**Reassessment of recommendations for processing mammal phosphate $\delta^{18}\text{O}$ data for
paleotemperature reconstruction**

Grzegorz Skrzypek^{a*}, Rohan Sadler^{b,c}, Andrzej Wiśniewski^d

^a *West Australian Biogeochemistry Centre, School of Plant Biology M090,
The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

^b *Bush Futures, 4 James Street, Guildford, WA 6055, Australia*

^c *School of Agricultural and Resource Economics M089, The University of Western Australia,
35 Stirling Highway, Crawley, WA 6009, Australia*

^d *Institute of Archaeology, The University of Wrocław, Szewska Str. 48, 50–139 Wrocław, Poland*

*Corresponding author. Tel.: +61 8 6488 4584 E-mail address: grzegorz.skrzypek@uwa.edu.au,
gskrzypek@yahoo.com

Keywords: stable isotope; oxygen; teeth; phosphate; paleotemperature; calibration

28 **ABSTRACT**

29 The stable oxygen isotope composition of phosphates ($\delta^{18}\text{O}_\text{P}$) extracted from vertebrate bone and
30 teeth material is widely used for paleotemperature reconstruction over the Quaternary period.
31 However, inconsistency in analytical and statistical procedures used by different researchers for
32 recalculation of mean air temperatures from measured $\delta^{18}\text{O}_\text{P}$ values makes integration of different
33 data sets challenging. Therefore, consistency in both stable isotope data processing and the
34 application of appropriate statistical procedures is highly desirable. This paper evaluates the
35 uncertainty associated with two most commonly used regression methods, least-squares inverted fit
36 and transposed fit. We assessed the performance of both fits by designing 'real-life' experiments
37 using various data sets. We used both regression models to reversely calculate the differences
38 between actual recorded in 'real-life' values and values calculated from the regression models. Our
39 calculations clearly show that the mean absolute errors are always substantially higher for the
40 inverted fit and the transposed fit always returned mean values closer to the values measured in 'real
41 life'. In all studied scenarios, the predictive models performed better than causal models, returning
42 mean absolute errors by 12 to 65% lower. In the paper, we also examined the other main sources of
43 uncertainty, concerns about calculation of $\delta^{18}\text{O}_\text{P}$ of phosphates from $\delta^{18}\text{O}_\text{C}$ of carbonates, averaging
44 of $\delta^{18}\text{O}_\text{W}$ in precipitation and highlight major challenges in the field of phosphate stable oxygen
45 isotope biogeochemistry. The estimation of paleotemperatures from isotope records could be
46 improved by employing globally consistent calculation procedures.

47

48

49 **1. Introduction**

50 In order to compare the results of different paleotemperature studies, consistency in both stable
51 isotope data processing and the application of appropriate statistical procedures is highly desirable.
52 This is particularly true with regard to the stable oxygen isotope composition of phosphates ($\delta^{18}\text{O}_\text{P}$)
53 and carbonates ($\delta^{18}\text{O}_\text{C}$) extracted from mammalian teeth and bone samples (e.g., D'Angela and

Longinelli, 1990). Recently, Pryor et al. (2014) attempted to verify regression methods used for paleotemperature calculations. However, finding a suitable published data set for testing a theoretical statistical hypothesis is challenging because relatively few studies publish entire raw data sets or provide the whole step-by-step calculation procedures. Pryor et al. (2014) chose to use a published data set to compare different regression methods and concluded their paper offering recommendations for calculating paleotemperature from $\delta^{18}\text{O}_\text{P}$ data. In this paper, we present calculations for both discussed models, 'inverted fit' and 'transposed fit', based on published $\delta^{18}\text{O}_\text{P}$ results (Pryor et al., 2014) and a large data set for Europe from the Global Network of Isotopes in Precipitation/International Atomic Energy Agency (GNIP/IAEA). After reassessing uncertainty introduced by each of the models, we offer our opinions about the procedures advocated by Pryor et al. (2014) and we suggest an alternative approach, while also highlighting other major concerns associated with the application of $\delta^{18}\text{O}_\text{P}$ for paleoclimate studies.

66

67

68 **2. Causal or predictive linear regression model?**

The calculation of paleotemperatures requires application of multiple regression models to translate $\delta^{18}\text{O}_\text{P}$ recorded in tooth phosphates to mean air temperature (T_air given in $^\circ\text{C}$) reflecting the lifespan of the animal when the tooth grew. In brief, this procedure consists of three steps (Skrzypek et al., 2011): 1) direct analysis of $\delta^{18}\text{O}_\text{P}$ in phosphate extracted from the sample; 2) recalculation of $\delta^{18}\text{O}_\text{P}$ to $\delta^{18}\text{O}_\text{W}$ of water consumed by the animal (the so-called environmental water) using one of the species-specific equations calculated independently by various calibration studies (e.g., D'Angela and Longinelli, 1990; Hoppe, 2006; Huertas et al., 1995); and 3) calculation of T_air from $\delta^{18}\text{O}_\text{W}$ using the relationship between air temperature and $\delta^{18}\text{O}_\text{W}$ in precipitation obtained from independent long-term observations — this relation is specific for geographic location, longitude, altitude and climate variables.

Pryor et al. (2014) highlighted that the least squares regression models required for calculations of $\delta^{18}\text{O}_\text{W}$ in step 2 and T_air in step 3 can be designed in two different ways — as a forward fit or as a transposed fit — depending which variable is assigned as the x- or y-axis (Eqs. 1,2,4 and 5, Table 1). Hence in step 2 the forward fit estimates a simple linear model of the regression of $\delta^{18}\text{O}_\text{P}$ on $\delta^{18}\text{O}_\text{W}$, before expressing $\delta^{18}\text{O}_\text{W}$ in terms of $\delta^{18}\text{O}_\text{P}$ (termed inverted forward regression by Pryor et al. 2014). In contrast, the transposed fit estimates $\delta^{18}\text{O}_\text{W}$ directly in terms of $\delta^{18}\text{O}_\text{P}$. The same fitting method as in step 2 is applied again at step 3 for expressing T_air in terms of $\delta^{18}\text{O}_\text{W}$. Importantly, in the forward fit the regression equations need to be inverted before $\delta^{18}\text{O}_\text{W}$ and T_air may be expressed as a response variable that can be predicted from the regression equation (inverted fit Eqs. 3 and 6, Table 1). A problem arises when the forward fit and the transposed fit models are far from ideal and the linear correlation is relatively weak. In this case, equations 2 and 3 ($c \neq 1/a$ and $d \neq b/a$), as well as equations 5 and 6 ($g \neq 1/e$ and $h \neq f/e$), will frequently differ and return significantly different results (where a, c, e and g are slopes and b, d, f and h are intercepts of the forward fit and transposed fit regression equations). In consequence the equation returned as the inverted fit will not be the same as the transposed fit (Fig. 1). Pryor et al. (2014) advocated the least-squares regression inverted fit (e.g. $\delta^{18}\text{O}_\text{W} = 1/a \times \delta^{18}\text{O}_\text{P} - b/a$ and $T_\text{air} = 1/e \times \delta^{18}\text{O}_\text{W} - f/e$) as the most appropriate method for calibrating causal isotopic relationships. We argue instead that the superiority of an inverted fit was not actually determined experimentally by Pryor et al. (2014), but was simply a philosophical point of view that a model using a causal link is better than a predictive one. From a causal perspective, in the natural environment, the water stable isotope composition ($\delta^{18}\text{O}_\text{W}$) obviously influences the phosphate stable isotope composition ($\delta^{18}\text{O}_\text{P}$), and not the other way around. T_air also influences the water stable isotope composition in precipitation ($\delta^{18}\text{O}_\text{W}$), and not the opposite. However, from a purely statistical perspective, the predicted (calculated) variables are $\delta^{18}\text{O}_\text{W}$ (for $\delta^{18}\text{O}_\text{P}$ - $\delta^{18}\text{O}_\text{W}$ model) and T_air (for T_air - $\delta^{18}\text{O}_\text{W}$ model). The best-case scenario, of course, is when the direction of a causal link matches that of the predictive link. However, if this is not the case, in our opinion, the investigator primarily needs to consider what is measured and what is calculated from this

105 measurement to determine the predicted value from the predictor. Therefore, from a logical point of
106 view the transposed fit is not inferior to the inverted fit and can be used for all calculations applying
107 a Least Squares regression model (LSM).

108 From a practical point of view there is always going to be uncertainty in the relationship between
109 the predicted value and the predictor. Pryor et al. (2014) claim that “ *follows that values of ‘x’*
110 *calculated using a transposed regression fit of $x(y)$ will be consistently higher than those produced*
111 *from the inverted forward regression fit of $y(x)$ for the range of values below the mean (x, y) , and*
112 *consistently lower for those above (x, y) ”. However, we would argue that the slope of the linear
113 relationship is important. Note that if the slope of an inverted fit is greater than the slope of a
114 forward fit (as in Table 2 of Pryor et al. 2014) then residuals, as a measure of the difference
115 between true/observed values and their corresponding model predictions, will be larger in scale for
116 the inverted fit. In this case, the inverted fit will lead to poorer measures of predictive accuracy than
117 the transposed fit when most of the variability is assigned to the y-axis of the forward fit as Pryor et
118 al. (2014) have argued. *A priori* we do not know whether the slope of the transposed relationship is
119 greater than or less than the inverted fit). It simply happens in this instance (or realisation) that the
120 regressions considered by Pryor et al. (2014; Table 2) had a lower slope value for the transposed fit
121 than for the inverted fit for both the $\delta^{18}\text{O}_\text{P}$ - $\delta^{18}\text{O}_\text{W}$ and $\delta^{18}\text{O}_\text{W}$ - T_air models, and so temperature
122 predictions were demonstrably different between the inverted and transposed fits. It does not,
123 however, demonstrate the superiority of the inverted fit over the transposed fit against some
124 objective criterion.*

125 From a statistical point of view the difference between inverted fits (termed the 'classical'
126 estimator) and transposed fits (termed the 'inverse' estimator) has been subject to debate for almost
127 50 years since the matter was first raised by Krutchkoff (1967). The key issue with inverted fits has
128 been that estimates of $x = (y - b)/a$ have in theory an infinite variance and mean, given that both y
129 and a are uncertain variables. Hence, the classical estimator of the inverted fit represents
130 asymptotically a ratio of normal variables that can be described by a Cauchy distribution under

131 common assumptions. Any variance based performance criterion applied to the classical estimator
132 that measures predictive 'success', such as the Mean Squared Error (MSE), will also be poorly
133 defined unless well designed (such as the 'standard error of the inverse prediction'; Demidenko, et
134 al., 2013). More successfully, a measure of Pitman Closeness (PMC) has been applied (Halperin,
135 1970; Kannan et al., 2007), as has a derivative of a Bayesian ratio of posterior probabilities (Brown
136 1982). Although Halperin (1970) found that there was no conclusive dominance with regard to the
137 PMC criterion of the inverted fit (i.e., classical estimator) *vis á vis* the transposed fit (i.e., inverse
138 estimator), both Kannan et al. (2007) and Brown (1982) have stated that the transposed fit was
139 preferred in most test cases.

140 We assessed the performance of both fits presented by Pryor et al. (2014) (i.e. the inverted and
141 the transposed) in most common applications for isotope paleotemperature studies by designing a
142 simple 'real-life' experiment. We used the data sets from Global Network of Isotopes in
143 Precipitation of The International Atomic Energy Agency (GNIP/IAEA) for the monthly mean
144 stable oxygen isotope composition ($\delta^{18}\text{O}_\text{w}$) and air temperature (T_air). We selected data sets for the
145 entire continental Europe (13k records) and for Krakow (334) and Vienna (516), as these stations
146 were used by Pryor et al. 2014 for their calculations. Next, we calculated the inverted and the
147 transposed regression models (Table 2, supplementary file mmc1.xls). We then used $\delta^{18}\text{O}_\text{w}$ and
148 these regression models to reverse calculate T_air . Finally, we calculated the differences between
149 values recorded by weather stations and those calculated from the regression model monthly mean
150 T_air . We assessed the performance of each model by calculating a robust measure of prediction
151 accuracy in the mean absolute error (MAE), calculated as the mean of absolute differences between
152 the model and observed temperatures for each month.

153 The MAE calculated for Krakow using the inverted fit (4.9°C) was 24% higher than that
154 calculated using the transposed fit (3.9°C). Similar results were obtained for Vienna, where MAE
155 for the inverted fit was 4.6°C, or 20% higher than the transposed fit (3.8°C). For the whole of
156 Europe, the total difference was even larger, where the MAE for the inverted fit (6°C) was 40%

157 higher than that calculated using the transposed fit (4.3°C). We also compared sums of residual
158 errors and MAE for several other stations used in the study by Skrzypek et al. (2011): Bern,
159 Karlsruhe, Ljubljana, Minsk, Moscow, St. Petersburg, Stuttgart, Weil am Rhein and Zagreb. The
160 transposed models always returned MAE values lower by 12 to 65% when compared to the inverted
161 fit (see supplementary file mmc1.xls).

162 This simple experiment clearly displays that the sums of residual errors and MAE for all tested
163 locations are substantially higher for the inverted fit, showing that the mean monthly temperatures
164 (T_{air}) calculated using the inverted fit risk being significantly different from true values, especially
165 when sample size is low. The transposed fits therefore performed in a predictive sense far better
166 than the inverted fits recommended by Pryor et al. (2014), despite the statistical agnosticism of this
167 approach towards the representation of a causal link within the model.

168 We used a similar test as above to assess the performance of the inverted and the transposed
169 model for calculation of $\delta^{18}\text{O}_w$ from $\delta^{18}\text{O}_p$. We used the data sets as republished by Pryor et al.
170 (2014) and calculated $\delta^{18}\text{O}_w$ using these two models, as well as the sum of absolute differences
171 between the calculated and the originally measured $\delta^{18}\text{O}_w$ values in sampling locations for three
172 commonly used genera: horse (Huertas et al. 1995), elephant (Ayliffe et al. 1992) and bison (Hoppe
173 et al. 2006). We again found that the MAE was always much lower, for all three species (by 13 to
174 22%), when the transposed fit was used compared to the inverted fit (Table 3, supplementary file
175 mmc1.xls). Note that the errors from these two predictions ($\delta^{18}\text{O}_w - \delta^{18}\text{O}_p$ and $\delta^{18}\text{O}_w - T_{\text{air}}$) will
176 compound when predicting T_{air} from $\delta^{18}\text{O}_p$.

177 More generally, the statistical approach should be to treat any calibration data as 'training' data
178 for developing a predictive model, and to then apply the model that predicts 'best' according to some
179 specified criterion. For example, a cross-validation criterion can to an extent account for model
180 misspecification as well as uncertainty in the predictions of a single model as discussed here. As the
181 dominance of one predictive model over another may vary from one context to another, it is best to
182 evaluate the predictive accuracy of the model so as to demonstrate the suitability of the model in the

183 given context. This should be coupled with the recommendation of Pryor et al. (2014) to quantify as
184 much as possible measures of parameter and prediction uncertainty.

185

186

187 **3. Carbonate or phosphate $\delta^{18}\text{O}$ for paleotemperature calculations?**

188 The $\delta^{18}\text{O}_\text{P}$ value of enamel phosphates is considered to be the most resistant to decomposition of
189 the biological material to be found within vertebrate bodies. Therefore, the use of $\delta^{18}\text{O}_\text{P}$ was
190 proposed for calculation of $\delta^{18}\text{O}_\text{W}$ of water available to animals during their lifetimes (D'Angela
191 and Longinelli 1990). However, a direct relationship has also been noted between $\delta^{18}\text{O}_\text{P}$ and the
192 coexisting biocarbonates forming the bone and tooth structural $\delta^{18}\text{O}_\text{C}$. A few similar equations
193 describing this relationship have been published (Iacumin et al., 1996; Zazzo et al., 2004; Bryan and
194 Froelich 1995; Lécuyer et al., 2010) and have been used for calculation of $\delta^{18}\text{O}_\text{P}$ from $\delta^{18}\text{O}_\text{C}$ and
195 finally paleotemperatures.

196 According to Pryor et al. (2014), this step introduces so small error that it was not even assessed
197 in their study. However, the $\delta^{18}\text{O}_\text{C}$ is much more prone to geochemical and biological alterations
198 than is $\delta^{18}\text{O}_\text{P}$. Arppe et al. (2010) proposed the use of $\delta^{18}\text{O}_\text{C}$ to assess the preservation of $\delta^{18}\text{O}_\text{P}$,
199 confirming that if carbonates were not significantly altered, then the especially more resistant $\delta^{18}\text{O}_\text{P}$
200 would preserve the original signatures for the animal's lifetime. Analysis of $\delta^{18}\text{O}_\text{C}$ alone does not
201 offer any controls for preservation of the original isotope signatures and may lead to incorrect
202 interpretation if the original $\delta^{18}\text{O}_\text{C}$ signatures are not well preserved. Moreover, the procedure of
203 pre-treatment of carbonates prior to stable isotope analyses may lead to significant discrepancies
204 between data sets, creating an additional challenge, despite the usually high precision of $\delta^{18}\text{O}_\text{C}$
205 measurements being $\leq 0.10\text{‰}$ (Pryor et al., 2013). Hence, in principle, the $\delta^{18}\text{O}_\text{P}$ of phosphates
206 should be analysed in the first place, even though the analytical procedure is much more time
207 consuming and despite the potentially higher analytical uncertainty than for carbonates. The
208 carbonate $\delta^{18}\text{O}_\text{C}$ could then be used for verification of the $\delta^{18}\text{O}_\text{P}$ preservation.

209

210

211 **4. Averaging the stable oxygen isotope composition in precipitation?**

212 The strength of correlation between the monthly mean $\delta^{18}\text{O}_\text{W}$ and T_air depends on geographic
213 location and frequently is weak, as can be calculated from the GNIP/IAEA data set. The question of
214 how strongly the mean T_air value reflects the mean annual temperature is an additional challenge.
215 This is particularly true in locations where precipitation is highly seasonal and therefore could
216 reflect $\delta^{18}\text{O}_\text{W}$ in only one season (either warm or cold). The precipitation $\delta^{18}\text{O}_\text{W}$ value also depends
217 on additional factors which may be linked with air mass temperature (e.g. direction from which the
218 rains come and distance from coastlines, orogenic, continental and amount effects, e.g. Róžański et
219 al., 1993). This problem can be partially solved, by using mean $\delta^{18}\text{O}_\text{W}$ weighed by precipitation
220 volume. However, precipitation patterns could have changed over millennia, reflecting different
221 distributions of baric lows and highs at a continental scale due, for example, to changes in the extent
222 of glaciation. Taking into account these precautions, Skrzypek et al. (2011) applied equations from
223 several GNIP water collection stations up to 1000 km distance to calculate T_air based on the
224 obtained $\delta^{18}\text{O}_\text{W}$ values. This allows assessment of the potential influence of changes in the past
225 precipitation, when the climate in the sampling location could potentially be either more a
226 'Continental' or 'Atlantic' type. Assessment of these differences arising from seasonality seems to
227 require the use of the most detailed data available (e.g. at least monthly GNIP/IAEA data sets). In
228 contrast, Pryor et al. (2014) suggest using annual means for both $\delta^{18}\text{O}_\text{W}$ and T_air . This approach
229 results in a serious reduction in the number of records (twelve fold), which eventually makes the
230 $\delta^{18}\text{O}_\text{W}$ - T_air regression model stronger but all seasonality is lost due to averaging. This is in
231 contradiction with recently proposed approaches in other precipitation studies. For example, Baldini
232 et al. (2010) and Fisher and Baldini (2011) proved that daily or single-event sampling increases the
233 strength of the $\delta^{18}\text{O}_\text{W}$ - T_air correlation. Unsurprisingly, $\delta^{18}\text{O}_\text{W}$ better reflects the mean T_air measured
234 exactly during the precipitation events. This leads to the conclusion that temperatures during days

without precipitation are not reflected in $\delta^{18}\text{O}_\text{W}$ at all and the actual mean annual temperature could be either higher or lower depending on location and season. Hence, the use of all (not averaged) data seems to be a better alternative, even though the regression model is weaker. However, the fact that the obtained temperatures mainly reflect a mean temperature during precipitation should be borne in mind.

240

241

242 **5. Propagation of uncertainties in paleotemperature studies based on $\delta^{18}\text{O}_\text{P}$**

243 The calculation of uncertainty propagation in geosciences is frequently challenging, particularly
244 in complex geochemical analyses that have used various data sets obtained with different extractive
245 methods, referencing strategies and regression approaches. The approach reported by Pryor et al.
246 (2014) of aggregating uncertainty across the estimated parameters into a single uncertainty measure
247 (i.e., standard error) for the predicted variable is one of the possibilities. The approach enables an
248 assessment of how the number of replicates and samples in single-species analyses can influence
249 aggregate uncertainty in linear regression models ($\delta^{18}\text{O}_\text{P}$ - $\delta^{18}\text{O}_\text{W}$ and $\delta^{18}\text{O}_\text{W}$ - T_air). Reporting
250 aggregate uncertainty, although valuable, can be misleading if all probable alternate scenarios are
251 not analysed (e.g., various climate scenarios and multiple species). Note that Pryor et al. (2014)
252 have provided an incorrect expression for δ_x when defining the inverted fit $x=(y-b)/a$ using
253 estimated parameters from the forward fit: the parameters a and b have been treated as constants
254 rather than as random terms with their own associated uncertainty (see, for example, Demidenko et
255 al. 2013) and variance in parameter estimates has not been taken to account. Monte Carlo methods
256 can be applied in the first instance to estimate compound variance terms in what is essentially a
257 two-stage least squares regression (Freedman, 1984), given analytical solutions may be largely
258 intractable for easy use.

259 Alternatively, practitioners can report the range of values and standard deviation for the results
260 for each calculation step, instead of one general combined uncertainty (Skrzypek et al., 2011). This

261 is more in keeping with a hierarchical or mixed-models approach that treats parameters as random
262 variables, but without defining a distribution for the parameters. An additional advantage of this
263 approach could be that it can focus on more than just one species and instead on a few species
264 whose remains were buried together and that likely coexisted in the past. The advantage of the use
265 of various species is that it allows cross-referencing the results obtained using different $\delta^{18}\text{O}_\text{P}$ - $\delta^{18}\text{O}_\text{W}$
266 regression models independently published for different species (Skrzypek et al., 2011).
267 Furthermore, multiple climate scenarios reflecting possible differences in precipitation pattern
268 differences in the past can be considered using different regression models for the $\delta^{18}\text{O}_\text{W}$ - T_air
269 relation. Aggregated measures of error should not be the single representation of 'error', especially
270 in increasingly complex models with multiple components and stages of estimation across which
271 errors should be propagated and quantified.

272

273

274 **6. Other sources of uncertainties**

275 All of these suggested improvements, however, seem to have limitations. The use of the stable
276 oxygen isotope composition of phosphates actually continues to be challenging mainly for two
277 reasons which were not considered in the discussion above. These two reasons are far more
278 fundamental: 1) inconsistency in the reference materials used; and, 2) differences in phosphate
279 extraction techniques. Several different reference standards might be used as no official phosphate
280 standard with certified δ -values has been established nor is any normalisation procedure commonly
281 accepted.

282 The most commonly used TC/EA technique is characterized by relatively high analytical
283 uncertainty (Brand et al., 2009; Skrzypek and Sadler, 2011) when compared to other analytical
284 techniques used in continuous flow mode (e.g., EA, Coplen et al. 2006; Skrzypek et al., 2009). This
285 high uncertainty in the TC/EA technique is reflected in the high uncertainty in calibration of the
286 primary reference materials. Consequently, selection of different standards and numbers of

287 replicates of standards may lead to differences in the propagation of errors associated with the
288 standards used (Skrzypek and Sadler, 2011) and to substantial differences between laboratories
289 (Brand et al., 2009). Some researchers suggest overcoming this problem by using organic primary
290 reference materials for all samples (USGS601 and USGS602) and the 'unofficial' commonly
291 available phosphate standard (NBS120c) as a checking standard (Halas et al., 2013). However, even
292 for the quite well-calibrated NBS120c, different $\delta^{18}\text{O}_\text{P}$ values have been reported (O'Neil et al.,
293 1994; Vennemann et al., 2002).

294 The uncertainty due to inconsistency in standard selection is amplified by the uncertainty due to
295 differences in the phosphate extraction techniques (Chang and Blake, 2015; Puceat et al., 2010).
296 The TC/EA technique yields CO from Ag_3PO_4 . In contrast, the fluorination technique (conventional
297 or laser) can be used for extracting O_2 from Ag_3PO_4 or BiPO_4 (Vennemann et al., 2002). According
298 to Chang and Blake (2015), an offset of $>0.9\text{‰}$ occurs between techniques using different chemical
299 precipitates Ag_3PO_4 or BiPO_4 . Therefore, it is critical that all details of analytical procedures,
300 selected reference materials and the normalization procedure be clearly reported to allow the future
301 recalculation of data if needed (Coleman and Meier-Augenstein, 2014).

302

303

304 **7. Conclusions**

305 We welcome the systematic approach to the error calculation presented by Pryor et al. (2014)
306 and agree that LSM is more appropriate than RMA for the calculation of the environmental water
307 stable oxygen isotope composition ($\delta^{18}\text{O}_\text{W}$) from phosphate signatures ($\delta^{18}\text{O}_\text{P}$) as well as for the
308 calculation of air temperature (T_air) from $\delta^{18}\text{O}_\text{W}$. However, results of our study do not concur with
309 the recommendations of Pryor et al. (2014), who suggested an 'inverted fit model' is the best
310 available model for calculation of the $\delta^{18}\text{O}_\text{W}$ of environmental water from $\delta^{18}\text{O}_\text{P}$ of phosphates and
311 the calculation of air temperature from $\delta^{18}\text{O}_\text{W}$. In summary, the following easy-to-implement

312 suggestions can be considered while processing the stable oxygen isotope results for
313 paleotemperature reconstructions:

- 314 - the transposed fit introduces a lower overall error than the inverted fit for $\delta^{18}\text{O}_\text{W}$ calculation and
315 calculation of T_air ; therefore, the predictive models are more statistically efficient than the causal
316 models in this instance;
- 317 - preferably, both phosphate $\delta^{18}\text{O}_\text{P}$ and carbonate $\delta^{18}\text{O}_\text{C}$ should be analysed and $\delta^{18}\text{O}_\text{C}$ used for
318 cross-verification of the original preservation signature in $\delta^{18}\text{O}_\text{P}$
- 319 - unnecessary averaging of precipitation $\delta^{18}\text{O}_\text{W}$ data should be avoided, as this reduces seasonal
320 differences; if possible, higher resolution sampling of precipitation should be used;
- 321 - all raw data should be reported, along with a transparent step-by-step set of calculations to allow
322 future verification and recalculation of results.

323 In addition, applying directly equations estimating $\delta^{18}\text{O}_\text{P}$ of phosphates based on analysed $\delta^{18}\text{O}_\text{C}$
324 may lead to serious discrepancies in temperature calculations. This is because analyses of $\delta^{18}\text{O}_\text{C}$
325 alone do not allow verification of the stage of preservation of the original isotope signatures in the
326 samples. Furthermore, a unified normalization procedure applied world-wide is highly desirable for
327 standards selection and phosphate extraction procedures in $\delta^{18}\text{O}$ analyses using the TC/EA
328 technique. Only a fully unified approach will allow future re-evaluation and recalibration of tooth
329 phosphates in commonly used mammalian species.

330

331

332 **Acknowledgments**

333 This research was supported by the Australian Research Council, a Future Fellowship awarded to
334 Grzegorz Skrzypek (FT110100352).

335

336 **Appendix. Supplementary material**

337 Supplementary material related to this article can be found online
338 mmc1_SkrzypekSadlerWisniewski2015_3Palaeo_supp.xlsx

339

340 **References**

341 Arppe, L.M., Karhu, J.A., 2006. Implications for the Late Pleistocene climate in Finland and
342 adjacent areas from the isotopic composition of mammoth skeletal remains. *Palaeogeogr.*
343 *Palaeoclim. Palaeoecol.* 231, 322–330.

344 Ayliffe, L.K., Lister, A.M., Chivas, A.R., 1992. The preservation of glacial–interglacial climatic
345 signatures in the oxygen isotopes of elephant skeletal phosphate. *Palaeogeogr. Palaeoclimatol.*
346 *Palaeoecol.* 99, 179–191.

347 Baldini, L.M., McDermott, F., Baldini, U.L., Fischer, M., Mollhoff, M., 2010. An investigation of
348 the controls on Irish precipitation $\delta^{18}\text{O}$ values on monthly and event timescales. *Clim. Dyn.* 35,
349 977–993.

350 Brand, W.A., Coplen, T.B., Aerts-Bijma, A.T., Böhlke, J.K., Gehre, M., Geilmann, H., Gröning,
351 M., Jansen, H.G., Meijer, H.A.J., Mroczkowski, S.J., Qi, H., Soergel, K., Stuart-Williams, H.,
352 Weise, S.M., Werner, R.A., 2009. Comprehensive inter-laboratory calibration of reference materials
353 for $\delta^{18}\text{O}$ versus VSMOW using various on-line high-temperature conversion techniques. *Rapid*
354 *Commun. Mass Spectrom.* 23, 999–1019.

355 Brown, P.J., 1982. Multivariate calibration. *Journal of the Royal Statistical Society - Series B* 44,
356 287–321.

357 Bryan, J.D., Froelich, P.N., 1995. A model of oxygen isotope fractionation in body water of large
358 mammals. *Geochim. Cosmochim. Acta* 59, 4523–4537.

359 Chang, S.J., Blake, R.E., 2015. Precise calibration of equilibrium oxygen isotope fractionations
360 between dissolved phosphate and water from 3 to 37°C. *Geochim. Cosmochim. Acta* 150, 314–329.

361 Coleman, M., Meier-Augenstein, W., 2014. Ignoring IUPAC guidelines for measurement and
362 reporting of stable isotope abundance values affects us all. *Rapid Commun. Mass Spectrom.* 28,
363 1953–1955.

364 Coplen, T.B., Brand, W.A., Gehre, M., Gröning, M., Meijer, H.A.J., Toman, B., Verkouteren, R.M.,
365 2006. New guidelines for $\delta^{13}\text{C}$ measurements. *Analytical Chemistry* 78, 2439–2441.

366 D'Angela, D., Longinelli, A., 1990. Oxygen isotopes in living mammal's bone phosphate: further
367 results. *Chem. Geol.* 86, 75–82.

368 Demidenko, E., Williams, B.B., Flood, A.B., Swartz, H.M., 2013. Standard error of inverse
369 prediction for dose-relationship: approximate and exact statistical inference, *Statistical Medicine* 32,
370 2048-2061.

371 Fischer, M., Baldini, L.M., 2011. A climate-isotope regression model with seasonally-varying and
372 time-integrated relationships. *Clim. Dyn.* 37, 2235–2251.

373 Freedman, D., 1984. On bootstrapping two-stage least-squares estimates in stationary linear models.
374 *The Annals of Statistics* 12, 827-842.

375 Halas, S., Skrzypek, G., Meier-Augenstein, W., Pelc, A., Kemp, H., 2011. Inter-laboratory
376 calibration of new silver orthophosphate comparison materials for stable oxygen isotope analysis of
377 phosphates. *Rapid Commun. Mass Spectrom.* 25, 579-584.

378 Halperin, M., 1970. On inverse estimation in linear regression. *Technometrics* 12, 727-736.

379 Hoppe, K., 2006. Correlation between the oxygen isotope ratio of North American bison teeth and
380 local waters: implication for paleoclimatic reconstructions. *Earth Planet Sci. Lett.* 244, 408–417.

381 Huertas, A.D., Iacumin, P., Stenni, B., Chill, B.S., Hill, N., Longinelli, A., 1995. Oxygen isotope
382 variations of phosphate in mammalian bone and tooth enamel. *Geochim. Cosmochim. Acta* 59,
383 4299-4305.

384 Iacumin, P., Bocherens, H., Mariotti, A., Longinelli, A., 1996. Oxygen isotope analyses of co-
385 existing carbonate and phosphate in biogenic apatite: a way to monitor diagenetic alteration of bone
386 phosphate. *Earth Planet. Sci. Lett.* 142, 1-6.

387 IAEA/WMO Global Network of Isotopes in Precipitation/ International Atomic Energy Agency.
388 The GNIP Database. http://www-naweb.iaea.org/napc/ih/IHS_resources_gnip.html

389 Kannan, N., Keating, J.P., Mason, R.L., 2007. A comparison of classical and inverse estimators in
390 the calibration problem. *Communications in Statistics–Theory and Methods* 36, 83–95.

391 Krutchkoff, R.G., 1967. Classical and inverse regression methods of calibration. *Technometrics* 9,
392 425–439.

393 Lécuyer, C., Balter, V., Martineau, F., Fourel, F., Bernard, A., Amiot, R., Gardien, V., Otero, O.,
394 Legendre, S., Panczer, G., Simon, L., Martini, R., 2010. Oxygen isotope fractionation between

395 apatite-bound carbonate and water determined from controlled experiments with synthetic apatites
 396 precipitated at 10–37°C. *Geochim. Cosmochim. Acta* 74, 2072–2081.

397 O'Neil, J.R., Roe, L.J., Reinhard, E., Blake, R.E., 1994. A rapid and precise method of oxygen
 398 isotope analysis of biogenic phosphate. *Isr J. Earth Sci.* 43, 203–212.

399 Pryor, A.J.E., Stevens, R.E., O'Connell, T.C., Lister, J.R., 2014. Quantification and propagation of
 400 errors when converting vertebrate biomineral oxygen isotope data to temperature for palaeoclimate
 401 reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology* 412, 99–107.

402 Pryor, A.J.E., O'Connell, T.C., Wojtal, P., Krzemińska, A., Stevens, R.E., 2013. Investigating
 403 climate at the Upper Palaeolithic site of Kraków Spadzista Street (B), Poland, using oxygen
 404 isotopes. *Quat. Int.* 294, 108–119.

405 Puceat, E., Joachimski, M.M., Bouilloux, A., Monna, F., Bonin, A., Motreuil, S., Moriniere, P.,
 406 Henard, S., Mourin, J., Dera, G., Quesne, D., 2010. Revised phosphate-water fractionation equation
 407 reassessing paleotemperatures derived from biogenic apatite. *Earth Planet. Sci. Lett.* 298, 135–142.

408 Róžański, K., Araguás-Araguás, L., Gonfiantini, R., 1993. Isotopic patterns in modern global
 409 precipitation. In: Swart, P.K., Lohmann, K.C., McKenzie, J., Savin, S. (Eds.), *Climate Change in*
 410 *Continental Isotopic Records*. American Geophysical Union, Washington DC, pp. 1–36.

411 Skrzypek, G., Sadler, R., 2011. A strategy for selection of reference materials in stable oxygen
 412 isotope analyses of solid materials. *Rapid Commun. Mass Spectrom.* 25, 1625–1630.

413 Skrzypek, G., Sadler, R., Paul, D., 2010. Error propagation in normalization of stable isotope data:
 414 A Monte Carlo analysis. *Rapid Commun. Mass Spectrom.* 24, 2697–2705.

415 Skrzypek, G., Wiśniewski, A., Grierson, P.F., 2011. How cold was it for Neanderthals moving to
 416 Central Europe during warm phases of the last glaciation? *Quaternary Science Reviews* 30, 481–
 417 487.

418 Vennemann, T.W., Fricke, H.C., Blake, R.E., O'Neil, J.R., Colman, A., 2002. Oxygen isotope
 419 analysis of phosphates: a comparison of techniques for analysis of Ag₃PO₄. *Chem. Geol.* 185, 321–
 420 336.

421 Zazzo, A., Leuyer, C., Mariotti, A., 2004. Experimentally-controlled carbon and oxygen isotope
 422 exchange between bioapatites and water under inorganic and microbially-mediated conditions.
 423 *Geochim. Cosmochim. Acta* 68, 1–12.

424

425

426 Table 1. Equations of the linear models for the different fitting methods, and associated parameters.

Fitting method	Equations for step 2	Equations for step 3
forward fit	$\delta^{18}\text{O}_p = a \times \delta^{18}\text{O}_w + b$ (Eq.1)	$\delta^{18}\text{O}_w = e \times T_{\text{air}} + f$ (Eq.4)
transposed fit	$\delta^{18}\text{O}_w = c \times \delta^{18}\text{O}_p + d$ (Eq.2)	$T_{\text{air}} = g \times \delta^{18}\text{O}_w + h$ (Eq.5)
inverted fit	$\delta^{18}\text{O}_w = 1/a \times \delta^{18}\text{O}_p - b/a$ (Eq.3)	$T_{\text{air}} = 1/e \times \delta^{18}\text{O}_w - f/e$ (Eq.6)

427

428

429 Table 2. Equations for transposed and inverted fits for GNIP/IAEA locations compared by Pryor et al. (2014), length of
 430 records as in Skrzypek et al. (2011). MAE calculated as sum of absolute differences between mean monthly temperature
 431 recorded by weather station and calculated based on the water stable oxygen isotope composition. The error (MAE)
 432 always was higher for Inverted fit.

Location	Years	Fit	Equation	MAE [°C]	MAE for inverted is higher than transposed by [%]
Krakow	1975- 2002	Transposed	$T_{\text{air}} = 1.804 \times \delta^{18}\text{O}_w + 26.020$	3.9	24
		Inverted	$T_{\text{air}} = 3.079 \times \delta^{18}\text{O}_w + 38.602$	4.9	
Vienna	1961- 2005	Transposed	$T_{\text{air}} = 1.544 \times \delta^{18}\text{O}_w + 25.304$	3.8	20
		Inverted	$T_{\text{air}} = 2.585 \times \delta^{18}\text{O}_w + 35.432$	4.6	
Europe	variable	Transposed	$T_{\text{air}} = 1.413 \times \delta^{18}\text{O}_w + 21.636$	4.3	40
		Inverted	$T_{\text{air}} = 2.878 \times \delta^{18}\text{O}_w + 33.796$	6.0	

433

434

435 Table 3. Transposed and inverted fits calculated based on data quoted from original publications by Pryor et al. (2014).
 436 MAE calculated as sum of absolute differences between the monthly stable oxygen composition observed at sampling
 437 locations and calculated based on the stable oxygen isotope composition of phosphates extracted from bones and teeth
 438 of mammals. The error (MAE) always was higher for Inverted fit.

Data set	Fit	Equation	MAE [‰]	MAE for inverted is higher than transposed by [%]
Huertas et al., 1995 – Horse	Transposed	$\delta^{18}\text{O}_w = 1.084 \times \delta^{18}\text{O}_p - 25.816$	1.5	13
	Inverted	$\delta^{18}\text{O}_w = 1.403 \times \delta^{18}\text{O}_p - 31.723$	1.7	
Ayliffe et al., 1992 - Elephant	Transposed	$\delta^{18}\text{O}_w = 0.906 \times \delta^{18}\text{O}_p - 21.933$	1.0	16
	Inverted	$\delta^{18}\text{O}_w = 1.063 \times \delta^{18}\text{O}_p - 24.802$	1.2	
Hoppe et al., 2006 – Bison	Transposed	$\delta^{18}\text{O}_w = 1.208 \times \delta^{18}\text{O}_p - 27.101$	1.4	22
	Inverted	$\delta^{18}\text{O}_w = 1.451 \times \delta^{18}\text{O}_p - 30.405$	1.8	

439

440

Figure 1. Comparison of regression models based on: A and B - real data set with a fit far from perfect (after Huertas et al., 1995) and C and D - a hypothetical perfect fit. Note that the equation for the Inverted fit (dashed line) calculated from the Forward fit is very different from the Transposed fit (solid line) for the real data set with imperfect fit (B). In contrast, the equation for the Inverted fit is identical with Transposed fit in the case of a perfect fit and both lines overlap (D). See also Table 1 for explanation of equation transformations.

448 Bullet points (5 points maximum 85 characters long):

449

450

451- the transposed fit introduces a lower overall error than the inverted fit

452- inverted fit model therefore is not appropriate for paleotemperature reconstruction

453- unnecessary averaging of precipitation data should be avoided

454- all raw data and a transparent step-by-step set of calculations need to be reported

455- a unified world-wide normalization and laboratory procedures are highly desirable

456

1 Highlights - bullet points (5 points maximum 85 characters long):

2

3

4- the transposed fit introduces a lower overall error than the inverted fit

5- inverted fit model therefore is not appropriate for paleotemperature reconstruction

6- unnecessary averaging of precipitation data should be avoided

7- all raw data and a transparent step-by-step set of calculations need to be reported

8- a unified world-wide normalization and laboratory procedures are highly desirable

9

Figure 1
[Click here to download high resolution image](#)