Chifeng Surface Sherd Density Data Screening

Analysis of raw sherd densities from systematic collections

The values explored here are densities (in sherds/m^2) of all pre-modern sherds (including all sherds except those identified as post-Liao). Since period-by-period densities are later calculated as a percentage of this overall density, the effect of the unclassified pre-modern sherds is divided between the periods represented according to the proportions of the identified pre-modern sherds. Thus a collection that has, for example, a large number of unknown sherds because they were difficult to sort between the two periods represented or a collection with small badly-eroded sherds that were difficult to identify does not get a low density because relatively few sherds were identified. This procedure should also help counteract random noise introduced by unavoidable variation between ceramic analysts who were conservative about classification to varying degrees. (That is, some analysts inevitably resort to the "unclassified" category more often than others.) Surface densities can be calculated readily for systematic collections, and the frequency distribution of these densities (sherds/m^2) is illustrated below.

```
Stem and Leaf Plot of variable:      DENSITY, N = 294
         Minimum:        0.000
         Lower hinge:        1.982
         Median:         3.149
         Upper hinge:        4.671
         Maximum:       35.527

               0   01112223344
               0   55556667777778999999
               1   0111112222222234444444444
               1 H 5555566666788889999999
               2   011111111122222444
               2   555566666666666688888888888899999999
               3 M 0111111111111111222222233333333
               3   5555555566666666888899
               4   111111111222233333
               4 H 5555555556666666666688899
               5   0000002233
               5   5566888899
               6   002223333
               6   5699
               7   2223
               7   6
               8   003
               9   02234
               9   779
              10   1444
              10   7
              11   013
              11
              12   4
              12   78
              13
              13
              14   1
              14
              15
              15   7


              18   8
              22   3
              34   6
              35   5
```

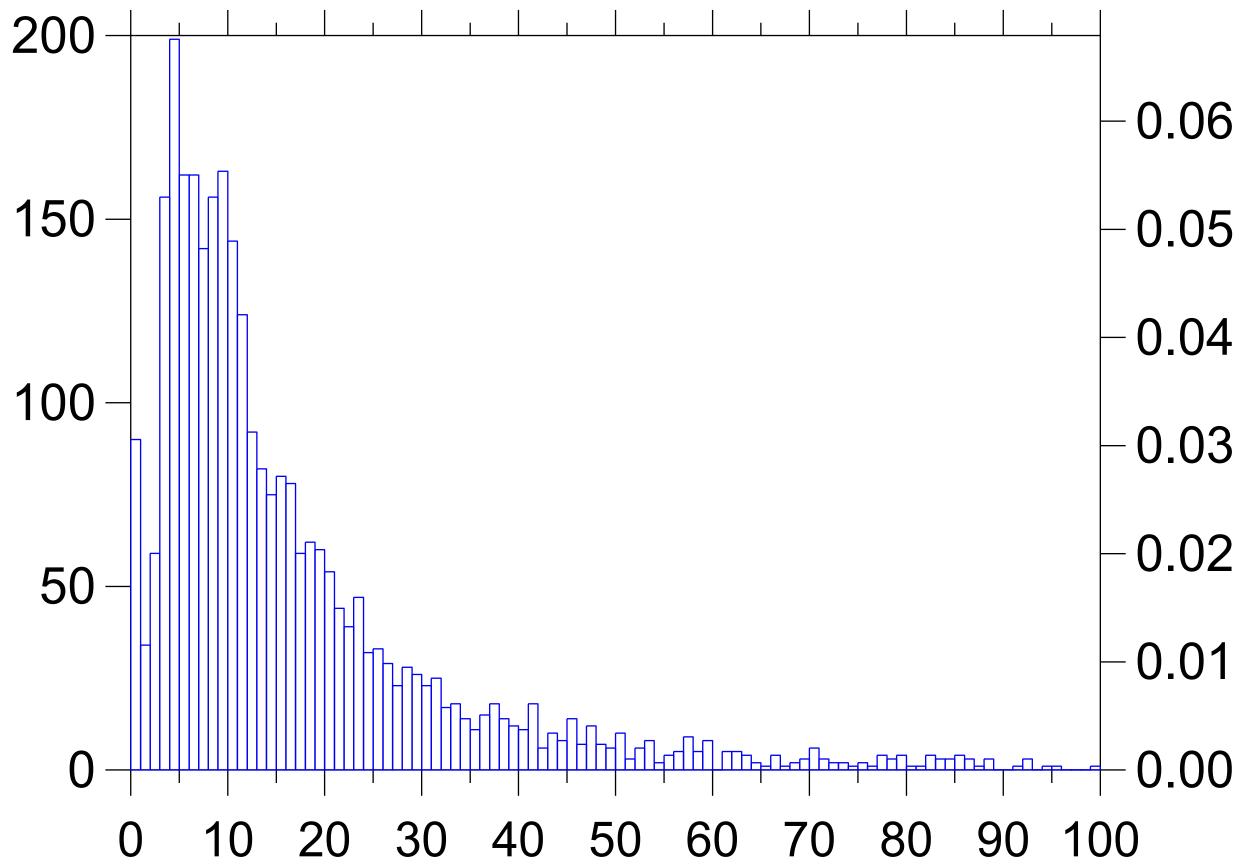
If outliers are identified as values more than two midspreads above the median, values greater than about 8.5 sherds/m^2 (two midspreads = 5.378, plus the median = 8.527) should be examined. These 25 possible outlier values are listed below:

Key	Value	Comments
04P013	35.53	identified in comments as "refuse pit?"--must be an artifact hot spot
01E147	34.68	3 of 4 collections in site are potential outliers
98A003	22.36	first site on survey--must be an artifact hot spot
04P012	18.83	identified in comments as "refuse pit?"--must be an artifact hot spot
99B127	15.71	1 other general collection in site--must be an artifact hot spot
07C021	14.15	all Zhaobaogou--all other collections in site general--must be hot spot
04P076	12.88	nothing unusual in notes
99D037	12.74	nothing unusual in notes
01C243	12.46	nothing unusual in notes
99D014	11.32	adjacent to mound
06A026	11.18	nothing unusual in notes
99D065	11.04	1 other general collection in site--could be an artifact hot spot
01E148	10.76	3 of 4 collections in site are potential outliers
01C183	10.47	nothing unusual in notes
01E145	10.47	3 of 4 collections are in site are potential outliers
99C036	10.47	nothing unusual in notes
07C173	10.19	nothing unusual in notes (in the Liao town)
01A229	9.91	nothing unusual in notes
99D046	9.77	nothing unusual in notes
00C051	9.77	nothing unusual in notes
99A101	9.48	nothing unusual in notes
00D009	9.34	site was excavated in 1992 by Jida
07B066	9.20	nothing unusual in notes
06A025	9.20	nothing unusual in notes
99B008	9.06	nothing unusual in notes

The highest four values are unquestionable outliers, and there are reasons in the field notes to suspect three of these should be taken as artifact hot spots in any event. In the fourth instance (01E147) the collection is one of four systematic collections in a Zhanguo-Han site, of which three are high enough values to be possible outliers. The fifth highest value (for 99B127) is for a collection in a small site where the only other collection was a general one. This also seems likely to be an artifact hot spot. The sixth highest value is for a collection (07C021) consisting of 100 Zhaobaogou sherds, and is surely a product of collecting intensively for Zhaobaogou sherds in a small area once they had been noticed. From here on down there is at most only occasional reason to believe that artifact hot spots are involved. Collection 99D046, with a density of 9.71 sherds/m^2; is from excavated site 342. The 26 collection circles later collected in this area had a mean density of 11.31 sherds/m^2. This demonstrates that the value of 9.71 is not just a reflection of an artifact hot spot and may even slightly underestimate the density in this part of the site. On the other hand this is clearly an unusually high density in a part of a site occupied very intensively for a long period of time. The resulting surface density must be near the upper limit for reliable density values for entire collection units. Considering overall pre-modern sherd densities, then, only the six values greater than 13 seem suspicious as outliers just because they are so high. As it turns out when area-density values for each period are explored, these very high density values mostly do not cause problems for several reasons, mentioned below.

Analysis of raw sherd counts from general collections

These values are the counts of sherds from general collections, including all sherds except those identified as modern. Sherds in the "unclassified" category are included here, following the same logic as for the analysis of raw densities from systematic collections, above. The histogram on the next page illustrates their frequency distribution (since for this number of collections a stem-and-leaf plot is impractical). As is to be expected, the frequency distribution straggles far off in the upward direction (even more than it appears in the histogram below, since of the 2,984 general collections the 39 with sherd counts >100 are omitted). There are no sherds at all from 90 general collections, so these disappear from density calculations (although they are present in the histogram). These include several kinds of things: collection units where there were features to record but no surface artifacts, collection units where the only sherds recovered turned out to be modern (or not sherds at all), and collection units whose sherd bags were lost or misnumbered.



On the basis of this histogram, the distribution can be divided into four categories: >35 sherds, 23–35 sherds, 8–22 sherds, and 1–7 sherds. This division follows at least small suggestions of multi-modality as a rationale for the exact cut points, although the decision to break the distribution into four parts (as opposed to, say, three or five is essentially arbitrary). On the assumption that larger collections usually come from denser sherd scatters, different density values have been assigned to each of these four categories. Results of surface collecting at site 342 before stratigraphic testing provide support for this notion as well as a starting point for thinking about density values. In the area delineated for testing as site 342 (which includes what was originally called site 343 as well), one systematic collection and three general collections were made during the initial survey visit. The systematic collection (99D046) had a density of 9.71 sherds/m^2, as discussed above. The 26 collection circles later collected in this area had a roughly similar mean density of 11.31 sherds/m^2, suggesting that the very high density of 9.71 was an accurate assessment for the collection unit as a whole. The areas where general collections were made during the initial survey visit, yielded the following results in systematic collection circles prior to the excavation of stratigraphic tests:

Key	No. of sherds from survey	No. of systematic circles	Mean density of systematic circles
99D045	79	12	4.59 sherds/m^2
99D044	45	12	3.59 sherds/m^2
99D043	15	21	1.19 sherds/m^2

Since 3 or 4 sherds or more per circle is the criterion for making a systematic collection, the theoretical minimum density value for a systematic collection would be about 0.5 sherds/m^2 (3.5 sherds/7.065 m^2 = 0.495 sherds/m^2). This would simultaneously be the theoretical maximum density value for a general collection. Since in practice, deciding on survey whether to make a systematic or general collection is a hasty subjective judgment, one would expect mistakes: sometimes systematic collections would be made where the density turned out to be less than 0.5 sherds/m^2; and sometimes, general collections where the density turned out to be greater than 0.5 sherds/m^2. The first kind of mistake is visible in the systematic collection data, since 10 of the 294 systematic collections had non-zero values of less than the theoretical minimum of 0.5 sherds/m^2. (One systematic collection had a density of 0 sherds/m^2 because the sherds were lost.) There simply must have been far more errors than this in judging whether to make a systematic collection or not. Most of those errors were probably in deciding not to make one where the density really was greater than 0.5 sherds/m^2, and the number of this kind of error seems likely to be substantial. Precisely this kind of error was made in the cases of all three general collections originally made at site 342 (see above). It is easy to understand why the subjective decision might be strongly biased in this direction. In the first place, this subjective judgment about density is based on glancing around. Careful collecting almost inevitably produces more sherds than seen by just glancing around, so the systematic collection densities usually will turn usually out to be somewhat higher than initially estimated. A second factor may be that much of a general collection is already made while entering a site before realizing that the density is high enough for a systematic collection; a general collection already begun creates an inertia that discourages starting over with a systematic collection. The density represented by a fairly large general collection is, thus, on average probably well above the theoretical 0.5 sherds/m^2 value.

At the lower end of the scale, some general collections produced only a very few sherds because after the survey team collected for a while, they simply ceased to find any more sherds. A general collection of only 10 sherds, though, cannot be taken to indicate that there were only 10 sherds on the surface in an entire hectare. The careful examination characteristic of a systematic collection would certainly yield more than a general collection would if both were practiced across the same area. The first stage of intensive collecting at Fushanzhuang (Peterson 2006) is probably broadly comparable to systematic collection from survey. Since Fushanzhuang, overall, is a site with low surface densities, the first stage of intensive collection there may provide an idea of what overall sherd densities small general collections truly represent:

Fushanzhuang Grid	No. of sherds collected	Area	Density
F09B	322	400 m^2	0.81 sherds/m^2
F11	288	400 m^2	0.72 sherds/m^2
F08A-D	1046	1600 m^2	0.65 sherds/m^2
F12	235	400 m^2	0.59 sherds/m^2
F14	214	400 m^2	0.54 sherds/m^2
F15	207	400 m^2	0.52 sherds/m^2
F09A	202	400 m^2	0.51 sherds/m^2
F06	147	400 m^2	0.37 sherds/m^2
F16A-B	177	500 m^2	0.35 sherds/m^2
F10B	127	400 m^2	0.32 sherds/m^2
F10A	116	400 m^2	0.29 sherds/m^2
F13B	83	300 m^2	0.28 sherds/m^2
F13A	72	300 m^2	0.24 sherds/m^2
F01	82	400 m^2	0.21 sherds/m^2
F04	77	400 m^2	0.19 sherds/m^2
F02	72	400 m^2	0.18 sherds/m^2
F07	62	400 m^2	0.16 sherds/m^2
F03	63	400 m^2	0.16 sherds/m^2
F05	49	400 m^2	0.12 sherds/m^2

The lower range of these densities seems to fall near the lower limit for sites regularly detectable on survey. That is, if sherds were much sparser than the lowest of these densities, it is entirely possible that they could be walked across at the normal pedestrian pace of survey without noticing any sherds. Imagine, for example, a density of 0.1 sherds/m^2 across an area of 100 by 100 m. This amounts to a total of 1,000 sherds on the surface of the entire area of 1 ha. If surveyors walk transects about 50 m apart, on average two surveyors would cross through this area. If they were able to observe carefully strips 1 m to either side of their path, they would be looking carefully at 4% of the surface area. Thus, on average, some 40 sherds would fall within their view. As the material recovered from Fushanzhuang makes clear, the vast majority of these sherds would be so small as not usually to be noticed on moving pedestrian survey (or even on the somewhat more careful examination of a general collection—it would probably take a systematic collection to discover that the density was this high). On survey, at least three sherds would have to be noticed for identification as an occupied area. Occasionally, of course, areas with even lower densities would fortuitously be discovered, just as occasionally areas with densities this high would be missed. On average, though 0.1 sherds/m^2 seems a plausible minimum density value to use for very small general collections.

According to this logic, then, the following approximate density values have been assigned to the four categories of sherd collections (large, medium, small, and very small):

1. The 342 collections with >35 sherds = 2.00 sherds/m^2. There are large numbers of sherds in these general collections, and some, like those at site 342 (above) surely represent real densities higher than 2.00. Site 342 is, however, an unusually dense surface scatter, and it is to be expected that densities in areas mistakenly given general collections were not usually that high.
2. The 326 collections with 23–35 sherds = 0.50 sherds/m^2. This is the theoretical maximum value for general collections (see above). A substantial number of general collections must represent areas whose density values reach this high. Fushanzhuang was taken above as an example of a fairly low-density sherd scatter; of its 19 grids, 7 reach densities >0.50 sherds/m^2. It is important to remember, though, that these densities are for the higher-density artifact clusters taken to represent household units; such high values do not represent average densities across an entire hectare. It does not seem likely that survey collections from Fushanzhuang would have produced as many as 23–35 sherds, and so the value of 0.50 sherds/m^2 seems comfortably above Fushanzhuang general densities.
3. The 1,512 collections with 8–22 sherds = 0.25 sherds/m^2. At least the denser sectors at Fushanzhuang might plausibly have produced numbers of sherds in this range for general collections. Considering that overall densities for 1-ha collections units at Fushanzhuang would be lower than those obtained in the grids, which were located in artifact concentrations, the value of 0.25 sherds/m^2 seems a plausible overall collection unit density for the denser sectors of the Fushanzhuang site—ones where the densities within household clusters ranged up to between 0.50 and 0.81 sherds/m^2.
4. The 914 collections with 1–7 sherds = 0.10 sherds/m^2. This is probably about the minimum reliably detectable surface-scatter density, as discussed above. (Collection units with no pre-modern sherds, of course, receive an overall density value of 0.00 sherds/m^2.)

These values have been assigned as overall densities for general collections, according to the number of pre-modern sherds they contained, again including the "unclassified" category.

Special treatment of collection units from the intensive collection area at Fushanzhuang

There are 20 collection units for the area of the Fushanzhuang site, which had been intensively surface collected before regional-scale survey was completed in its area. These 20 regional survey collection units were thus created after the fact and assigned values consistent with the values they would have received on regional survey. The intensive collections at Fushanzhuang (Peterson 2006) were the basis for assigning these values. Sherd-count data for these 20 collections are simply the sherd counts from the intensive collection units they correspond to spatially These are far larger numbers of sherds than would have been recovered by regional survey, and this must be allowed for in any analysis that departs from sherd counts. The calculation of density values discussed here is based on proportions of sherds of different periods, and these can be calculated accurately from the larger numbers of sherds produced by the intensive collections. An overall surface density value, however, must be assigned to these 20 collection units in a way compatible with the way these values were assigned in the regional survey. A density value of 0.25 sherds/m^2 was assigned to the collection including grids whose first-stage densities within household clusters were >0.50 sherds/m^2 (collection units 04P148, 04P151, 04P159, 04P163, and 04P165). A density value of 0.10 sherds/m^2 was assigned to the other collections in this set (04P149, 04P150, 04P152–04P158, 04P160–04P162, 04P164, and 04P166–04P167).

Exploration of each period's density-area values by collection unit

After assigning the above density values to general collections, and without changing the outliers discussed in the first section, each collection's overall density value was multiplied by the proportion of identified pre-modern sherds in each period to obtain a separate density value for each collection for each period. These density values were multiplied by the total area represented by the collection to arrive at a density-area value for each collection for each period. These sets of values are explored below. Density-area values of 0 are omitted from the following analyses.

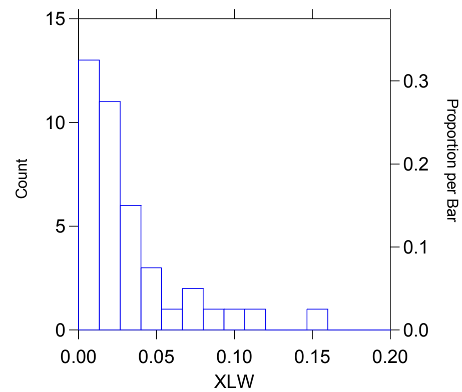
Xinglongwa

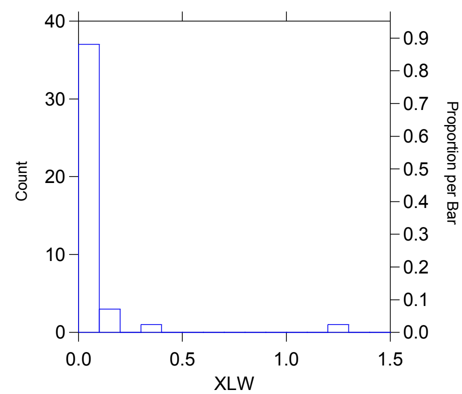
```
Stem and Leaf Plot of variable:          XLW, N = 42
         Minimum:         0.000
         Lower hinge:         0.012
         Median:          0.022
         Upper hinge:         0.041
         Maximum:         1.225

              0    00346779
              1  H 01123346799
              2  M 1225589
              3    2345
              4  H 01
              5    1
              6    3
              7    99
              8    9
              9
             10    59
             11
             12
             13
             14    7

Extreme outliers:

             33    7
            122    5
```





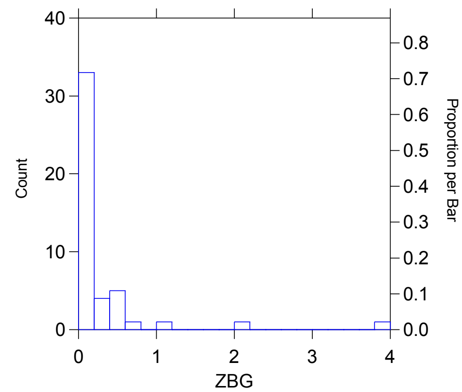
There are clearly two extreme outliers in this distribution. If these two are omitted, the distribution of values at or below 0.150 (lower histogram) is unremarkable. The highest value of 1.225 is from 00E023, an usually large collection unit (2.35 ha), which also has sherds of Upper Xiajiadian and Zhanguo-Han. It seems likely that the 12 Xinglongwa sherds do not represent such a high density across the entire area, so this value has been reduced to 0.150 (the maximum non-outlier value). The next value of 0.337 is from 07B087, where the collection was made at a burial eroding out at the surface—clearly a hot spot, so this value has also been reduced to 0.150. The next highest value for Xinglongwa is 0.148 from a systematic collection with no indication of anything unusual, which is why this has been taken to mark the upper limit of non-outlier values.

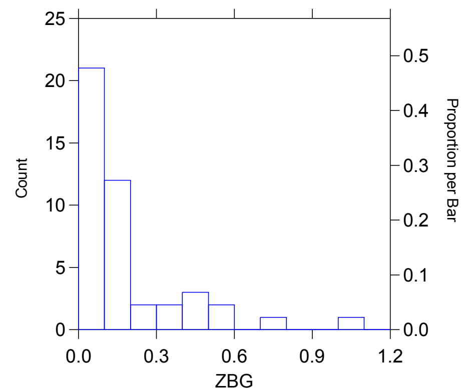
Zhaobaogou

```
Stem and Leaf Plot of variable:          ZBG, N = 46
         Minimum:         0.005
         Lower hinge:         0.036
         Median:          0.105
         Upper hinge:         0.240
         Maximum:         3.847

               0 H 001111122333444
               0   566889
               1 M 0000001244
               1   58
               2 H 24
               2
               3   13
               3
               4   14
               4   7
               5   4
               5   6
               6
               6
               7
               7   9
               8
               8
               9
               9
              10
              10   7
Extreme outliers:

              20   0
              38   4
```





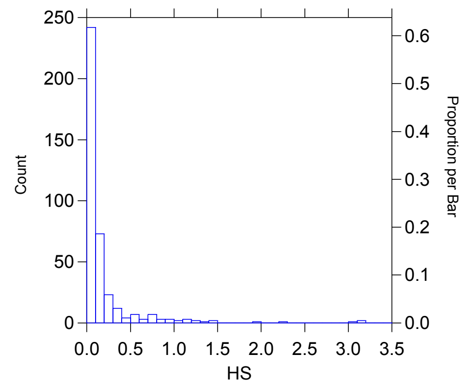
There are clearly two extreme outliers in the Zhaobaogou distribution as well. If these two are omitted, the distribution of values at or below 1.100 (lower histogram) is unremarkable. The highest value of 3.847 comes from 07C021, a systematic collection (noted above as an outlier in general) with 100 Zhaobaogou sherds and nothing else. This collection unit must have been placed on a surface artifact hot spot, and this value has been reduced to 1.100. The next highest value of 2.000 comes from 99A014, a systematic collection with 21 Zhaobaogou and 3 Liao sherds. It seems likely that this was also placed on a hot spot, so it has also been reduced to 1.100. The next highest value for Zhaobaogou (1.075) is from the same site: collection 99A013. It might be a slight exaggeration of the Zhaobaogou density, but this is clearly a site with a good bit of Zhaobaogou material, and it is not too much higher than the next value down, which is from a systematic collection whose sherds are mostly not Zhaobaogou. Values from 1.075 down, then, have not been modified.

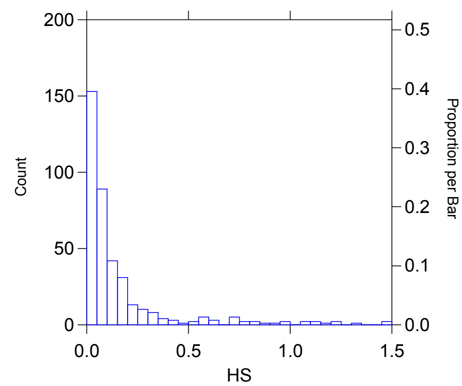
Hongshan

```
Stem and Leaf Plot of variable:          HS, N = 392
         Minimum:        0.001
         Lower hinge:        0.034
         Median:         0.074
         Upper hinge:        0.167
         Maximum:        3.132

               0    00000000000000000000001111111111111111111111111111111111111
               0 H  22222222222222222222222222222222333333333333333333333333333333
               0    44444444444444444444444444444444555555555555555555555
               0 M  66666666666666777777777777777777777777
               0    888888888888888889999999999999
               1    0000000001111111111
               1    222222222222333333
               1    444445555555
               1 H  66667777777
               1    8888889999999
               2    1
               2    22222233
               2    444455
               2    677
               2    88999
               3    00
               3    2333
               3    445
               3    67
               3    9
               4    011
               4
               4
               4    6
               4
               5    01
               5
               5    55
               5    77
               5    9
               6    114
               7    00
               7    33
               7    4
               7    6
               7    8
               8    0
               8    2
               8    5
               8
               8
               9
               9    3
               9
               9    7
               9    8
              10
              10
              10    5
              10
              10    8
              11    1
              11
              11    4
              11
              11    8
              12
              12    33
              12
              12
              12
              13    0
              13
              13
              13
              13
              14
              14
              14
              14    66
Extreme outliers:

              19    1
              22    6
              30    9
              31    03
```





Without the five extreme outliers, the distribution of values below 1.500 (lower histogram) is unremarkable. The following values have been reduced to 1.500:

–00B015 (3.132), a general collection with 35 Hongshan sherds and an area of 1.57 ha; two adjacent units have high but not outlier Hongshan values

–06A015 (3.108), a general collection with 25 Hongshan and 1 Upper Xiajiadian sherds in an area of 1.62 ha; three adjacent units have high but not outlier Hongshan values

–98A019 (3.093), a systematic collection with 17 Hongshan and 6 Lower Xiajiadian sherds; three adjacent lots have high Hongshan values, but this must be a hot spot

–98A018 (2.266), a systematic collection with 51 Hongshan sherds; three adjacent lots have high Hongshan values, but this must be a hot spot

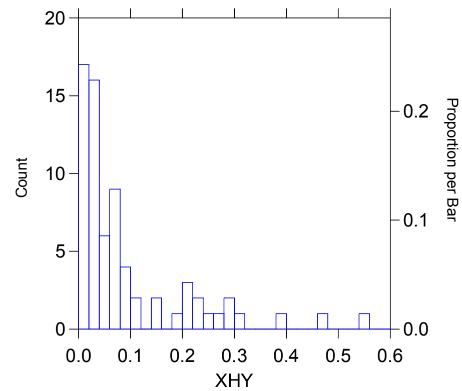
–99C008 (1.920), a general collection with 33 Hongshan and 4 other sherds; likely a bit high, and reducing it to 1.500 lowers it a little

Values from 1.500 down have not been modified.

Xiaoheyan

```
Stem and Leaf Plot of variable:         XHY, N = 70
         Minimum:         0.000
         Lower hinge:         0.021
         Median:          0.047
         Upper hinge:         0.104
         Maximum:         0.558

               0    00000000000111111
               0  H 2222222222233333
               0  M 445555
               0    666667777
               0    8899
               1  H 00
               1
               1    44
               1
               1    9
               2    001
               2    22
               2    5
               2    7
               2    88
               3    1
               3
               3
               3
               3    9
               4
               4
               4
               4    7
               4
               5
               5
               5    5
```



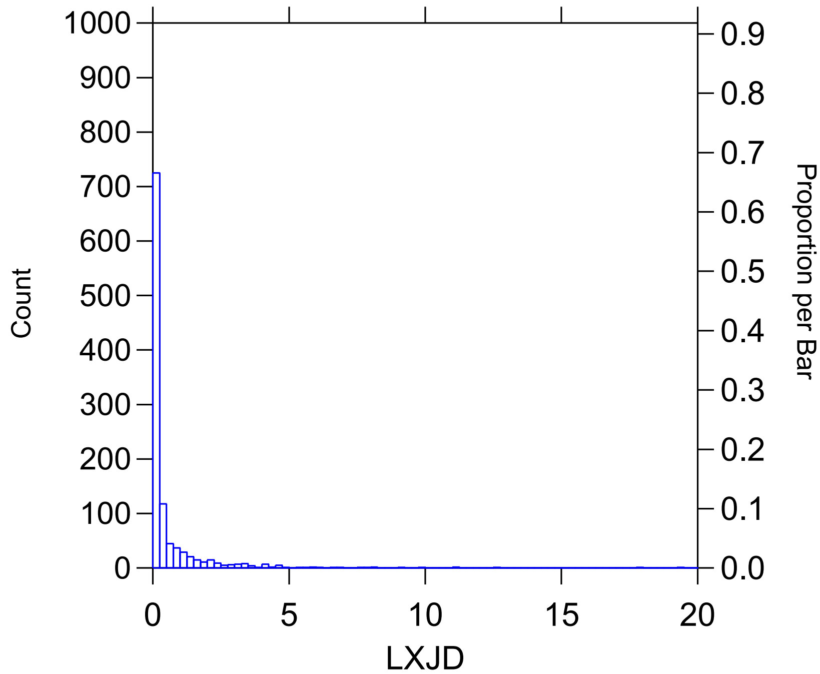
No outliers or anything else about this distribution calls for remedial action, and the values are modest compared with those of adjacent periods. No changes have been made.

Lower Xiajiadian

```
Stem and Leaf Plot of variable:        LXJD, N = 1089
         Minimum:         0.000
         Lower hinge:         0.045
         Median:          0.118
         Upper hinge:         0.397
         Maximum:        19.396

  0 H 00000000000000000000000000000000000000000000000000000011111111111111111111111111111111111111111111111111>>>>>>>>
  1 M 000000000000000000111111111111111111111111111111112222222222222222222222222223333333333333344444444444>>>>>>>>
  2   000000111111111111122222222222222222223333333333333344455555666666778888888888889999999999
  3 H 00000000111111222223333333333333444445556666666667777788999999999
  4   01111222333444556677789
  5   0133345555577778889
  6   001112234455566778999
  7   001335556899
  8   01233356666688
  9   001344566678899
 10   122334677888999
 11   33556789
 12   014444556678
 13   335556778
 14   134668
 15   35889
 16   00122349
 17   2257777
 18   49
 19   4578
 20   22578
 21   122468
 22   0023668
 23   567
 24   048
 25   03447
 26
 27   8
 28   5557
 29   2
 30   347
 31   6789
 32   668
 33   06
 34   016
 35   5
 36   4
 37   016
 38
 39
 40   1578
 41   7
 42   12
 43   27
 44
 45   059
 46   3
 47   2
 48   2
 49
 50
 51
 52
 53   5
 54
 55
 56   9
 57
 58   1
 59   5
 60   8
 61
 62
 63
 64
 65   4
 66
 67
 68
 69   9
 70
 71
 72
 73
 74
 75   9
 76
 77   8
 78
 79
 80   1
 81   8
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91   7
 92
 93
 94
 95
 96
 97   8
 98
 99
110   5
111
112   3
113
114
115
116
117
118
119
120
121
122
123
124
125   8
```

```
CONTINUED
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179   6
180
181
182
183
184
185
186
187
188
189
190
191
192
193   9
```



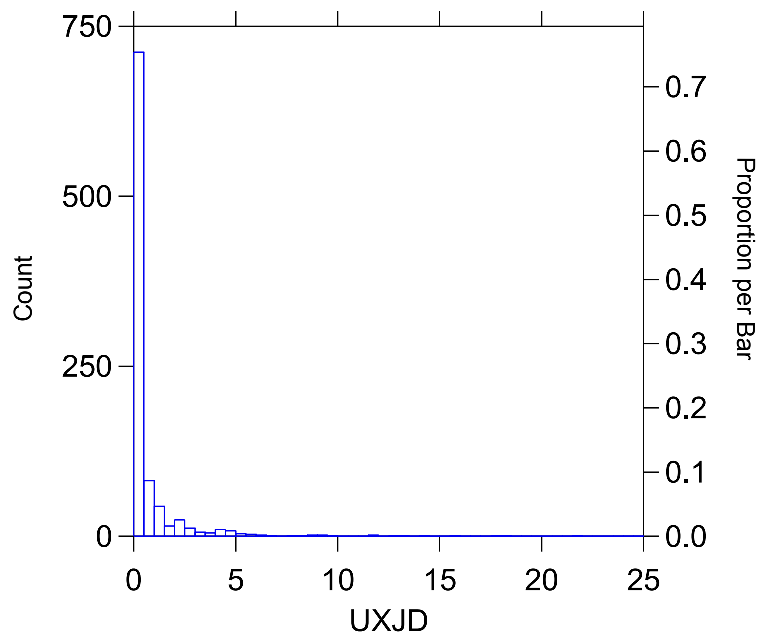
This distribution has a very strong upward straggle without an easy break to facilitate identification of outliers. The two highest values are set off by a break in the distribution, but these values are high largely because the collection units were large; their densities are not outliers. They thus probably should be unusually large area-density values. Some of the high values are attributable to collecting hot spots, but the high densities from Lower Xiajiadian hot spots noted above combine with small areas to yield area-density values that are not outliers. Thus no modifications to the Lower Xiajiadian area-density values seem warranted.

Upper Xiajiadian

```
Stem and Leaf Plot of variable:        UXJD, N = 944
         Minimum:        0.000
         Lower hinge:        0.064
         Median:         0.171
         Upper hinge:        0.492
         Maximum:       21.712

  0 H 000000000000000000000000000000000011111111111111111111111111111111111111111112222222222222222222222222222222222222222>>>
  1 M 0000000000000000000000000011111111111111111222222222222222222222222222222333333333333333333444444444444444444444
  2   0000000000000000011111111111111112222222222222222222222333333344444445555555555555666666777777777777788889999
  3   000000111111222222333344455555566666666667777778888999999
  4   0001111122333334444455566666677777889999
  5   12223333333334444455667788889999
  6   0001222335679
  7   000111334456699
  8   000011455666789
  9   12234788
 10   333667799
 11   00112334668999
 12   45679
 13   1266789
 14   123445789
 15   4458
 16   236
 17   1469
 18   3
 19   368
 20   47
 21   166
 22   00133445
 23   234557
 24   00339
 25   9
 26   48
 27   1
 28   279
 29   13449
 30   45
 31   1
 32   0
 33
 34   13
 35
 36   07
 37   7
 38   4
 39   8
 40   24
 41   023
 42   99
 43   23
 44   4
 45   133
 46   38
 47   0
 48
 49   59
 50   69
 51
 52   4
 53
 54   0
 55   7
 56
 57   0
 58
 59   1
 60   6
 61
 62
 63   1
 64
 65   8
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79   2
 80
 81   7
 82
 83
 84
 85   7
 86
 87
 88
 89   5
 90   6
 91   7
 92
 93
 94
 95
 96
 97   3
 98
 99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116   3
117
118   1
119
120

CONTINUED
121
122
123
124
125
126
127
128
129   6
130   1
131
132
133
134
135
136
137
138
139
140
141
142   2
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159   5
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178   9
179
180
181
182   1
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217   1
```



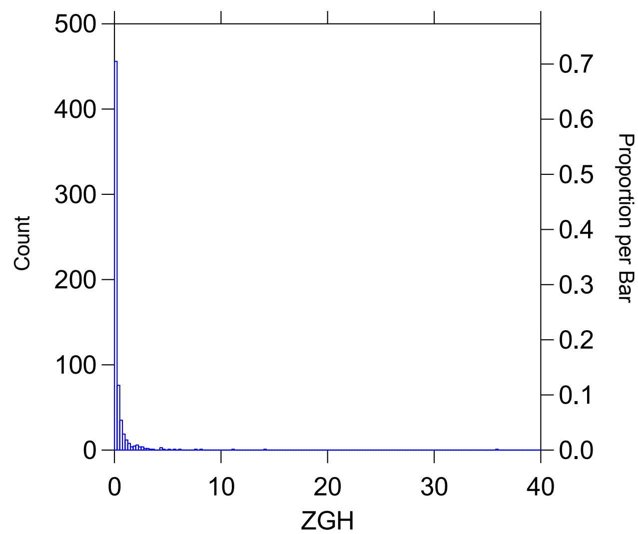
This distribution, like Lower Xiajiadian, has a very strong upward straggle without an easy break to facilitate identification of outliers. The highest values are systematic collections in units with large areas; their densities are not outliers. They thus probably should be unusually large area-density values. No modifications to the Upper Xiajiadian area-density values seem warranted.

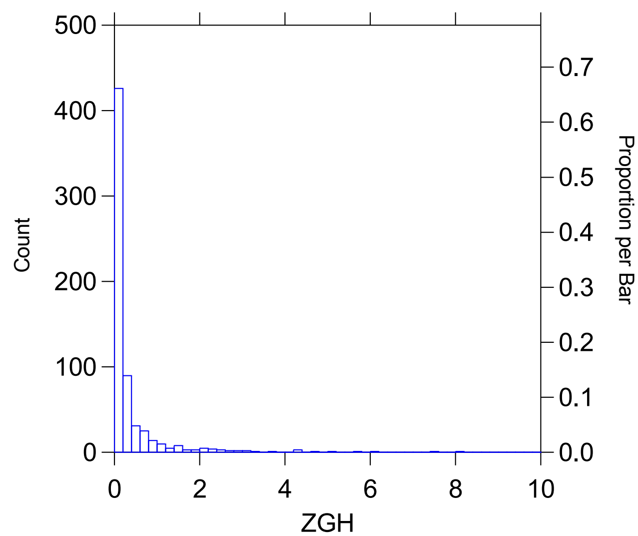
Zhanguo-Han

```
Stem and Leaf Plot of variable:          ZGH, N = 647
         Minimum:        0.000
         Lower hinge:        0.030
         Median:         0.093
         Upper hinge:        0.307
         Maximum:       35.773

          0 H 0000000000000000000000000000000000000000000000000111111111111111111111111111111111111111111111111111111111111>>>
          1   00000000000000000000111111111111111111222222222222333333444444444444445555555666666677778889999999999
          2   000000001111122233333444444444455555556666677777788888899
          3 H 000000011112223333444455567777899
          4   0022345778888999
          5   112222344556778
          6   002223344556888
          7   122456689
          8   0335568
          9   0123368
         10   011239
         11   0122
         12   1159
         13   5
         14   14567
         15   579
         16   7
         17   66
         18   79
         19   2
         20   12
         21   168
         22   48
         23   49
         24   6
         25   35
         26   45
         28
         29   79
         30   14
         31
         32
         33   6
         34
         35
         36   0
         37
         38
         39
         40
         41
         42   79
         43   4
         44
         45
         46   4
         47
         48
         49
         50   2
         51
         52
         53
         54
         55
         56   0
         57
         58
         59
         60   0
         61
         62
         63
         64
         65
         66
         67
         68
         69
         70
         71
         72
         73
         74
         75   2
         76
         77
         78
         79
         80
         81   1
EXTREME OUTLIERS:

        111   2
        141   2
        357   7
```

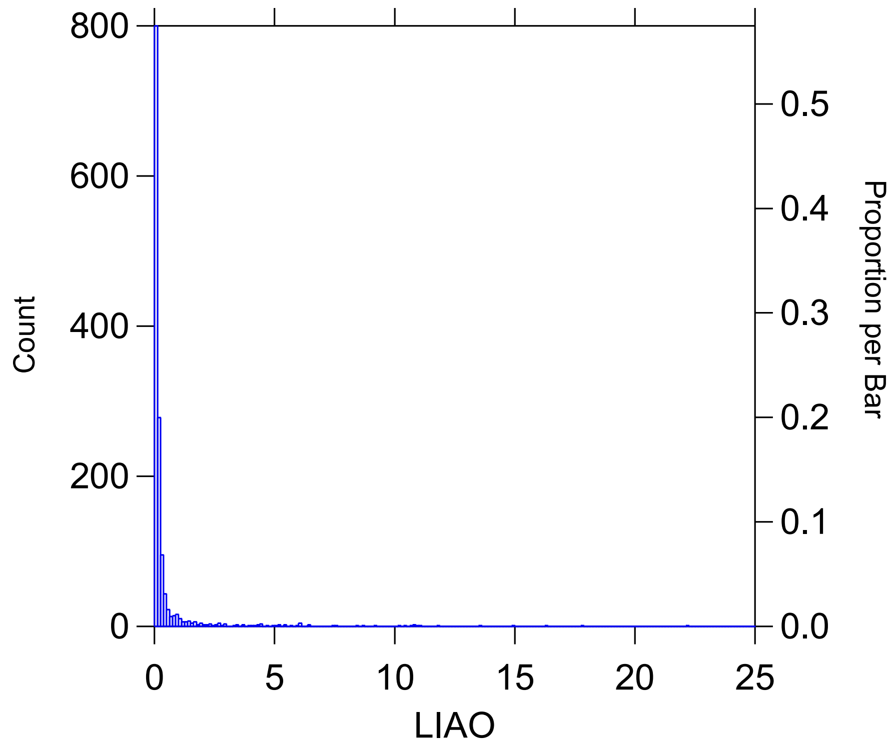




Zhanguo-Han provides another distribution with very strong upward straggle. A break in the distribution suggests three extreme outliers, and their elimination would leave a better distribution (lower histogram). The highest two values are 35.773 and 14.122 for 01E147 and 01E148, respectively. These collection units are not especially large (1.14 ha and 1.43 ha), but have systematic collections with overall densities noted as outliers in the first section. A third collection from this same site has a density only slightly lower. Surely these are hot spots, although clearly this is a site with very dense Zhanguo-Han sherds. 98A003 is third highest, with 11.122, again surely a hot spot (in the first site of the survey). Other hot spot collections correspond to small areas and do not produce such extreme area-density values. The three highest values, then, have been reduced to 8.115 (the same as the fourth highest).

Liao

```
Stem and Leaf Plot of variable:        LIAO, N = 1392
         Minimum:        0.000
         Lower hinge:        0.047
         Median:         0.101
         Upper hinge:        0.225
         Maximum:       22.244
               0 H 00000000000000000000000000000000000000111111111111111111111111111111111111111111111111111111111111111111111111>>>
               1 M 00000000000000000000000000000000000000000000111111111111111111111111111111111111111111111112222222222222222222>>>
               2 H 00000000000000000000011111111111111111111112222222222222222222222222222233333333333344444444444444455555555555>>>
               3   000000011222222223333344444555566667777899999999
               4   0000011223333345566666777777889
               5   0122233444667789
               6   00111256789
               7   01122223566777889
               8   001267789
               9   123345666899
              10   2445899
              11   001577
              12   02358
              13   03477
              14   112367
              15   012
              16   17799
              17   2479
              18
              19   0288
              20   35
              21   7
              22   26
              23   56
              24   3
              25   59
              26   34
              27   13
              28   3
              29   149
              30
              31
              32
              33   1
              34   39
              35
              36   5
              37   3
              38   7
              39
              40   6
              41
              42   1
              43   15
              44   224
              45
              46   4
              47
              48
              49   0
              50   2
              51   25
              52
              53
              54   07
              55
              56
              57   2
              58
              59   6
              60   999
              61   2
              62
              63
              64   89
              65
              66
              67
              68
              69
              70
              71
              72
              73
              74   9
              75   1
              76
              77
              78
              79
              80
              81
              82
              83
              84   5
              85
              86
              87   4
              88
              89
              90
              91
              92   4
              93
              94
              95
              96
              97
              98
              99
             100
             101
             102   3
             103
             104   2
             105
             106
             107   3
             108   46
             109   4
             110
             111   0         EXTREME OUTLIERS?:
             112
             113                        136   1
             114                        149   0
             115                        163   0
             116                        177   8
             117   7                    222   4
```



Liao provides yet another distribution with a very strong upward straggle, accentuated here by very large numbers of collections with quite low densities. Five values might be separated as extreme outliers by a bit of a break in the distribution, but this break is not as clearly marked as those utilized for other periods. Moreover, not only these five possible extreme outliers, but the 14 highest area-density values are in a historically known Liao town. They are all systematic collections, some made in 2007, some in 2000. This pattern suggests that these are not outliers for which some correction is in order but rather an accurate reflection of the extremely high density values in this town. Thus no Liao area-density values were modified.

Final Adjustment of Density Figures

The outliers listed above, then, for which corrective action was taken to bring the area-density index into line with the overall frequency distribution required correction of density figures so that the area-density analysis embedded in maps is not affected by those outliers. The densities for these 12 collections have been reduced in the data files so that the area-density index multiplies out to the corrected values as discussed above. Density corrections were made as follows:

Collection No.	Period	Uncorrected Density	Corrected Density
98A003	ZGH	10.68	7.7950
98A018	HS	7.22	4.7800
98A019	HS	2.41	1.1670
99A014	ZBG	3.22	1.7700
99C008	HS	1.78	1.3940
00B015	HS	2.00	0.9580
00E023	XLW	0.52	0.0638
01E147	ZGH	31.42	7.1280
01E148	ZGH	9.91	5.6935
06A015	HS	1.92	0.9280
07B087	XLW	0.50	0.2231
07C021	ZBG	14.15	4.0480

These corrected density values are the basis for the settlement analysis. If the outlier was created by the special combination of an unrealistically high density value together with a large area across which such a high density could not have been maintained, the corrected density value may be quite small. Since it is used in such cases to characterize a large area, though, the resulting relative population indicated for the locality is as it should be.

Reference Cited

Peterson, Christian E.

2006 *"Crafting" Hongshan Communities? Household Archaeology in the Chifeng Region of Eastern Inner Mongolia, PRC*. Ph.D. Dissertation, University of Pittsburgh.