# GROUPS OF TOO CLOSE RADIOCARBON DATES

## ADAM WALANUS

*University of Rzeszów, Rejtana 16, 35-310 Rzeszów*
*(e-mail: walanus@univ.rzeszow.pl)*

**Key words:**
RADIOCARBON,
CHI-SQUARED,
STATISTICS

**Abstract:** In the archaeological and palaeogeographical literature, it is relatively frequent to find groups of radiocarbon dates of too close values. Too close means here that the dates exhibit no other variability than that of random origin, quantified by the given measurement errors. The chi-squared statistic seems to be appropriate to test, if the given group of dates is of random variability, of larger variability (what is typical) or of too small variability. The last case is hard to explain.

## 1. PREREQUISITES TO THE PROBLEM

Taking into consideration any group of radiocarbon dates, one logical constraint is to be fulfilled in every case; this constraint is of statistical nature.

As a "group of dates" can be considered dates connected with one site, one problem, archaeological culture, etc. In principle, however, from the logical point of view, it is not necessary for dates to be connected by some link. It may be important, since the question what does it mean "one site" or "one problem" may not have an unequivocal answer.

The mentioned constraint and, at the same time, the subject of this paper is relevant to the case of group of dates which cover a relatively narrow time span.

In connection with the meaning of the notion of a group of dates, one remark about some important point of the typical statistical reasoning needs to be made here. Namely, it is not a correct way to search for special piece of reality (eg. very close dates), and then to test statistically whether it is not too special. In principle, a randomly taken group of radiocarbon dates can be tested for some statistical property. However, it is difficult to strictly fulfil the necessary statistical assumptions. Anyway, obtaining some quantification of some statistical parameters can be informative and prompt to further thinking.

An example of a group of dates close one to another is: 8760±50, 8720±70, 8880±60, 8850±50 and 8880±50 BP (**Table 1**, column No. 3). The dates are given here in the same order as in the original paper (Dahl *et al.*, 2002). Considered here are "groups", not "sequences" of dates; therefore, the order of dates is irrelevant. With the stratigraphical sequences of dates another interesting statistical question is connected: how probable is the strict resemblance of the stratigraphical order by the radiocarbon dates.

Here conventional radiocarbon dates are considered. The technique used below is well applicable to the typical, normally distributed variables, while it would not be so straightforward to repeat the considerations in the case of calibrated dates. However, the general idea is relevant to both types of data presentation.

## 2. IMPLICATIONS OF THE STATISTICAL MEANING OF A CONVENTIONAL RADIOCARBON DATE

The sentence age of 5000±50 BP means no more, and no less, that the age of the given sample is given by the normal (Gaussian) probability distribution (with the expected value 5000 and dispersion 50). It means that, for example: probability that the true age of the sample is 5000 BP is equal to 0.00798, probability of age 4950 is 0.00484, and so on (see **Fig. 1**). The values of probability are low because the total probability (equal to 1) is distributed (unequally) among many years (in principle, among infinitely many).

On the one hand, age of 5000±50 BP means that the real age of the sample may be 5050, or 5100 etc. with the lower or higher probability. On the other hand, however, dating a sample of the real radiocarbon age 5000 BP with a precision ±50, one can say that the dating result (to be obtained) is normally distributed, according to 5000±50. This second point of view is a theoretical one, since we don't know what is the real age of the sample, while we do know the date (5000±50 BP).
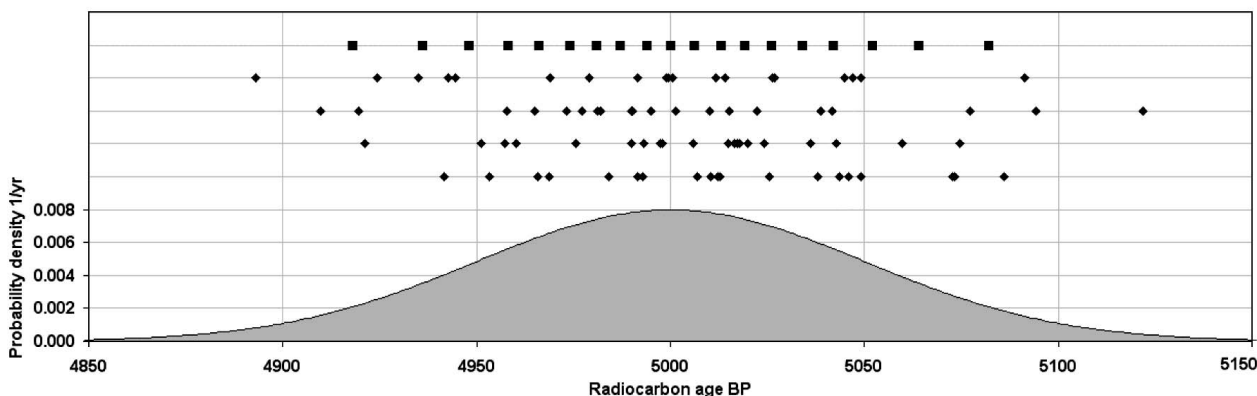
**Fig. 1.** *Probability distribution of the true sample age, in the case of date 5000±50 BP (the bell-curve). The squares in the first row indicate the "theoretical" distribution of 20 dates, according to the given probability distribution. The diamonds in four rows below are examples of groups consisting of 20 "real" ages. (All diamonds can be treated as an illustration of one group of 80 dates, as well.)*

In **Fig. 1,** models of date are given. The diamonds represent dates (organized in four groups of 20 dates). The points can equally well represent the real ages of samples, for which, the dating result is 5000±50, or results of repeated dating of sample of age 5000 (with the dating precision ±50). The second approach despite being abstract is useful to the following considerations.

### 3. GROUP OF DATES

Two mutually exclusive cases are possible when considering the group of dates: all samples are of equal true age, or not all samples are of equal age. The first case, of course, is a rare one. Logically simple, the proposed classification of groups needs, in fact, some clarification. The question is in the definition of the equality of ages. In principle, an age is given by the real number (for example 5017.37530452 years BP). As a rule we confine ourselves to the precision of single years or 5 or 10 years. It is connected with the dating precision. The ages of two samples can be treated as equal if the difference between them is small in comparison with the dating precision. For example, if the dating error is 50 yr, than the ages 4980, and 4990 can be treated as equal. It is valid for both points of view: true ages which differ less than 10 yr cannot be expected to be different in radiocarbon dating with the precision ±50, and, as well, two dates (4980±50 and 4990±50 BP) shouldn't be treated as different.

Repeated dating of one sample (example: Damon *et al.*, 1989) will not give the same results. For example, the probability that the difference between two repeated dates will be less than 10 yr (for dating precision ±50) is only 0.11. Of course, it is assumed that the dating precision ($\sigma$) is well estimated by the radiocarbon laboratory. It means (in some sense), that no other than statistical errors are expected.

In the already mentioned group of dates (**Table 1**, column 3), two are equal (8880 BP), within the 10 yr precision. Such a situation is possible. The probability that two dates will be equal in a group of dates, increases with the group size. However, the reasonable observer expects not too many equal dates. The group of dates: 8880±50, 8880±60, 8880±40, 8880±50 and 8880±70 BP is certainly dubious.

All what can be done in order to rescue such group of dates is to accept that all five samples were of equal age. Unfortunately, this is not enough. Anyway, we can do nothing more; the samples can't have ages more equal, than simply equal.

It is a very low probability, equal to 0.00004, that five samples would give equal ages (assuming 50 yr dating error, and 10 yr round-off). It means that an event happened which can happen about once per 24,000 groups of samples. This is the statistical reason for doubts. Such an event is possible; however, it should be very rare. In the typical statistical hypothesis testing, the so-called significance level is assumed. It is the value of arbitrarily chosen low probability, frequently 0.05.

The example of five equal ages is extreme, however; a group of only too close dates also can be dubious. For example such an (artificial) group of dates: 8890±50, 8840±70, 8880±60, 8850±50 and 8880±50 BP (**Fig. 2**), is dubious. Why?
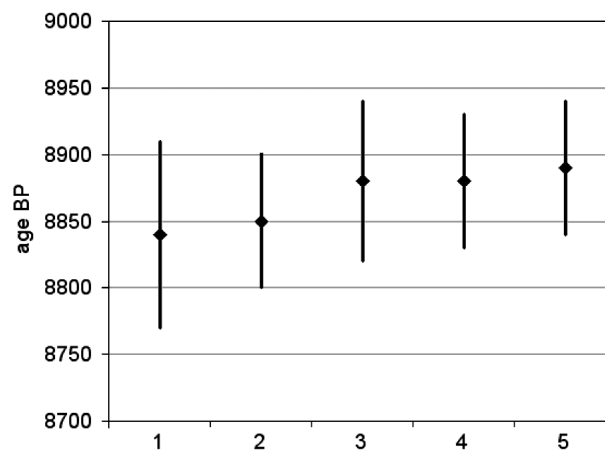


**Fig. 2.** *The group consisted of five dates of too close values (plotted in ascending order). The group is dubious even if it is assumed that all samples were of equal age.*

## 4. THE CHI-SQUARED STATISTICS

Dates are necessarily attributed with errors; in this example the measurement error is about ±50 yr. It is possible, even most probable, for a date to have its value equal to that of the true sample age. However, it is far from probable that all 5 dates will have such a good luck. The spread connected with the dating error, must manifest itself.

The problem is standard in mathematical statistics. The following procedure can be applied here: (1) calculate the weighted average for the group of dates, (2) subtract the obtained average from the dates, (3) divide resulted values by the respective errors, (4) take squares of the obtained values, (5) sum up all squares. The obtained sum, it is the so-called chi-squared statistic. The probability distribution of chi-squared is well known (**Fig. 3**).

The normalized normal random variables have values of the order of 1. So do squares of these variables. The sum of 5 of such squares may be expected to be about 5. However, in the case of the group of 5 dates, the average value has been calculated and subtracted from the dates. As a result, the "efficient" number of dates is only 4, it the so-called degrees of freedom of chi-squared.

The group of dates from **Fig. 2** gave chi-squared 0.6, the value below 0.71 which is the boundary of the area of too low chi-squares (see **Fig. 3**). It means that an event of low probability (0.05) happened. It is a strange and dubious situation, like in the case of throwing 6 coins and obtaining all heads (or tails). It is not very strange, but is expected, on average, only once per 20 times.

The chi-squared distribution has two tails. The right one, however, is not interesting in the case of considering a group of radiocarbon dates. If the chi-squared is too large (for 5 dates, larger than 9.5), the simple answer is that samples were of different age – the typical case.

However, too low value of chi-squared really causes a problem. The following two answers are possible:

1. The errors of dates were overestimated in the laboratory. The given values (±50, ±70, ...yr) are too high, and in fact they are lower. It is, however, hard to be expected, that laboratory sells dates as not very precise, while they are, in fact, more precise.

2. The considered group of dates was found, as an extreme from among about 20 similar cases of groups of samples of equal ages (within group).

In either case, the corollary is strongly supported, that the dated samples are of equal age.

The already mentioned group of real dates (**Table 1**, column 3) has chi-squared 6.3. It is not the most probable value 2, see **Fig. 3**), however, it is well within the "allowed" area or reasonable probabilities. The value 6.3 is higher than the expected value of the probability distribution, what makes weak (very weak) indication that samples can be, in fact, of slightly different ages.

To the group of dates which has the chi-squared value below the respective upper statistical limit (**Fig. 3**), the following line of reasoning applies. The simplest possible hypothesis considering the group of samples of unknown age is that all of them have the same age. In a typical case, such a hypothesis is rejected under strong evidence coming with the radiocarbon dates, namely the dates are too different within the group (chi-squared is much higher than the upper statistical limit). However, if no such evidence exists, because the chi-squared is below the upper limit, the hypothesis about the equality of sample ages cannot be rejected. In fact, the scientist is forced to accept, that all samples are of the same age. In particular, any inferences about the time span of the culture have no grounding.

## 5. THE ARCHAEOLOGICAL CULTURE TIME SPAN

The example number 5 in **Table 1** has chi-squared (11.5) well within the statistical boundaries. The probability to obtain higher value is 0.4, consequently, the probability to obtain lower value is 0.6. So, it is evident that the obtained value is in very good agreement with the assumption of the equality of all sample ages. In other words, there is absolutely no reason to reject that hypothesis. If so, how it is possible to calculate the time span for the culture dated by the group of given samples? The answer is that the time span can't be calculated if all samples have the same age. Of course, it is unreasonable to accept that the time span is equal to zero; probably the samples are of slightly different age. Anyway, there are no grounds to calculate the time span.
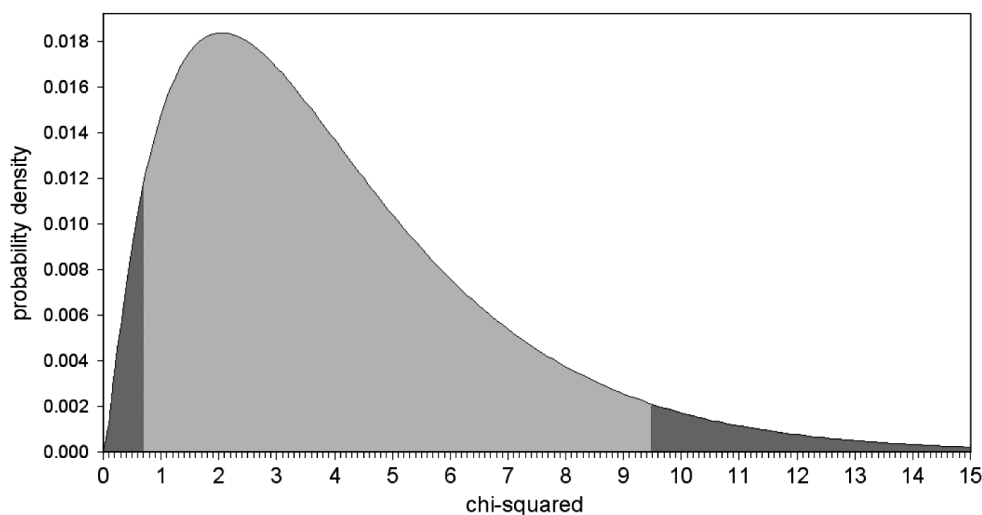


**Fig. 3.** *Chi-squared probability distribution (in the case of four degrees of freedom). The marked tails have a probability 0.05 (the lower, and the higher; both have probability 0.1).*

In opposition to the time span, the time position of the culture, in such a case, is easy to be calculated, it is simply the weighted average of the dates. In principle, the calculation of the average value in the other case, i.e. when samples are expected to be of different ages, is not strictly correct. The meaning of such an average is not clear.

For the mentioned example 5 (**Table 1**), the difference between the oldest and the youngest date is 115 years. The radiocarbon dates does not disproof the hypothesis that the time span of the culture is, say, 100 or 120 yr. However, the origin of such a hypothesis is to be independent of the dates. Its origin should be somewhere else. Such a hypothesis, by no means comes from the given radiocarbon dates.

In the next example group (**Table 1**, Column 6), there are as many as 28 dates. The chi-squared is slightly above the upper limit. The difference between the oldest and the youngest date here is 220 years. In comparison to example 5, the chi-squared is really relatively higher (41.9 (16.2-40.1), 11.5 (3.0-18.3)), however, what is more important, the number of samples is higher. For 28 random numbers, the expected range is higher than in the case of 11 samples. It is not strange, that the lottery player has won after many trials; it is rare to win in the first game. The 28 radiocarbon dates indicate *a posteriori* nothing else but one (average) age for all samples.

Once again it is worth mentioning, that since samples cannot have ages more equal than simply equal, the case from the column 1 (**Table 1**) is to be treated as strange. It is hard to explain a chi-squared below the lower statistical limit.

**Table 1.** Examples (not randomly chosen from the literature) of groups of dates with low values of chi-squared. Given limits of chi-squared (**Fig. 3**) indicate range of chi-squared values, which are statistically accepted, under the assumption that all dated samples has been of equal age. For both boundaries (lower and upper), independently, the significance level 0.05 is used.

| No | 1 | 2 | 3 | 4 | 5 | 6 | |
|----|---|---|---|---|---|---|---|
| Dates | 3545±40 | 2970±35 | 8760±50 | 3680±100 | 2935±40 | 4040±50 | 3920±60 |
| | 3540±45 | 2960±40 | 8720±70 | 3670±60 | 2930±40 | 4020±50 | 3915±40 |
| | 3540±40 | 2955±40 | 8880±60 | 3650±80 | 2910±35 | 3990±65 | 3910±40 |
| | 3540±30 | 2950±40 | 8850±50 | 3560±110 | 2890±30 | 3980±80 | 3905±45 |
| | 3525±30 | 2925±40 | 8880±50 | 3540±80 | 2860±30 | 3960±45 | 3900±60 |
| | 3520±35 | | | 3580±60 | 2875±40 | 3930±60 | 3895±45 |
| | 3515±30 | | | | 2850±35 | 3920±60 | 3890±50 |
| | 3505±35 | | | | 2845±35 | 3890±60 | 3880±50 |
| | 3490±45 | | | | 2870±50 | 4030±60 | 3870±60 |
| | | | | | 2820±30 | 4040±35 | 3870±45 |
| | | | | | 2830±40 | 4010±100 | 3850±50 |
| | | | | | | 3990±35 | 3820±80 |
| | | | | | | 3980±45 | 3900±35 |
| | | | | | | 3950±60 | 3860±40 |
| chi-squared | **1.8** | 0.80 | 6.3 | 2.9 | 11.5 | 41.9 | |
| limits of chi | 2.7 - 15.5 | 0.71 - 9.5 | 0.71 - 9.5 | 1.15 - 11.1 | 3.9 - 18.3 | 16.2 - 40.1 | |
| Reference | *Makarowicz, 2001* | *Makarowicz, 2001* | *Dahl et al., 2002* | *Kadrow and Machnik, 1997* | *Ignaczak and Ślusarska-Michalik, 2003* | *Czebreszuk and Szmyt, 2001* | |

**Table 2.** Examples of pairs of groups of dates with low chi-squared (description, see **Table 1**). Re-arranged data are in bold.

| No | 1a | 1b | 1c | 2a | 2b | 2c |
|----|----|----|----|----|----|----|
| Dates | 4510±55 | 4525±55 | **4510±55** | 3895±45 | **3895±45** | **4380±40** |
| | 4450±55 | 4525±55 | **4450±55** | 3910±40 | **3910±40** | **4370±50** |
| | 4470±55 | 4490±55 | **4405±55** | 3990±35 | **3990±35** | **4385±45** |
| | 4405±55 | | **4525±55** | 4380±40 | | **4400±50** |
| | | | **4525±55** | 4370±50 | | **4415±45** |
| | | | **4490±55** | 4385±45 | | **4420±55** |
| | | | | 4400±50 | | **4480±40** |
| | | | | 4415±45 | | **4490±40** |
| | | | | 4420±55 | | **4520±45** |
| | | | | 4480±40 | | **4525±45** |
| | | | | 4490±40 | | |
| | | | | 4520±45 | | |
| | | | | 4525±45 | | |
| chi-squared | 1.9 | 0.3 | 3.8 | *374* | 16.1 | 3.7 |
| limits of chi | 0.35 - 7.8 | 0.10 - 6.0 | 1.63 - 12.6 | 5.2 - 21 | 3.3 - 16.9 | 0.10 - 6.0 |
| Reference | *Hügi and Michel-Tobler, 2004* | | | *Szmyt, 2000* | | |

## 6. TWO GROUPS OF DATES

Typical statistical reasoning in comparing two groups of data can be summarised in two steps:

1. Check the *a priori* assumption that the groups are homogeneous, i.e. that they are groups of measurements of the same quantity (of the same age). It can be done with the use of chi-squared, calculated within groups.

2. If the answer to the question (1) is 'yes' then check the hypothesis that the quantities represented by the groups are equal. Typically the averages are compared, whether they are significantly different or not.

However, in step (2), as well as in (1), the chi-squared value calculated for all measurements (from both groups) can be used. As an example, dates given in column 1a and 1b from **Table 2** can be used. There are four and three dates of two objects. Chi-squared, in both cases is pretty well within the statistically "allowed" range. It means that the groups are homogeneous; 4 samples and 3 samples were of equal age, respectively. Moreover, the group obtained by joining both groups together (1c), is homogeneous, as well. It means that both ages, for both groups are equal. In other words: all 7 samples are of equal age.

Not so clear is the following reasoning applied to the dates from column 2a (**Table 2**). Here the evidently non-homogenous group is represented (chi-squared 374). It is easy, however, to find two subgroups: the first three dates and the remaining dates. They are really subgroups, because the chi-squared for both (2b and 2c) are within the range of reasonable probability. However, the original group of dates (2a) has an objective reason to be the group, since the samples are taken from one object. Division of that group into two smaller, is, in principle subjective. It is not a good way of statistical reasoning to find firstly the proper point of division, than to divide the group, and finally to be happy that the division is constructive, i.e. really diminishes the total chi-squared. However, what is typical in statistics, strongly depends on the level of statistical significance. In the given example (2a-2c), the chi-square starts from 374, to get $16.1+3.7=19.8 \approx 20$, as a result of the decision that in fact there are two separate groups of dates.

## 7. SUMMARY

The error reported with conventional radiocarbon date has a clear statistical meaning, it is an estimate of the standard deviation of the probability distribution of true radiocarbon age. It is not typical situation in technical or physical measurements, and it results from the "simplicity" of the statistical character of counting of $^{14}C$ disintegrations or directly, $^{14}C$ atoms. That good opportunity is to be exploited, and it is. For example, precise calculation of the complicated distribution of calibrated date, depend deeply on the exact value of error. From the point of view of the date user, as well as date producer, it is much better for the date error to be lesser than larger. However, if given value of error have been obtained in the measurement, it does operate in two directions. Error not only gives the possibility to the true age to be different from the reported date, by one or two (say) errors, but also gives kind of duty to the true age to be different. While, for the true age it is, of course, the most probable to be equal to the reported date, it is not the most probable case that the absolute value of the deviation will be zero. The expected value of deviation is about 80% of error. It means that in case of more dates, the average deviation can be relatively precisely estimated. And, the deviation cannot be too small. While there are many reasons for deviation of true ages from reported dates to be higher than those related to the statistical error, there is no reason why deviation would be too small. Simple statistical procedure, with chi-squared distribution is proposed to test statistical significance of small deviations of dates. If the obtained chi-squared value is too small (according to some significance level) it means that things are too good (to be true), i.e. dates are too "consistent".

## REFERENCES

**Czebreszuk J. and Szmyt M., 2001:** The 3rd Millennium BC in Kujawy in the Light of $^{14}C$ Dates. In: Czebreszuk J. and Müller J., eds, *Die absolute Chronelogie in Mitteleuropa, 3000-2000 v. Chr.*, Poznań/Bamberg/Rahden: 177-208.

**Dahl S.O., Nesje A., Oyvind L., Fjordheim K. and Matthews J.A. 2002:** Timing, equlibrium-line altitudes and climatic implications of two early-Holocene glacier readvances during the Erdalen Event at Jostedalsbreen, western Norway. *The Holocene* 12(1): 17-25.

**Damon P.E., Donahue D.J., Gore B.H., Hatheway A.L., Jull A.J.T., Linick T.W., Sercel P.J., Toolin L.J., Bronk C.R., Hall E.T., Hedges R.E.M., Housley R., Law I.A., Perry C., Bonani G., Trumbore S., Woelfli W., Ambers J.C., Bowman S.G.E., Leese M.N. and Tite M.S., 1989:** Radiocarbon Dating of the Shroud of Turin, *Nature* 337: 611-615.

**Hügi U. and Michel-Tobler Ch., 2004:** Oberrieden ZH-Riet - eine frühhorgenzeitliche Siedlung (Oberrieden ZH-Riet - an early Horgen settlement). *Jahrbuch der Schweizenschen Gesellschaft für Ur- und Frühgeschiche* 87: 7-31 (in German).

**Ignaczak M. and Ślusarska-Michalik K., 2003:** The radiocarbon chronology of the Urnfield Complex and the dating of cultural phenomena in the Pontic Area (late Bronze Age and early Iron Age). *The foundations of radiocarbon chronology of cultures between the Vistula and Dnieper: 4000-1000 BC, Baltic-Pontic studies* 12: 382-395.

**Kadrow Sł. and Machnik J., 1997:** Kultura mierzanowicka, chronologia, taksonomia i rozwój przestrzenny (The Mierzanowice Culture. Chronology, taxonomy and spatial evolution). *Prace Komisji Archeologicznej PAN, Oddział w Krakowie* 29: 7-196 (in Polish).

**Makarowicz P., 2001:** The Second Half of the Third and Second Millennium BC in Kujawy, Northern Poland, in the Light of $^{14}C$ Determinations. In: Czebreszuk J. and Müller J., eds, *Die absolute Chronolegie in Mitteleuropa, 3000-2000 v.Chr.*, Poznań/Bamberg/Rahden 209-270.

**Szmyt M., 2000:** Osadnictwo społeczności kultury amfor kulistych (Settlement of Globular Amphora Culture communities). In: Kośko A., ed., *Archeologiczne badania ratownicze wzdłuż trasy gazociągu tranzytowego*. (The archeological rescue research along the transit gas pipeline route) Wyd. Poznańskie, Poznań: 135-329 (in Polish).