Online Resource 1 to the paper entitled '*The copper of the Avars. The early medieval origins of copper ore extraction in the Carpathian Mountains*'
by Mozgai et al.

Problems with statistical methods during data evaluation

It must be noted that a frequent misconception is that artefacts can be traced to ore deposits using purely statistical criteria. The first grave mistake, M1, is to assume that all samples from a given population are absolutely identical, and that variations in the measured values of some quantity (e.g., isotopic compositions of ore minerals) are exclusively due to random measurement uncertainties. Mistake M1 is based on the assumption that there do not exist any systematic reasons (such as polyphase evolution, protracted ore-forming processes, secondary alteration) that could modify the measured quantity. Mistake M2 is the assumption that all possible samples of all possible ore deposits have been analyzed, and that three exist no unsampled ores anywhere. Mistake M3 is applying Gaussian statistics to a non-Gaussian data set: the division by √n results in very narrow confidence levels (i.e., "kernel densities") for the reservoir with which the unknown sample is compared. Mistake M4 is related to M2 and M3, which both require the nonexistence of systematically biased outliers, as Gaussian data-sets by definition do not have points that lie outside the random distribution.

In reality, sufficiently detailed investigations of real-world ore fields (e.g., Mondillo et al. 2018, Kang et al. 2020) usually focus on successive growth phases, on redox changes, and on spatial patterns of mineralogical and compositional variations. The observations contradict the homogeneity assumed by M1. This had been clearly laid out by Stos-Gale and Gale (2009). As redox changes have been described at the meter scale (Kang et al. 2020), it is not possible to envisage that all possible variations have been taken into account: as ore exploitation progresses and the mining front moves by tens or hundreds of meters, it is well documented that "uncharted" ore assemblages become available to miners, contradicting M2. The calculation of a kernel density (M3) for a given ore deposit is based on the assumption of "how far" an ore deposit extends. One could be misled to believe that the Mississippi Valley, or Sardinia, are single homogeneous deposits, as they were produced in a single ore-forming event (Stos-Gale and Gale 2009). By choosing arbitrarily small subsets one may get the incorrect impression that the chosen subset is homogeneous, but this is not a legitimate approach. The reality check is provided by the late Bronze Age ore deposits in Mitterberg, Austria (Pernicka et al. 2016). In an area around 100 km^2 copper ore was probably formed in "one" event, whereby "one" is a vague and inaccurate definition (see below). The original Mitterberg data are replotted in Fig. S1., and their mean and standard deviation calculated. The 2 sigma KDE ellipse is shown in Fig. S1. If one were to apply the "Euclidean distance from the KDE" philosophy, then six of the analyzed samples can impossibly belong to Mitterberg. Yet they were collected precisely there. The model is unable to predict the observed facts, because the Mitterberg data-set with which the Mitterberg samples are compared is not Gaussian. In reality, ore deposits are formed by processes described in detail by ore geologists (e.g., Kang et al. 2020, and references therein). Ores are formed by a series of heterogeneous sulfides- and/or oxides-bearing fluids (whereby redox changes occur at different times). If cursory sample collections fail to recover each and every possible, obvious and hidden, mineral paragenesis, then the reconstruction is incomplete. Returning to the Mitterberg data: if only the western, or only the eastern samples had been collected, the Mitterberg kernel density would be different from that calculated. Note also that any calculated KDE of the Mitterberg ore field based on samples cropping out today is certainly incomplete, because not all samples mined from 1500 BCE to the present day were available for the calculation.

Instead of the minimum size defined by the KDE, the correct assignment of an object to an ore deposit must rely on the maximum extension of all ore samples of a candidate deposit collected so far. It is also extremely important to stress that all samples had exactly the same likelihood of being mined by prehistoric or modern miners. Mitterberg sample MA081378, because it was collected, had exactly 100% likelihood of being collected; it would be a gross mistake (M3) to assign it a lower likelihood only because it is far from the KDE. A hypothetical bronze artefact with that isotopic composition could be from Mitterberg even if its Euclidean distance is too far for the KDE mathematical model.

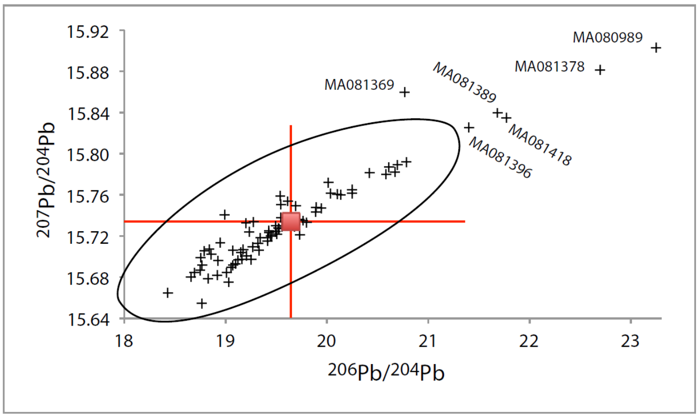

Fig. S1. Ore samples from Mitterberg, Austria, replotted from the data by Pernicka et al (2016). The average isotopic composition is shown by the red square and its 2-sigma uncertainty by the red error bars. The kernel density estimate is shown by the solid ellipse. Six ores sampled in Mitterberg lie outside the ellipse and would be misedentied as "not compatible with Mitterberg" by a purely statistical interpretation.

References

Kang JS, Wille M, Hofmann BA, Strauss H, Villa IM (2020) Heterogeneous lead isotopic compositions of sulfide minerals from a hydrothermal replacement deposit (Janggun mine, South Korea). Ore Geol Rev 122: 103527. https://doi.org/10.1016/j.oregeorev.2020.103527

Mondillo N, Arfè G, Boni M, Balassone G, Boyce A, Joachimski M, Kang JS, Villa IM (2018) The Cristal Zinc prospect (Amazonas region, northern Peru). Part I: New insights on the sulfide mineralization in the Bongará province. Ore Geol Rev 94: 261–276. https://doi.org/10.1016/j.oregeorev.2018.01.021

Pernicka E, Lutz J, Stöllner T (2016) Bronze Age Copper Produced at Mitterberg, Austria, and its Distribution. Archaeol Austriaca Band 100/2016: 19–55. https://doi.org/10.1553/ARCHAEOLOGIA100S19

Stos-Gale Z, Gale NH (2009) Metal provenancing using isotopes and the Oxford archaeological lead isotope database (OXALID). Archaeol Anthropol Sci 1: 195–213. https://doi.org/10.1007/s12520-009-0011-6