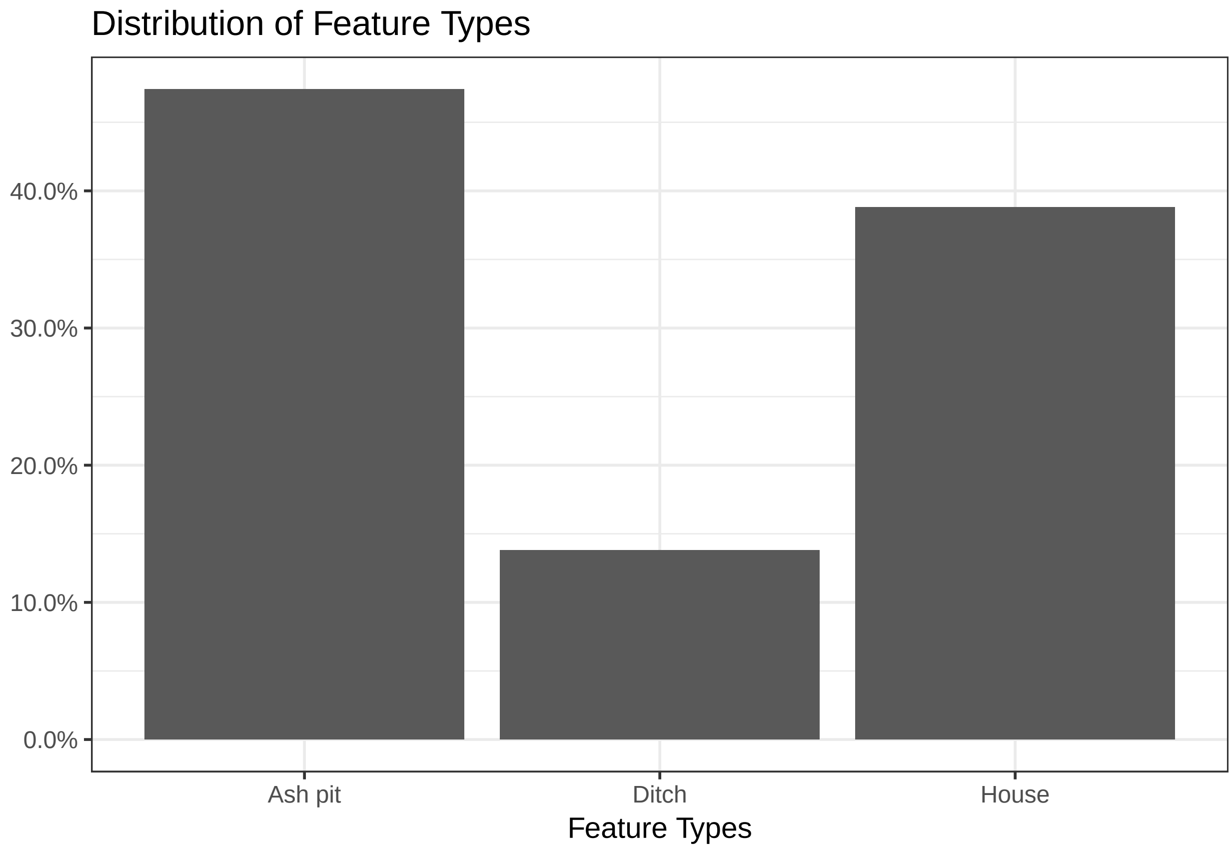
When analyzing on the data, I tried to visualize the univariate distribution of variables and the bivariate associations between variables. For univariate distributions, I created barplots for Period and FeatureType, as well as histograms showing the densities of total seed, millet, luxury cereal (LC), and bean. For bivariate plots, I made boxplots showing the densities of the four groups (total seed, millet, LC, and bean) by period and densities of the four groups by feature types. I hope these descriptive statistics and visualizations could reveal some patterns and characteristics of Huizui seed samples among various periods and feature types. Then I tried to build models. I chose the quasi-poisson model because poisson distribution is the simplest distribution used for modeling count data, and the quasi-poisson model adjusts for over-dispersion (Zeileis et al.). I created models with different set(s) of independent variables for the densities of total seed, millet, LC, and bean as dependent variables respectively. The results and interpretations of these models are shown below.

Results

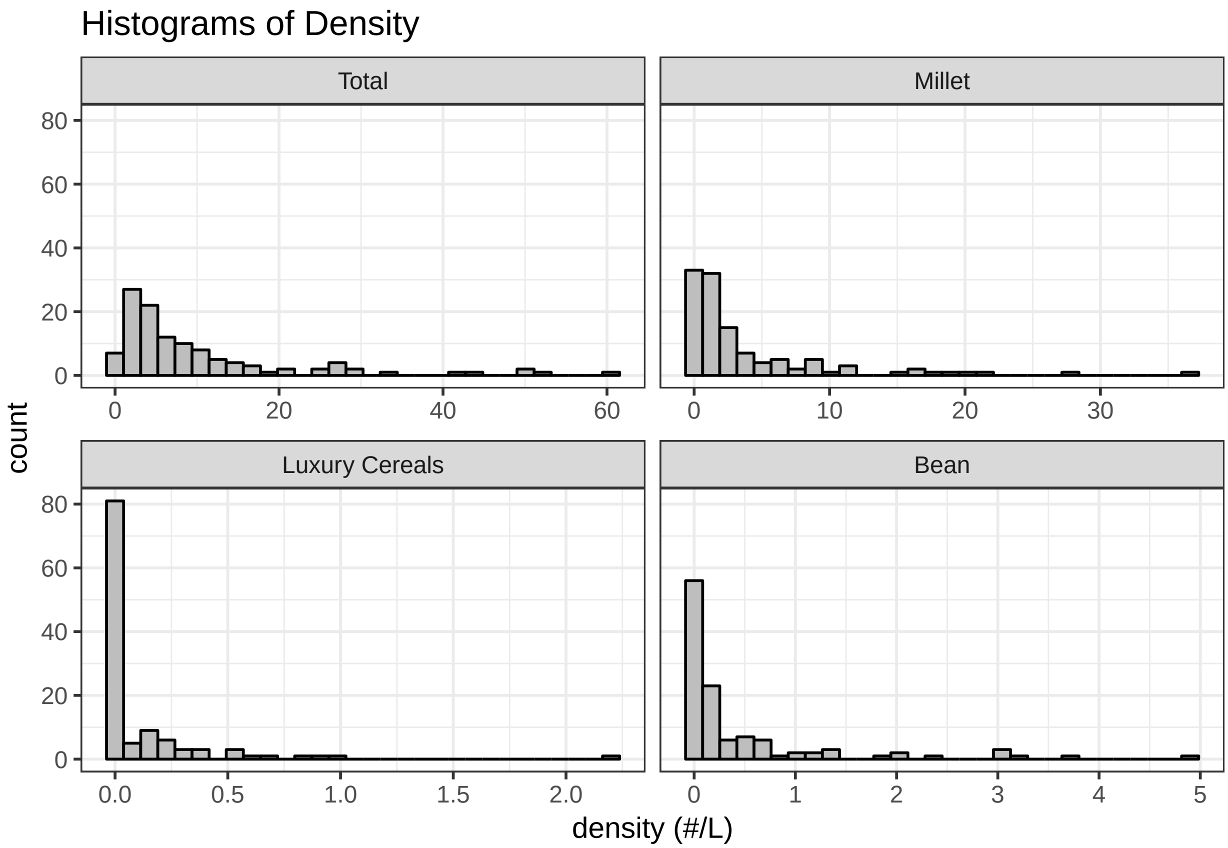
Univariate distributions



This barplot shows the proportions of the three periods where the seeds are from–Yangshao, Longshan, and Erlitou. As mentioned, they are chronologically organized. The majority of seed samples is from the Erlitou period (more than 80%). Yangshao and Longshan seeds constitute less than 10% of total seeds, respectively, and the earliest Yangshao has the smallest proportion among the three periods.

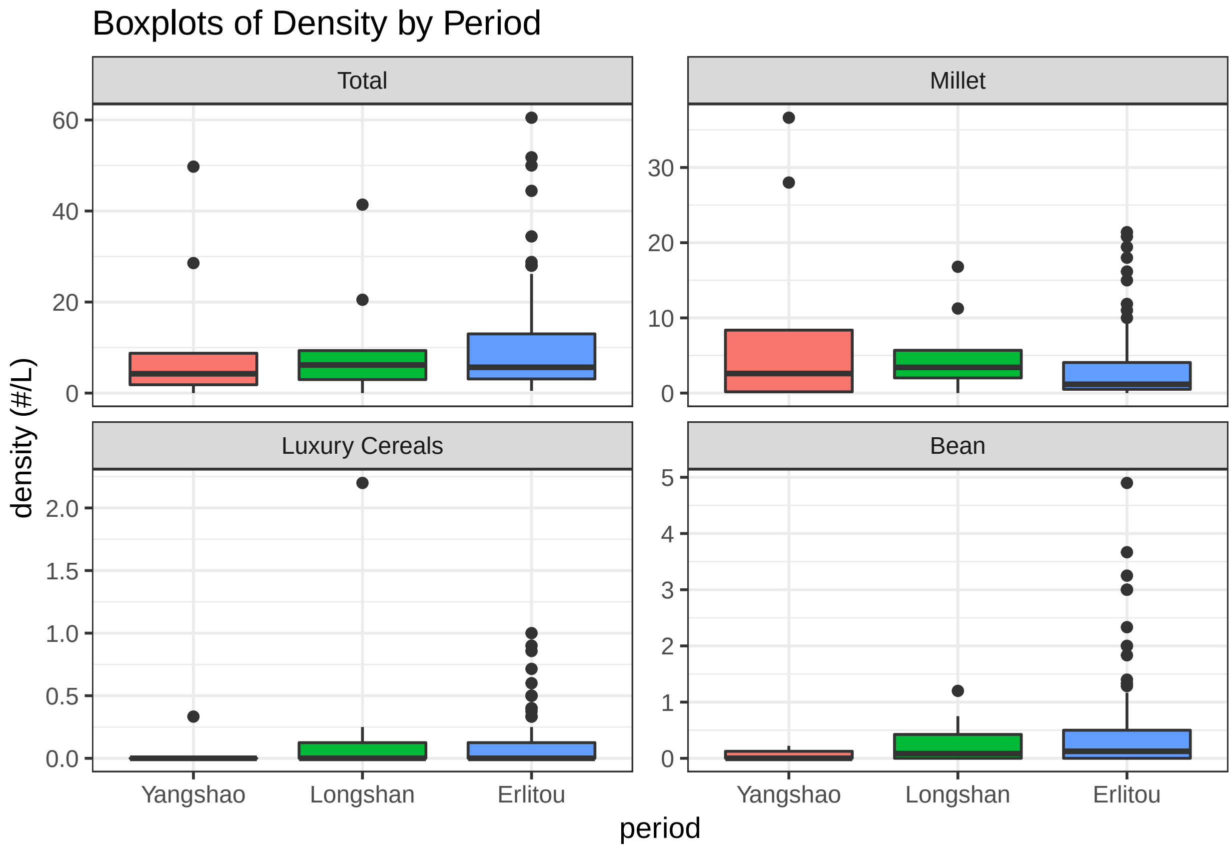


It could be seen from this barplot that among the three feature types, ash pits are the most common one, representing more than half of all seed samples. Approximately 40% of seed samples are found in house context that is the second largest feature type. Slightly less than 15% of all seed samples are from ditches.

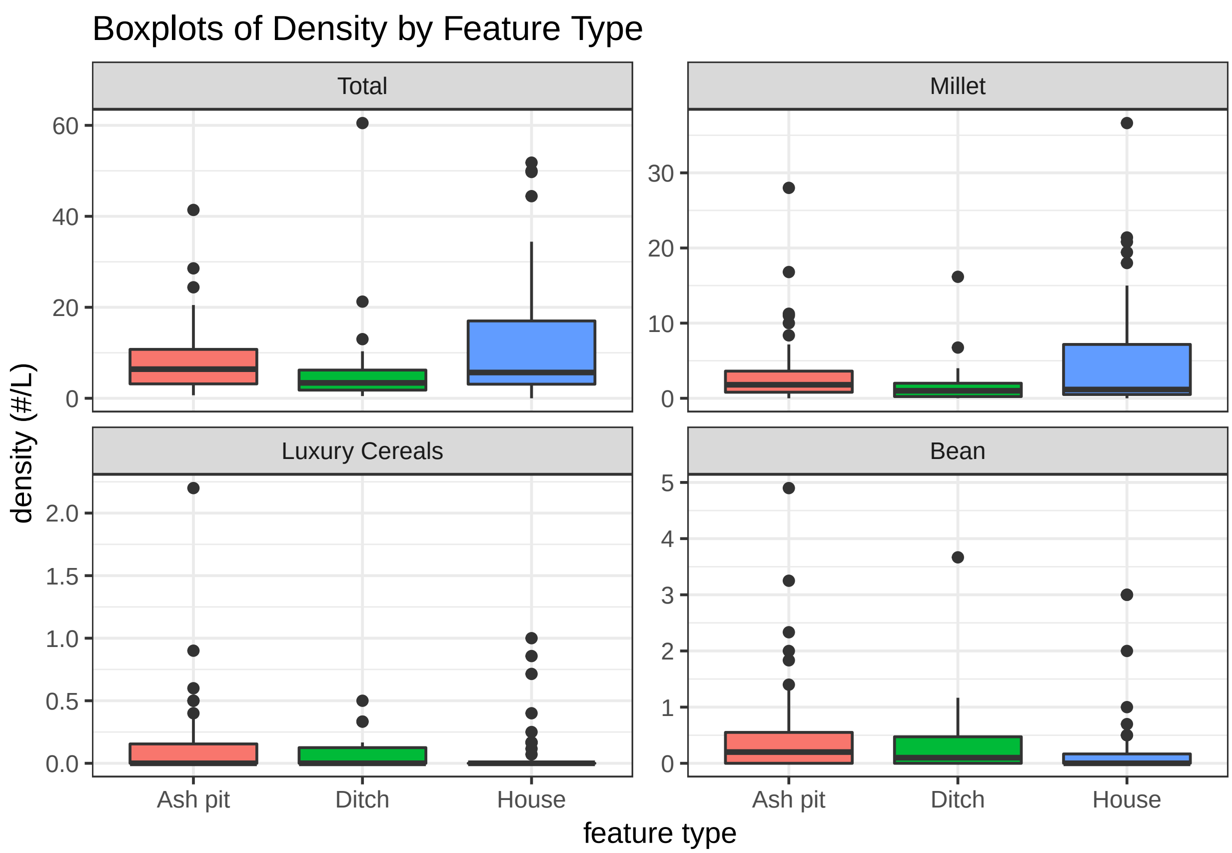


These four histograms show the densities of total seeds, millets, luxury cereals, and beans. Collectively, all four histograms have a minimum density of 0 #/L, which means there are soil samples with no seed/millet/luxury cereal/bean found. The histograms also have considerably different x scales. The densities of total seeds and millets are much higher than those of luxury cereals and beans. The total seed density histogram has a bin width of ~2 #/L. Most samples have total seed densities within the 0-20 #/L range, and the largest outlier reaches 60 #/L. The millet density histogram has a bin width of 1.25 #/L. Most samples with millets fall into the 0-10 #/L range, and the largest outlier has a millet density of more than 35 #/L. The luxury cereal density histogram has the lowest densities among the four. The bin width is ~0.07 #/L, and most samples have 0 or less than 0.07 #/L of LC densities. However, there's one sample with a LC density of more than 2.3 #/L, which is distinctively larger than other samples with LC seeds. The bean density histogram has the second lowest bin width of ~0.17 #/L. Most samples have bean densities within the 0-1 #/L range, and the largest outlier almost reaches 5 #/L. It is very important for us to have a sense of the density distributions among the four groups before delving into the associations among them.

Bivariate associations



These boxplots show the densities of total seed, millet, luxury cereals, and bean by period. The y (density) scales of the four plots are very different, which means millets are the most commonly found, and there are more beans found among the samples than luxury cereals. The total seed density boxplot has the highest densities among the three periods. The Yangshao period samples have the lowest 1st quartile and median, while the Erlitou period has the highest 3rd quartile and most outliers. This is understandable because the Erlitou samples constitute the largest proportion among the three periods. This is also the case for the luxury cereals and beans. However, the Yangshao period seems to have the highest millet densities among the three.



These boxplots show the densities of total seed, millet, luxury cereals, and bean by feature type. The y (density) scales of the four plots are very different, which means millets are the most commonly found, and there are more beans found among the samples than luxury cereals. The total seed density boxplot has the highest densities among the three periods. House seems to be the feature from which most seeds especially millets are found. However, luxury cereals and beans are more commonly found in ash pits and ditches than in houses.

Quasi-Poisson models

Total seed density as dependent variable

For total seed densities as the dependent variable, I created four quasi-poisson models with different independent variables (Table 1).

Model 1 has Period as the only independent variable. It shows that the Erlitou period soil samples have a total seed density of 12.93 #/L on average; the Longshan period samples have a total seed density of 29% less than Erlitou samples, on average; and the Yangshao period samples have a total seed density of 30% less than Erlitou samples, on average. Only the intercept is statistically significant.

Model 2 has feature types as the only independent variable. It shows that soil samples from ash pits have a total seed density of 8.5 #/L on average; samples from ditches have a total

Table 1: Quasi-poisson models of total seed densities by period and feature types

	Model 1	Model 2	Model 3	Model 4
Intercept	2.56^{***}	2.14^{***}	2.50^{***}	2.61^{***}
	(0.31)	(0.16)	(0.34)	(0.42)
Longshan Period	-0.34		-0.33	-0.24
	(0.51)		(0.51)	(0.57)
Yangshao Period	-0.35		-0.41	-0.57
	(0.34)		(0.34)	(0.45)
Ditch		-0.05	-0.09	-2.18
		(0.38)	(0.39)	(1.72)
House		0.28	0.30	0.55
		(0.23)	(0.23)	(0.60)
Yangshao:Ditch				2.43
				(1.76)
Longshan:House				-1.88
				(1.76)
Yangshao:House				-0.21
				(0.65)
AIC				
BIC				
Log Likelihood				
Deviance	6909.27	6854.40	6746.87	6301.35
Num. obs.	116	116	116	116

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

seed density of 5% less than ash pit samples, on average; and samples from houses have a total seed density of 32% more than ash pit samples, on average. Only the intercept is statistically significant.

Model 3 has both perid and feature types as independent variables. It shows that Erlitou soil samples from ash pits have a total seed density of 12.18 #/L on average. Within the same feature type, the Longshan period samples have a total seed density of 29% less than Erlitou samples, on average; and the Yangshao period samples have a total seed density of 34% less than Erlitou samples, on average. Within the same period, samples from ditches have a total seed density of 9% less than ash pit samples, on average; and samples from houses have a total seed density of 35% more than ash pit samples, on average. Only the intercept is statistically significant.

Model 4 added the interaction between perid and feature types in addition to Model 3. It shows that Erlitou soil samples from ash pits have a total seed density of 13.6 #/L on average. Among samples from ash pits, the Longshan period samples have a total seed density of 22% less than Erlitou samples, on average; and the Yangshao period samples have a total seed density of 44% less than Erlitou samples, on average. Among the Erlitou samples, samples from ditches have a total seed density of 89% less than ash pit samples, on average; and samples from houses have a total seed density of 73% more than ash pit samples, on average. Within the Yangshao period, ditch samples had a 28% higher total seed density than ash pit samples. Within the Longshan period, house samples had a 74% lower total seed density than ash pit samples. Within the Yangshao period, house samples had a 40% higher total seed density than ash pit samples. No coefficient in Model 4 is statistically significant.

Millet density as dependent variable

For millet densities as the dependent variable, I created four quasi-poisson models with different independent variables (Table 2).

Model 1 has Period as the only independent variable. It shows that the Erlitou period soil samples have a millet density of 10.17 #/L on average; the Longshan period samples have a millet density of 55% less than Erlitou samples, on average; and the Yangshao period samples have a millet density of 69% less than Erlitou samples, on average. Only the intercept and the coefficient of Yangshao period are statistically significant.

Model 2 has feature types as the only independent variable. It shows that soil samples from ash pits have a millet density of 3.42 #/L on average; samples from ditches have a millet density of 35% less than ash pit samples, on average; and samples from houses have a millet density of 49% more than ash pit samples, on average. Only the intercept is statistically significant.

Model 3 has both perid and feature types as independent variables. It shows that Erlitou soil samples from ash pits have a millet density of 9.77 #/L on average. Within the same feature type, the Longshan period samples have a millet density of 57% less than Erlitou samples, on average; and the Yangshao period samples have a millet density of 93% less than Erlitou samples, on average. Within the same period, samples from ditches have a millet

Table 2: Quasi-poisson models of millet densities by period and feature types

	Model 1	Model 2	Model 3	Model 4
Intercept	2.32^{***}	1.23^{***}	2.28^{***}	2.44^{***}
	(0.28)	(0.22)	(0.31)	(0.34)
Longshan Period	-0.78		-0.84	-0.80
	(0.52)		(0.51)	(0.54)
Yangshao Period	-1.16^{***}		-1.30^{***}	-1.61^{***}
	(0.33)		(0.32)	(0.43)
Ditch		-0.43	-0.61	-2.88
		(0.62)	(0.55)	(1.99)
House		0.40	0.48	0.41
		(0.31)	(0.27)	(0.51)
Yangshao:Ditch				3.01
				(2.07)
Longshan:House				-1.26
				(1.54)
Yangshao:House				0.24
				(0.61)
AIC				
BIC				
Log Likelihood				
Deviance	4349.47	4709.00	4054.35	3812.01
Num. obs.	116	116	116	116

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

density of 46% less than ash pit samples, on average; and samples from houses have a millet density of 61% more than ash pit samples, on average. Only the intercept and the coefficient of Yangshao period are statistically significant.

Model 4 added the interaction between perid and feature types in addition to Model 3. It shows that Erlitou soil samples from ash pits have a millet density of 11.47 #/L on average. Among samples from ash pits, the Longshan period samples have a millet density of 56% less than Erlitou samples, on average; and the Yangshao period samples have a millet density of 81% less than Erlitou samples, on average. Among the Erlitou samples, samples from ditches have a millet density of 95% less than ash pit samples, on average; and samples from houses have a millet density of 50% more than ash pit samples, on average. Within the Yangshao period, ditch samples had a 14% higher millet density than ash pit samples. Within the Longshan period, house samples had a 58% lower millet density than ash pit samples. Within the Yangshao period, house samples had a 91% higher millet density than ash pit samples. No coefficient in Model 4 is statistically significant.

Luxury cereal density as dependent variable

For luxury cereal (LC) densities as the dependent variable, I created four quasi-poisson models with different independent variables (Table 3).

Model 1 has Period as the only independent variable. It shows that the Erlitou period soil samples have a LC density of 0.04 #/L on average; the Longshan period samples have a LC density of 400% more than Erlitou samples, on average; and the Yangshao period samples have a LC density of 116% more than Erlitou samples, on average. Only the intercept is statistically significant.

Model 2 has feature types as the only independent variable. It shows that soil samples from ash pits have a LC density of 0.13 #/L on average; samples from ditches have a LC density of 46% less than ash pit samples, on average; and samples from houses have a LC density of 43% less than ash pit samples, on average. Only the intercept is statistically significant.

Model 3 has both perid and feature types as independent variables. It shows that Erlitou soil samples from ash pits have a LC density of 0.06 #/L on average. Within the same feature type, the Longshan period samples have a LC density of 335% more than Erlitou samples, on average; and the Yangshao period samples have a LC density of 116% more than Erlitou samples, on average. Within the same period, samples from ditches have a LC density of 34% less than ash pit samples, on average; and samples from houses have a LC density of 39% less than ash pit samples, on average. Only the intercept is statistically significant.

Model 4 added the interaction between perid and feature types in addition to Model 3. It shows that Erlitou soil samples from ash pits have a LC density of 0.1 #/L on average. Among samples from ash pits, the Longshan period samples have a LC density of 189% more than Erlitou samples, on average; and the Yangshao period samples have a LC density of 17% more than Erlitou samples, on average. Among the Erlitou samples, samples from ditches have an almost identical LC density as ash pit samples, on average; and samples from houses also have the same LC density as ash pit samples, on average. Within the Yangshao period,

Table 3: Quasi-poisson models of luxury cereal densities by period and feature types

	Model 1	Model 2	Model 3	Model 4
Intercept	-3.08^{**}	-2.02^{***}	-2.87^{**}	-2.34^{*}
	(1.02)	(0.27)	(1.04)	(1.00)
Longshan Period	1.61		1.47	1.06
	(1.13)		(1.15)	(1.10)
Yangshao Period	0.77		0.77	0.16
	(1.04)		(1.05)	(1.04)
Ditch		-0.61	-0.41	-15.70
		(0.83)	(0.78)	(2090.63)
House		-0.56	-0.49	-16.11
		(0.51)	(0.48)	(2562.90)
Yangshao:Ditch				15.48
				(2090.63)
Longshan:House				-0.51
				(3607.69)
Yangshao:House				15.81
				(2562.90)
AIC				
BIC				
Log Likelihood				
Deviance	223.31	226.63	219.59	212.65
Num. obs.	116	116	116	116

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Quasi-poisson models of bean densities by period and feature types

	Model 1	Model 2	Model 3	Model 4
Intercept	-2.79	-0.67^{**}	-2.51	-2.74
	(1.56)	(0.22)	(1.45)	(2.06)
Longshan Period	1.32		1.18	1.30
	(1.78)		(1.65)	(2.24)
Yangshao Period	1.90		2.00	2.24
	(1.57)		(1.45)	(2.07)
Ditch		-0.25	-0.21	-0.09
		(0.58)	(0.55)	(3.57)
House		-1.24^{*}	-1.32^{*}	-0.09
		(0.54)	(0.51)	(3.57)
Yangshao:Ditch				-0.13
				(3.61)
Longshan:House				-0.08
				(4.22)
Yangshao:House				-1.29
				(3.61)
AIC				
BIC				
Log Likelihood				
Deviance	641.71	609.56	569.28	566.59
Num. obs.	116	116	116	116

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

ditch samples had a 20% lower LC density than ash pit samples. Within the Longshan period, house samples had a almost identical LC density as ash pit samples. Within the Yangshao period, house samples had a 26% lower LC density than ash pit samples. Only the intercept is statistically significant.

Bean density as dependent variable

For bean densities as the dependent variable, I created four quasi-poisson models with different independent variables (Table 4).

Model 1 has Period as the only independent variable. It shows that the Erlitou period soil samples have a bean density of 0.06 #/L on average; the Longshan period samples have a bean density of 274% more than Erlitou samples, on average; and the Yangshao period samples have a bean density of 569% more than Erlitou samples, on average. No coefficient is statistically significant.

Model 2 has feature types as the only independent variable. It shows that soil samples from ash pits have a bean density of 0.51 #/L on average; samples from ditches have a bean density of 23% less than ash pit samples, on average; and samples from houses have a bean density of 71% less than ash pit samples, on average. Only the intercept and the coefficient of House are statistically significant.

Model 3 has both perid and feature types as independent variables. It shows that Erlitou soil samples from ash pits have a bean density of 0.08 #/L on average. Within the same feature type, the Longshan period samples have a bean density of 225% more than Erlitou samples, on average; and the Yangshao period samples have a bean density of 639% more than Erlitou samples, on average. Within the same period, samples from ditches have a bean density of 19% less than ash pit samples, on average; and samples from houses have a bean density of 74% less than ash pit samples, on average. Only the coefficient of House is statistically significant.

Model 4 added the interaction between perid and feature types in addition to Model 3. It shows that Erlitou soil samples from ash pits have a bean density of 0.06 #/L on average. Among samples from ash pits, the Longshan period samples have a bean density of 267% more than Erlitou samples, on average; and the Yangshao period samples have a bean density of 839% more than Erlitou samples, on average. Among the Erlitou samples, samples from ditches have a bean density of 9% less than ash pit samples, on average; and samples from houses have a bean density of 9% less than ash pit samples, on average. Within the Yangshao period, ditch samples had a 20% lower bean density than ash pit samples. Within the Longshan period, house samples had a 16% lower bean density than ash pit samples. Within the Yangshao period, house samples had a 75% lower bean density than ash pit samples. No coefficient in Model 4 is statistically significant.

Conclusions

From the analysis above, we gained a better understanding of dryland farming patterns during Neolithic period at the Huizui site. We learned that among the samples we got, a majority of them (80%) are from the Erlitou period. More than a half of them are from ash pits and 40% of them from houses. Among the three large groups of taxa (millet, luxury cereal, and bean), the density of millets is the largest, followed by beans, and luxury cereal is the rarest among our soil samples. It's interesting that millets are commonly found in the Yangshao period houses, while there is an obvious increase of luxury cereals and beans during the later Longshan and Erlitou periods from ash pits and ditches. These findings are consistent with some previous archaeobotanical studies, revealing millets as the main staple throughout the Neolithic period while exotic cereal (rice, wheat, and barley) and bean consumptions gradually increased. The quasi-poisson models also confirm these findings and provide new information. For millet density as the dependent variable, the coefficients of Yangshao period are significant in three models. This means the Yangshao period samples have a statistically significantly less millet density than Erlitou samples, both alone and within the same feature type. I think this result is to some extent affected by the unbalanced sample sizes of the three periods. There are many samples with millets found during the Yangshao and Erlitou periods while the other two groups of taxa are often absent. We have more millet data to be analyzed so it becomes significant. The quasi-poisson models also show that samples from houses have less bean densities than ash pit samples, both alone and within the same period. These findings are still preliminary. More data is needed in the future for more convincing conclusions. Future research could gain more soil samples and construct a better dataset with balanced data across periods and feature types.

References

Ford, Anne. 2004. “Ground Stone Tool Production at Huizui, China: An Analysis of A Manufacturing Site in the Yilou River Basin.” Bulletin of the Indo-Pacific Prehistory Association 24: 71–78.

Lee, Gyoung-Ah, and Sheahan Bestel. 2007. “Contextual Analysis of Plant Remains at the Erlitou-Period Huizui Site, Henan, China.” Bulletin of the Indo-Pacific Prehistory Association 27: 49–60.

Liu, Li, Shaodong Zhai, and Xingcan Chen. 2013. “Production of Ground Stone Tools at Taosi and Huizui: A Comparison.” In A Companion to Chinese Archaeology Pp. 278–299. Hoboken, NJ: John Wiley & Sons Inc.

Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2008. “Regression Models for Count Data in R.” Journal of Statistical Software 27(8).