Journal of Agricultural Science, Cambridge (1998), **131**, 465–476. *Printed in the United Kingdom*

The sensitivity of n-alkane analysis to measurement error: implications for use in the study of diet composition

J. A. NEWMAN[1]*, F. CRIBARI-NETO[2] AND M. J. JENSEN[3]

[1] *Department of Zoology, Southern Illinois University, Carbondale IL 62901-6501, USA*
[2] *Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brazil*
[3] *Department of Economics, University of Missouri, Columbia MO 65211, USA*

(*Revised MS received 5 May 1998*)

SUMMARY

It is possible to estimate diet composition from an analysis of n-alkanes in the faeces of ruminant animals. The method requires the estimation of the concentrations of n-alkanes in the plants and faeces and then the solving of a system of simultaneous equations. There are at least three places in which significant measurement error may be introduced. First, there may be error in the determination of the concentrations of the n-alkanes in the herbage. This error may be the result of analytical error in the chemical analysis, or in the gathering of the representative sample of herbage. In either case, error in this estimate may be particularly important, since this estimate is not independently repeated for each animal in the study, but is conducted once and used throughout the study. Error may also be introduced in the estimates of digestibility of the n-alkanes themselves. The n-alkane method might be ideal if in fact the n-alkanes were completely indigestible – they are not and, furthermore, they are differentially digestible. Lastly, there may be measurement error in the estimate of the n-alkane concentrations in the faeces, which utilize the same analytical procedures that are used on the herbage. That is, if measurement error exists in the herbage estimates, it is quite possible that it also exists in the faeces estimates. We address these issues through the use of Monte Carlo simulation to investigate the likely effects of measurement error on diet composition and digestibility estimates obtained using the n-alkane method. Our results suggest the following conclusions: (1) in the face of any sort of measurement error, estimates of digestibility are likely to be unreliable; (2) when measurement error exists, one of the diet components will usually be under-estimated and the other will usually be over-estimated; (3) any sort of progressive bias in the n-alkane recovery estimates will probably have large and very significant effects on the results; and (4) if measurement error in the estimates of the n-alkane concentrations in the herbage and in the faeces are similar in expectation, then their effects tend to cancel each other out.

INTRODUCTION

Researchers have been using n-alkanes, variable length carbon chains found in the wax cuticle of plants, as markers in digestibility and diet selection studies on grazing animals for over 10 years. Since Dove & Mayes' (1991) excellent review article, in which they pointed out the advantages of the use of plant wax alkanes to estimate diet selection, many researchers have adopted this method (e.g. Dove 1992; Malossini *et al*. 1994; Salt *et al*. 1994; Vulich & Hanrahan 1995; Vulich *et al*. 1995).

There are actually two different aspects to the analysis. The first involves the methodology of taking samples, preparing samples and analysing the samples for n-alkane concentrations. The second aspect involves the use of the information on n-alkane concentrations to construct estimates of diet composition and diet digestibility. At least four methods have been used. Dove (1992) used simultaneous equations. Newman *et al*. (1995) suggested the use of a technique for solving an over-determined system of equations to improve on Dove's method. Using this method, more information can be utilized to construct

* To whom all correspondence should be addressed. Email: newman@zoology.siu.edu

the diet estimate. Salt *et al.* (1994) used a complicated non-linear quasi-Newtonian method to form diet estimates. Finally, Dove & Moore (1996) used an iterative method called non-negative least squares.

It is easily demonstrated (Draper & Smith 1981) that the method discussed by Newman *et al.* (1995) is always the least squares estimate, and if the error is normally distributed it is the maximum likelihood estimate as well. Because of the simplicity of this method, its mathematically desirable properties and its ease of implementation, we concentrate on this method here. How bias might influence the quasi-Newtonian approach of Salt *et al.* (1994) or the iterative method of Dove & Moore (1996) is not known, but we suspect that these other methods would give similar results. For example, the method of Dove & Moore (1996) relies on an iterative algorithm which performs least squares estimations and, as we show below, this type of estimation is quite sensitive to measurement errors. Further research in this area is warranted.

There are two places in which errors may enter the estimated solutions: in the herbage n-alkane concentrations and in the estimates of the faecal recoveries. Errors in the estimates of herbage n-alkane concentrations have two main sources. The first is analytical error; that is, error in the chemical analysis that leads to the estimates of concentrations of each n-alkane in each plant species. The second source of error is in sampling the plant material. Dove & Mayes (1991), Dove (1992) and Newman *et al.* (1995) have pointed out that it is critical that the plant sample be characteristic of the plant species being consumed. This may be difficult to achieve for a number of reasons. We know, from Laredo *et al.*'s (1991) work, that alkane concentrations vary between different plant parts. More generally, due to the inherent spatial variation in the distribution and abundance of plant species in both the pasture in general and specifically in the grazed horizon, pastures exhibit considerable variation in n-alkane concentrations (Malossini *et al.* 1994). Errors in herbage n-alkane concentrations may be one of the most important classes of errors since the same matrix is used for all animals in the study (which is also true of faecal recoveries).

Errors in the faecal recoverability estimates may also come from two sources. Like errors in herbage n-alkane concentrations, errors in faecal recoverability estimates may result from analytical errors involved in the chemical analysis of the n-alkane concentrations. The other source of error comes from the method itself. There are a number of different methods for determining faecal recoveries, but generally they are all based on the theory that adjacent even length n-alkanes have similar digestibilities to the odd-length chains of interest. Recovery rates are estimated by dosing the animal with an adjacent chain of even length and using this estimate for the recovery rate of the odd length chain. There may be systematic differences in the recovery rates for adjacent chain lengths, and recovery in general may be related to chain length. This relationship between chain length and recovery differs from study to study. Casson *et al.* (1990) and Mayes *et al.* (1986*a*, *b*) found increasing recoverability of n-alkanes as a function of chain length. However, Vulich *et al.* (1991) found no significant relationship (although there was much variance in the estimates and a downward trend in the data). In either case, it is clear that there is variability in the estimated recoveries that is ignored in the estimation of diet composition.

The faecal concentrations are taken to be the dependent variable in these analyses. They will always have an element of variability, quite apart from measurement error. Multiple estimates from individual animals are variable (e.g. Vulich *et al.* 1991; Malossini *et al.* 1994). Some of the variability in faecal concentrations probably comes from diurnal patterns of diet selection such as those reported by Coates *et al.* (1987), Newman *et al.* (1994) and Parsons *et al.* (1994).

How the naturally occurring variability in the faecal concentrations interacts with measurement errors in herbage concentrations and in the recoveries is the focus of this study. We quantify the effects of measurement error on the bias in the estimates of diet composition and diet digestibility using computer simulations.

BASIC PROBLEM

The matrix algebra needed to utilize the n-alkane concentrations to predict diet composition and digestibilities is contained in Newman *et al.* (1995). Throughout this example, we will use the concentrations of five different n-alkanes to estimate the diet comprising two herbage species. Throughout the paper, matrices are shown in bold type.

Diet composition is found using the following formula:

$$\mathbf{H}d = \mathbf{R}f \tag{1}$$

where $\mathbf{H}$ is a 5×2 matrix of the five n-alkane concentrations in the two herbage species, $\mathbf{R}$ is a 5×5 diagonal matrix where the entries are the inverse of the digestibilities for each of the five n-alkanes, f is a 5×1 vector of the concentrations of each of the five n-alkanes in the animal's faeces, and $d = (d_1, d_2)'$ is a 2×1 vector of the unknown quantities of each herbage species.

The least squares estimate of d, $\hat{d}$, is used to estimate the proportion of each herbage species in the diet either using:

$$\hat{\rho}_1 = \frac{\hat{d}_1}{\hat{d}_1 + \hat{d}_2} \quad \text{or} \quad \hat{\rho}_2 = 1 - \frac{\hat{d}_1}{\hat{d}_1 + \hat{d}_2} \tag{2}$$

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

Table 1. *Values used in simulations and results for the unbiased problem using* 10 *animals for each experiment. Values in parentheses are the estimated variances*

Parameters	Case 1	Case 2
d_1	1·75	0·35
$\hat{d}_1$	1·758	0·360
	$(3{\cdot}4\times10^{-5})$	$(3{\cdot}4\times10^{-5})$
d_2	1·75	3·15
$\hat{d}_2$	1·729	3·125
	$(7{\cdot}6\times10^{-5})$	$(7{\cdot}6\times10^{-5})$
ρ_1	0·5	0·1
$\hat{\rho}_1$	0·504	0·103
	$(3{\cdot}93\times10^{-6})$	$(3{\cdot}93\times10^{-6})$
ρ_2	0·5	0·9
$\hat{\rho}_2$	0·496	0·897
	$(3{\cdot}93\times10^{-6})$	$(3{\cdot}93\times10^{-6})$
ϕ	0·714	0·714
$\hat{\phi}$	0·713	0·713
	$(1{\cdot}9\times10^{-7})$	$(1{\cdot}9\times10^{-7})$

Throughout we will refer to $\hat{\rho}_1$ as ratio 1 and $\hat{\rho}_2$ as ratio 2. $\hat{d}$ is also used to estimate digestibility as:

$$\hat{\phi} = 1 - \frac{1}{\hat{d}_1 + \hat{d}_2} \tag{3}$$

Sensitivity analysis

We simulate the calculations implied in Eqn (1) for a hypothetical experiment in which there are 10 animals and the experiment is replicated 10000 times. For each of the 10000 trials, the matrices **H** and **R** are the same for all animals, but a new realization of f is generated for each animal. Throughout, we consider f to be a random variable:

$$\hat{f} = f + \epsilon_f \tag{4}$$

where $\hat{f}$ is the vector of observed values, f is the vector of true values, and ϵ_f is a vector of random errors, where each element is a zero mean normal random variable. Thus, Eqn (1) is actually

$$\mathbf{H}d = \mathbf{R}\hat{f} \tag{5}$$

Throughout, we will develop two cases, these values are shown in Table 1.

It is important to note that, although the effects of measurement errors on $\hat{d}$ are well known for Eqn (5), the effects of measurement error on $\hat{\rho}$ are not well known. Indeed, it is not even clear that the expectations of $\hat{\rho}_i$ are finite. To see this, let us simplify the problem for the purpose of illustration. Suppose that $\check{\rho} = \hat{d}_1/\hat{d}_2$. Assuming that the $\hat{d}_i$ are independent standard normal random variables, then it can be shown that $\check{\rho}$ is a Cauchy random variable (e.g. DeGroot 1986) which does not have a well defined mean and has infinite variance. Our problem with $\hat{\rho}$ is slightly more complex than this example, since the numerator and denominator may not be independent; but the essence of the argument is the same. Therefore the use of Monte Carlo simulations is justified for investigating the effects of measurement error on estimates of diet composition.

RESULTS

Variance in $\hat{f}_i$

When the only source of variance in Eqn (5) is in the elements $\hat{f}_i$, then Eqn (5) is a very reliable predictor; the results for both cases are shown in Table 1. We see that the method very reliably estimates the dietary proportions and diet digestibility.

Measurement error in $\mathbf{H}_{ij}$

We introduce normal error to the measurement of the elements $\mathbf{H}_{ij}$, as:

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}_{\mathbf{H}} \tag{6}$$

where

$$\mathrm{E}[\mathbf{E}_{\mathbf{H}}] = \alpha\mathbf{H} = \boldsymbol{\mu}_{\mathbf{H}} \tag{7}$$

$\mathbf{E}_{\mathbf{H}}$ is a matrix of error values added to the elements of **H**. Each element of $\mathbf{E}_{\mathbf{H}}$, say $\mathbf{E}_{\mathbf{H}}^{ij}$, is normally distributed with mean $\boldsymbol{\mu}_{\mathbf{H}}^{ij}$ and variance ${}_{\mathbf{H}}\Omega_{ij}^2$. α is the amount of contamination in the estimate. Positive or negative values of $\boldsymbol{\mu}_{\mathbf{H}}^{ij}$ imply that $\hat{\mathbf{H}}_{ij}$ tends to over- or under-estimate the values of $\mathbf{H}_{ij}$.

Now, we modify Eqn (5) as

$$\hat{\mathbf{H}}d = \mathbf{R}\hat{f} \tag{8}$$

and use Eqn (8) to estimate the effects of the distribution of $\mathbf{E}_{\mathbf{H}}^{ij}$ on the estimates of diet composition and digestibility. Figures 1*a* and 1*b* show the percentage relative bias of these estimates plotted against the percentage contamination. Percentage relative bias is defined as

$$\frac{\hat{d}_i - d_i}{d_i} \times 100 \quad \text{for} \quad i = 1, 2 \tag{9}$$

where $\hat{d} = (\hat{d}_1, \hat{d}_2)'$ are the values of the diet composition estimated by Eqn (8), and d is the vector of the true values of the diet composition. Percentage contamination refers to the value of $\boldsymbol{\mu}_{\mathbf{H}}^{ij}$. A value of $-20\,\%$ contamination would imply that if the true value were $\mathbf{H}_{ij} = 10$, then the mean estimated value would be $\hat{\mathbf{H}}_{ij} = 8$.

Table 1 shows the true values for case 1 and case 2. Figure 1*a* clearly shows that although the elements of $\hat{d}$ can become very biased in the face of measurement error in **H**, the estimates of ratio 1 ($\hat{\rho}_1$ from Eqn (2)) and ratio 2 ($\hat{\rho}_2$ from Eqn (2)) remain fairly reliable. However, when the true proportion is different from 1:1, then the estimate of diet proportion becomes more biased, with the effect becoming larger the

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

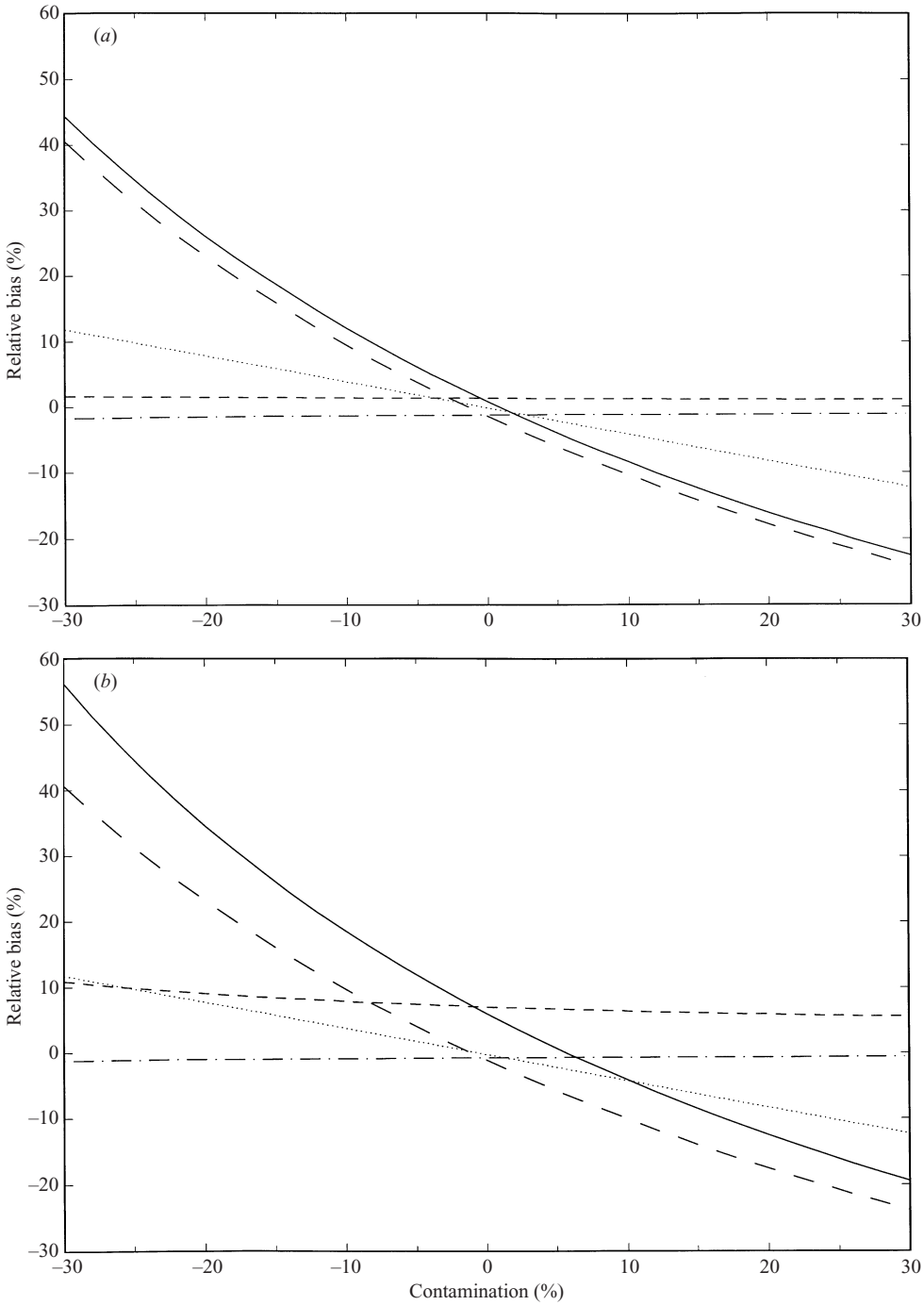

Fig. 1. Measurement error in the herbage concentrations of n-alkanes for (*a*) case 1 and (*b*) case 2. The cases are defined in Table 1. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (- - -), ratio 2 by (–·–) and digestibility by (···).

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

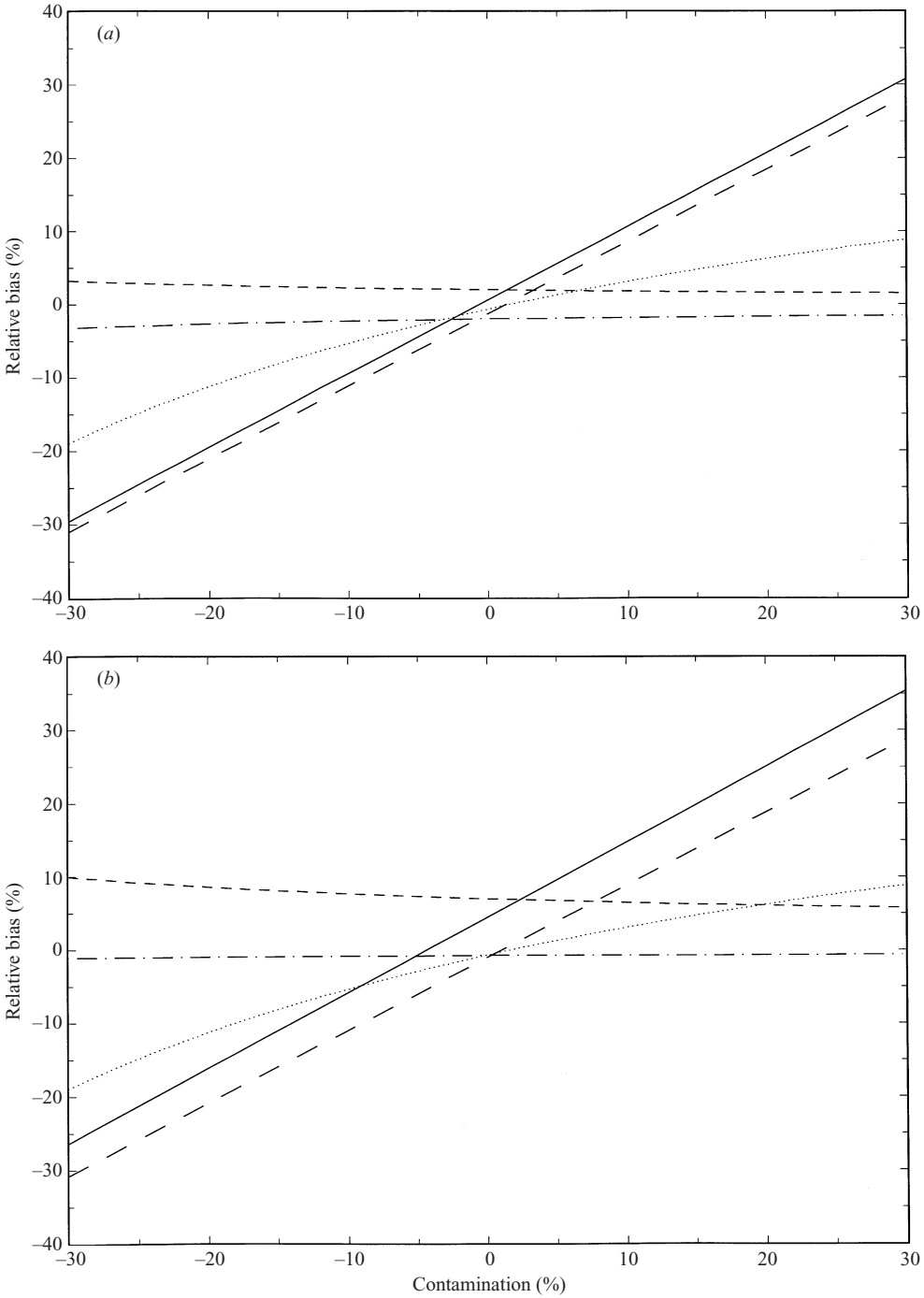

Fig. 2. Measurement error in the n-alkane faecal recoverability estimates for (*a*) case 1 and (*b*) case 2. The cases are defined in Table 1. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (---), ratio 2 by (–·–) and digestibility by (···).

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

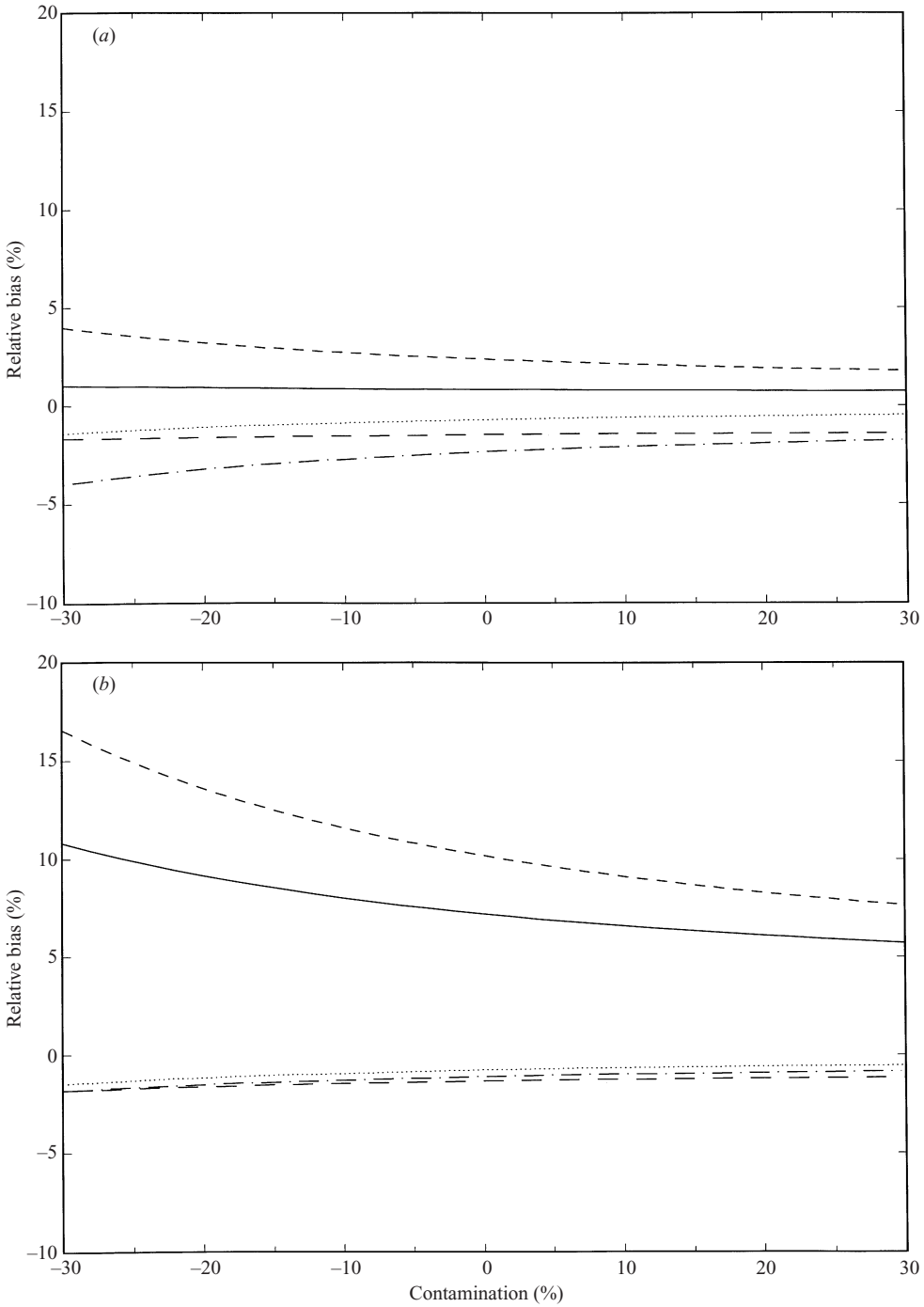

Fig. 3. Simultaneous error in the herbage concentrations of n-alkanes, and the faecal recoveries of those n-alkanes for (*a*) case 1 and (*b*) case 2. The cases are defined in Table 1. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (---), ratio 2 by (–·–) and digestibility by (···).

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

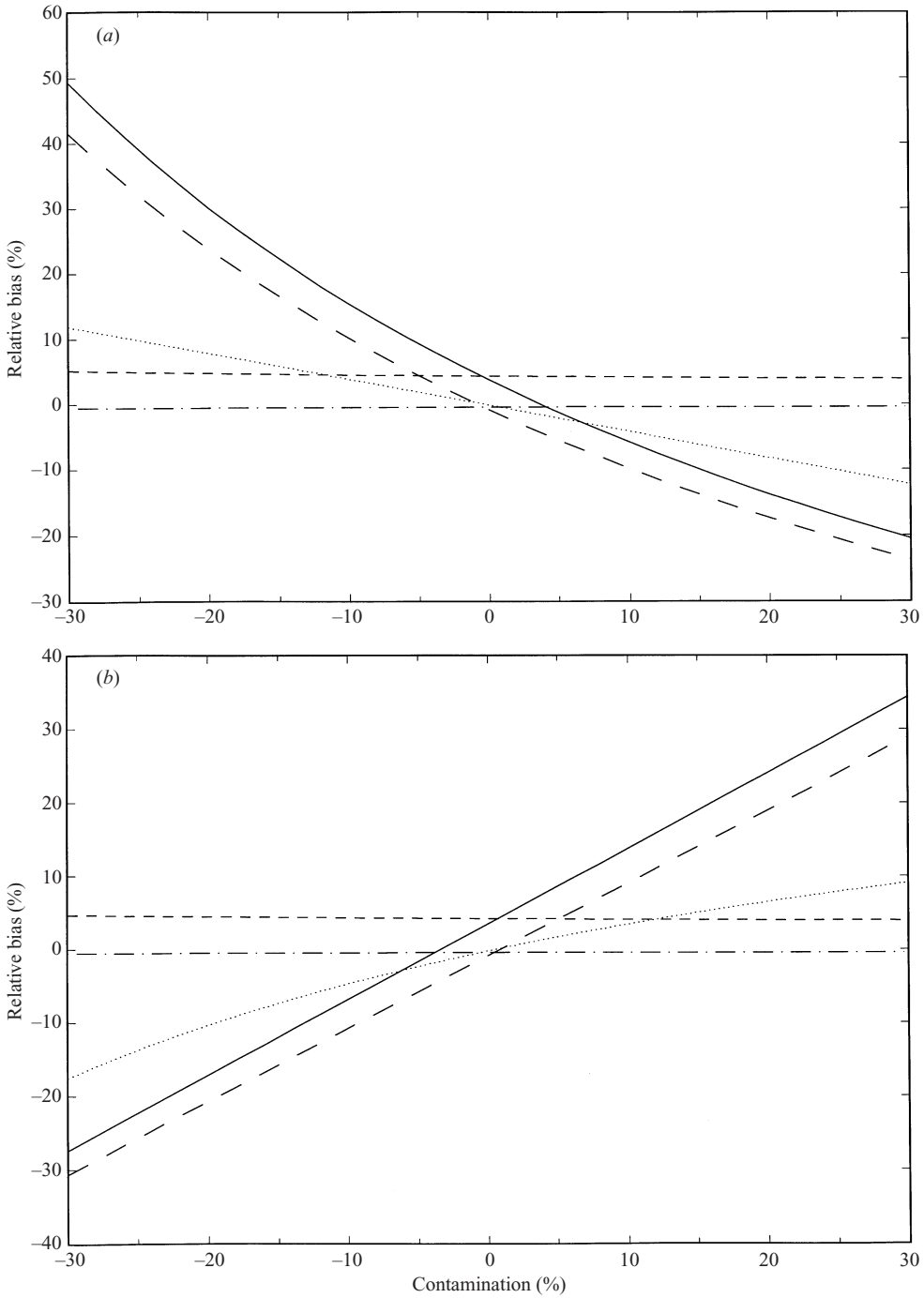

Fig. 4. Measurement error in (*a*) the herbage concentrations and (*b*) the faecal concentrations of n-alkanes for case 2; the effects of variance. The cases are defined in Table 1. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (---), ratio 2 by (–·–) and digestibility by (···).

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

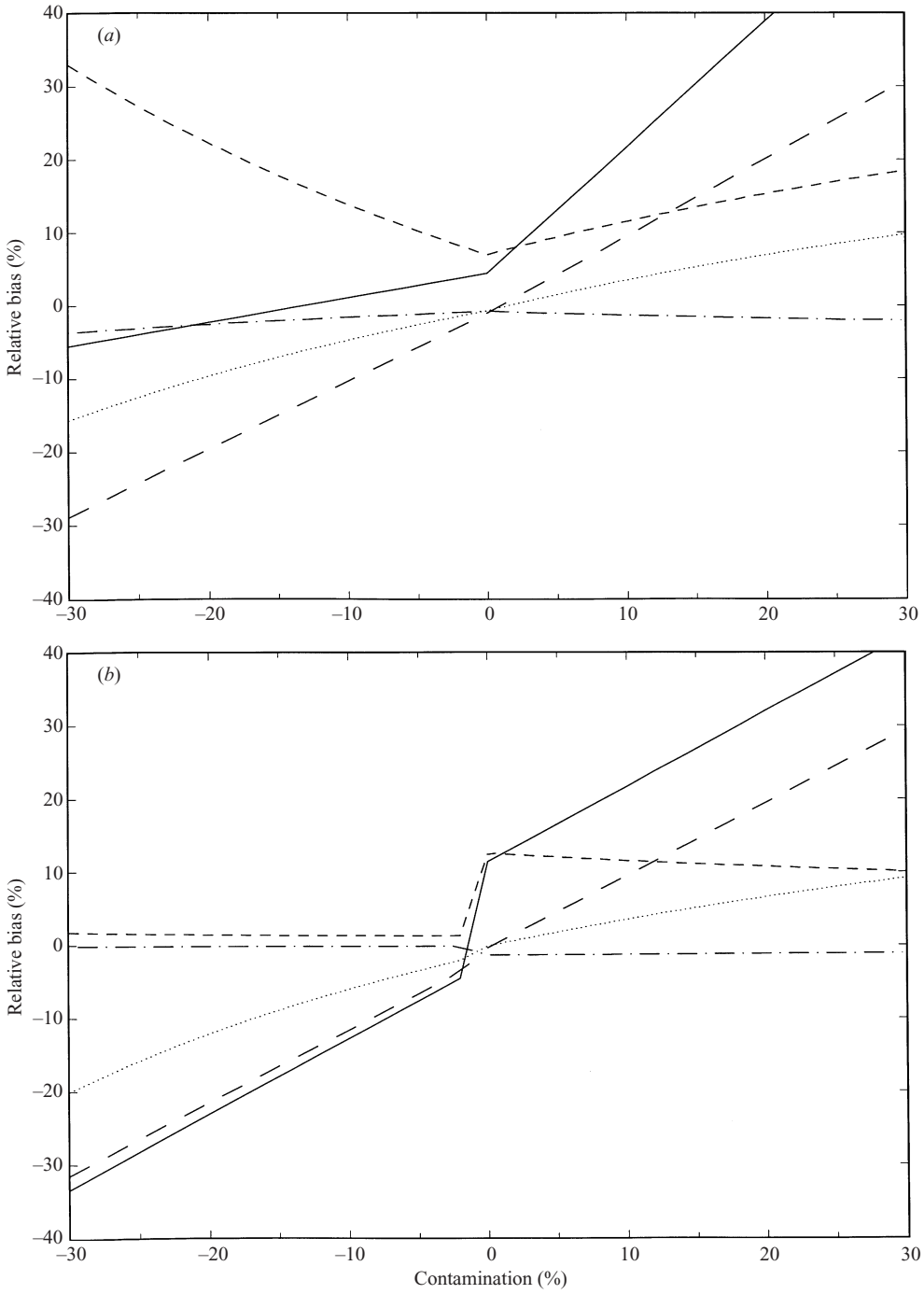

Fig. 5. Progressive measurement error in the faecal n-alkane recoveries, case 2; for (*a*) version I and (*b*) version II. The cases are defined in Table 1. See text for definition of versions. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (---), ratio 2 by (–·–) and digestibility by (···).

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

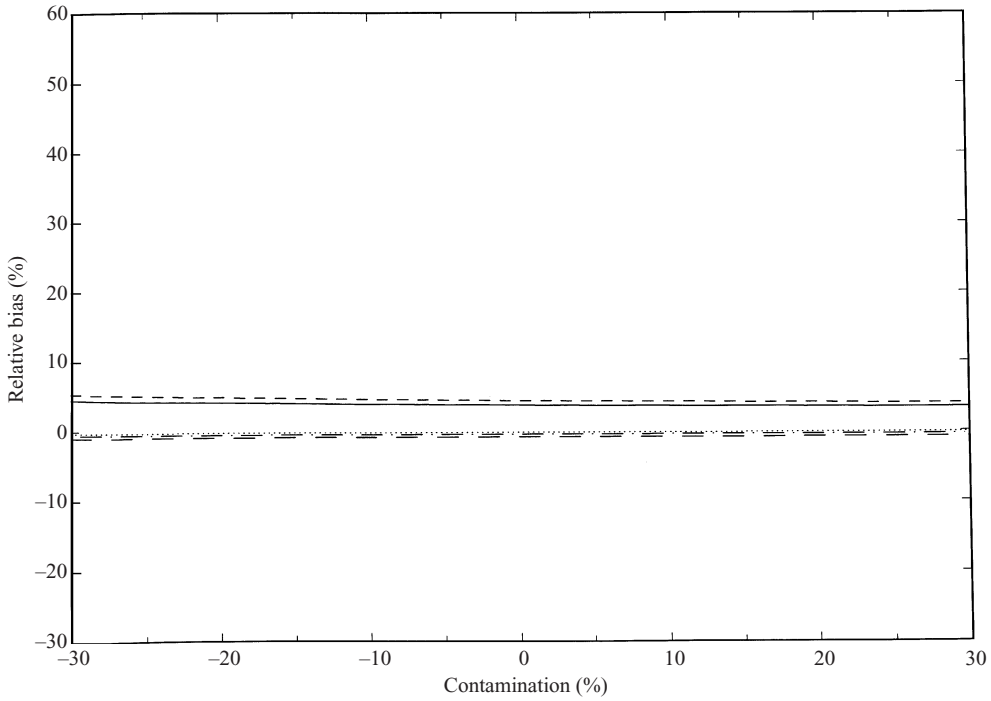

Fig. 6. Simultaneous measurement error in the herbage and faecal concentrations of n-alkanes, case 2. The cases are defined in Table 1. See text for definition of versions. This figure shows the % relative bias as a function of % contamination. $\hat{d}_1$ is the estimated quantity of species 1 in the diet, $\hat{d}_2$ is the estimated quantity of species 2 in the diet, ratio 1 is $\hat{\rho}_1$, ratio 2 is $\hat{\rho}_2$ and digestibility is calculated according to Eqn (3). $\hat{d}_1$ is denoted by (—), $\hat{d}_2$ by (– –), ratio 1 by (---), ratio 2 by (–·–) and digestibility by (···).

further the deviance from 1:1. This effect is seen in Fig. 1*b*. Figures 1*a* and 1*b* also show that the estimate $\hat{\phi}$ is subject to bias whenever the contamination differs from zero.

Measurement error in $\mathbf{R}_{ii}$

We introduce normal error to the measurement of the elements $\mathbf{R}_{ii}$ as:

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{E}_{\mathbf{R}} \tag{10}$$

where

$$\mathrm{E}[\mathbf{E}_{\mathbf{R}}] = \alpha\mathbf{R} = \boldsymbol{\mu}_{\mathbf{R}} \tag{11}$$

$\mathbf{E}_{\mathbf{R}}$ is a 5×5 diagonal matrix of error values added to the elements $\mathbf{R}_{ii}$. In Eqn (11) α is the amount of contamination. The elements $\mathbf{E}_{\mathbf{R}}^{ii}$ are distributed as normal random variables with mean $\boldsymbol{\mu}_{\mathbf{R}}^{ii}$ and variance ${}_{\mathbf{R}}\Omega_{ii}^2 = 0{\cdot}04$. Positive or negative values of $\boldsymbol{\mu}_{\mathbf{R}}^{ii}$ imply that $\hat{\mathbf{R}}_{ii}$ tends to over- or under-estimate the elements of $\mathbf{R}_{ii}$.

Now, we modify Eqn (5) as

$$\mathbf{H}d = \hat{\mathbf{R}}\hat{f} \tag{12}$$

Figures 2*a* and 2*b* show the relative bias for cases 1 and 2, respectively. Once again, the estimates of digestibility seem to be unreliable, and the estimates of diet composition become more unreliable as the true fractional diet composition differs from 1:1.

Simultaneous measurement error in $\mathbf{H}_{ij}$ *and* $\mathbf{R}_{ii}$

Here we combine the previous two sections and modify Eqn (5) as

$$\hat{\mathbf{H}}d = \hat{\mathbf{R}}\hat{f} \tag{13}$$

Figures 3*a* and 3*b* show the relative biases for cases 1 and 2, respectively. Notice that simultaneous error tends to lower the bias in the estimates $\hat{d}$, but increases the bias in the estimates of $\hat{\rho}_i$. In this case, the estimates of digestibility tend to be more reliable.

Effects of variance in measurement error in $\mathbf{H}_{ij}$ *and* $\mathbf{R}_{ii}$

Here we consider, for case 2 separately, the effect of variance in the measurement errors of $\mathbf{H}_{ij}$ and $\mathbf{R}_{ii}$, respectively. That is, we use ${}_{\mathbf{H}}\Omega_{ij}^2 = \mathbf{H}_{ij}/4$ instead of ${}_{\mathbf{H}}\Omega_{ij}^2 = \mathbf{H}_{ij}$, and ${}_{\mathbf{R}}\Omega_{ii}^2 = 0{\cdot}01$ instead of ${}_{\mathbf{R}}\Omega_{ii}^2 = 0{\cdot}04$. Figure 4*a* shows the results of the former (cf. Fig. 1*b*)

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

and Fig. 4*b* shows the results of the latter (cf. Fig. 2*b*). These two figures show that decreasing the variance in the measurement error tends to decrease the relative bias. In other words, this shows that increasing the precision in measuring n-alkane concentrations decreases the relative bias.

Effects of progressive measurement error in $\mathbf{R}_{ii}$

If recoveries are related to chain length, then contamination in the elements of $\mathbf{R}_{ii}$ will get bigger or smaller as the chain length increases. We develop two possibilities. In the first we add 1% contamination for each increase in chain length; so if $\mathbf{R}_{11} = 10\%$, then $\mathbf{R}_{22} = 11\%$ and so on until $\mathbf{R}_{55} = 14\%$. In the second possibility, we increase the contamination by 10% for each alkane; so if $\mathbf{R}_{11} = 30\%$, then $\mathbf{R}_{22} = 33\%$, and so on until $\mathbf{R}_{55} = 42\%$.

Figures 5*a* and 5*b* show the results for each possibility. We see that one estimate (ratio 2) of diet proportion remains reliable throughout, but the other estimate (ratio 1) of diet proportion becomes very unreliable. Note also that the estimate of digestibility continues to be unreliable.

Simultaneous measurement error in $\mathbf{H}_{ij}$ *and* f_i

It is possible that any bias introduced into the estimate $\hat{\mathbf{H}}_{ij}$ due to the analytical procedures will also be introduced into the estimates of f_i since the procedures are identical. Here we repeat case 2 for measurement error in $\mathbf{H}_{ij}$ (Fig. 1*b*), but this time we simulate the same measurement error in f_i as is present in $\mathbf{H}_{ij}$. That is, we repeat our analysis of Eqn (8), but where

$$\mathrm{E}[\epsilon_f] = \alpha f = \mu_{\epsilon_f} \tag{14}$$

The results are shown in Fig. 6. This figure suggests that as long as $\mu^i_{\epsilon_f}$ and $\mu^{ij}_{\mathbf{H}}$ are obtained as the same fractions of the true values (i.e. α is the same), the error tends to cancel out and the estimates of diet composition and digestibility are once again reliable. See the Appendix for a short proof of this result.

DISCUSSION

We chose to use a sample size of 10 animals for our simulated experiments. This is reasonable given the sample sizes used in published experiments. It is worth pointing out that as the sample size becomes large, the bias becomes small. So, since bias can be a problem with this technique, in the face of measurement error, using large sample sizes (> 30) would help reduce the bias. This is easily demonstrated by simulation but the results are not presented in the interest of brevity.

In the face of any sort of contamination, the estimate of diet digestibility tends to be unreliable. Given the current uncertainty associated with these measures, digestibility calculated using the alkane technique should be considered suspect.

When measurement error is present, ratio 1 and ratio 2 are differently biased. One of these, ratio 1 in our examples, will over-estimate the true proportion of species *i* in the diet, while the other, ratio 2 in our examples, will under-estimate the true proportion of species *j* in the diet. This effect becomes more pronounced the more the true proportions differ from 1:1 (cf. Figs 1*a* and 1*b* on 2*a* and 2*b*). Not only do the signs of the relative bias differ, but the magnitude of the relative bias differs as well. For example, in Figs 1*b* and 2*b*, ratio 2 yielded less relatively biased results. That is, while ratio 2 always produces negative relative bias, the absolute value of the relative bias is always less than the relative bias in ratio 1. There is, however, no general rule about which ratio will produce the more reliable result; it depends on the elements of the matrices. When the estimated proportions of each species in the diet differ greatly from 1:1, researchers should consider applying a bias correction technique such as the bootstrap (see comments at the end of this section).

We introduced bias as large as 30%. It is clear from our results that, in some cases, very small amounts of bias can have very large effects. It is usually impossible to establish the amount of bias in published research papers. In the light of our results, researchers should do two things routinely. First, they should take multiple samples of the herbage on offer and report the standard dispersion measures of these samples. Second, where possible, they should make multiple estimates of the alkane concentrations in a single sample of herbage or faeces and again report the standard dispersion measures of these estimates. Reports of these statistics could give the reader an indication of the potential magnitude of measurement error-induced bias.

Any sort of progressive bias in the measurements of recoveries is likely to have large and very significant effects on the results. Given the current techniques for estimating recoverability values, it is very possible that progressive bias is present. This probably represents the most significant drawback to alkane analysis. Until it can be satisfactorily demonstrated that the method of estimating recoveries is unbiased, or at least not progressively biased, results reported using this technique should be considered suspect.

The results of simultaneous bias in the herbage and faecal estimates are interesting. Figure 6 suggests that these two sources of bias cancel each other out (a result which we prove in the Appendix). It is important to point out two aspects of this result before we jump to the conclusion that this technique is safe and relatively bias-free. First, we note that the bias in the

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press

estimates of herbage and faecal concentrations will only be similar when the source of this bias is the analytical procedure itself. If there is bias in the gathering of the herbage sample, for reasons discussed by Newman *et al*. (1995), then it is unlikely that this bias will be compensated for by bias in the faecal samples. Secondly, we point out that the results shown in Fig. 6 do not consider error in the estimates of the recoveries. Indeed, these results suggest that error in the recoveries may be the most significant source of bias in the final estimates, since this will pervade the results when bias in the herbage and faeces cancel each other out.

Given the strong possibility of consistent under- or over-estimation in determining the elements of **H** and **R**, it seems that the application of a bias correction technique such as the bootstrap will be a worthwhile modification of the present statistical techniques (Efron & Tibshirani 1993).

We thank Southern Illinois University's Office of Research Development Administration for financial support. J. A. Newman also thanks Illinois State's Council for Food and Agricultural Research for financial support while conducting this research. F. Cribari-Neto gratefully acknowledges financial support from CNPq/Brazil.

REFERENCES

Casson, T., Rowe, J. B., Thorn, C. W. & Harris, D. (1990). The use of natural n-alkanes in medic and clover as indigestible markers. *Proceedings of the Australian Society of Animal Production* **18**, 462.

Coates, D. B., Schachenmann, P. & Jones, R. J. (1987). Reliability of extrusa samples collected from steers fistulated at the oesophagus to estimate the diet of resident animals in grazing experiments. *Australian Journal of Experimental Agriculture* **27**, 739–745.

DeGroot, M. H. (1986). *Probability and Statistics*, *2nd Edn*. Reading MA: Addison-Wesley Publishing Company.

Dove, H. (1992). Using n-alkanes of plant cuticular wax to estimate the species composition of herbage mixtures. *Australian Journal of Agricultural Research* **43**, 1711–1724.

Dove, H. & Mayes, R. W. (1991). The use of plant wax alkanes as marker substances in studies of the nutrition of herbivores: a review. *Australian Journal of Agricultural Research* **42**, 913–952.

Dove, H. & Moore, A. D. (1996). Using a least-squares optimization procedure to estimate botanical composition based on the alkanes of plant cuticular wax. *Australian Journal of Agricultural Research* **46**, 1535–1544.

Draper, N. R. & Smith, H. (1981). *Applied Regression Analysis*. New York: John Wiley & Sons.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Laredo, M. A., Simpson, G. D., Minson, D. J. & Orpin, C. G. (1991). The potential for using n-alkanes in tropical forages as a marker for the determination of dry matter intake by grazing ruminants. *Journal of Agricultural Science*, *Cambridge* **117**, 355–361.

Malossini, F., Bovolenta, S., Piasentier, E. & Valentinotti, M. (1994). Variability of n-alkane content in a natural pasture and in faeces of grazing dairy cows. *Animal Feed Science and Technology* **50**, 113–122.

Mayes, R. W., Lamb, C. S. & Colgrove, P. M. (1986*a*). The use of dosed and herbage n-alkanes as markers for the determination of herbage intake. *Journal of Agricultural Science*, *Cambridge* **107**, 161–170.

Mayes, R. W., Lamb, C. S. & Colgrove, P. M. (1986*b*). Determination of herbage intake of sucking lambs using long-chain n-alkanes as markers. *Animal Production* **42**, 457.

Newman, J. A., Parsons, A. J. & Harvey, A. (1992). Not all sheep prefer clover: diet selection revisited. *Journal of Agricultural Science*, *Cambridge* **119**, 275–283.

Newman, J. A., Penning, P. D., Parsons, A. J., Harvey, A. & Orr, R. J. (1994). Fasting affects intake behaviour and diet preference by grazing sheep. *Animal Behaviour* **47**, 185–193.

Newman, J. A., Thompson, W. A., Penning, P. D. & Mayes, R. W. (1995). Least-squares estimation of diet composition from n-alkanes in herbage and faeces using matrix mathematics. *Australian Journal of Agricultural Research* **46**, 793–805.

Parsons, A. J., Newman, J. A., Penning, P. D., Harvey, A. & Orr, R. J. (1994). Diet preference in sheep: effects of recent diet, physiological state and species abundance. *Journal of Animal Ecology* **63**, 465–478.

Salt, C. A., Mayes, R. W., Colgrove, P. M. & Lamb, C. S. (1994). The effects of season and diet composition on the radiocaesium intake by sheep grazing on heather moorland. *Journal of Applied Ecology* **31**, 125–136.

Vulich, S. A. & Hanrahan, J. P. (1995). Faecal sampling for the estimation of herbage intake using n-alkanes: evaluation of sample pooling and the use of rectal grab samples. *Journal of Agricultural Science*, *Cambridge* **124**, 79–86.

Vulich, S. A., O'Riordan, E. G. & Hanrahan, J. P. (1991). Use of n-alkanes for the estimation of herbage intake in sheep: accuracy and precision of the estimates. *Journal of Agricultural Science*, *Cambridge* **116**, 319–323.

Vulich, S. A., Hanrahan, J. P. & Crowley, B. A. (1995). Modification of the analytical procedures for the determination of herbage and faecal n-alkanes used in the estimation of herbage intake. *Journal of Agricultural Science*, *Cambridge* **124**, 71–77.

APPENDIX

Here, we show that when the error structures in $\mathbf{H}$ and $\hat{f}$ have the same structure, the errors tend to cancel out. We start by writing:

$$\mathbf{H}d = \mathbf{R}\hat{f}$$

and since $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}_{\mathbf{H}}$ and $\hat{f} = f + \epsilon_f$, we have

$$(\mathbf{H} + \mathbf{E}_{\mathbf{H}})\, d = \mathbf{R}(f + \epsilon_f)$$

Taking expectations and assuming that $\mathrm{E}[\mathbf{E}_{\mathbf{H}}] = \alpha\mathbf{H}$ and $\mathrm{E}[\epsilon_f] = \alpha f$, we obtain

$$\mathbf{H}d + \{\mathrm{E}[\mathbf{E}_{\mathbf{H}}]\}\, d = \mathbf{R}f + \mathbf{R}\mathrm{E}[\epsilon_f]$$
$$\mathbf{H}d + \alpha\mathbf{H}d = \mathbf{R}f + \mathbf{R}\alpha f$$
$$(1+\alpha)\,\mathbf{H}d = (1+\alpha)\,\mathbf{R}f$$
$$\mathbf{H}d = \mathbf{R}f$$

https://doi.org/10.1017/S0021859698005930 Published online by Cambridge University Press