Journal of Agricultural Science (2002), **138**, 425–434. © 2002 *Cambridge University Press*
DOI: 10.1017/S0021859602002046 *Printed in the United Kingdom*

Assessment of the use of n-alkanes as markers to describe the complex diets of herbivores

H. MARTINS[1,3]*, D. A. ELSTON[2], R. W. MAYES[1] AND J. A. MILNE[1]

[1] *Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK*
[2] *Biomathematics and Statistics Scotland, Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK*
[3] *Applied Ecology Centre Prof. Baeta Neves, Agronomy Institute of Lisbon, Tapada da Ajuda, 1349-018 Lisboa Codex, Portugal*

(*Revised MS received 3 January 2002*)

SUMMARY

Previous approaches to the description of complex diets, based on n-alkanes and optimization techniques, have grouped the plant species to reduce the number of components. Diet estimates have been obtained with least-squares routines by minimizing the discrepancy between faecal alkane concentrations calculated from herbage concentrations and actual faecal alkane concentrations. The effect of diet selection within groups can only be assessed by using sensitivity tests or by giving subjective weights to the individual plants. In the current study, a new optimization algorithm was developed that selects weightings that lead to consistent estimates of group proportions. The diet of the wild rabbit in a southern Portuguese montado was used as a case study. Estimates of the diet composition obtained using the new algorithm were compared with those of a conventional routine. The new algorithm was shown to provide, on average, more accurate estimates of the proportions of the groups in the diet. The effect of grouping plant species according to criteria other than similarity in n-alkane pattern on the accuracy of estimates was shown to be non-significant.

INTRODUCTION

Particular patterns of concentrations of n-alkanes in cuticular wax are specific to individual plant species (Dove *et al*. 1996; Bugalho 1999; Chen *et al*. 1999) or parts of plants (Dove *et al*. 1996). Therefore, there is potential for them to be used as markers in the estimation of diet composition from the pattern of n-alkane concentrations in the faeces, assuming that they are fully recovered or, if faecal recoveries are incomplete, that relative recoveries are known (Dove & Mayes 1991).

The usual method for estimating diet composition from the pattern of n-alkanes in faeces and in plant species is based on a least-squares procedure using the available n-alkanes (Salt *et al*. 1994; Dove & Moore 1995; Mayes *et al*. 1995). One of the most commonly used approaches is the non-negative least-squares algorithm implemented in the software EATWHAT (Dove & Moore 1995).

The n-alkane technique has been used recently to study the diet composition of wild herbivores, such as chamois (Pérez-Barbería *et al*. 1997), red deer (Bugalho 1999), mountain hare (Hulbert 1993; Hulbert & Andersen 2001; Hulbert *et al*. 2001), roe deer (Lang *et al*. 2000; Hulbert & Andersen 2001) and hairy-nosed wombat (Woolnough 1998). In all of these studies only a small number of dietary components (between three and six) was considered, either because the environment was a simplified one in terms of the number of food items on offer (see Rosenzweig 1995) or the studied animal had a simplified diet in terms of composition (see Hulbert *et al*. 2001).

In the savannah-like southern Portuguese montados, which are ecosystems of high diversity in terms of vegetation structure and presence of plant species (Joffre *et al*. 1999), the diet of the wild rabbit is likely to be complex. By using the microhistological analysis technique, M. C. Reis *et al*. (unpublished) found 14 plant species to be the most abundant in the diet of a coastal population of wild rabbits in the south of Portugal.

* To whom all correspondence should be addressed. Present address: Applied Ecology Centre Prof. Baeta Neves, Agronomy Institute of Lisbon, Tapada da Ajuda, 1349-018 Lisboa Codex, Portugal.
Email: ceabn@ip.pt

The maximum number of diet components that can be discriminated by using n-alkanes is limited to the number of n-alkane markers available, which are between 8 and 15 (Dove & Mayes 1991). One way of dealing with complex diets in the face of a limited number of n-alkanes is by pooling plant species into groups which are then treated as individual dietary components (Bugalho 1999; Mayes & Dove 2000). However, there could be differential selection of species within groups. Thus in order to obtain accurate estimates of diet composition at the group level, appropriate weighting of species within groups will be necessary. Such weighting could be based on estimates of the relative amounts of different species available for consumption although different species preferences would not be considered. An alternative approach for establishing appropriate weighting factors is to use an iterative algorithm which selects weightings that lead to consistent estimates of group proportions. A disadvantage of the EATWHAT least-squares routine is that computations using different diet component data sets cannot be repeated automatically. Moreover, questions arise as how best to organize the plant species into groups and how that affects the accuracy of the estimates.

In the current study, further developments on the use of n-alkane analysis to describe complex diets are assessed, using as a case study the diet of wild rabbit in a southern Portuguese montado. A new least-squares algorithm is developed, which bases the least-squares optimization procedure on calculating a weighted n-alkane concentration for each dietary component, with the weights reflecting the different estimated proportions of the plant species in the diet. The diet estimation is undertaken, therefore, at the level of the groups and at the level of the individual plant species. The accuracy of the diet estimates obtained is compared with that derived after giving equal weighting to individual species within a group (simple means of species alkane concentrations). The effect on accuracy of following different approaches to grouping the plant species, based on similarity either of n-alkane concentrations or of usefulness for management purposes, is also analysed.

MATERIALS AND METHODS

Study area

The study area was a 270-ha hunting estate in south-east Portugal (38°47′N, 7°25′W), comprising a rolling landscape between 300 and 420 m in altitude. The climate is a Mediterranean type with a strong seasonality, being characterized by hot and dry summers (average temperature of 25 °C and average rainfall of 2 mm in August) and rainy and mild winters (average temperature of 9 °C and average rainfall of 100 mm in January) (Rosário *et al.* 1983).

The vegetation consisted mainly of cork oak (*Quercus suber* L.) and holm oak (*Quercus rotundifolia* L.) stands, some with crops of triticale (× *Triticosecale* Wittmack) and oats (*Avena sativa* L.). The remaining understorey was dominated either by patches of gum cistus (*Cistus ladanifer* L.) or natural pasture. Amongst the 120 different herbage species identified in the natural pasture, the most abundant were *Agrostis pourretii* Willd., *Bromus hordeaceus* L., *Echium plantagineum* L., *Leontodon taraxacoides* (Vill.) Mérat, *Ornithopus compressus* L., *Rhagadiolus stellatus* (L.) Gaertner, *Trifolium angustifolium* L. and *Vulpia geniculata* (L.) Link.

N-alkane analysis

Sixteen groups of warrens were selected as being evenly spread across the study area. A circle with a radius of 100-m was drawn from the peripheral warrens of each group. The area within the circle was defined as a collection site as it was likely to reflect the home-range covered by the individuals belonging to the warrens. Gibb (1993) suggested this distance as being, on average, the maximal dispersion distance at night for both sexes. Movements recorded during crepuscular observations on rabbit behaviour using the focal point technique (Martin & Bateson 1993), also provided support for the radius chosen. Fresh faecal samples were collected from the collection site in each of three seasons. The seasons were those used by Alves & Moreno (1996) in a study on the reproductive cycle of the same population and were defined as the breeding season (between January and June), the post-breeding season (between July and September) and winter (between October and December). From each collection site, two sets of faecal samples were collected in the middle of each season, with an interval of 2 weeks between collections of a set. A total of 96 faecal samples was collected.

The presence and frequency of plant species were assessed at each collection site in each season, using 20×40 cm random vegetation quadrats (Frame 1993) in a number related to the area of each of the vegetation patches within a collection site, but constrained by the time available for sampling. Between 6 and 24 quadrats were used to sample vegetation patches whose surface areas varied from 0·07 ha to 7 ha, making a total of 1080 vegetation quadrats sampled.

Vegetation samples for n-alkane analysis were collected randomly over the study area and contemporaneously with the faecal samples, as suggested by Dove & Mayes (1991). The 30–50 plant species collected, depending upon season, were selected on the basis of (1) being abundant at each collection site, (2) showing signs of being consumed and (3) being a known seasonal feed resource, such as tree shoots and acorns.

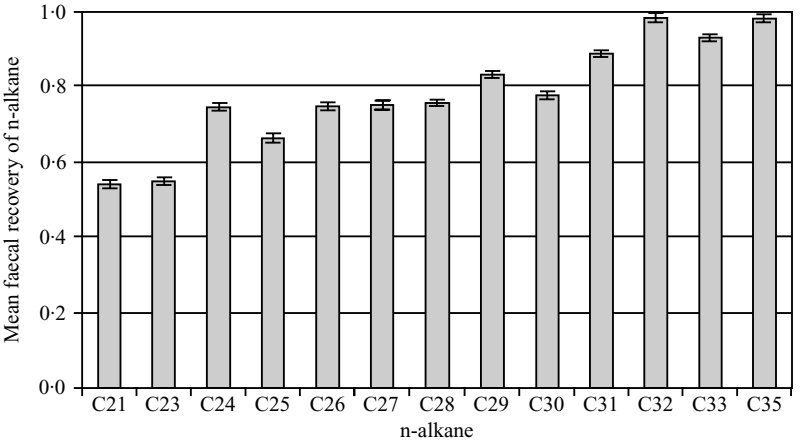

Fig. 1. Mean faecal recoveries (±S.E. mean) of each n-alkane (means of eight rabbits).

Herbage and faecal samples were oven-dried at 60 °C for 48 h and milled through a 1 mm size mesh. The ground samples were analysed for n-alkane concentrations using gas chromatography, according to the method described in Mayes *et al.* (1986) with modifications reported in Salt *et al.* (1992). The strategy used in the current study was to use as many alkane markers as possible. Thus the concentrations of n-alkanes between eicosane (C_{21}) and pentatriacosane (C_{35}) were measured, with the exception of the concentration of docosane (C_{22}) and tetratiacontane (C_{34}) because these were used as internal standards. C_{22} and C_{34} alkanes are not found in most plant material (Mayes *et al.* 1986).

The faecal recovery of the n-alkanes has to be taken into consideration since it might affect the accuracy of the diet estimates by biasing diet estimates towards dietary components with a predominance of longer-chain n-alkanes (Dove & Mayes 1996). Previous studies have shown that faecal recovery of their n-alkanes increases with the chain length in ruminants, but not in horses, pigs and mountain hares (Dove & Mayes 1991). Analyses of data from caged domestic rabbits fed dried grass pellets also showed that faecal recoveries increased with the chain length (Letso 1995). Data were also produced for fresh grass and used in the current study to confirm this effect. To establish the pattern of faecal recoveries in rabbits, a regression model was fitted to the mean of the angular-transformed faecal recoveries obtained from eight individually caged domestic rabbits fed fresh perennial ryegrass (not prevented from practising copraphagy). Use of transformed data was required to address the observed variance/mean relationship. Faecal samples were collected daily during a period of 5 days, followed by an analysis of the n-alkane content of herbage and faecal samples (M. Letso, unpublished). The calculations for faecal recoveries were made for chain lengths of C_{21}–C_{35}. On the transformed scale the faecal recoveries were found to increase approximately linearly with increasing carbon-chain length ($t_8 = 2{\cdot}95, r^2 = 0{\cdot}86, P < 0{\cdot}001$) (Fig. 1). Consequently, the observed mean faecal recovery values for the n-alkanes of different chain length were applied to faecal samples used in the current study before estimating the diet composition.

Estimation of the diet composition

The absolute concentrations of the n-alkanes, from both faecal and vegetation samples, were compared iteratively within a least-squares optimization routine. The procedure is similar to that used by Salt *et al.* (1994), and makes use of the NAG routine, E04NCF (Numerical Algorithm Group 1987) which performs a constrained quadratic optimization. A small custom-built programme (in FORTRAN) was written in order to apply appropriate weighting factors to plant species within groups and to organize the data for computation by the NAG routine.

The core of the algorithm operates in a very similar way to the one implemented in the software EATWHAT by Dove & Moore (1995). Consequently, the same nomenclature is used:

i, j are indices for n-alkane and dietary component respectively;
p is the N-vector of diet proportions;
H is a M × N matrix containing the concentrations of M alkanes in N dietary components;
f is a M-vector containing the concentrations of each one of the n-alkanes in the faecal sample;
r is a M-vector containing faecal recovery rates for each n-alkane;
b is a M-vector containing faecal n-alkane concentrations corrected for recovery $b_i = f_i/r_i$;
x is a N-vector containing the quantity of each species that is consumed to produce 1 kg of faeces.

The problem consists of finding the x, under that constraint of being positive or zero, to minimize

$$S^2 = |Hx - b|^2 = \sum_{i=1}^{M} \left(\sum_{j=1}^{N} h_{ij}x_j - b_i \right)^2, \; x \geqslant 0$$

The process of adjusting the different intakes of the diet components from initial values is carried out until the error sum of squares reaches its minimum, S^2_{min}.

Having x, it is possible to calculate the proportion of the dietary components as:

$$p = \frac{1}{\sum_{j=1}^{N} x_j} x$$

The algorithm allows for different proportions of plant species within a group, which is more realistic since different plant species are consumed in different proportions. An iterative least-squares analysis finds the intake, and consequently the proportion of the diet components, that minimizes the squared deviations between the observed faecal concentrations and the expected faecal concentrations calculated from the n-alkane content in the herbage samples.

The procedure implemented in the new algorithm described in the current paper is presented in a simplified way in Fig. 2, considering the hypothetical case of a diet with two components, Component I and Component II. Therefore N, the number of diet components, is 2 and p is a 2-vector. M is the number of n-alkane markers used, which covers the full range C_{21}–C_{35}, with the exception of C_{22} and C_{34}, and, therefore, is equal to 13. It is assumed that Component I of the diet has two plant species, $S_{1.1}$ and $S_{1.2}$, and Component II has three plant species $S_{2.1}$, $S_{2.2}$ and $S_{2.3}$. To achieve an estimate of the diet composition the algorithm performs the following steps:

1st step – For each group, estimate a weighted mean alkane profile using as weights the proportion of each species in each group (for the first iteration, these proportions were set to be equal within each group). Then, using these weighted means, estimate the amounts of each group in the diet.

2nd step – Keep the amounts of Component I and II constant, but allow the proportions of species in Component I to vary. Thus re-use the least squares algorithm to estimate $p^*_{1.1}$ and $p^*_{1.2}$ with the constraint $p^*_{1.1} + p^*_{1.2} = 1$.

3rd step – Keep the amounts of Components I and II constant, but allow the proportions of species in Component II to vary. Thus re-use the least squares algorithm to estimate $p^*_{2.1}$, $p^*_{2.2}$ and $p^*_{2.3}$ with the constraint $p^*_{2.1} + p^*_{2.2} + p^*_{2.3} = 1$.

4th step – Return to step 1.

These four steps are repeated iteratively until convergence is achieved. Note that, when the number of species is large relative to the number of markers, the problem is ill-conditioned and so a 1-step optimization treating all species independently will not have a unique solution. The above algorithm is an attempt to work round this problem by iterating between estimation at the group level and estimation of the composition of the different groups. However, where the number of species is small relative to the number of markers, the iterative algorithm is not guaranteed to find the optimal solution and so will be sub-optimal.

Simulation of estimates of diet composition

Between 30 and 50 different plant species were collected at the study area in each season, and organized into groups according to the similarity in the n-alkane concentrations. This approach was selected for being the one that has been followed in previous studies when the diet components outnumber the markers (Bugalho 1999). These groups were defined using cluster analysis performed in a hierarchical way, using the proportion of concentration of each alkane in the total n-alkane concentration of the sample, after log-ratio transformation to remove linear dependency in the variables used (see Elston *et al.* 1996). Ward's method was used to form the clusters by maximizing within-cluster homogeneity (Sharma 1996). The cut-off point to define the clusters was chosen in order to achieve a maximum of seven clusters. The patterns of n-alkane concentrations made it possible to isolate well-defined groups for the breeding, post-breeding and winter seasons.

Thirty mixtures of these plant species were used to simulate faecal n-alkane concentrations. Their allocation to the clusters previously defined was respected and the concentration of each n-alkane was calculated across the groups using species mean n-alkane concentrations, both considering equal and different proportions for the individual plant species within a group. Estimates of diet composition were obtained in both cases using equal weighting of individual species within a group and the new algorithm (see above). The accuracy of the estimates obtained with the two methods was compared by plotting the simulated proportions of the dietary components against the estimated ones and by calculating a standard deviation of the estimates (S.D.E.) as follows:

$$\text{S.D.E.} = \sqrt{\text{mean (estimated} - \text{simulated)}^2}$$

The effect of introducing error into the measurement of the n-alkane concentrations of the plant mixtures on estimates of diet composition was also assessed. This required the estimation of the random variation for each combination of plant species and n-alkane concentrations. These variances were estimated using the following modelling approach. The mean and the variance of each n-alkane concentration was calculated across four replicates of each of 60 different species and parts of plants. The variance was related

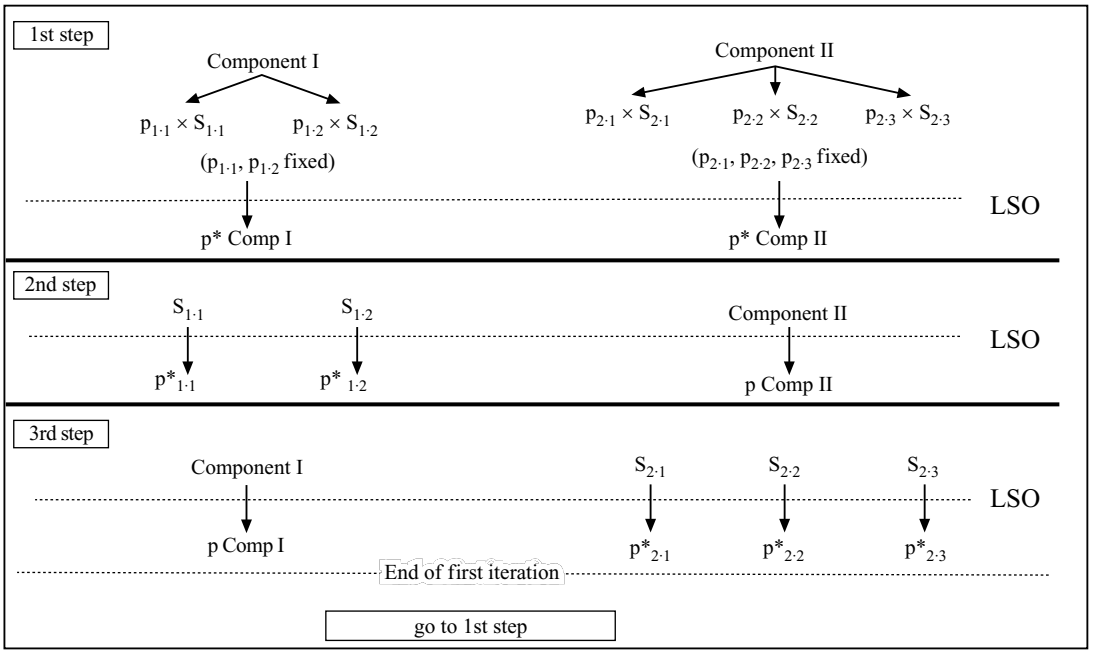

Fig. 2. Illustration of the stepwise procedure followed by the least-squares optimization (LSO) algorithm implemented in this study, in the case of a simplified diet with only two diet components, Component I and Component II; the first consists of two plant species, $S_{1\cdot1}$ and $S_{1\cdot2}$ and the second of three plant species, $S_{2\cdot1}$, $S_{2\cdot2}$ and $S_{2\cdot3}$. The proportions of the diet components and plant species are iteratively estimated by repeating the steps until convergence is achieved (estimated values marked with *, as opposed to fixed values).

to the mean by fitting a generalized linear model with gamma errors and a log link function using the statistical package Genstat 5 (Genstat Committee 1993). The logarithm of the mean of the n-alkane concentrations was considered to be the independent variable. The fitted model describing the variance-mean relationship was:

$$\text{variance} = 0{\cdot}044 \times \text{mean}^{1{\cdot}4028}$$

with the exponent significantly different from zero ($t = 33{\cdot}13$, $df = 778$, $P < 0{\cdot}001$).

Random contributions from each of the species to the 30 plant mixtures were simulated from the fitted gamma distributions. The random contributions had, as their mean, the species n-alkane concentration and, as their variance, the modelled variance adjusted by a multiplicative scale factor taking the values 0·016, 0·031, 0·060, 0·125 and 0·250. The new iterative algorithm was used to estimate the diet composition for each simulated diet, and to obtain a minimized sum of squared discrepancies, S^2_{min}. These values of S^2_{min} were plotted against the variance scale factor.

Assessment of the effect of different approaches to grouping plant species on the accuracy of the diet composition estimates

A mixture of 13 plant species found in the montado was used as a representative diet, which was different for each collection site and in each season. The plant species were selected on the basis of the information on the frequency obtained from the vegetation quadrats, and the likelihood of being present in the diet of rabbits as inferred from the microhistological results of M. C. Reis *et al.* (unpublished).

Four different approaches to grouping the plant species into diet components were tested in this study, namely:

Approach 1. Groupings of plant species constituted in terms of their similarity in the relative proportions of their n-alkane concentrations using cluster analysis.
Approach 2. Functional groupings of plants – trees (cork oak and holm oak), shrub (gum cistus), legumes, grasses, oats, triticale, dicotyledons and others.
Approach 3. Broad groupings of plants with interest for management such as grass-forb, cork oak, holm oak, olive tree and gum cistus.
Approach 4. Individual plant species.

Approach 4 was included as a way of assessing the effect on accuracy of considering as many dietary components as the number of markers allows, in comparison with situations when the plant species were aggregated to reduce the number of components.

The effect of the four different approaches of grouping the plant species on the accuracy of diet estimates was evaluated in terms of the values of S^2_{min}

obtained using equal weighting of species within a group and the new iterative algorithm.

The comparison between approaches was made using ANOVA (Zar 1984) with 'treatments' being the type of approach (clusters, functional groups, management groups and individual plant species) and the method followed to estimate the diet composition. Collection site, season (breeding, post-breeding and winter), and subsample (two every season) were treated as blocks. The independent variable was the S^2_{min} log-transformed for normality.

All the statistical analyses were performed using Genstat 5 (Genstat 5 Committee 1993).

RESULTS

Simulation of estimates of diet composition

When diets were simulated such that equal proportions of plant species occurred within each group,

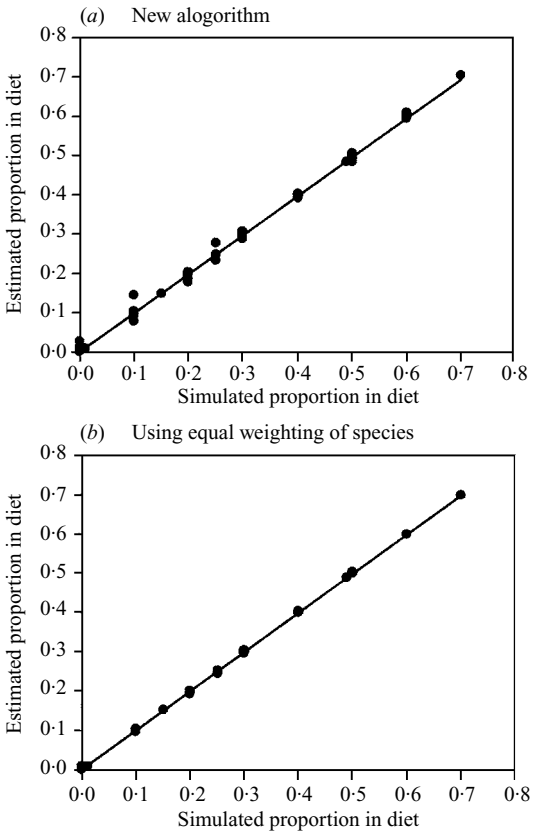

Fig. 3. Relationship between simulated and estimated group proportions in diet estimates obtained with (*a*) the new algorithm ($r^2 = 0{\cdot}99$, S.E. $= 0{\cdot}007$) and (*b*) by using equal weighting of plant species ($r^2 = 1{\cdot}0$, S.E. $= 0{\cdot}001$), when individual plant species are present in equal proportions within a group.

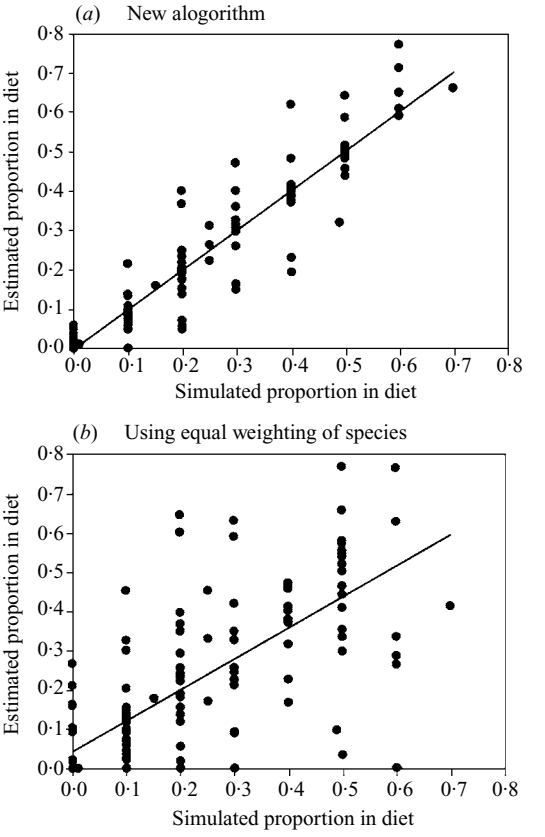

Fig. 4. Relationship between simulated and estimated group proportions in diet estimates obtained with (*a*) the new algorithm ($r^2 = 0{\cdot}91$, S.E. $= 0{\cdot}062$) and (*b*) by using equal weighting of plant species ($r^2 = 0{\cdot}52$, S.E. $= 0{\cdot}148$), when individual plant species are present in different proportions within a group.

both the new algorithm and that using equal weighting of species within a group performed well in accurately estimating the proportion of the groups in the diet (Fig. 3). The intercept of the regression line was not significantly different from zero in both cases ($t_{140} = 1{\cdot}11$, $P = 0{\cdot}27$, $t_{140} = 0{\cdot}52$, $P = 0{\cdot}60$). The standard deviations of the estimates were 0·007 and 0·001, respectively. However, when diets were simulated with unequal proportions of plant species within each group, the new algorithm performed significantly better than that using equal weighting of species within a group (Fig. 4). In the case of the new algorithm, 91% of the variance in group proportions was explained by the regression line, the intercept was not significantly different from zero ($t_{140} = -0{\cdot}06$, $P = 0{\cdot}95$) and the standard deviation of the estimates was 0·062. The regression line associated with estimates obtained by using equal weighting of species within a group only explained 52% of the variance in

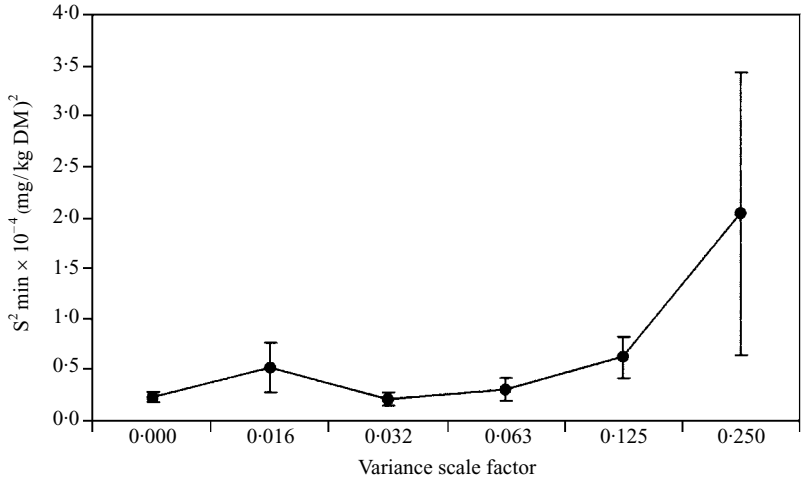

Fig. 5. Change in the mean S^2_{min} ($\pm$S.E. mean $\times 10^{-4}$) associated with the diet composition estimates obtained with the new algorithm, in relation to increases in the variance scale factor.

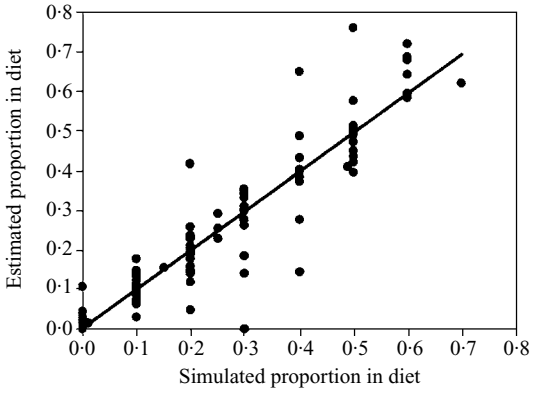

Fig. 6. Relationship between simulated and estimated proportions of groups in diet obtained with the new algorithm and considering the influence of the error (variance scale factor equal to 1·0) ($r^2 = 0{\cdot}90$, S.E. = 0·064).

the results, the intercept was significantly different from zero ($t_{140} = 2{\cdot}52$, $P = 0{\cdot}05$) and the standard deviation of the estimates was 0·153.

Neither the new algorithm nor the one using equal weighting of species within a group produced accurate estimates of the proportion of individual plant species. For the latter, the regression line between the estimated and the simulated proportions of the individual plant species ($y = 0{\cdot}19 + 0{\cdot}47x$) only explained 47% of the variance, with the intercept also being significantly different from zero ($t_{388} = 14{\cdot}35$, $P < 0{\cdot}001$) and the standard deviation of the estimates equal to 0·260. The regression line between the estimated and the simulated proportions of the individual plant species ($y = 0{\cdot}07 + 0{\cdot}80x$) for the new algorithm explained 61% of the variance, the intercept was significantly different from zero ($t_{388} = 4{\cdot}38$, $P < 0{\cdot}001$) and the standard deviation of the estimates was 0·241.

The values of S^2_{min} remained low between a variance scale factor of 0 and 0·125 but increased 0·5 to 2×10^{-4} (mg/kg DM)^2 between a variance scale factor of 0·016 and 0·250 (Fig. 5). The standard deviation of the estimates ranged between 0·042 and 0·086. Even when the error is multiplied by a factor of 0·250, the new algorithm still showed the ability to estimate accurately the average proportion of the groups in the diet (Fig. 6). The regression line fitted between the simulated and the estimated proportions in diet explained 90% of the variance and the intercept was

Table 1. *Results of the ANOVA to rest the effect of the approach followed to grouping the plant species on the log S^2_{min} in the two methods for diet estimation, in comparison with other sources of variation, namely collection site, season and sampling replicate*

Source of variation	D.F.	S.S.	M.S.
Collection site (CS)	15	570·707	38·05
Season (S)	2	389·144	194·57
Subsample (SS)	1	12·278	12·28
CS × S	30	628·676	20·96
CS × SS	15	52·644	3·51
S × SS	2	171·113	85·56
CS × S × SS	30	213·673	7·12
Method	1	168·991	168·99
Approach	3	191·991	63·99
Method × Approach	3	119·610	39·87
Residual	665	276·292	0·42
Total	767	2795·119	

S.S. = sums of squares; M.S. = mean square.

not significantly different from zero ($t_{140} = 0{\cdot}28$, $P = 0{\cdot}777$).

Assessment of the effect of different approaches to grouping plant species on the accuracy of diet composition estimates

In spite of the considerable amount of variance explained by collection site and season, and their interactions (see Table 1), there was a large and significant effect of the method used to estimate the diet composition ($F_{1,665} = 406{\cdot}74$, $P < 0{\cdot}001$), and smaller effects of both the approach followed to group the plant species ($F_{3,665} = 154{\cdot}03$, $P < 0{\cdot}001$) and their interaction ($F_{3,665} = 95{\cdot}96$, $P < 0{\cdot}001$).

The lowest value of S^2_{min} obtained with the new algorithm was associated with the approach of considering the individual plant species (transformed value 1·689, $P < 0{\cdot}001$). The values of least-squares sum of errors obtained by the other three approaches in grouping the plant species were significantly higher but did not differ significantly from each other (cluster analysis groups: 1·950, functional groups: 2·106, management groups: 1·992; S.E.D. = 0·093, $P < 0{\cdot}05$). Using the original algorithm, the values obtained with the different approaches in grouping the plant species were all significantly different from each other (S.E.D. = 0·093, $P < 0{\cdot}001$). The lowest value of S^2_{min} was also given by the approach with the individual plant species (1·689), followed by the approach with cluster analysis groups (2·661) and the approach with functional groups (2·964). The highest value was obtained with the approach with management groups (4·176).

DISCUSSION

The new algorithm represents a further development in the approach followed previously. It estimates the proportions of dietary components that minimize the discrepancy between the calculated and observed n-alkane concentrations of faecal samples not only at the level of the plant group, as the previous approach, but also at the level of the individual plant species included in the group. The simulations presented demonstrate that the estimates of the average diet composition obtained with the new algorithm are more accurate on average in terms of groups. The calculation of a weighted average alkane concentration for each group, taking into consideration the different proportions of the individual plant species within the group, performs better in the estimation procedure than a simple average. Inspection of the results revealed that high discrepancies between the estimates obtained by using equal weighting of species within a group and the new algorithm are also likely when the most selected plant species are spread across different groups.

The inability of the algorithm to estimate accurately the proportion of individual plant species is explained partly by the fact that the least-squares procedure is a sub-optimal one and partly by the estimation problem being ill-conditioned at the species level when the number of species is large relative to the number of markers. The sub-optimality arises because the iterative optimization scheme involves separate optimization between and within groups whereas a fully optimal scheme would simply consider each species separately. However, for complex diets with a limited number of markers, there is insufficient information to estimate the proportion of each species in the diet, hence the proposed algorithm. The difficulty in determining the best fit also arises when the groups are not well resolved, with higher variance in n-alkane concentration within them than between them, and when plant species with higher n-alkane concentrations mask the presence of other species (Dove & Mayes 1991).

The low standard deviation of the estimates and the small increase in the least-squares error with the variance scale factor suggest that the new algorithm is robust in obtaining diet estimates at the group level. Sampling errors in faeces and herbage samples collection were not considered since the assumption was made that the approach to the collection of vegetation and faecal samples minimized them (see Dove & Mayes 1991).

The method of grouping did not affect significantly the accuracy of the diet estimates with the new algorithm and, therefore, should be decided according to the purpose of the study. The approach of defining groups using cluster analysis on the basis of similarity in alkane concentrations is analytically advisable because the groups are better resolved, facilitating the convergence of the least-squares procedure towards the optimal proportions of the dietary components. However, it has the drawback that plant species belonging to very different taxa can have similar patterns of n-alkanes and can be included in the same cluster (see also Bugalho 1999). As a consequence, this approach might be of limited value for habitat management purposes and an approach of organizing the plant species in functional or management groups might be more adequate.

In conclusion, the new algorithm presented has potential for the study of complex diets. In the particular case of the diet of wild rabbit in a montado, it was demonstrated that it could incorporate the fact that plant species within groups are consumed in different proportions in order to improve the accuracy of diet estimates. It was also successful in capturing seasonal and spatial variation in the diet composition. However, this was only possible in terms of groupings with interest for the management of the plant species and habitats, rather than in terms of individual plant species. Further research to increase the number of

cuticular plant wax chemical components as markers is needed if an accurate estimation of the proportion of all the plant species found in complex diets is desired (Mayes & Dove 2000). These could be either markers found in the urine, such as plant phenolic compounds (Mayes *et al*. 1995), or other components of plant wax, such as alkenes (e.g. Dove *et al*. 1992), long-chain fatty acids, long-chain fatty alcohols (C_{19}–C_{32}) and β-diketones (Tulloch *et al*. 1980) found in the faeces.

H. Martins was funded by European Commission Programme PRAXIS XXI (BD9396/96). The other authors were funded by the Scottish Executive Rural Affairs Department. We are grateful to Direcção Geral de Florestas for allowing the data collection at Tapada Pequena de Vila Viçosa, to Stuart Lamb for the technical support and to Prof. E. R. Ørskov, Dr I. J. Gordon and Dr F. J. Pérez-Barbería and an anonymous referee for kindly commenting on drafts of the manuscript.

REFERENCES

Alves, P. C. & Moreno, S. (1996). Estudo da reprodução do Coelho-bravo (*Oryctolagus cuniculus*) em Portugal. *Revista Florestal* **9**, 149–166.

Bugalho, M. (1999). *The foraging behaviour of Red deer* (Cervus elaphus) *in a Mediterranean environment*. Ph.D. thesis, University of Aberdeen, Scotland.

Chen, W., Scott, J. M., Blair, G. J. & Lefroy, R. D. B. (1999). Using plant cuticular alkanes to study plant–animal interactions on pastures. *Canadian Journal of Animal Science* **79**, 553–556.

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. & Mayes, R. W. (1996). Plant wax components: a new approach to estimating intake and diet composition in herbivores. *Journal of Nutrition* **126**, 13–26.

Dove, H. & Moore, A. D. (1995). Using a least-squares optimisation procedure to estimate botanical composition based on the alkanes of plant cuticular wax. *Australian Journal of Agricultural Research* **46**, 1535–1544.

Dove, H., Siever-Kelly, C., Leury, B. J., Gatford, K. L. & Simpson, R. J. (1992). Using plant wax alkanes to quantify the intake of plant parts by grazing animals. *Proceedings of the Nutrition Society of Australia* **17**, 149.

Dove, H., Mayes, R. W. & Freer, M. (1996). Effects of species, plant part and age on the n-alkane concentrations in the cuticular wax of pasture plants. *Australian Journal of Agricultural Research* **47**, 1333–1347.

Elston, D. A., Illius, A. W. & Gordon, I. J. (1996). Assessment of preference among a range of options using log-ratio analysis. *Ecology* **77**, 2538–2548.

Frame, J. (1993). Herbage mass. In *Sward Measurement Handbook* (Eds A. Davies, R. D. Baker, S. A. Grant & A. S. Laidlaw), pp. 39–67. Reading: British Grassland Society.

Genstat Committee (1993). *Genstat 5 Reference Manual*. Oxford: Clarendon Press.

Gibb, J. A. (1993). Sociality, time and space in a sparse population of rabbits (*Oryctolagus cuniculus*). *Journal of Zoology* **229**, 581–607.

Hulbert, I. A. R. (1993). *The effects of upland afforestation on the ecology of lagomorphs with different feeding strategies*. Ph.D. thesis, University of Edinburgh, Scotland.

Hulbert, I. A. R. & Andersen, R. (2001). Food competition between a large ruminant and a small hind but fermentor: the case of the roe deer and mountain hare. *Oecologia* **128**, 499–508.

Hulbert, I. A. R., Iason, G. R. & Mayes, R. W. (2001). The flexibility of an intermediate feeder: dietary selection by mountain hares measured using faecal n-alkanes. *Oecologia* **129**, 197–205.

Joffre, R., Rambal, S. & Ratte, J. P. (1999). The dehesa of southern Spain and Portugal as a natural ecosystem mimic. *Agroforestry Systems* **45**, 57–59.

Lang, D., Lechner-Doll, M. & Barleben, H. (2000). Using n-alkanes as marker substances in nutritional studies of roe deer (*Capreolus capreolus*). *Advances in Ethology* **35**, 70.

Letso, M. (1995). *A study into the use of n-alkanes to determine dietary intake and digestibility in grazing rabbits*. M.Sc. Thesis in Animal Nutrition, University of Aberdeen, Scotland.

Martin, P. & Bateson, P. (1993). *Measuring Behaviour. An Introductory Guide*. 2nd Edition. Cambridge: Cambridge University Press.

Mayes, R. W. & Dove, H. (2000). Measurement of dietary nutrient intake in free-ranging mammalian herbivores. *Nutrition Research Reviews* **13**, 107–138.

Mayes, R. W., Lamb, C. S. & Colgrove, P. M. (1986). 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., Dove, H., Chen, X. B. & Guada, J. A. (1995). Advances in the use of faecal and urinary markers for measuring diet composition, herbage intake and nutrient utilisation in herbivores. In *Recent Developments in the Nutrition of Herbivores* (Eds M. Journet, E. Grenet, M. T. Farce, M. Thériez & C. Demarquilly), pp. 318–406. Paris: INRA Editions.

Numerical Algorithm Group (1987). *NAG Fortran Library Manual*. Mark 14, Vol. 314.

Pérez-Barbería, F. J., Oliván, M., Osoro, K. & Nores, C. (1997). Sex, seasonal and spatial differences in the diet of the Cantabrian chamois *Rupicapra pyrenaica parva*. *Acta Theriologica* **42**, 37–46.

Rosário, L., Marques, A., Bugalho, M. & Lopes, F. (1983). *Plano de utilização da Tapada Pequena de Vila Viçosa*: Ministério da Agricultura Comércio e Pescas–Direcção Geral das Florestas, Lisboa.

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

Salt, C. A., Mayes, R. W. & Elston, D. A. (1992). Effects

of season, grazing intensity and diet composition on the radiocaesium intake by sheep on re-seeded hill pasture. *Journal of Applied Ecology* **29**, 378–387.

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.

Sharma, S. (1996). *Applied Multivariate Techniques*. Chichester: John Wiley & Sons.

Tulloch, A. P., Baum, B. R. & Hoffman, L. L. (1980). A survey of epicuticular waxes among genera of Triticeae. 2. Chemistry. *Canadian Journal of Botany* **58**, 2602–2615.

Woolnough, A. P. (1998). *The feeding ecology of the northern hairy-nosed wombat*, Lasiorhinus krefftii (*Marsupialia: Vombatidae*). Ph.D. thesis, Australian National University, Australia.

Zar, J. H. (1984). *Biostatistical Analysis*. Upper Saddle River, NJ: Prentice-Hall International Editions.