

Concerning calculation methods and limitations of proxy-estimates of Proteins, Carbohydrates and Lipids in crustacean zooplankton from CHN analyses

Richard A. VOLLENWEIDER

Senior Scientist Emeritus, NWRI-CCIW, Burlington, Ontario, Canada

The present comments resulted from a discussion I had with Dr. Riccardi about the validity of proxy estimates of protein, carbohydrate, and lipid in various zooplankton species of Northern Italian lakes (Riccardi & Mangoni 1999). For these estimates the authors used a calculation method proposed by Gnaiger & Bitterlich (1984; below referred to as G&B). The reason why I questioned the application of the unmodified G&B model to zooplankton was that carbohydrate in crustaceans is mostly present in form of chitin. The argumentation grew rapidly beyond a simple dispute about modes of calculation, leading to some more principle considerations about the limits of proxy estimates.

Background. G&B's thoughtful and well documented paper treated the question of how to estimate the proxy composition in proteins (PR), carbohydrates (CH) and lipids (LP) in fish tissues and of other aquatic animals, and their energetic content, from CHN measured weight fractions of carbon (W_C), nitrogen (W_N) and hydrogen (W_H) in dry weight. Unaware of this publication I proposed a similar approach using Spoehr & Milner's method (Spoehr & Milner 1949; Vollenweider 1985). Either approach, G&B's and mine, was based on standard stoichiometry of appropriately selected reference compounds acting as dummy variables for PR, CH, and LP, with the assumption that proteins, carbohydrates and lipids are solely composed of carbon, nitrogen, hydrogen and oxygen, neglecting all other elements, such as phosphorus, sulphur, etc. Accordingly, the problem to solve was to partition the measured mass fraction W_C , W_N , and W_H to the dummy variables in proportion to their assumed stoichiometry.

To this end G&B developed a set of loosely connected linear equations, one each for estimating the weight fractions W_{PR} , W_{CH} , and W_{LP} , respectively (cf. Appendix). The numeric values of the constants required were obtained by progressive substitution and rearrangement of the terms, and reducing the input values apparently to two, carbon and nitrogen (cf. Tab. 2 in G&B). Hydrogen was treated separately and introduced (implicitly also oxygen) over the estimation of residual water, W_{H_2O} , in the ash free dry weight (cf. Eq. (A14); Appendix in G&B), and the results integrated into the first constant of the equations. To elucidate the implications of this for the reader somewhat difficult to follow mathematical treatment a kind of nomogram was then constructed that illustrates the importance of residual

water in ash free dry matter for the correct estimate of W_{PR} , W_{CH} , and W_{LP} (G&B Fig. 1; cf. also G&B Newsletter 1985). To have drawn attention to this question is a most valid contribution by G&B, although, as will be shown below, the problem can be solved differently.

I myself addressed the problem instead by first calculating Spoehr & Milner's R values. R is an aggregated value of C, H, and O ranging from 0 to 100 that measures the oxidation/reduction state of an organic compound at the basis of the theoretical amount of oxygen required to oxidize the compound to CO_2 and H_2O . Nitrogen is ignored, but appears indirectly in the estimate of oxygen ($O = 100 - (C + N + H)$). R(PR), R(CH), and R(LP) values are easily computed from the presumed stoichiometry of the relative dummy variables. R(BM) of the composite biomass, instead is obtained from the analytically measured percent composition in C, N, H, and O of the ash-free dry weight. R(BM) equals $R(PR) \times W_{PR} + R(CH) \times W_{CH} + R(LP) \times W_{LP}$. The latter relation offers a convenient way to check for correctness of the W_j estimates. Hence, also in this case, the solution reduces to a partition problem. Yet, a shortcoming of the Spoehr-Milner model is that residual water cannot be estimated from the model itself because $R(H_2O) = 0$.

I draw then attention to the fact that if the model is applied to crustacean zooplankton the stoichiometric coefficients for CH need be changed since carbohydrate in zooplankton is primarily present in form of chitin (an amino polysaccharide) containing some 6 to 7% of nitrogen. Accordingly measured organic nitrogen is to be distributed between two components, PR & CH. To achieve this I proposed an iterative procedure: in step one all nitrogen is allocated to protein, and PR, CH & LP are estimated; in step two an aliquot of nitrogen (6.65% of the estimated CH in step one) is subtracted from the total nitrogen, and the difference used as the new input value, repeating the estimates, a.s.o. With this procedure stability of distribution of the respective estimates was normally reached after 4 to 6 steps. In a recent unpublished revision were I used G&B's mass fraction coefficients, further allowing for a reasonable estimate of residual water, and correcting the input values accordingly, the first loop of the iterative model gives estimates that are close to those obtained with the G&B model. Yet, after 4 to 6 loops the estimates may be considerably at variance, since G&B used glycogen as reference compound for CH, while I used chitobiose.

Modification of the calculation procedure. In attempting to modify G&B's model to be applicable to crustacean zooplankton I found their substitution method, though correct except for a small error¹⁾, to be unnecessarily cumbersome. On the other hand also the iterative model cannot be satisfactorily modified because of the impossibility to estimate residual water concurrently. Both problems can be overcome at once using instead matrix algebra. Matrix notation reduces the implied distribution problem to one of solving a system of simultaneous linear equations, what considerably simplifies calculation. The beauty of linear systems is that they can be dealt with on common spreadsheets (like LOTUS, EXCEL etc.) involving but a few operations, and moreover requires from the operator but a minimum of knowledge about the mathematics behind the programme. Most important, matrix algebra facilitates adaptation of the model to any set of reference compounds with no further need to having to recalculate, as with the G&B procedure, the constants of all the equations. The only conditions are that the number of equations must much the number of available input values; in the present case 4, and the determinant of the mass coefficient matrix must not be = 0.

In short, in matrix notation the problem has the following form: $|a_{ij}| \times |W_j| = |W_i|$, with the following meaning of the subscripts: $i = C, N, H, O$; $j = P$ (for PR), K (for CH), L (for LP), W (for H₂O); its solution is $|W_j| = |a_{ij}|^{-1} \times |W_i|$. The development of $|a_{ij}|$ gives a 4×4 (quadratic) coefficient matrix, that of $|W_i|$ gives the input matrix, and that of $|W_j|$ the resulting solution matrix, both being (vertical) 1×4 matrices (or column vectors). $|a_{ij}|^{-1}$ is the inverted matrix of $|a_{ij}|$, denominated in the following c_{ji} , which is another 4×4 matrix. The c_{ji} are the partition coefficients, which can be positive or negative, however. As to the meaning of the matrix components consider the following examples: a_{CL} is the mass fraction of carbon in lipids, W_H the mass fraction of hydrogen in the ash-free dry weight, W_K the mass fraction of carbohydrates estimated, etc. Further, matrix conditions for the problem in question require that the sum of the components of either vector, W_i and W_j , equals 1 (or 100 if input values are in percent), while the sum of each column vector of the a_{ij} matrix equals always 1. The same applies to the c_{ji} matrix. To notice: the input values W_i are the ash corrected data according to G&B's formula (1). The procedural details will become clear with the example discussed below, and the notes in the Appendix.

Apart from using as dummy variable for CH chitobiose units, (C₈H₁₄O₅N)₂O, the main difference between the G&B method and matrix notation is that in the latter oxygen (O) and water (H₂O) are explicit variables of the system.

¹⁾ Gnaiger & Bitterlich use for the fraction of hydrogen in water the value of .1006; the more correct value instead is .1119 (=2.016/18.016).

Accordingly, the estimate of the mass fraction of residual water, W_w , becomes part of the output matrix. This makes G&B's Eq.(A14) - their equation for estimating residual water - expendable. The residual water fraction estimated by either procedure are indeed identical, conditional only that for both the same stoichiometric mass coefficients are used. Nonetheless, G&B's precautionary remarks about the effects on PR, CH, and LP estimates due to possible alteration of the H₂O content after determination of dry weight and during sample manipulation remain valid (cf. also Newsletter 2, 1985). This is not a mathematical but an operational problem; the H₂O estimate is that of the sample at the moment of CHN measurement.

Examples. Table 1 reports step by step calculation of %PR, %CH, %LP and %H₂O for *Cyclops abyssorum* in two versions: **A**) - using G&B's coefficients given in G&B Tab. 1 (marked small rectangle in matrix); also the fractional value for hydrogen in water in the expanded matrix (large rectangle) is theirs). **B**) - using the modified coefficient matrix to account for chitin in CH. The ash corrected input values %C, %N, %H, used for both matrix versions, the original G&B and the iterative models have been provided by Dr. N. Riccardi; %O is 100-(%C+%N+%H). Finally, because in her work Riccardi sets G&B's factor XNP (fraction of protein-N per total organic nitrogen) = .93 - supposedly to account for N in chitin - also her estimates are reported. These different estimates can therefore directly be compared. Notice: For consistency all estimates with exception of the first W_j matrix have been H₂O adjusted.

Versión A). Apart from Riccardi's estimates, the remarkable identity of values obtained by the three calculation method means: 1) that the expanded a_{ij} matrix correctly sizes the fractional coefficients of the respective reference compounds justifying the assumption that the estimated ash corrected dry matter essentially consists of carbon, nitrogen, hydrogen and oxygen only, and therefore the respective oxygen coefficients and input values can be estimated by difference. However, as evidenced below, any measurement uncertainty in C, N, and H will propagate to the O estimates. 2) Although not explicitly introduced as variable in G&B, oxygen appears as a hidden variable in their equations.

Versión A) against versión B) estimates, instead, differ substantially, particularly with regard to the CH estimates. CH estimates obtained with the chitobiose modified coefficient matrix almost double in value. The decrease of PR estimates is a consequence of the partition of N between PR and CH, of course; yet, whereas the LP estimates remain almost unaltered the difference in PR estimates goes almost quantitatively to CH. This pattern has been corroborated by many more examples calculated with both models. Further to notice: Although residual water contents were found to vary from

Tab. 1. demonstrates: 1) that the matrix solution and G&B Eq.(A14) give the same estimate for residual water; 2) that PR, CH, and LP estimates obtained with the original G&B equations and the matrix solution adjusted for residual water are equal within a few one hundredths; 3) that the same values are obtained with the Vollenweider iterative model adjusted for residual water using the respective matrix or G&B estimates; 4) that altering of the CH mass fraction vector leads to substantially different PR, CH, and LP estimates; 5) that Riccardi's estimates of PR, CH, and LP are at variance with the estimates of other models.

A) Coefficient Matrix a_{ij} : original G&B values, also for H₂O

a_{ij} Matrix	PR	CH (glycogen)	LP	H ₂ O	
C	0.5290	0.4440	0.7760	0.0000	G&B values
N	0.1730	0.0000	0.0000	0.0000	
H	0.0700	0.0620	0.1140	0.1006	
O	0.2280	0.4940	0.1100	0.8994	
SUM	1	1	1	1	

Inverted Matrix

c_{ij} Matrix	C	N	H	O
PR	0.0000	5.7803	0.0000	0.0000
CH	2.5476	-2.7899	-19.4393	2.1743
LP	-0.1690	-2.3442	11.1225	-1.2441
H ₂ O	-1.3786	0.3537	9.3168	0.0697
SUM	1	1	1	1

Example: *Cyclops abyssorum* (nauplii)

Data provided by Dr. N. Riccardi:

CNH measured = % C, N, H, ash corrected

%	W_i Matrix		Riccardi correction
C	50.774	X(np)=.93 →	50.774
N	11.800		10.974
H	7.464		7.464
O	29.962		30.778
SUM	100		

%	W_i Matrix Estimates	Adjusted for water	G&B original model	Vollenw. iterative Model	Riccardi estimate N-adjusted
PR	68.208	72.414	72.41	72.41	67.177
CH	16.481	17.497	17.48	17.5	21.796
LP	9.503	10.089	10.11	10.09	11.026
H ₂ O	5.808	-	*)	-	-
SUM	100	100	100	100	100

*) Notice: H₂O with G&B Eq. (A14) gives 5.8 (Loop 0)

B) Coefficient Matrix a_{ij} : adjusted for chitin & corrected H₂O

a_{ij} Matrix	PR	CH (chitobiose)	LP	H ₂ O	
C	0.5290	0.4528	0.7760	0.0000	corrected values
N	0.1730	0.0660	0.0000	0.0000	
H	0.0700	0.0665	0.1140	0.1118	
O	0.2280	0.4147	0.1100	0.8882	
SUM	1	1	1	1	

Inverted Matrix

c_{ij} Matrix	C	N	H	O
PR	-1.4546	7.5377	11.2707	-1.4187
CH	3.8129	-4.6063	-29.5429	3.7186
LP	0.0554	-2.4506	9.5552	-1.2027
H ₂ O	-1.4137	0.5193	9.7170	-0.0972
SUM	1	1	1	1

Same example

%	W_i Matrix		Riccardi correction
C	50.774	X(np)=.93 →	50.774
N	11.800		10.974
H	7.464		7.464
O	29.962		30.788
SUM	100		

%	W_i Matrix estimates	Adjusted for water	Vollenw. iterative Model	Riccardi estimate N-adjusted
PR	56.704	59.04	58.95	51.07
CH	30.153	31.40	31.52	38.35
LP	9.180	9.56	9.54	10.58
H ₂ O	3.962	-	-	-
SUM	100	100	100.01	100

(Loop 5)

data set to data set, those obtained with chitobiose modified coefficient matrices were generally lower than those obtained with the original G&B procedure.

Which of the two models gives results closer to reality? Without comparison of model data with direct biochemical measurements the answer remains as yet open. G&B could show that for their kind of material (muscle, liver, gut, food, feces) correlations between biochemical

measurements of carbohydrate, protein and lipid and model data were sufficiently high to lend confidence in model predictions, at least as far as averages were concerned; individual biochemical measurements, particularly for carbohydrates in gut, food, and feces, instead, showed considerable variations. This is of interest in the context of model application to zooplankton. Unfortunately, the material studied by Riccardi lacks direct biochemical measurements. What is clear, however, is that

because of the presence of chitin in the cuticula body protein must be lower relative to the measured total organic nitrogen than the values predicted by the G&B's model. Still, the assumption that the cuticula of zooplankton is only chitin is also questionable; it also contains other components, such as protein (for pertinent literature about the chitonoproteic microstructure cf. e.g., Goffinet & Jeuniaux 1994; a.o.). Besides, from a physiological point of view there is no reason to assume that carbohydrates like glycogen have no role in the energy metabolism of zooplankton, wherefore glycogen is likely present in zooplankton. Hence, one may hypothesize that the more likely composition of crustacean plankton in protein, carbohydrate and lipid will be intermediate between G&B's and the modified model predictions: lower in protein, but higher in carbohydrate, distributed between glycogen and chitin.

While at present the actual glycogen/chitin relationship is unknown, one can play with the a_{iK} coefficients to explore the question what the effect of various assumption about the respective ratio would be on the estimates of PR, CH, and LP. To do this one may calculate the a_{iK} coefficients as a mixture of $a_{i(K=ct)}$ and $a_{i(K=gl)}$, varying the glycogen fraction but keeping the sum of the partial mixing factors constant. This has been done, always for the *Cyclops abyssorum* example, in figure 1. With a glycogen fraction = 0 one obtains the version B) solution (at left of the diagram). Increasing the glycogen portion the total CH (chitin+glycogen) estimates decrease, and the PR estimates increase. But as long as the assumed glycogen fraction is modest, say <20%, the effect on the respective PR, CH and LP estimates is modest. Yet at higher glycogen fractions the divergence in estimates becomes more apparent. Finally, with a glycogen fraction = 100% one obtains the G&B solution (at the right of the diagram).

Similarly, one could play with introducing a protein component into the a_{ij} matrix, but the question is where. As long as the choice was between glycogen and chitin the right place was modifying the CH coefficients; both substances are in essence carbohydrates. In regard to proteins, instead, one would have to distinguish between protein in different compartments, in the cuticula and in other body parts. As dissection of small zooplankton bodies is hardly a practical possibility such a distinction would require the availability of distinctive markers. As at present non is available, the matter remains conjectural.

Measurement errors; sensitivity analysis. For the mathematically inclined reader matrix solutions of these and other problems offer a broad field for further exploration. However, said this, it's easy to lose sight of others, perhaps more burdensome problems with proxy-estimates. Whichever calculation model is used, and whatever assumptions are made, estimates are only as good as the confidence one can have in the input data.

Recognizing this turns the attention back to the question about the reliability of the initial CHN analyses. In fact, in the previous calculations we have assumed that CHN measurements are error free. This is obviously not true. An indication about the extent of analytical uncertainties encountered in her work is given in Riccardi (2000). Thus, let's ask how errors in the input data affect the PR, CH, LP, and H₂O estimates.

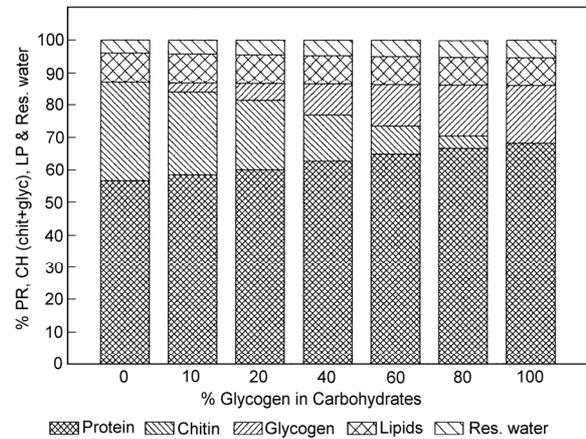


Fig. 1. Illustrates the effect of increasing glycogen/chitobiose ratios used in lieu of fixed CH mass fraction coefficients on the final estimates of PR, CH, LP, and residual water.

To answer the question one has to understand how the $|c_{ji}|$ matrix enters into the calculation of the W_j matrix. In short: W_j for each j is the sum of the products of all W_i 's multiplied with the corresponding factor c_{ji} (cf. also Appendix). Previously we called the c_{ji} "partition coefficients"; this is not quite correct. A better term would be "influence coefficients" because the c_{ji} values can either be positive or negative. It means that the influence of any one c_{ji} on the respective estimates of PR, CH, LP and H₂O is either positive (increasing W_j), or negative (decreasing W_j), and its magnitude is proportional to the absolute value of the c_{ji} 's. Thus, even without considering the actual error distribution around the means of the input data one recognizes immediately that some c_{ji} affect the estimates only marginally, others more, and still others very substantially. Also, comparing the c_{ji} within and between the matrices of the two versions of the above example one sees that in either case the major influence exerted on the PR, CH, LP and H₂O estimates comes from the vector c_{jH} , and further that its influence will be highest in version B).

Hydrogen, the critical factor. Once the coefficient matrix a_{ij} has been determined the c_{ji} is also determined, to say its values remain invariant in any further calculation. Therefore, variations in W_j depend only on variations in the W_i matrix. Let's now study the effect of uncertainty in W_H measurement on the estimates of PR, CH, LP, and H₂O for the *C. abyssorum* example. For this let's assume that $H=7.46$ was the correct mean value

but that measurement uncertainty around the mean would cause H to fluctuate by .1 and .2 units. Recalculating the PR, CH, LP, and H₂O estimates accordingly, and plotting these estimates against H (cf. Fig. 2) shows modest increases of PR (dPR/dH=12.7) and LP (dLP/dH=10.8) over the total range of H. H₂O increases relative to the other two are about the same (dH₂O/dH=9.8), but extrapolation of the slope beyond the range indicates that below H<7.2 would become negative. CH, instead decreases dramatically over the range from 36.8 to 23.5 (dCH/dH=-33.3).

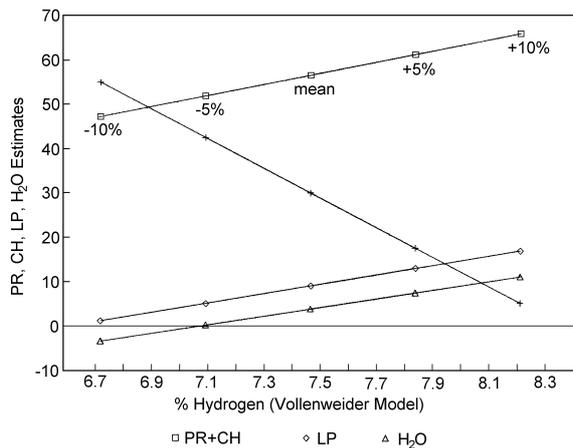


Fig. 2. Illustrates the effect of uncertainties of H determinations around the mean in CHN measurements on the PR, CH, LP, and residual water. Based on the above *Cyclops abyssorum* data, and indications provided by Riccardi about potential analytical uncertainties according to her experience. Notice: corresponding diagrams could be constructed for the other components.

Considering that ± 2 relative to the mean represents an uncertainty of only $\pm 2.7\%$, the question arises: are the 5 sets of estimates different? Intuitively, one would say yes; yet, statistically speaking the answer is less clear, depending entirely on the circumstances surrounding the measurements. If these are known, the question can be resolved using appropriate statistical techniques. Let's assume that both, the mean of H and the STD s_H were obtained from $N=3$ measurements (which is often the case), and s_H , was estimated to $=.1$. Then one may ask: Does the difference, dM, between any arbitrary selected couple among the H's, say 7.36 and 7.46, significantly deviate from 0, and if, at what level of confidence. The test involves estimating Student's t. Essentially there are three slightly varied test models of the same type: a) a one sample test; b) a two sample test; c) evaluating the STD around dM.

In the *one sample test* one of the terms of the couple is assumed to be error-free, e.g., the mean 7.46. To test $dM=7.46-7.36=.1$ one calculates $t=(dM/STD)\times\sqrt{N}=(.1/.1)\times\sqrt{3}=1.732$. 1.732 is then compared with Student's $t_{(\alpha/2, N-1)}=4.304$ (for $3-1=2$ degrees of freedom at a

p(95%) level of confidence, equivalent to $\alpha/2=.025$; the respective t values are tabulated in statistical text books). If instead the difference $7.46-7.56=-.2$ is tested then t calculated would be -3.464 , a.s.o. Both the absolute values calculated are less the 4.304; hence, the differences do not significantly deviate from 0 at the p(95%) level, to say variations of ± 2 around the mean are undistinguishable from the mean.

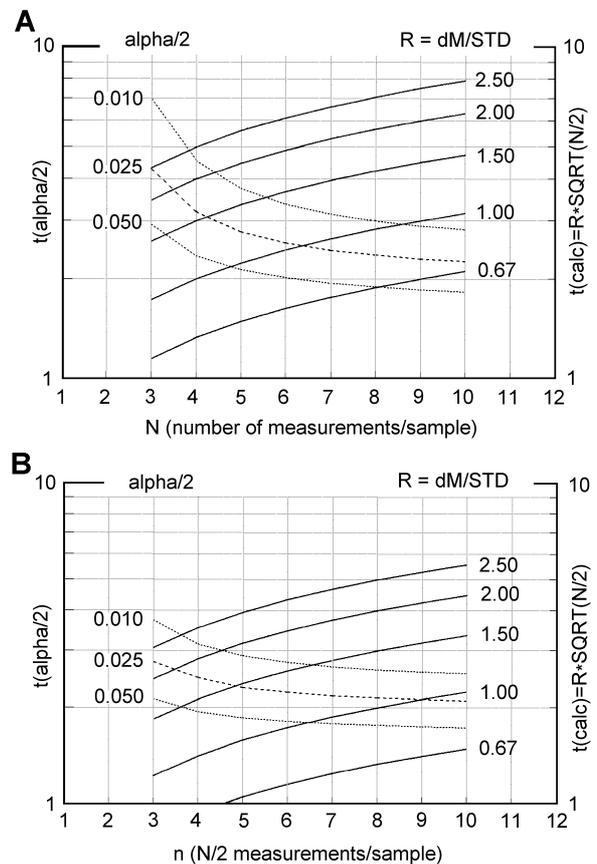


Fig. 3. Student's *versus* N & confidence levels (expected and calculated t-values). **A:** one sample test. **B:** two sample test. The figures illustrate the number N of replicate measurements necessary to meet the criterion of significance at various levels of probability (alpha) of standardized differences between two means (dM/STD). N necessary is determined by the intersection of the two respective curves.

To simplify matters and to generalize the case the upper half of figure 3 shows calculated ts for some fixed R ($R=(dM/s)$) (solid lines), over which Student's $t_{(\alpha/2, N-1)}$ are superposed for three levels of $\alpha/2$ (dotted lines), both sets plotted against N. [Notice: any specific R can be the result of any combination of dM and s that give the same ratio, e.g., $R=1.5=.3/.2=.463/.309$, etc.; therefore the figure is not specific to the above example].

In the *two sample test* differences of two means, each of which was obtained from n_1 and n_2 measurements, and having STDs s_1 and s_2 , are compared. The

respective model is $t = dM/\sqrt{((s_1^2/n_1)+(s_2^2/n_2))}$. This is the more common situation. However, a generalized diagram can only be plotted for the specific condition that $n_1=n_2=n$, and $s_1=s_2=s$, which simplifies the calculation of t , $t = (dM/s)\times\sqrt{(n/2)}$. This has been done in figure 3 (lower part). To notice that in this case t is to be compared with Student's $t_{(n/2, 2n-2)}$. Comparison between the two models show that the second, in which both terms are assumed to have errors, is less stringent than the first one, because of common variance and that Student's t refers to double the number of degrees of freedom $((2n-2)/(n-1)=2)$.

Finally one could calculate the STD of the difference dM , so that $dM = dM \pm t_{(n/2, 2n-2)}\times s/\sqrt{(n/2)}$. If the $dM \pm \text{STD}$ includes 0, then dM cannot be distinguished from 0.

Inspection of the diagrams shows that with only 3 measurements/sample for H distinction of Rs at the p(95%) level is only possible if R is at least 2.5. Applied to the *C. abyssorum* example this means that estimated Hs in the range between 7.21 and 7.71 are statistically undistinguishable. As a consequence also the derived estimates of PR, CH, LP, and H₂O over this range will be undistinguishable at the p(95%) level of confidence. Lessen expectation, instead, one may say that the outer values may be different but at a lower then p(95%) level of confidence. Whatever position one is taking, this paragraph shows how critical the precision of H measurements are for the proxy-estimates here in question.

Hydrogen versus Carbon and Nitrogen. In principle one can extend the considerations made about hydrogen to the other input variables. However it is not necessary to enter into much details. The most simple way is to compare the slopes dX/dY ($X = \text{PR, CH, LP, and H}_2\text{O}$; $Y = \text{H, C, N}$), while only the factor considered is varied, keeping the others constant. This gives for the above example the following values;

	dX/dH	dX/dC	dX/dN
PR	+12.69	- .036	+8.96
CH	-33.26	+ .094	-8.33
LP	+10.76	+1.258	-1.25
H ₂ O	+ 9.81	-1.316	+ .62

Clearly, the most sensitive factor remains hydrogen, followed by nitrogen that affects still PR and CH more noticeably and somewhat LP, while variation in carbon has almost no effect on PR and CH estimates, and only minor on LP and H₂O.

STD of PR, CH, LP and H₂O estimates. STDs of estimates are more familiar to most readers then the above sensitivity analyses. However, because the comprehensive STDs of estimates are a function of the STDs of all the input components W_i the mode of error propagation

is not immediately evident. Yet, as long as the function is a linear combination of the components the matter is relatively simple, but would be more complicated if the function includes product and ratios, as in G&B derivations. In fact, The matrix solution contains only linear terms, and, besides, products are multiplications by scalars only. To simplify matters further $W_i \pm s_i$ can be written as $W_i \times (1 \pm cv_i) = W_i + (\pm W_i \times cv_i)$, for $cv_i = s_i/W_i$. Accordingly, the problem reduces to a problem for which we have already a partial solution that requires some modification, however. This involves the component products of each W_j be multiplied by the corresponding cv_i , neglecting the sign, and adding the products according to pythagorean rules (for further details cf. Appendix).

Choosing for instance the following values for cv : $cv_C=.015$; $cv_N=.01$; and varying cv_H from .05 to .01, and setting $cv_W = 1 - \sqrt{(cv_C^2 + cv_N^2 + cv_H^2)}$, then the following PR, CH, LP, and H₂O estimates with their STD are:

	$cv_H=.05$	$cv_H=.025$	$cv_H=.01$
PR = 56.70	± 4.85	± 2.80	± 1.85
CH = 30.15	± 12.5	± 6.98	± 4.31
LP = 9.18	± 3.94	± 2.07	± 1.07
H ₂ O = 3.96	± 3.61	± 2.03	± 1.27

This tabulation underlines that any variation in cv_H affects all STD, but that the most sensitive is STD(CH). With a $cv_H=.05$ (with $H=7.5$, $s_H \approx .375$), which analytically is not exceptional according to Riccardi's experiences, a STD of 12.5 makes the CH estimate of 30.15 almost meaningless. Even the estimate of H₂O becomes questionable. A $cv_H=.025$ ($s_H \approx .187$) is already better but still not fully satisfactory; only if cv_H is brought below .25 all estimate assume some consistency.

In a corresponding way one could also vary the cv_C and/or the cv_N , but at this point a further demonstration becomes pointless. In any case, as the previous table shows, the effects would be far less, particularly regarding C, while variation in N would be moderately be noted in STD(PR) and STD(CH) estimates.

Discussion. From a strict computational point of view the matrix model is preferable to that of the original Gnaiger-Bitterlich procedure for its higher flexibility compared to the latter. Mass fraction coefficient matrices can easily be exchanged with others, as in the case of crustacean zooplankton that requires replacing the glycogen coefficients with those of the chitobiose, or mixtures of the two. Yet, to subtract just an arbitrary value from the CHN measured nitrogen to account for nitrogen in chitin, as Riccardi did, is not satisfactory, but to use a factor $X_{PN} < 1$ may be necessary for other reasons, as discussed in G&B. The iterative model proposed by me has also its limitations. While it gives the same results as the matrix calculation, this is only true if

the residual water content can be assumed = 0, or is either arbitrarily estimated, or obtained from either G&B or the matrix solution; thus, there is no real advantage using it. Not least, only the matrix model allows to explore with ease the effect of various scenarios on the internal factor configuration, and/or to compute the STDs of the estimates in a simple way because no ratios between terms are involved.

However, whether one uses a matrix model, or the G&B model in its original version (cf. Table 2) these kind of models are very sensitive to measurement uncertainties in the input values, of which H has been singled out as the most critical one. While all model versions produce similar trends in all estimates, the slopes of PR and CH relative to variations in H or H₂O are more pronounced with the chitin modified matrix solution than with the original G&B model. This applies also to the estimated caloric content. Still, G&B's equation for estimating energy content appears to be less sensitive to changes in H or H₂O than if the energy content is estimated according to $Kcal = 5.5 \times PR + 4.1 \times CH + 9.3 \times LP$.

Unfortunately, analytical problems afflicting H measurements coincide unfavourably with structural the properties of the models (large negative values in the c_{jH} vector), which also inflate the STD of CH estimates uncomfortably. G&B have recognized the problem in the few examples listed in their Newsletter, but relate it primarily to residual water content, which, of course, is part of the problem. Yet, G&B's equations could easily be rewritten in terms of W_C , W_N , and W_H by substituting W_{H_2O} in their equations with the terms of Eq. 14. E.g., the constants b for W_K would then change to 2.1741, .3727, -4.9637, plus an additional constant -21.61 to be multiplied by W_H , and the matter could be discussed terms of H. Thus, e.g., the high negative value of the latter corroborates the strong negative dependency of W_K on W_H . However, this change in perspective should not be confounded with the concern expressed here about measurement uncertainties. In fact, whether one uses G&B original equations, or those W_{H_2O} substituted, the end results are (and most be) the same. Measurement uncertainties in the input values, on the other hand, propagate, as the matrix solution shows, through the whole system affecting all estimates to various degrees.

A coincidental reason that the problem about uncertainties in the input values, particularly regarding hydrogen, had not been sufficiently recognized earlier may also have been that some older statistical texts list as the error of the mean = $\pm s/\sqrt{N}$; this gives a false impression about the confidence interval around the means when N is small, say 3. In this case the 95% confidence interval around the mean would be estimated to $\pm s/\sqrt{3} = \pm s \times .577$, whereas the real confidence band is $\pm s \times t_{(\alpha/2, N-1)}/\sqrt{3} = \pm s \times 4.303/1.732 = \pm s \times 2.484$, - which is drastically different.

What to do? Logically the first thing to undertake would be to improve the precision of CHN measurements. However, this may not be as easy. Thus, the next best thing to do is playing with statistical properties. There are two alternatives: either to increase the number of measurements/sample and/or increase the number of replicate samples to determine a more reliable mean; or to be satisfied with a lower level of confidence, whichever is possible or acceptable. E.g., to reduce the band width of the confidence interval at the p(95%) level to lie within $\pm s$ the number of measurement/sample should be at least 6 to 7, while discarding clear outliers. Yet, this should not be done arbitrarily, but according to well established statistical rules and procedures. On the other hand, it would be unwise to accept a confidence level of less than p(90%). If for any reason this cannot be achieved, then data published should at least be accordingly flagged as questionable.

Conclusions. The above discussion indicates that every effort should be made to improve the precision of CHN measurements, particularly with regard to H measurements. Otherwise proxy estimates of PR, CH, and LP may become doubtful. Yet, prior to bluntly discard proxy-estimates as too uncertain the matter should also be evaluated in the light of what one can expect from direct biochemical measurements. Neither is free of large uncertainties; thus, the question becomes one of weighing the operational advantages against the disadvantages in the choice between indirect and direct estimates. Direct biochemical analyses permit to explore the presence of an array of different biochemical species (cf. Zamer 1989), which with indirect estimates are indiscriminately lumped into broad but physiologically and structurally somewhat questionable categories. On the other hand, indirect estimates offer the advantage to explore relatively rapidly possible compositional variations in these broad categories over a wide range of either species or development stages of species, or species grown under varied environmental conditions, what then further can be pursued using direct biochemical techniques.

Thus, while a critical look on the inherent weaknesses of proxy-estimates may temper expectations, carefully interpreted results can nevertheless be quite useful in routine surveys, on the condition that they are sufficiently supported by occasional key biochemical analyses. Applied to zooplankton this still needs verifying. Also, such models may be helped if besides the three input parameters (C,N,H) also oxygen could be estimated analytically, at least occasionally to check the correctness of oxygen estimates by simple difference. Seen in a broader perspective it may also be desirable to find in addition to the four input values at least one independent fifth, or, alternatively, some specific biochemical markers that would permit to discriminate, e.g., between protein and carbohydrate compartments, say proteins in the cuticula and proteins in the rest of the body, or between chitin and other carbohydrates, etc.

Acknowledgment. I wish to thank Dr. N. Riccardi for providing me her original data material of which I have made an arbitrary selection discussed in the present paper, and supplying me with additional literature. Regarding further implication resulting from the study consult her paper in this volume (Riccardi 2000).

REFERENCES

- Gnaiger, E. & G. Bitterlich. 1984. Proximate biochemical composition and caloric content calculated from elemental CHN analysis: a stoichiometric concept. *Oecologia* (Berlin), 62: 289-298.
- Gnaiger, E. & G. Bitterlich. 1985. Cyclobios Newsletter 2. Mimeography: 5 pp.
- Goffinet, G. & Ch. Jeuniaux. 1994. Le tégument: Morphologie et Biochimie. In: P. Grassé (Ed.), *Traité de Zoologie*. Tome VII (Ed. J. Forest). Masson, Paris.
- Riccardi, N. 2000. Comparison of different stoichiometric methods for the estimation of proximate biochemical composition of crustacean zooplankton and some considerations on energy transfer to planktophagous fish. *J. Limnol.*, 59(2): 179-185.
- Riccardi, N. & M. Mangoni. 1999. Considerations on the biochemical composition of some zooplankton species. *J. Limnol.*, 58(1): 58-65.
- Vollenweider, R.A. 1985. Elemental and biochemical composition of plankton biomass; some comments and explorations. *Arch. Hydrobiol.*, 105: 11-29.
- Wadsworth, H.M. 1990. *Statistical Methods for Engineers and Scientists*. McGraw-Hill Publishing Company New York
- Zamer, W.E., J.M. Shik & D.W. Tapley. 1989. Protein measurement and energetic considerations: Comparisons of biochemical and stoichiometric methods using bovine serum albumin and protein isolated from sea anemones. *Limnol. Oceanogr.*, 34: 256-263.

Appendix

Abbreviations and Symbols used

PR, CH, LP stay for protein, carbohydrate, lipid in a general sense
 STD for Standard Deviation; CV for Coefficient of Variation in percent
 W for mass fraction, either as value $0 < 1$, or in percent

Specifically:

W_i = ash content corrected mass fractions for C, N, H, resulting as mean from N CHN analyses;
 W_o for oxygen by difference $(1 - (C+N+H))$, or, $100 - (\%C + \%N + \%H)$

$|W_i|$ = INPUT MATRIX

W_j = estimated mass fractions of PR, CH, LP and H_2O ; abbreviated $j = P, K, L, W$, respectively

$|W_j|$ = OUTPUT MATRIX

a_{ij} = stoichiometric mass fraction coefficients for the selected reference compound (dummy variables for PR, CH, LP, H_2O)

c_{ji} = corresponding coefficients of the inverted $|a_{ij}|$ matrix

$|a_{ij}|$ = coefficient matrix

$|c_{ji}|$ = inverted coefficient matrix

s_i = STD of W_i estimated from N CHN analyses

cv_i = CV as fraction of 1.

Gnaiger & Bitterlich Equations (cf. Table 2 in G&B)

$Y_j = b_{jo} \times (1 - W_{H_2O}) + b_{jC} \times W_C + b_{jN} \times W_N \times X_{PN}$, with the following values of the b's for

$$W_K: 2.337, -3.012, -4.300$$

$$W_L: -1.337, 3.012, -1.480$$

$$W_P: 0, 0, 5.780$$

and for residual water (Eq. 14, Appendix in G&B)

$$W_{H_2O} = .0697 - 1.4483 \times W_C + .284 \times W_N \times X_{PN}$$

X_{PN} being the fraction of measured N in protein

Matrix multiplication; conventions used in the present context. To clearly distinguish between the a matrix and the c matrix as the inverse of the former, the sequence of the indices have been inverted. This is justified because the relationship between rows and columns in the c matrix is in effect also inverted. Multiplication of the two matrices, $|c_{ji}| \times W_i$, is then done summing the products of the elements in the j th row of the first matrix with the i th element in the second one, which gives:

$c_{PC} \times W_C + c_{PN} \times W_N + c_{PH} \times W_H + c_{PO} \times W_O$	$= W_P$
$c_{KC} \times W_C + c_{KN} \times W_N + c_{KH} \times W_H + c_{KO} \times W_O$	$= W_K$
$c_{LC} \times W_C + c_{LN} \times W_N + c_{LH} \times W_H + c_{LO} \times W_O$	$= W_L$
$c_{WC} \times W_C + c_{WN} \times W_N + c_{WH} \times W_H + c_{WO} \times W_O$	$= W_W$
Column sums	$W_C \quad W_N \quad W_H \quad W_O \quad \quad 1$

Further to notice that the matrix is complementary in the sense that each column sum equals the respective input values W_i because the sum of the respective c_{ji} 's = 1. Yet, how the inversion of the a matrix is done is not considered here because matrix inversion can easier be accomplished by computer.

STD of PR, CH, LP, H₂O. Referring to the formula for the total differential of the function $f(x, y, \text{etc.})$, $\Delta f = \delta f/\delta x \times \Delta x + \delta f/\delta y \times \Delta y$ etc., one recognizes that by partial differentiation of the W_j 's with respect to C, N, H, H₂O the coefficients c_{ji} remain unaltered. Accordingly, for calculating the STD of the W_j 's one can simply replace W_i in the above equations by the appropriate $\pm s_i$ or its correlate, $\pm cv_i \times W_i$. However, the \pm sign changes the addition rule, since $\pm m \pm n \pm \dots = \sqrt{(m^2 + n^2 + \dots)}$, (pythagorean rule of addition). Hence, the STD of each W_j becomes the square root of the sum of the squared terms: $c_{ji} \times s_i$, or $c_{ji} \times cv_i \times W_i$.

Whether one uses s_i or $cv_i \times W_i$ depends on the information base. Theoretically, s_i is the correct value, but it means that s_i would have to be determined for each individual analysis run. In praxis s_i 's are often only periodically determined to check the performance and stability of the instrument. From this one may then estimate an average CV, but to calculate from CV the STD is only justified if the range of variation of W_i remains relatively small; otherwise s_i would become proportional to W_i . On the other hand to use cv_i instead of s_i can be useful in exploratory studies in the absence of precise values for s_i , as in the present case. (In the material analyzed by Riccardi *et al.* all the W_C data scattered around the 50% mark (STD= ± 2.5), the W_N around the 10.7% mark (STD= ± 1.1), and the W_H around the 7.8% mark (STD= ± 4)).

Examples of calculation

a) Matrix multiplication, for second row of the *C. abyssorum* example (cf. Tab. 1)

$$(3.81 \times 50.77) - (4.61 \times 11.8) - (29.54 \times 7.46) + (3.72 \times 29.96) = 30.15\%CH$$

b) STD. Using $cv_C = .015$, $cv_N = .01$, $cv_H = .05$, $cv_W = .053$, and adding the products according to pythagorean rule gives:

$$\sqrt{[(\pm 3.81 \times 50.77 \times .015)^2 + (\pm 4.61 \times 11.8 \times .01)^2 + (\pm 29.54 \times 7.46 \times .05)^2 + (\pm 3.72 \times 29.96 \times .053)^2]} = 12.5.$$

Thus, %CH = 31.15 \pm 12.5

(Notice: Small differences between the reported end results and those that may result by calculating the MEAN and STD directly using the above values will be due to rounding errors; computer calculated data use 9 decimals).