Using doubly-labelled water to measure free-living energy expenditure:
Some old things to remember and some new things to consider

John R. Speakmana,b, Catherine Hamblyb

Abstract
The doubly-labelled water (DLW) method provides the ability to measure the energy expenditure of free-living animals based only on the injection of two isotopes in water (one of oxygen and one of hydrogen) and traditionally the collection of 2 blood samples. We review here the fundamental basis of how the method works, and highlight how the choice of the appropriate calculation equation can have a large impact on the resultant estimates, particularly in species where the difference between the isotope elimination constants is small. This knowledge is not new, but is worth reiterating given the potential for error by making the wrong choice. In particular, it is important to remember that for mammals weighing less than 5 kg, and birds weighing less than 2 kg, the single pool model performs best in validation studies, while in mammals above 15 kg the two-pool models perform best. Above 2 kg in birds and between 5 and 15 kg in mammals, however, the model superiority is uncertain. Even where the choice based on body mass would appear clear, the decision may need to be tempered by species specific information regarding potential additional sources for hydrogen turnover, such as de novo lipogenesis or methanogenesis. Recent advances in the technique have included attempts to make the method less invasive by using innovative methods for dosing and sample collection. In addition, the advent of laser spectroscopy, as a replacement technology for mass spectrometry, may open up many new opportunities in the field. These potentially include direct sampling of breath in the field and tracking background isotope drift using 17Oxygen levels.

1. DLW 101
When cells aerobically respire they utilise oxygen to liberate energy from diverse substrates generating CO2 in the process. The CO2 needs to be removed from the body and in multicellular organisms it is transported to the respiratory organs (e.g. lungs, gills or tracheae) where it is expired. During their journey from the cells to the outside world however the CO2 molecules are not static entities. They continuously combine with water under the action of carbonic anhydrase to form bicarbonate and other ionic species (Chinard et al., 1958). During the process the oxygen molecules from the water become mixed up with the oxygen molecules from the CO2. This was discovered in 1949 when isotopically heavy water (containing 18Oxygen) was first injected into a mouse and it was found that the isotope transferred from the water to the CO2 (Lifson et al., 1949). This discovery led to the remarkable, and not immediately obvious, insight that it might be possible to utilise this exchange to measure CO2 production of an animal (and hence infer its energy expenditure). Moreover, this might be achieved without the need to confine it in a respiration chamber, or require it to wear a respiratory mask.

The theoretical basis of the idea was that because the oxygen molecules get exchanged between H2O and CO2 during CO2 transport then if a dose of 18Oxygen is introduced into the body it will be gradually eliminated because of the combined flow of water and O2/CO2 through the body. In fact the gradual decline depends only on the flow of the materials and the size of the pool that they are transported in. Hence, if we convert the isotope enrichments over time using logs they fall on a straight line that we can characterise by the steepness (k0) minus the reference isotope enrichment. If we define the size of the body water pool as N then the combined flow of water (rH2O) and CO2 (rCO2) is equal to

$$rH_2O + rCO_2 \times 2 = k_0 \cdot N.$$  (1)
on 2 occasions to take a sample of its blood to get a measure of the
rH₂O + rCO₂. Second, most people are not interested in the combined
rH₂O + rCO₂, but these quantities separately. However, if you introduce
an isotope of hydrogen into the body water then by the same logic it will
be eliminated exponentially, but only by rH₂O.

\[
rH_2O = k_d \cdot N
\]  

(2)

hence substituting 2 into 1 gives

\[
rCO₂ = N/2 \left( k_b - k_d \right)
\]  

(3)

Thus an estimate of rCO₂ can be found by simultaneously introduc-
ing both isotopic oxygen and hydrogen into the body and tracing their
elimination tracks, requiring only that a sample of the body water
pool (blood) is collected on two occasions for the isotopes to diverge
sufficiently. Because deuterium (chemical symbol D) and ¹⁸O are used
Lifson christened the method the D¹⁸O method in a seminal validation
paper published in 1955 (Lifson et al., 1955), but later because two iso-
topic labels in water are used the technique became known as ‘dou-
ubly-labelled water’ or the DLW method. In the initial validation work
the technique was compared to the standard method of indirect calorim-
etry and shown to give an estimate of CO₂ production that on average
matched that by indirect calorimetry with an error of around 3%. This
was followed by several additional validations in the late 1950s
(e.g. McClintock and Lifson, 1958a, 1958b). The scene was set to mea-
sure the energy demands of truly free-living organisms by simply cap-
turing them, injecting them with isotopes, taking a small blood
sample after the isotopes had equilibrated with the body water pool, re-
leasing them and capturing them sometime later for a second sample.

You might think that the dawn of a new method able to measure en-
ergy demands of free-living animals would have sparked an enormous
response, but progress was painfully slow. After the seminal validation
in 1955 it was another 9 years before Lefebvre (1964) used the tech-
nique to measure the free-living energy demands of flight in homing pi-
geons (Columba livia). The first application to a free-living mammal had
to wait 15 years (Mullen, 1970). The key problem was that the technol-
ygy able to measure isotope enrichments was very expensive, and it re-
quired large doses of stable isotopes that were also very costly. In 1975 it
was speculated whether the method would ever be used to measure a
human (Lifson et al., 1975). They estimated that at a cost of 50,000 US
dollars per subject the answer was probably no. In fact, the first humans
were measured only 7 years later (Schoeller and Santen, 1982). The key
changes in the interim were that mass spectrometry technology ad-
vanced significantly, reducing the required isotope doses, and commer-
cial production of the isotopes massively reduced their cost. Currently
the largest animal on which the method has been applied (and pub-
lished) is the walrus (Odobenus rosmarus) weighing almost a metric
ton (998 kg) (Aquarone et al., 2006), although even larger animals have
been measured in unpublished work (D. Costa, pers. comm.).

2. Some complexities

The simple model elaborated above hides a host of assumptions
about the way isotopes are distributed in the body and the reactions
they become engaged in. These assumptions have been elaborated in
detail by several authors (Lifson and McClintock, 1966; Nagy, 1980;
Speakman, 1997). Briefly they include 1) that the body water pool size
remains constant during the measurement period, 2) that the rates of
water and CO₂ flux are constant during the measurement period,
3) that the isotopes label only the water and CO₂ in the body, 4) that
the isotopes only leave the body as water and CO₂, 5) that the enrich-
ments of the isotopes leaving the body are the same as those left behind,
6) that isotopes do not re-enter the body once they have left and 7) that
the background levels of the isotopes remain constant over the mea-
surement interval.

We will explore here in more detail some of the consequences
of some of these assumptions. When isotopes leave the body the enrich-
ments in the compounds that leave is slightly different from the enrich-
ments in the compounds left behind (fractionation) because lighter
isotopes have slightly lower energies of evaporation. Lifson et al.
(1955) incorporated these effects into their calculations but the infor-
mation they had about these effects was imperfect — fractionation ef-
facts were only known at 25 °C rather than at body temperature, and
the proportion of water that is evaporated was also unknown, and
was hence assumed to be 50%. These effects were later refined by vari-
ous authors (e.g. Nagy, 1980) culminating in 1997 in the following
equation

\[
rCO₂ = N/2 \cdot 0.007k_d \cdot N
\]  

(4)

[Eq. (7.17) in Speakman, 1997 which is currently the standard equation
used in the majority of studies of free-living small animals]. You can eas-
ily see in this equation the basic structure of Eq. (3). For details of how
the additional factors were derived, refer to Nagy (1980) and
Speakman (1997).

The advent of studies in humans brought several refinements in the
technique and a period of intense debate about the fundamental basis of
the methodology. The first of these debates centred around how we
should measure N — the body water pool. Prior to the 1980s this was
often measured by killing the animal at the end of the study and desic-
cating it to measure the water content directly — or using an approxi-
mation from the measured body mass based on desiccation of other
individuals (e.g. Nagy, 1980). Clearly both these approaches were un-
suitable for humans. However, it was noted that if the exact amount of
the dose is known and the first isotope enrichment in the body is mea-
sured after the dose has spread through the entire pool, then it is possi-
ble to estimate N from the dilution principle (Halliday and Millar, 1977).
It had been known since the early 1970s that the dilution space for oxy-
gen (N₀) is smaller than the dilution space for deuterium (Nₓ) by be-
tween 1 and 8%, averaging about 3% (reviewed in Sheng and Huggins,
1979) [Nₓ/N₀ = 1.03]. So what should be used for N: N₀ or Nₓ? In fact
desiccation studies of many animals suggested that N₀ was a faithful
representation of the body water pool (e.g. Nagy, 1980), but that Nₓ
over-estimates it primarily because a small amount of hydrogen also ex-
changes onto amino acids in proteins (Culebras and Moore, 1977), and
therefore Nₓ was the best estimate. A revolutionary suggestion however
(Coward and Prentice, 1985), was that we should use both N₀ and Nₓ
and multiply the elimination constants by their respective dilution
spaces. Ignoring fractionation then Eq. (3) would become

\[
rCO₂ = \frac{1}{2} \left( k_b - N₀ - k_d \cdot N_d \right)
\]  

(5)

This equation (and derivatives) became known as the ‘two pool
model’, and Eq. (3) plus derivatives, the ‘single pool model’. Because
Nₓ > N₀ the single pool model equations always generate higher esti-
mates than their equivalents based on two-pool models. The key ques-
tion is what equation is better: single pool or two-pool? The answer is it
depends on body size (Speakman, 1993). It turns out that validation
studies, where DLW is compared to indirect calorimetry in mammals
that weigh more than 15 kg, suggest that the two-pool model gives the
closest estimate. In mammals weighing less than 5 kg, however, the
single pool model works best (reviewed in Speakman, 1997 and con-
firmed in more recent validations e.g. Speakman et al. 2001; Sparling et
al., 2008). Presently there are no validation studies for mam-
als weighing between 5 and 15 kg, so the actual location of the transi-
tion between the models is uncertain. Validation studies in this
size range are needed. In birds validation studies for species weighing
less than 2 kg indicate that the single pool model is best (reviewed in
Speakman, 1997 and subsequent validation studies; Visser and
Schekkerman, 1999; Van Trigt et al., 2002; Shirai et al., 2012). Above
this body mass the situation is less clear. In eider ducks (Somateria
Please cite this article as: Speakman, J.R., Hambly, C., Using doubly-labelled water to measure free-living energy expenditure: Some old things to remember and some new things to consider, Comp. Biochem. Physiol., A (2016), http://dx.doi.org/10.1016/j.cbpa.2016.03.017
mollissima), weighing 2 to 3.8 kg the two pool model was superior (Hawkins et al., 2000). However, in bald eagles (Haliaeetus leucocephalus) weighing 5 kg the single pool model was most accurate (Dykstra et al., 1997). So the clear body mass effect in mammals is less clearly defined in birds, perhaps because of the smaller body size range in the birds, and because validations in mammals have been performed in two distinct groups (<5 kg and >15 kg). As we fill in this gap the situation in mammals will likely also become less clear.

A theoretical reason explaining this body size effect was proposed in 1987 (Speakman, 1987) when it was noted that labels of hydrogen not only participate in extra exchange, creating an elevated dilution space, but they also participate in extra elimination processes — e.g. via methane production (Midwood et al., 1989) or de novo lipogenesis (Haggarty et al., 1991). These phenomena have opposite effects on the final CO2 estimate to that of the excess hydrogen space. However, while the extra dilution space is independent of body mass, being always around 3%, the extra elimination is not — as a proportion of total hydrogen flux it is greater in smaller animals. This means that in smaller animals the extra pool size effect is completely offset by the extra flux effect and it is better to use the single pool model, while in larger animals the two-pool model works best — because it assumes that the extra flux is zero, and compared with the dilution space difference it effectively is in large mammals.

In practice the difference between the pool models depends on the ratio of the elimination constants (k_{1}/k_{2}). In theory, for example if the k_{1}/k_{2} ratio was only 1.03 and the pool size ratio (N_{1}/N_{2}) was also 1.03, then the CO2 using the two pool model would be zero! Hence, the smaller the k_{1}/k_{2} ratio the larger the difference between the two calculations. These differences can be enormous. We illustrate this effect using data from two of our recent studies of large animals: the giant panda (Ailuropoda melanoleuca) which has a very low k_{1}/k_{2} of 1.077 (Nie et al., 2015) and the Shetland pony (Equus caballus) which has a higher k_{1}/k_{2} of 1.29 (Brinkmann et al., 2014). Since both of these animals exceed by far the recommended cut-off of 15 kg, then on the basis of the validation studies, the two pool model should be used (Speakman, 1993, 1997). For both these animals we therefore used a two-pool model equation, (based on Eq. (4)) but incorporating more recent fractionation assumptions, specifically Eq. (7.43) in Speakman, 1997). Using this equation the metabolic rate of the panda averages 5.14 MJ/day (sd = 0.76, n = 8: Nie et al., 2015) which is exceptionally low compared to the allometric prediction for a 92 kg mammal (13.8 MJ/day: Speakman and Krol, 2010). The DEE of the much larger pony averages 63.4 MJ/day (sd = 14.5, n = 10): Brinkmann et al., 2014). If we use the inappropriate single pool equation (Eq. (7.17) in Speakman, 1997) for these large animals, the estimate for the pony increases to 74.3 MJ/day (sd = 16.8) which is 14.7% higher. But because of the much lower k_{1}/k_{2} ratio (and a slightly higher N_{1}/N_{2} ratio) the estimated metabolism of the panda when using the single pool equation increases to 21.5 MJ/day — about 4× greater than the two-pool calculation!! Clearly model selection can have an enormous impact on the result when k_{1}/k_{2} is low, with attendant impacts on the interpretation. For example, in this instance, if we had used the single pool rather than the two-pool model, we would have concluded the complete opposite result to that published. That is we would have claimed that the panda has a much higher metabolic rate (about 1.5× greater) than predicted from the allometric relationship based on other mammals.

How much faith then can we place in the model selection based on body size? To reiterate: validation studies suggest in mammals using the two-pool model above 15 kg, and the single pool below 5 kg. Between these limits the situation is uncertain. In birds, below 2 kg the single pool model is best, but above that the situation is also unclear. However, even were the suggestion based on body mass is clear, where the k_{1}/k_{2} ratio is low we should take extra precautions because the impact of the wrong choice makes such a big difference. For example, one explanation for the very low measured DEE of the panda, when using the two pool model, could be that the two pool model assumption of no excess hydrogen elimination is violated because of a large extra hydrogen flux via methane (Midwood et al., 1989), potentially as a result of the panda’s bamboo diet. We were aware of this possibility, but it seems highly unlikely, given that previous studies have indicated that pandas do not produce appreciable amounts of methane (Hackstein, 1997). Moreover, to be more confident in the figures we independently validated the DEE estimates using DLW for 5 captive pandas by measuring their DEE based on their metabolisable energy intake and urine production and these estimates for DEE did not differ significantly from the DLW estimates (Nie et al., 2015). Finally, one needs to also apply a bit of common sense. It is inconceivable for the panda to have a metabolic rate that is 1.5× the allometric prediction for other terrestrial mammals, when it also has chronically depressed thyroid hormone levels that are indicative of low basal metabolism (Johnstone et al., 2005; Welcker et al., 2013), spends most of its time at rest, and when it does move, it is glacially slow covering on average 15.5 m per hour when foraging and 26.9 m per hour when commuting (Nie et al., 2015).

This comparison emphasises four key points. First, that when k_{1}/k_{2} is low, correct model selection (hence equation) based on body mass is crucial. Second, although body mass provides a strong guide for model selection, around the transition masses we have relatively little (or no) validation data. In this situation calculating the data using both approaches to see how it influences the interpretation seems the most cautious and sensible approach (see e.g. Briscoe et al., 2016). Third, one should always be aware that this model selection could be compromised by violations of the model assumptions — in particular for the two-pool model which assumes that extra hydrogen flux via e.g. methane or de novo lipogenesis is effectively zero. Finally, if the estimates of metabolism don’t make any sense relative to other information at hand, be aware that this could be because you are using the wrong equation.

3. Recent advances and future perspectives

3.1. Behavioural impacts of the method and refinements in the methodology

Since very early in its inception there have been concerns over the impact of the method on the behaviour of the animals being measured and hence whether the technique provides useful representative estimates (e.g. Speakman et al., 1991). In 1997 this evidence was reviewed, concluding that the method was probably benign in most circumstances except perhaps when using intramuscular injections (Speakman, 1997). However, concerns over the impact of the method continue, particularly as there does appear to be a less invasive approach available compared to the standard method — called the ‘single sample’ method (not to be confused with the single pool model), which was originally developed by Williams and Nagy (1984). Normally in the classical application of the method an animal is captured, injected and held for an equilibration period to allow isotopes to completely equilibrate with the water pool before being bled, and is then released back to the wild before a second capture and blood sample. In the single sample method the animal is captured and injected but then released immediately. The initial enrichment is estimated from the weight of the injection and weight of the animal and an equation derived from a sample of animals where the initial sample had been collected. Because the initial value in the single sample method is a prediction rather than a measurement, the single sample method gives more variable results. In streaked shearwaters (Calonectris leucomelas) it was found the single sample method gave an estimate on average 12% different (+/−12.1%) from the two-sample method, when applied to the same individuals i.e. calculated by using the measured or predicted initial enrichment for each bird (Nizuma and Shirai, 2015). Hence the estimates may be more variable, but if they involve less disturbance, maybe they are more valuable. Schultner et al. (2010) evaluated this question in black-legged kittiwakes (Rissa tridactyla) and came to the conclusion that the initial
blood sample (or holding the bird in captivity for a period) does disturb the animals flight and feeding behaviour when raising chicks, and that using the single sample method is therefore preferable. Our own feeling is that the reduced precision with the single sample method indicates that wherever possible the classical two sample method should be used — which you need to do anyway on a sample of animals when using the single sample approach, to get the prediction equation for the initial enrichments. If observations suggest the initial holding and bleeding do cause significant behavioural disturbance (as in Schultner et al., 2010), then it may be possible to alter the field application protocol to account for this. For example, in the kittiwakes the birds stop feeding and flying for several hours after being blood sampled. However, they naturally stop flying and feeding at night. By timing the capture and blood sampling at dusk one could ensure that the disturbance coincided with a period when the animals would be inactive anyway and hence minimise its impact.

It is also important to note that the behavioural impact of the method is sometimes complex, and probably species specific. For example, in great tits (Parus major) raising nestlings, we found that the application of the method had no impact on the feeding behaviour of the animal being measured, when compared to its behaviour the day prior to the DLW measurement day. But the feeding behaviour of the partner of the measured individual (who wasn’t being measured by DLW) was significantly impacted (Hinsley et al., 2011). This might for example have occurred because when the target bird was being held for isotope equilibrium for an hour, the partner was stimulated to elevate their delivery rate of food to the nestlings to compensate.

Reducing disturbance caused by the method is nevertheless always preferable, particularly when studying endangered species. One potential approach to this is to try and dispense altogether with blood sampling, once the initial sample has been taken, but rely instead on collecting the sample of body water from some other bodily fluid. The most obvious sources are urine and faeces (Williams et al., 1997; Speakman, 1997). Less obvious sources are breath (Krol and Speakman, 1999; Mitchell et al., 2015) and saliva (Speakman, 1997; Guidotti et al., 2015). These different sources have their own particular problems. For urine, for example, it seems that once water enters the bladder it is no longer in isotopic exchange with the rest of the body. Hence the bladder acts as a storage vessel integrating the isotope enrichment over the period since the previous urination. This can be further complicated if the bladder is not emptied completely on each urination. Hence assigning a time to the sample involves an error. Despite these problems, in humans, urine is the routine sample of choice. This is only possible, however, because one is able to give human subjects precise instructions on when to urinate and which sample to collect. The normal advice for sample collection is to empty the bladder completely on the first void of the day, and then collect the sample from the second void about 30 min later. This way the time of the sample is known to an accuracy of about 30 min. Even so failure to completely empty the bladder can be an issue in other subjects. Clearly issuing such instructions to animals is unfeasible. For breath sampling the main issue is that breath is highly fractionated and our knowledge of these fractionation processes is incomplete. For saliva sampling the animal needs to be captured and held, so its advantages over blood are negligible. Hence, probably the most feasible completely non-invasive sample source in wild animals is faeces (Williams et al., 1997). The idea for such non-invasive field studies of animals would then be that you follow the animal and observe it until it defecates, and then recover this sample from the field as soon as possible after its production. This approach has been championed by our colleague Michael Scantlebury, and we have tried to use it in several different circumstances. To date, however, we have only been successful in applying it to the cheetah (Acinonyx jubatus) (Scantlebury et al., 2014). The key for success in this approach may be doing this in a dry habitat where the potential for faecal contamination between production and collection is minimised (see also the early use of this approach by Williams et al., 1997).

These issues emphasise that the best and most reliable sample of body water is blood. Is it possible to collect blood non-invasively? The answer seems to be yes. Researchers have used highly innovative sampling strategies based on blood sucking insects to collect ‘stress free’ blood samples. This technique was developed by Cristian Voigt (Voigt et al., 2003), but although this approach has been known for nearly 2 decades, and is established to be less stressful (Voigt et al., 2013) it has not been widely adopted. However, see Bauch et al. (2013) for an application of this approach that doesn’t even require capturing the subject animal, because the bugs are hidden in a bird’s nest in a dummy egg! Finally, other researchers have employed novel approaches to dose the animals by injecting the dose into a food item placed out in the field that is then consumed by the target animal. This will obviously be easier for some species than others — hummingbird feeders provide an example of an easy to use possible route for quantitative non-invasive dosing.

With some imagination therefore one could envisage a completely non-invasive application of the method involving dosing the target individual via a food item and getting the follow-up sample(s) from faecal deposits (or strategically located bugs), accepting some loss of precision because these will not be optimally timed. As far as we are aware this has not been successfully achieved yet, and the potential for failure may make it a very hit and miss strategy. Potential failure points are legion: wrong target animal eats the dose, dose is incompletely consumed or fractionates while waiting to be consumed, dose not consumed at all and therefore costly financial loss, weight of the dose animal is unknown so may be given too much or too little isotope, good quality uncontaminated faecal samples cannot be recovered, etc. Once achieved it will be important to know if the derived measurements are substantially divergent from the more traditional application of the method, and hence if the extra effort, elevated risk of failure and reduced precision are worth it.

3.2. Laser spectroscopy

The workhorse of isotope studies for the last 60 years has been the mass spectrometer, developed by Neir in the 1940s (Nier, 1947), who was incidentally a co-author on the 1949 paper by Lifson and colleagues that discovered the isotopic exchange of 18Oxygen between water and respiratory CO2. Over the last decade, however, there has been a revolution in the field of isotope geochemistry (where measurements are made at background levels of enrichment), and that has been the replacement of mass spectrometry by laser technology. These technologies are now starting to migrate into the medical and ecological fields where the interest is in tracer levels of measurement. The technology is based on the principle that different chemicals absorb light at very precise frequencies. Hence using finely tuned lasers it is possible to measure the levels of different isotopes from the absorption spectrum of light over a very narrow wavelength range. There are currently two slightly different technologies that exploit the same principle — ‘cavity ring-down spectroscopy’ used by machines produced by the Picarro company, and ‘off-axis integrated cavity output spectroscopy’ (Baer et al., 2002) used by machines produced by the Los Gatos company (now part of the ABB group). There have been several attempts over the past 5 years to use these different machines to apply the DLW method in humans (Thorsen et al., 2011). Our own group has been consulting and collaborating with Los Gatos to achieve this. We first showed that the off axis machine is capable of measuring standard enriched waters and also urine samples at background levels and spiked with isotopes (Berman et al., 2013). Most recently we have been involved in an indirect calorimetry chamber validation study where the performance of the laser machine is compared directly with mass spectrometry, and estimated energy expenditure derived by DLW from both isotope methodologies is compared to indirect calorimetry in humans. The work is currently (Jan 2016) still in progress, but initial data from 12 subjects suggests that the laser machine is comparable in terms of accuracy and
precision to the mass spectrometer method (Melanson et al., 2015). It is worth noting that long before these recent laser spectroscopy applications, Eric Kerstel at the University of Groningen built a customised laser spectroscopy device (Kerstel et al., 1999) and validated its use for DLW studies in birds (van Trigt et al., 2001). Moreover in the 1970s Ken Nagy developed, validated (Nagy, 1980) and used an alternative method to measure \(^{18}\)O enrichments using cyclotron based activation (Wood et al., 1975). In both cases, however, the methods did not spread to other laboratories because the devices were not commercially available.

These studies bode well for the future of the method because laser technology has several advantages over mass spectrometry. First, it does not require high performance vacuum technology, liquid nitrogen, or pure noble gases for sample transfer, and is hence cheaper to operate than standard mass spectrometry. Second, our work has suggested that samples may need minimal pre-processing before analysis (Berman et al., 2013). Third, the laser machines are very much plug and play type devices compared to mass-spectrometers that require highly specialised and trained individuals to operate them effectively. Fourth, the machines can measure both isotopes in a single sample. This potentially increases throughput and it reduces the sample volume required to make a measurement. This might contribute to reducing the invasive-ness of the procedure, and could extend downwards the size of the animals that it is feasible to measure. At present the smallest animals measured using the DLW technique were small groups of Drosophila melanogaster (Piper et al., 2014) individually weighing about 1 mg. Fifth, the laser machines are cheaper than mass spectrometers (although at 80,000 US$ probably beyond the range of most labs wanting to make single or a small number of applications of the method). Sixth, it may be possible to set them up in the field (Berman et al., 2009), enabling rapid turnaround in data acquisition. Indeed with machines based out in the field this may open up an additional non-invasive sampling methodology, by collecting breath samples directly off animals after capture (Mitchell et al., 2015). Although if the primary stress on the animal is due to the capture and confinement rather than the blood sample the ‘non-invasiveness’ of this approach is dubious. The main downsides are that the machines have very bad memory effects. That is there is a carryover from one measurement to the next. This necessitates multiple sample measurements until the readings reach stability, and this offsets the potential time gains that accrue due to both isotopes being measured in one sample. Plus they need more copious analysis of reference samples to ensure reliability and comparability of analyses. Nevertheless, the fact that isotope geology and geochemistry now rely almost exclusively on this technology indicates that the days of the mass spectrometer in DLW and other medical applications are probably numbered.

The extent to which laser technology will simply replace mass spectrometry or truly open up new possibilities remains uncertain. We detail as follows one such potential impact.

One thing that we have not mentioned so far is the fact that the stable isotopes of \(^{18}\)Oxygen and deuterium are naturally occurring. In your body about 2000 out of every million oxygen atoms is \(^{18}\)Oxygen and about 150 from every million hydrogen atoms is deuterium. Hence the exponential elimination of the isotope dose does not return to 0, but returns to the baseline isotope enrichment. So before we plot the logged enrichment to get the elimination line, we must subtract the background enrichment. Specifically for example for the oxygen elimination constant \(k_0\),

\[ k_0 = \frac{\log_{10}(E_t - E_{fb}) - \log_{10}(E_{fi} - E_{fb})}{t} \tag{6} \]

where \(E_t\) is the initial isotope enrichment after dosing, \(E_{fb}\) is the initial background isotope level, \(E_{fi}\) is the enrichment of \(^{18}\)Oxygen when the animal is recaptured (final) and \(E_{fb}\) is the background level at the final sample, and \(t\) is the time between the initial and final samples.

A problem is that when we take the final sample we don’t know what the background is (i.e. \(E_{fb}\) is unknown) and indeed it cannot be measured because we injected the dose and some of the dose is necessarily left at the end masking any change in the background. The only available solution at the moment is to assume that it is equal to the initial background (\(E_{fb}\)), which is normally measured on a sample taken before the animal is dosed. However, the background isotope enrichment is not constant (e.g. Horvitz and Schoeller, 2001; Harbison et al., 2015: Fig. 1), so this assumption introduces an error (Ritz et al., 1996).

For small animals this is not a serious issue. First, the dose is relatively large, so even after 48 h the remaining isotope is generally 100s of ppm higher than the background — and therefore a 1 ppm error in the assumed final background (\(E_{fb}\)) will introduce an error of maximally 1%. Plus because the elapsed time (\(t\)) is short, the chance the background will have changed so much is remote. Indeed this robustness to background issues means that for small animals it is often sufficient to estimate background levels by measuring a group of animals that have not been dosed and using that estimate for both \(E_{fb}\) and \(E_{fi}\) (method C: Speakman and Racey, 1987), or even a sample of the local drinking water (method B: Speakman and Racey, 1987).

In larger animals, however, this becomes a more serious problem. First, because of the high costs of the \(^{18}\)Oxygen isotope the doses that are used as animals get larger are proportionately lower (Speakman, 1997). Hence, the initial enrichment is often only 100 or 150 ppm above background. If this is left to wash out over 2 to 3 half lives as recommended (Nagy, 1983) then the final enrichment ends up being only around 10–15 ppm above background. Now a background drift of 1 ppm introduces potentially a 7–10% error. Plus, since it takes maybe 10–20 days for this amount of elimination, a background change of at least 1 ppm is highly likely (Fig. 1). To minimise this error one could shorten the measurement — but this is a flawed approach because as the measurement is shortened another error is introduced; the inability to distinguish divergence of the labels. It has been elegantly shown that this background drift error can be mitigated by clever manipulation of the dosing ratio of the isotopes (Horvitz and Schoeller, 2001) and is minimised by roughly putting into the dose solution about twice as much \(^{18}\)Oxygen as deuterium. This mitigates the problem, but doesn’t remove it. However the laser technology opens up a potential novel approach to solve this problem.

Eleven years ago we outlined a potential solution to the ‘background drift’ issue (Speakman, 2005). The basis of this potential solution is that when isotope enrichments change they do so in synchrony, because the processes that affect background levels of deuterium (like evaporation

![Fig. 1. Example time track of background isotope enrichments of deuterium. \(^{18}\)Oxygen and \(^{18}\)Oxygen in a single human subject tracked daily for 30 days. The data are presented in delta units per mil. To translate these into ppm 10 in deuterium is equal to 0.155 ppm and 10 in \(^{18}\)Oxygen is equal to 2.0 ppm. Two things are apparent from this figure. First, the isotopes track each other very closely. Second, there can be quite large changes over relatively short time periods. For example between day 12 and day 15 the \(^{18}\)Oxygen background changed by 1.5 ppm (see also Berman et al., 2015 for similar tracks in other individuals).](image-url)
and fractionation) are broadly the same as those driving background levels of 18-oxygen. So if we know the change in one, we can predict the change in the other. This is no help of course because we don’t know either of them because the dose contains both labels. However, if there was a third isotope affected by the same processes, that had not been artificially enriched, we could use changes in it to predict the changes of the other 2 isotopes. In fact there is a third isotope — 17-oxygen. The suggestion (Speakman, 2005) was that tracking levels of 17-oxygen might allow us to predict background drift in 18-oxygen and deuterium. This was never adopted (even by ourselves) because measuring 17-oxygen by mass spectrometry is difficult, because the signal in CO2 (the molecule in which oxygen is normally measured by mass spectrometry) is clouded by variation in 16-carbon levels. However, the laser method opens up this opportunity because measuring 17-oxygen alongside 18-oxygen and deuterium by laser spectroscopy is relatively easy. In fact commercial machines already exist that do this in the field of isotope geochemistry. In collaboration with Los Gatos, and Ed Melanson (University of Colorado), we have been using such machines to track the simultaneous patterns of background variation in all three isotopes in a population of adult humans living in Denver, Colorado, USA, and the results look extremely promising in terms of machines to track the simultaneous patterns of background variation.

References


