A function describing the regularity of changes of the potential relative error that occurs in the comparison of randomly selected assemblages of potsherds

Alexander Akulov
independent scholar; Saint Petersburg, Russia; e-mail: aynu@inbox.ru

Tresi Nonno
independent scholar; Chiba, Japan; e-mail: tresi_nonno@hotmail.com

Abstract

The error that occurs in estimating degrees of resemblance of randomly selected assemblages of potsherds depends on the ratio of the numbers of potsherds. Originally the error is determined by the following points: if the ratio of numbers of potsherds is 1 – 0.7 the error is 0.01 or less, if the ratio is 0.65 – 0.6 the error is about 0.045, if the ratio is 0.55 – 0.5 the error is about 0.075, if the ratio is 0.45 the error is 0.14, if the ratio is 0.4 and less the error is about 0.19. To derive a function from the aboveShown data was employed an online calculator that used several regression models to approximate an unknown function given by a set of data points. The function that fits the aboveShown data best way is the following quadratic regression: \( \delta = -0.88r^2 + 1.52r + 0.35 \), \( r \) is the ratio of numbers of potsherds.

Keywords: comparing assemblages of potsherds; mathematical semiotics; ornaments of pottery; semiotics

1. Introduction to the problem

If there are two randomly selected assemblages/collections of randomly broken potsherds with fragments of a certain ornament, it is possible to conclude about the most frequent ornamental imprints. The most frequent imprints are supposed to be the most characteristic imprints of a certain local ceramic tradition. And thus, comparing the frequency of different imprints, it is possible to conclude about the degree of resemblance of assemblages of potsherds and about the degree of relatedness of the corresponding groups of people. To estimate the degree of resemblance of two assemblages of potsherds should be done the following procedures: 1) to estimate the degree of correlation of sets of represented ornamental imprints, 2) to estimate the degree of correlation of frequencies of common imprints (imprints represented upon potsherds belonging to each of compared assemblages), 3) to estimate the degree of exclusion of potential deviation/error, 4) to take a superposition of these three parameters. The closer are certain ornamental traditions, the higher is the corresponding degree of resemblance.

For the sake of shortness and convenience, the method was named the Monte Carlo method since it is about comparing randomly selected collections of potsherds which in their turn were randomly broken. There are actually three moments of random here: potsherds were randomly broken, a random amount of potsherds was picked up, and a random amount of the picked up was published. The method can also be named the method of frequencies since its main point is the comparison of frequencies of ornamental imprints.
Originally the formula used for calculating the degree of resemblance of two assemblages of potsherds was the following:

\[
\frac{1}{2} \left( \frac{m}{N_{\text{imp}(A)}} + \frac{m}{N_{\text{imp}(B)}} \right) \times \frac{1}{m} \left( \frac{\text{lower frequency}_1}{\text{higher frequency}_1} + \cdots + \frac{\text{lower frequency}_m}{\text{higher frequency}_m} \right)
\]

where:

- \( N_{\text{imp}(A)} \) – the number of imprints represented in A (first set),
- \( N_{\text{imp}(B)} \) – the number of imprints represented in B (second set),
- \( m \) – the number of common imprints.

(see Akulov, Nonno 2019: 36).

And recently the formula has been modified in order to take into account the potential error, the modified formula has the following view:

\[
\frac{1}{2} \left( \frac{m}{N_{\text{imp}(A)}} + \frac{m}{N_{\text{imp}(B)}} \right) \times \frac{1}{m} \left( \frac{\text{lower frequency}_1}{\text{higher frequency}_1} + \cdots + \frac{\text{lower frequency}_m}{\text{higher frequency}_m} \right) \times (1 - \delta)
\]

where \( \delta \) is the relative error, and the \((1 - \delta)\) element shows how close the compared collections are in the quantitative aspect, and to what extent the possibility of deviation is excluded (see Nonno 2022: 17).

In a special work, the regularity of the change in the relative error depending on the number of potsherds in the compared sets was described (see Nonno 2019). The regularity of changes of the potential relative error was set by several points: if the ratio of numbers of potsherds in the compared assemblages is 1 – 0.7 then the potential error is 0.01 or even less, if the ratio is 0.65 – 0.6 then the potential error is about 0.045, if the ratio is 0.55 – 0.5 then the potential error is about 0.075, if the ratio is 0.45 then the potential error is 0.14, and if the ratio is 0.4 and less then the potential error is about 0.19 (for more details about the regularity see Nonno 2019: 54 – 55). And it was noted that there definitely could be a certain function behind this regularity (Nonno 2022: 17).

As far as it is more convenient to have a function that describes the regularity of changes of the error, so in the current paper we want to represent a function that can describe the changes of the potential error.

2. An approximation of the function

The regularity of changes of the error shown above is summarized in the table below (see Table 1).

<table>
<thead>
<tr>
<th>x</th>
<th>0.4</th>
<th>0.45</th>
<th>0.5</th>
<th>0.55</th>
<th>0.6</th>
<th>0.65</th>
<th>0.7</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>0.19</td>
<td>0.14</td>
<td>0.075</td>
<td>0.075</td>
<td>0.045</td>
<td>0.045</td>
<td>0.01</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. The line of \( x \) represents values of ratios of potsherds numbers in the compared assemblages, and the line of \( y \) represents the corresponding values of the error
To derive a function after the data shown in Table 1 we used an online calculator that used several regression models to approximate an unknown function given by a set of data points (see Function approximation with regression analysis).

We have received two functions that relatively well describe the data from the Table 1 (see Fig. 1).

The first function is a quadratic regression of the following view:

\[ \delta = 0.88r^2 - 1.52r + 0.65. \]

And the second function is a cubic regression of the following view:

\[ \delta = -2.17r^3 + 5.35r^2 - 4.42r + 1.24. \]

In both formulas r means the ratio of numbers of potsherds in the compared assemblages.

At the first sight they both relatively well describe the data of Table 1; however, we have to choose one of them. In order to choose the one that fits the data of Table 1 better, we test both these functions on the material of the standard set of assemblages.

3. The standard set of assemblages used in comparison of assemblages of the Pit-Comb Ware

In some papers devoted to the estimation of degrees of resemblance of different assemblages of potsherds of the Pit-Comb Ware is shown a table with values of the degree of resemblance
given by a standard set of assemblages (Table 1, also see Akulov 2019, Akulov 2020), i.e.: a set of assemblages that are supposed to have belonged to closely related groups. These assemblages are from sites that are very close geographically (Fig. 2) and also are close in time (all are supposed to belong to the Late Neolithic period, see Akulov 2019: 2).

Fig. 2. Locations of sites considered in the current paper; a triangle means sites with stone tools of late types, a circle means sites of the late Neolithic stage, and a square means Mesolithic site (the upper map was made after Bing maps screenshot, the lower image source – Gurina 1961: 415)
Table 2. The values of the degree of resemblance of the standard set of assemblages calculated without taking into account the potential error (R1 – Razliv 1, R2 – Razliv 2, R4 – Razliv 4, R5 – Razliv 5, R7 – Razliv 7, GR – Glinyanyi Ruchei, SG – Sosnovaya Gora, Tar – Tarkhovka)

When the formula for calculating the degree of resemblance was modified and took into account the potential error the table with the values of the degree of resemblance showed by the standard set of assemblages also was modified (see Table 3).

Table 3. The values of the degree of resemblance calculated with the formula that takes into account the potential error determined by a list of points

To understand which formula (quadratic regression or cubic regression) better fits the existing pattern, we calculate the values of the correlation index for the standard set of assemblages and compare the received values with the values shown in Table 3.

4. Calculating values of the degree of resemblance of the standard set of assemblages with quadratic regression

To calculate values of the degree of resemblance of the standard set of assemblages using quadratic regression we multiply the values shown in Table 2 by corresponding values of $(1 - \delta)$, in the current case $\delta$ is calculated according to the following formula: $\delta = 0.88r^2 - 1.52r + 0.65$. The information about numbers of potsherds of the compared assemblages for calculating ratios is taken from (Akulov 2019).
R1 and R2

The assemblage of R1 has 16 potsherds, the assemblage of R2 has 5 potsherds (Akulov 2019: 7, 8, 16 – 17), so their ratio is 5/16 = 0.31. According to the formula with quadratic regression, the error is the following: 0.88*0.31^2 – 1.52*0.31 + 0.65 = 0.26. And the corresponding value of the degree of resemblance of R1 and R2 is the following: 0.38 * (1 – 0.26) = 0.28.

R1 and R4

The assemblage of R4 has 5 potsherds (Akulov 2019: 9, 17), so the ratio of numbers of potsherds of R1 and R4 is 5/16 = 0.31. According to the formula with quadratic regression, the error is the following: 0.88*0.31^2 – 1.52*0.31 + 0.65 = 0.26. And the new value of the degree of resemblance of R1 and R4 is the following: 0.37 * (1 – 0.26) = 0.27.

R1 and R5

The assemblage of R5 has 8 potsherds (Akulov 2019: 9, 18), so the ratio of numbers of potsherds of R1 and R5 is 8/16 = 0.5. According to the formula with quadratic regression, the error is the following: 0.88*0.5^2 – 1.52*0.5 + 0.65 = 0.11. And the corresponding value of the degree of resemblance of R1 and R5 is the following: 0.25 * (1 – 0.11) = 0.22.

R1 and R7

The assemblage of R7 has 5 potsherds (Akulov 2019: 10, 20), so the ratio of numbers of potsherds of R1 and R7 is 5/16 = 0.31. According to the formula with quadratic regression, the error is the following: 0.88*0.31^2 – 1.52*0.31 + 0.65 = 0.26. And the corresponding value of the degree of resemblance of R1 and R7 is the following: 0.2 * (1 – 0.26) = 0.15.

R1 and GR

The assemblage of GR has 10 potsherds (Akulov 2019: 10, 19), so the ratio of numbers of potsherds of R1 and GR is 10/16 = 0.62. According to the formula with quadratic regression, the error is the following: 0.88*0.62^2 – 1.52*0.62 + 0.65 = 0.05. And the corresponding value of the degree of resemblance of R1 and GR is the following: 0.39 * (1 – 0.05) = 0.37.
R1 and SG

The assemblage of SG has 16 potsherds (Akulov 2019: 12 – 13, 21), so the ratio of numbers of potsherds of R1 and SG is 1.
According to the formula with quadratic regression, the error is the following:
$$0.88 \cdot 1^2 - 1.52 \cdot 1 + 0.65 = 0.01$$
And the corresponding value of the degree of resemblance of R1 and SG is the following:
$$0.26 \cdot (1 - 0.01) = 0.26.$$ 

R1 and Tar

The assemblage of Tar has 17 potsherds (Akulov 2019: 11, 20), so the ratio of numbers of potsherds of R1 and Tar is 16/17 ≈ 0.94.
According to the formula with quadratic regression, the error is the following:
$$0.88 \cdot 0.94^2 - 1.52 \cdot 0.94 + 0.65 = 0$$
And the corresponding value of the degree of resemblance of R1 and Tar is the following: 0.23 * (1 – 0) = 0.23.

R2 and R4

The assemblage of R2 has 5 potsherds, and the assemblage of R4 has 5 potsherds, so the ratio of numbers of potsherds of R2 and R4 is 1.
According to the formula with quadratic regression, the error is the following:
$$0.88 \cdot 1^2 - 1.52 \cdot 1 + 0.65 = 0.01$$
And the corresponding value of the degree of resemblance R2 and R4 is the following: 0.52 * (1 – 0.01) ≈ 0.51.

R2 and R5

The assemblage of R5 has 8 potsherds, so the ratio of numbers of potsherds of R2 and R5 is 5/8 ≈ 0.625.
According to the formula with quadratic regression, the error is the following:
$$0.88 \cdot 0.625^2 - 1.52 \cdot 0.625 + 0.65 \approx 0.04$$
And the corresponding value of the degree of resemblance R2 and R5 is the following: 0.28 * (1 – 0.04) ≈ 0.27.

R2 and R7

The assemblage of R7 has 5 potsherds, so the ratio of numbers of potsherds of R2 and R7 is 1.
According to the formula with quadratic regression, the error is the following:
$$0.88 \cdot 1^2 - 1.52 \cdot 1 + 0.65 = 0.01$$
And the corresponding value of the degree of resemblance of R2 and R7 is the following: 0.31 * (1 – 0.01) = 0.3
R2 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R2 and GR is 5/10 = 0.5. According to the formula with quadratic regression, the error is the following:

$$0.88 \times 0.5^2 - 1.52 \times 0.5 + 0.65 = 0.11$$

And the corresponding value of the degree of resemblance of R2 and GR is the following: 0.35 * (1 – 0.11) = 0.31.

R2 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R2 and SG is 5/16 ≈ 0.31. According to the formula with quadratic regression, the error is the following:

$$0.88 \times 0.31^2 - 1.52 \times 0.31 + 0.65 = 0.26$$

And the corresponding value of the degree of resemblance of R2 and SG is the following: 0.37 * (1 – 0.26) = 0.27.

R2 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R2 and Tar is 5/17 ≈ 0.29. According to the formula with quadratic regression, the error is the following:

$$0.88 \times 0.29^2 - 1.52 \times 0.29 + 0.65 = 0.28$$

And the corresponding value of the degree of resemblance of R2 and Tar is the following: 0.67 * (1 – 0.28) = 0.48.

R4 and R5

The assemblage of R4 has 5 potsherds, the assemblage of R5 has 8 potsherds, so the ratio of numbers of potsherds of R4 and R5 is 5/8 ≈ 0.625. According to the formula with quadratic regression, the error is the following:

$$0.88 \times 0.625^2 - 1.52 \times 0.625 + 0.65 = 0.04$$

And the corresponding value of the degree of resemblance of R4 and R5 is the following: 0.49 * (1 – 0.04) = 0.47.

R4 and R7

The assemblage of R7 has 5 potsherds, so the ratio of numbers of potsherds of R4 and R7 is 1. According to the formula with quadratic regression, the error is the following:

$$0.88 \times 1^2 - 1.52 \times 1 + 0.65 = 0.01$$

And the corresponding value of the degree of resemblance of R4 and R7 is the following: 0.38 * (1 – 0.01) = 0.37.
R4 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R4 and GR is 5/10 = 0.5.
According to the formula with quadratic regression, the error is the following:
0.88*0.5^2 – 1.52*0.5 + 0.65 = 0.11
And the corresponding value of the degree of resemblance is the following:
0.29 * (1 – 0.11) = 0.26.

R4 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R4 and SG is 5/16 = 0.31.
According to the formula with quadratic regression, the error is the following:
0.88*0.31^2 – 1.52*0.31 + 0.65 = 0.26
And the corresponding value of the degree of resemblance of R4 and SG is the following:
0.32 * (1 – 0.26) = 0.23.

R4 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R4 and Tar is 5/17 = 0.29
According to the formula with quadratic regression, the error is the following:
0.88*0.29^2 – 1.52*0.29 + 0.65 = 0.28
And the corresponding value of the degree of resemblance of R4 and Tar is the following:
0.57 * (1 – 0.28) = 0.41.

R5 and R7

R5 has 8 potsherds, R7 has 5 potsherds, so the ratio of numbers of potsherds of R5 and R7 is 5/8 = 0.625.
According to the formula with quadratic regression, the error is the following:
0.88 * 0.625^2 – 1.52*0.625 + 0.65 = 0.04
And the corresponding value of the degree of resemblance of R5 and R7 is the following:
0.5 * (1 – 0.04) = 0.48.

R5 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds or R5 and GR is 8/10 = 0.8.
According to the formula with quadratic regression, the error is the following:
0.88 * 0.8^2 – 1.52*0.8 + 0.65 = 0.
And so the corresponding value of the degree of resemblance of R5 and GR is the following:
0.25 * (1 – 0) = 0.25.
R5 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of potsherds of R5 and SG is $8/16 = 0.5$.
According to the formula with quadratic regression, the error is the following:
$$0.88 \times 0.5^2 - 1.52 \times 0.5 + 0.65 = 0.11$$
And the corresponding value of the degree of resemblance of R5 and SG is the following:
$$0.25 \times (1 - 0.11) = 0.22.$$ 

R5 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R5 and Tar is $8/17 \approx 0.47$.
According to the formula with quadratic regression, the error is the following:
$$0.88 \times 0.47^2 - 1.52 \times 0.47 + 0.65 = 0.13$$
And the corresponding value of the degree of resemblance of R5 and Tar is the following:
$$0.41 \times (1 - 0.13) = 0.35.$$ 

R7 and GR

The assemblage of R7 has 5 potsherds, the assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R7 and GR is $5/10 = 0.5$.
According to the formula with quadratic regression, the error is the following:
$$0.88 \times 0.5^2 - 1.52 \times 0.5 + 0.65 = 0.11$$
And the corresponding value of the degree of resemblance of R7 and GR is the following:
$$0.26 \times (1 - 0.11) = 0.23.$$ 

R7 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R7 and SG is $5/16 \approx 0.31$.
According to the formula with quadratic regression, the error is the following:
$$0.88 \times 0.31^2 - 1.52 \times 0.31 + 0.65 = 0.26$$
And the corresponding value of the degree of resemblance is the following:
$$0.18 \times (1 - 0.26) = 0.13.$$ 

R7 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R7 and Tar is $5/17 \approx 0.29$.
According to the formula with quadratic regression, the error is the following:
$$0.88 \times 0.29^2 - 1.52 \times 0.29 + 0.65 = 0.28$$
And the corresponding value of the degree of resemblance of R7 and Tar is the following:
$$0.43 \times (1 - 0.28) = 0.3.$$
GR and SG

The assemblage of GR has 10 potsherds, the assemblage of SG has 16 potsherds, so the ratio of numbers of potsherds of GR and SG is 10/16 = 0.625. According to the formula with quadratic regression, the error is the following: $0.88 \times 0.625^2 - 1.52 \times 0.625 + 0.65 = 0.04$.

And the corresponding value of the degree of resemblance of GR and SG is the following: $0.14 \times (1 - 0.04) = 0.13$.

GR and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of GR and Tar is 10/17 ≈ 0.59. According to the formula with quadratic regression, the error is the following: $0.88 \times 0.59^2 - 1.52 \times 0.59 + 0.65 = 0.06$.

And the corresponding value of the degree of resemblance of GR and Tar is the following: $0.34 \times (1 - 0.04) = 0.32$.

SG and Tar

The assemblage of SG has 16 potsherds; the assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of SG and Tar is 16/17 = 0.94. According to the formula with quadratic regression, the error is the following: $0.88 \times 0.94^2 - 1.52 \times 0.94 + 0.65 = 0$.

And the corresponding value of the degree of resemblance of GR and Tar is the following: $0.3 \times (1 - 0) = 0.3$.

All the values of the degree of resemblance of the assemblages of the standard set calculated with the formula with quadratic regression can be summarized in a table (see Table 4).

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.28</td>
<td>0.27</td>
<td>0.22</td>
<td>0.15</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.51</td>
<td>0.27</td>
<td>0.3</td>
<td>0.31</td>
<td>0.27</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>0.47</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td></td>
<td>0.48</td>
<td>0.25</td>
<td>0.22</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td></td>
<td></td>
<td>0.23</td>
<td>0.13</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The values of the degree of resemblance of the standard set of assemblages calculated with quadratic regression

5. Calculating values of the degree of resemblance of the standard set of assemblages with cubic regression

To calculate values of the degree of resemblance of the standard set of assemblages with cubic regression, we multiply the values shown in Table 2 by the corresponding values of $(1 - \delta)$, in the current case $\delta$ is calculated according to the following formula:

$\delta = -2.17r^3 + 5.35r^2 - 4.42r + 1.24.$
The information about numbers of potsherds of the compared assemblages for calculating ratios is taken from (Akulov 2019).

R1 and R2

The assemblage of R1 has 16 potsherds, the assemblage of R2 has 5 potsherds (Akulov 2019: 7, 8, 16 – 17), so their ratio is \( \frac{5}{16} = 0.31 \).

According to the formula with cubic regression, the error is the following:
\[-2.17*0.31^3 + 5.35*0.31^2 - 4.42*0.31 + 1.24 \approx 0.32\]

And the corresponding value of the degree of resemblance of R1 and R2 is the following:
\[0.38 \times (1 - 0.32) \approx 0.26.\]

R1 and R4

The assemblage of R4 has 5 potsherds (Akulov 2019: 9, 17), so the ratio of numbers of potsherds of R1 and R4 is \( \frac{5}{16} = 0.31 \).

According to the formula with cubic regression, the error is the following:
\[-2.17*0.31^3 + 5.35*0.31^2 - 4.42*0.31 + 1.24 = 0.32\]

And the corresponding value of the degree of resemblance of R1 and R4 is the following:
\[0.37 \times (1 - 0.32) \approx 0.25.\]

R1 and R5

The assemblage of R5 has 8 potsherds (Akulov 2019: 9, 18), so the ratio of numbers of potsherds of R1 and R5 is \( \frac{8}{16} = 0.5 \).

According to the formula with cubic regression, the error is the following:
\[-2.17*0.5^3 + 5.35*0.5^2 - 4.42*0.5 + 1.24 \approx 0.06\]

And the corresponding value of the degree of resemblance of R1 and R5 is the following:
\[0.25 \times (1 - 0.06) \approx 0.23.\]

R1 and R7

The assemblage of R7 has 5 potsherds (Akulov 2019: 10, 20), so the ratio of numbers of potsherds of R1 and R7 is \( \frac{5}{16} = 0.31 \).

According to the formula with cubic regression, the error is the following:
\[-2.17*0.31^3 + 5.35*0.31^2 - 4.42*0.31 + 1.24 \approx 0.32\]

And the corresponding value of the degree of resemblance of R1 and R7 is the following:
\[0.2 \times (1 - 0.32) \approx 0.13.\]

R1 and GR

The assemblage of GR has 10 potsherds (Akulov 2019: 10, 19), so the ratio of numbers of potsherds of R1 and GR is \( \frac{10}{16} = 0.62 \).

According to the formula with cubic regression, the error is the following:
\[-2.17*0.62^3 + 5.35*0.62^2 - 4.42*0.62 + 1.24 \approx 0.02\]

And the corresponding value of the degree of resemblance of R1 and GR is the following:
\[0.39 \times (1 - 0.02) \approx 0.38.\]
R1 and SG

The assemblage of SG has 16 potsherds (Akulov 2019: 12 – 13, 21), so the ratio of numbers of potsherds of R1 and SG is 1.
According to the formula with cubic regression, the error is the following:
\[-2.17*1^3 + 5.35*1^2 - 4.42*1 + 1.24 = 0.\]
And the corresponding value of the degree of resemblance of R1 and SG is the following:
\[0.26 \times (1 - 0) = 0.26.\]

R1 and Tar

The assemblage of Tar has 17 potsherds (Akulov 2019: 11, 20), so the ratio of numbers of potsherds of R1 and Tar is 16/17 ≈ 0.94.
According to the formula with cubic regression, the error is the following:
\[-2.17*0.94^3 + 5.35*0.94^2 - 4.42*0.94 + 1.24 ≈ 0.02.\]
And so the corresponding value of the degree of resemblance of R1 and Tar is the following:
\[0.23 \times (1 - 0.02) = 0.22.\]

R2 and R4

The assemblage of R2 has 5 potsherds, and the assemblage of R4 has 5 potsherds, so the ratio of numbers of potsherds of R2 and R4 is 1.
According to the formula with cubic regression, the error is the following:
\[-2.17*1^3 + 5.35*1^2 - 4.42*1 + 1.24 = 0.\]
And so the corresponding value of the degree of resemblance of R2 and R4 is the following:
\[0.52 \times (1 - 0) = 0.52.\]

R2 and R5

The assemblage of R5 has 8 potsherds, so the ratio of numbers of potsherds of R2 and R5 is 5/8 = 0.625. According to the formula with cubic regression, the error is the following:
\[-2.17*0.625^3 + 5.35*0.625^2 - 4.42*0.625 + 1.24 = 0.04.\]
And the corresponding value of the degree of resemblance of R2 and R5 is the following:
\[0.28 \times (1 - 0.04) = 0.27.\]

R2 and R7

The assemblage of R7 has 5 potsherds, so the ratio of numbers of potsherds of R2 and R7 is 1.
According to the formula with cubic regression, the error is the following:
\[-2.17*1^3 + 5.35*1^2 - 4.42*1 + 1.24 = 0.\]
And the corresponding value of the degree of resemblance of R2 and R7 is the following:
\[0.31 \times (1 - 0.0) = 0.31.\]

R2 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R2 and GR is 5/10 = 0.5. According to the formula with cubic regression, the error is the following:
-2.17*0.5^3 + 5.35*0.5^2 - 4.42*0.5 + 1.24 = 0.06.
And the corresponding value of the degree of resemblance of R2 and GR is the following:
0.35 * (1 – 0.06) = 0.33.

R2 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R2 and SG is 5/16 ≈ 0.31.
According to the formula with cubic regression, the error is the following:
-2.17*0.31^3 + 5.35*0.31^2 – 4.42*0.31 + 1.24 ≈ 0.32.
And the corresponding value of the degree of resemblance of R2 and SG is the following:
0.37 * (1 – 0.32) ≈ 0.25.

R2 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R2 and Tar is
5/17 ≈ 0.29. According to the formula with cubic regression, the error is the following:
-2.17*0.31^3 + 5.35*0.31^2 – 4.42*0.31 + 1.24 ≈ 0.32.
And the corresponding value of the degree of resemblance of R2 and Tar is the following:
0.67 * (1 – 0.32) = 0.45.

R4 and R5

The assemblage of R4 has 5 potsherds, the assemblage of R5 has 8 potsherds, so the ratio of
numbers of potsherds of R4 and R5 is 5/8 ≈ 0.625.
According to the formula with cubic regression, the error is the following:
-2.17*0.625^3 + 5.35*0.625^2 – 4.42*0.625 + 1.24 ≈ 0.04.
And the corresponding value of the degree of resemblance of R4 and R5 is the following:
0.49 * (1 – 0.04) = 0.47.

R4 and R7

The assemblage of R7 has 5 potsherds, so the ratio of numbers of potsherds of R4 and R7 is 1.
According to the new formula for error, we receive the following:
-2.17*1^3 + 5.35*1^2 – 4.42*1 + 1.24 = 0.
And the corresponding value of the degree of resemblance of R4 and R7 is the following:
0.38 * (1 – 0) = 0.38.

R4 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R4 and GR is
5/10 = 0.5. According to the formula with cubic regression, the error is the following:
-2.17*0.5^3 + 5.35*0.5^2 – 4.42*0.5 + 1.24 ≈ 0.06.
And the corresponding value of the degree of resemblance of R4 and GR is the following:
0.29 * (1 – 0.06) = 0.27.
R4 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R4 and SG is $5/16 \approx 0.31$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.31^3 + 5.35*0.31^2 - 4.42*0.31 + 1.24 \approx 0.32$$

And the corresponding value of the degree of resemblance of R4 and SG is the following:

$$0.32 * (1 - 0.32) \approx 0.22.$$  

R4 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R4 and Tar is $5/17 \approx 0.29$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.29^3 + 5.35*0.29^2 - 4.42*0.29 + 1.24 \approx 0.36.$$  

And the corresponding value of the degree of resemblance of R4 and Tar is the following:

$$0.57 * (1 - 0.36) = 0.36.$$  

R5 and R7

R5 has 8 potsherds, R7 has 5 potsherds, so the ratio of numbers of potsherds of R5 and R7 is $5/8 = 0.625$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.625^3 + 5.35*0.625^2 - 4.42*0.625 + 1.24 \approx 0.04.$$  

And the corresponding value of the degree of resemblance of R5 and R7 is the following:

$$0.5 * (1 - 0.04) = 0.48.$$  

R5 and GR

The assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds or R5 and GR is $8/10 = 0.8$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.8^3 + 5.35*0.8^2 - 4.42*0.8 + 1.24 \approx 0.02.$$  

And the corresponding value of the degree of resemblance of R5 and GR is the following:

$$0.25 * (1 - 0.02) = 0.24.$$  

R5 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of potsherds of R5 and SG is $8/16 = 0.5$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.5^3 + 5.35*0.5^2 - 4.42*0.5 + 1.24 \approx 0.06.$$  

And the corresponding value of the degree of resemblance of R5 and SG is the following:

$$0.25 * (1 - 0.06) = 0.23.$$  

R5 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R5 and Tar is $8/17 \approx 0.47$. According to the formula with cubic regression, the error is the following:

$$-2.17*0.47^3 + 5.35*0.47^2 - 4.42*0.47 + 1.24 \approx 0.06.$$  

And the corresponding value of the degree of resemblance of R5 and Tar is the following:

$$0.41 * (1 - 0.06) = 0.38.$$
R7 and GR

The assemblage of R7 has 5 potsherds, the assemblage of GR has 10 potsherds, so the ratio of numbers of potsherds of R7 and GR is 5/10 = 0.5. According to the formula with cubic regression, the error is the following:

\[-2.17*0.5^3 + 5.35*0.5^2 - 4.42*0.5 + 1.24 = 0.06\]

And the corresponding value of the degree of resemblance of R7 and GR is the following:

\[0.26 * (1 - 0.06) = 0.24\].

R7 and SG

The assemblage of SG has 16 potsherds, so the ratio of numbers of R7 and SG is 5/16 ≈ 0.31. According to the formula with cubic regression, the error is the following:

\[-2.17*0.31^3 + 5.35*0.31^2 - 4.42*0.31 + 1.24 ≈ 0.32\].

And the corresponding value of the degree of resemblance of R7 and SG is the following:

\[0.18 * (1 - 0.32) ≈ 0.12\].

R7 and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of R7 and Tar is 5/17 ≈ 0.29. According to the formula with cubic regression, the error is the following:

\[-2.17*0.29^3 + 5.35*0.29^2 - 4.42*0.29 + 1.24 ≈ 0.36\].

And the corresponding value of the degree of resemblance of R7 and Tar is the following:

\[0.43 * (1 - 0.36) ≈ 0.27\].

GR and SG

The assemblage of GR has 10 potsherds, the assemblage of SG has 16 potsherds, so the ratio of numbers of potsherds of GR and SG is 10/16 = 0.625. According to the formula with cubic regression, the error is the following:

\[-2.17*0.625^3 + 5.35*0.625^2 - 4.42*0.625 + 1.24 ≈ 0.04\].

And the corresponding value of the degree of resemblance of GR and SG is the following:

\[0.14 * (1 - 0.04) = 0.13\].

GR and Tar

The assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of GR and Tar is 10/17 ≈ 0.59. According to the formula with cubic regression, the error is the following:

\[-2.17*0.59^3 + 5.35*0.59^2 - 4.42*0.59 + 1.24 ≈ 0.05\].

And the corresponding value of the degree of resemblance of GR and Tar is the following:

\[0.34 * (1 - 0.05) = 0.32\].
SG and Tar

The assemblage of SG has 16 potsherds; the assemblage of Tar has 17 potsherds, so the ratio of numbers of potsherds of SG and Tar is $16/17 = 0.94$.

According to the formula with cubic regression, the error is the following:

$$-2.17 \times 0.94^3 + 5.35 \times 0.94^2 - 4.42 \times 0.94 + 1.24 = 0.02.$$ 

And the corresponding value of the degree of resemblance of SG and Tar is the following:

$$0.3 \times (1 - 0.02) = 0.29.$$ 

All the values of the degree of resemblance of the assemblages of the standard set calculated with the formula with cubic regression can be summarized in a table (see Table 5).

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.26</td>
<td>0.25</td>
<td>0.23</td>
<td>0.13</td>
<td>0.38</td>
<td>0.26</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.52</td>
<td>0.27</td>
<td>0.31</td>
<td>0.33</td>
<td>0.25</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>0.47</td>
<td>0.38</td>
<td>0.27</td>
<td>0.22</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>0.48</td>
<td>0.24</td>
<td>0.23</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td></td>
<td>0.24</td>
<td>0.12</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td>0.13</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td></td>
<td></td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. The set of values of the degree of resemblance of the standard set of assemblages calculated with cubic regression

6. Comparison of quadratic and cubic regressions

To understand which function (quadratic regression or cubic regression) better fits the regularity determined by points shown in the introduction, we now compare the distance between the table given by quadratic regression and the original table (the table given by the formula that takes into account the error determined by a list of points) with the distance between the table given by cubic regression and the original table.

The distance between two tables/matrixes is calculated by the following formula

$$\frac{1}{28} \sum_{i=1}^{28} |a_i - b_i|$$

where $a_i$ is an element of the first table, and $b_i$ is the corresponding element of the second table/matrix.
6.1. The distance between the table of quadratic regression and the original table

For the sake of convenience we re-show all the compared tables.

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.31</td>
<td>0.3</td>
<td>0.23</td>
<td>0.16</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.52</td>
<td>0.27</td>
<td>0.31</td>
<td>0.32</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>0.47</td>
<td>0.38</td>
<td>0.27</td>
<td>0.26</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>0.48</td>
<td>0.25</td>
<td>0.23</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td>0.24</td>
<td></td>
<td>0.14</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. The original set of values of the degree of resemblance of the standard set of assemblages

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.28</td>
<td>0.27</td>
<td>0.22</td>
<td>0.15</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.51</td>
<td>0.27</td>
<td>0.3</td>
<td>0.31</td>
<td>0.27</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>0.47</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>0.48</td>
<td>0.25</td>
<td>0.22</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td>0.23</td>
<td></td>
<td>0.13</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. The set of values of the degree of resemblance of the standard set of assemblages calculated with quadratic regression

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.01</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. The result of subtracting the table of quadratic regression from the original table

The arithmetic mean of Table 8 is the following:

\[
(0.03 + 0.04 + 0.01 + 0.03 + 0.03 + 0.08 + 0.18)/28 \approx 0.014.
\]
6.2. The distance between the table of cubic regression and the original table

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.26</td>
<td>0.25</td>
<td>0.23</td>
<td>0.13</td>
<td>0.38</td>
<td>0.26</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.52</td>
<td>0.27</td>
<td>0.31</td>
<td>0.33</td>
<td>0.25</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>0.47</td>
<td>0.38</td>
<td>0.27</td>
<td>0.22</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>0.48</td>
<td>0.24</td>
<td>0.23</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>0.24</td>
<td>0.12</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.13</td>
<td>0.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9. The set of values of the degree of resemblance of the standard set of assemblages calculated with cubic regression

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.05</td>
<td>0.05</td>
<td>0</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>0</td>
<td>0.02</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10. The result of subtracting the table of cubic regression from the original table

The arithmetic mean of Table 10 is the following:

\[
\frac{(0.05 + 0.05 + 0.03 + 0.03 + 0.11 + 0.23)/28 \approx 0.018.}
\]

6.3. Comparing the two distances

The distance between the original table and the table received with quadratic regression is 0.014.
And the distance between the original table and the table received with cubic regression is 0.018.
0.014 < 0.018, and so quadratic regression better fits the regularity of changes of the error that was described by a list of points.

7. Conclusion

Thus, now, having inserted the above used formula of quadratic regression into the formula that takes into account the error, we receive the following final formula:
where:

\[ N_{\text{imp}(A)} \] – the number of imprints represented in A (first set),

\[ N_{\text{imp}(B)} \] – the number of imprints represented in B (second set),

\[ m \] – the number of common imprints,

\[ r \] – is the ratio of numbers of potsherds of compared assemblages.

From now on this formula should be used in calculating degree of resemblance of assemblages of potsherds.

And the modified values of the degree of resemblance shown by the standard set of assemblages of potsherds of the Pit-Comb Ware are represented in Table 11.

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R4</th>
<th>R5</th>
<th>R7</th>
<th>GR</th>
<th>SG</th>
<th>Tar</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>0.27</td>
<td>0.22</td>
<td>0.15</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>0.51</td>
<td>0.27</td>
<td>0.3</td>
<td>0.31</td>
<td>0.27</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.47</td>
<td>0.37</td>
<td>0.26</td>
<td>0.23</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>0.25</td>
<td>0.22</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.23</td>
<td>0.13</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.13</td>
<td>0.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. The set of values of the degree of resemblance of the standard set of assemblages calculated with quadratic regression

Very important parameters for estimating degrees of resemblances of assemblages of potsherds of the Pit-Comb Ware are the diapason of values of the degrees of resemblance shown by the standard set of assemblage, and the arithmetic mean of the set of values of the degree of resemblance shown by the standard set.

The values, represented in Table 11, form a diapason from 0.13 to 0.51.

The arithmetic mean of values represented in Table 11 is the following:

\[
(0.28 + 0.78 + 0.96 + 1.3 + 1.42 + 1.24 + 2.39)/28 \approx 0.3.
\]

And, thus, if a certain particular value of the degree of resemblance is higher than 0.3 or close to 0.51 it means that the compared assemblages are rather close.
References


