

## FARMING SYSTEMS and ENVIRONMENT

# Using n-alkanes to estimate diet composition of herbivores: a novel mathematical approach

P. Barcia<sup>1</sup>, M. N. Bugalho<sup>2†</sup>, M. L. Campagnolo<sup>3</sup> and J. O. Cerdeira<sup>4</sup>

<sup>1</sup>Faculdade de Economia, Universidade Nova de Lisboa, Campus de Campolide, 1099-032 Lisboa, Portugal; <sup>2</sup>Centro de Ecologia Aplicada “Baeta Neves”, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal; <sup>3</sup>Departamento de Matemática, Instituto Superior de Agronomia and SQIG-IT, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal; <sup>4</sup>Departamento de Matemática, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal

(Received 7 February 2006; Accepted 5 October 2006)

*N-alkanes are long-chain saturated hydrocarbons occurring in plant cuticles that can be used as chemical markers for estimating the diet composition of herbivores. An important constraint of using n-alkanes to estimate diet composition with currently employed mathematical procedures is that the number of markers must be equal or larger than the number of diet components. This is a considerable limitation when dealing with free-ranging herbivores feeding on complex plant communities. We present a novel approach for the estimation of diet composition using n-alkanes which applies equally to cases where the number of markers is lower, equal or greater than the number of plant species in the diet. The model uses linear programming to estimate the minimum and maximum proportions of each plant in the diet, and avoids the need for grouping species in order to reduce the number of estimated dietary components. We illustrate the model with two data sets of n-alkane content of plants and faeces obtained from a sheep grazing experiment conducted in Australia and a red deer study in Portugal. Our results are consistent with previous studies on those data sets and provide additional information on the proportions of individual species in the diet. Results show that sheep included in the diet high proportions of white clover (from 0.25 to 0.37), and relatively high proportions of grasses (e.g. brome from 0.14 to 0.26) but tended to avoid Lotus spp. (always less than 0.04 of the diet). For red deer we found high proportions of legumes (e.g. Trifolium angustifolium and Vicia sativa reaching maximum proportions of 0.42 and 0.30 of the diet, respectively) with grasses being less important and Cistus ladanifer, a browse, also having relevance (from 0.21 to 0.42 of the diet).*

**Keywords:** grazing, linear models, linear programming, markers, ungulates

### Introduction

Knowing the diet composition of herbivores is important for understanding their foraging ecology and for mediating their effects on vegetation and ecosystems. Different methods can be used to estimate herbivore diets, although most of them have limitations and must be selected according to aims of study. N-alkanes have been used to estimate diet composition of domestic animals (Salt *et al.*, 1992; Duncan *et al.*, 1999; Hutchings *et al.*, 2000) and, less frequently, of wild herbivores (Bugalho *et al.*, 2001; Hulbert *et al.*, 2001; Rao *et al.*, 2003). N-alkanes are long-chain saturated hydrocarbons widespread in plant cuticles which can be recovered, after correcting for differential digestibility, in herbivore faeces as they are chemically inert (Dove and Mayes, 1996; Mayes and Dove, 2000). Additionally, since

plant species are characterised by different concentrations of n-alkanes, the chemical markers recovered in faeces can be used to identify and quantify the plants ingested by herbivores.

Dove and Mayes (1991) suggested modelling the problem of estimating diet composition based on n-alkanes with a constrained system of linear equations. A number of different least-squares procedures, including EATWHAT, developed in CSIRO, Australia (Dove and Moore, 1995), have been designed to determine diet composition using this model. The requirement that the maximum number of components that can be differentiated in the diet is limited by the number of available markers, is common to all those procedures. In the case where there are more plant components than n-alkanes, it is usual to group individual plant species, and express diet estimates in terms of the plant groups defined. Those groups can be defined either functionally (e.g. grasses, legumes, browse) or statistically

<sup>†</sup> E-mail: migbugalho@isa.utl.pt

(e.g. pooling of species with similar n-alkanes profiles). The groups however may not be suitable. For instance, the group of legumes is not adequate to estimate nitrogen intake since species within this group may have quite different nitrogen content (Mayes and Dove, 2000). Although Mayes and Dove (2000) suggest that n-alkanes can be used to differentiate plant parts in the diet of herbivores in our work 'dietary component' always refers to plant species (whole sample of plant leaf and stem). Indeed plant species accounts for more than 85% of the variation in n-alkane concentrations over plant parts (Dove *et al.*, 1996).

Our aim is to use n-alkanes to estimate diet composition of free-range herbivores feeding on complex plant communities, such as rangelands or other natural systems, where the number of plant species is generally high, usually outnumbering the number of reliable n-alkanes. In these situations, establishing coherent groups of plant species or limiting the potentially selected species to a pre-determined small subset may not be reliable.

We describe a model in which diet composition is identified with (possible) infinite bounded combinations of plant species. The upper and lower bounds on the concentrations of each plant species in the diets can easily be determined by linear programming. This approach can be applied equally to cases where the number of markers is lower, equal or greater than the number of species, thus avoiding the need for grouping species.

We tested the model on two data sets consisting of records of the n-alkane profiles of different plant species and faeces of sheep and deer collected in experiments conducted in Australia and Portugal. Each faecal sample has a distinct n-alkane content. By using a whole set of samples (four and nine samples of faeces for sheep and deer, respectively) our model incorporates information on the intrinsic variability of the data.

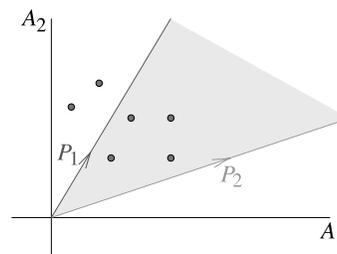
## Material and methods

### Mathematical description

The problem of estimating diet composition based on n-alkanes with a constraint system of linear equations, as introduced by Dove and Mayes (1991), can be stated as follows.

Suppose  $d$  is the number of n-alkanes in plants and faeces. Each faecal sample can be considered a point on the  $d$ -Euclidean space, which describes the concentration of each of the  $d$  n-alkanes. Each plant species can be interpreted as a vector in the  $d$ -Euclidean space, defining the value of each n-alkane in that plant. To illustrate this we use a fictitious example with two n-alkanes, two plant species, and six faecal samples (Figure 1).

Diets are linear combinations, i.e. weighted sums of the vectors representing the plant species. The weight or coefficient of each vector defines the quantity of the corresponding species in the diet. Each coefficient value, divided by the sum of the coefficients, gives the proportion of each plant species in the diet. If  $P_1, P_2, \dots, P_p$  are



**Figure 1** Example representing two n-alkanes (axes  $A_1$  and  $A_2$ ), two plant species (vectors  $P_1$  and  $P_2$ ), and six faecal samples (the six dots in the figure). The shaded area is the two-dimensional cone generated by vectors  $P_1$  and  $P_2$ .

the  $d$ -vectors representing  $p$  plant species, each vector indicating the concentrations of n-alkanes on that species, the linear combination with coefficients  $c_1, c_2, \dots, c_p$  is  $c_1 P_1 + c_2 P_2 + \dots + c_p P_p$ .

Clearly, linear combinations with negative coefficients are 'nonsensical diets'. Hence, we must restrict our attention to linear combinations having non-negative coefficients. The set of all non-negative linear combinations of vectors is called the cone generated by those vectors. In Figure 1 the cone generated by vectors  $P_1$  and  $P_2$  is the shaded region. We use  $C$  to denote the cone generated by the vectors corresponding to species, i.e.

$$C = \{c_1 P_1 + c_2 P_2 + \dots + c_p P_p, \text{ with } c_1, c_2, \dots, c_p \geq 0\}.$$

The estimation of diets from faecal samples  $F_1, F_2, \dots, F_q$  takes one sample  $F$  at a time and seeks whether either

- (i)  $F$  belongs to the cone  $C$ , or
- (ii)  $F$  is a point outside of  $C$ .

This is determined by solving the system of linear equations

$$Ax = F, \quad (S)$$

where  $A$  is the  $d \times p$  matrix whose columns are the vectors  $P_1, P_2, \dots, P_p$  representing the plant species. If all components of the solution  $x$  are non-negative then (i) occurs, and  $x$  specifies the composition of the corresponding species in the faeces.

If there is a negative component, we are faced with situation (ii), indicating that only 'nonsensical diets' can explain faeces  $F$ . This may result from errors in measuring the concentrations of n-alkanes, or from the existence of plant species which are included in the diet but not sampled in the field. The usual procedure in this case (e.g. Mayes *et al.*, 1994; Salt *et al.*, 1994; Dove and Moore, 1995) is to find a non-negative vector of coefficients which 'adequately approximates'  $x$ . A reasonable approach consists of

- (iii) identifying a point  $F'$  in the cone  $C$  considered to be 'similar' to  $F$ , and
- (iv) using  $F'$  instead of  $F$ , proceed as in (i) solving the system of equations (S).

Usually the projection of  $F$  onto  $C$ , i.e. the point in  $C$  which is closest to  $F$ , is selected to be  $F'$ . In the example of

Figure 1 there are two points in case (ii). Each of these points will be replaced by its projection onto  $C$ , which will lie in the ray of the cone defined by vector  $P_1$ . This implies that the herbage mixtures that will be derived from these two points consist only of  $P_1$ .

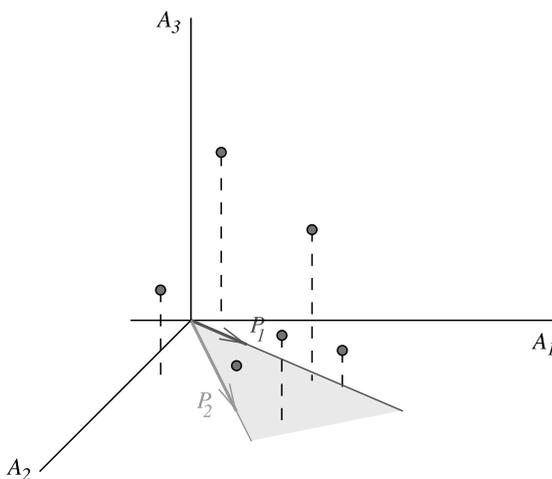
The estimated diet of the animal is finally computed from the solutions obtained with faecal samples  $F_1, F_2, \dots, F_q$ . Often, the diet is defined to be the average of the  $q$  solutions. Note that this diet is in fact the solution of the linear system of equations (S) with  $F$  equal to the average point of the projections of the faecal samples  $F_1, F_2, \dots, F_q$  onto  $C$ . (The projection onto  $C$  of a point in  $C$  coincides with the point.)

We implicitly assumed that the number of n-alkanes is equal to the number of plant species, i.e. that  $A$  is a square (and non-singular) matrix. If there are more alkanes than species the situation does not change substantially, apart from a larger percentage of faecal samples  $F$  in condition (ii) that may now be expected. Figure 2 is an example of three n-alkanes, two plant species, and six faecal samples. Only one point is in the cone.

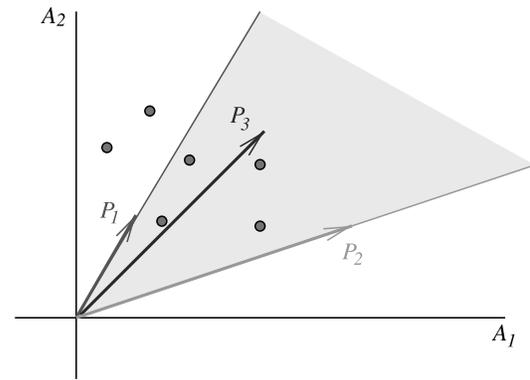
Consider now that there are more plant species than n-alkanes. Figure 3 is an example of two n-alkanes, three plant species, and six faecal samples.

The situation has now changed considerably. Now any non-negative linear combination of vectors  $P_1$  and  $P_3$  of Figure 3 defines a new vector  $P$  contained in the cone generated by  $P_1$  and  $P_3$ . Any point  $F$  inside the cone generated by  $P$  and  $P_2$  is a non-negative linear combination of  $P$  (and consequently of  $P_1$  and  $P_3$ ) and  $P_2$ . Since there may be infinitely many vectors  $P$  defined from  $P_1$  and  $P_3$  such that  $F$  is inside the cone generated by  $P$  and  $P_2$ , system (S) may have an infinite number of non-negative solutions. In other words, the concentration of n-alkanes in faecal samples may result from infinitely many different plant mixtures.

Whenever the number of species exceeds the number of n-alkanes, the usual approach is to group species into



**Figure 2** Example representing three n-alkanes (axes  $A_1, A_2$  and  $A_3$ ), two plant species (vectors  $P_1$  and  $P_2$ ), and six faecal samples (the six dots in the figure). The shaded area is the two-dimensional cone generated by vectors  $P_1$  and  $P_2$ .



**Figure 3** Example representing two n-alkanes (axes  $A_1$  and  $A_2$ ), three plant species (vectors  $P_1, P_2$  and  $P_3$ ), and six faecal samples (the six dots in the figure). The shaded area is the two-dimensional cone generated by vectors  $P_1, P_2$  and  $P_3$ .

relevant categories (e.g. Hulbert *et al.*, 2001; Rao *et al.*, 2003). This is settled defining a partition of the set of vectors  $P_1, P_2, \dots, P_p$  into  $d$  (the number of n-alkanes) subsets, and identifying each subset with a certain vector. Once this is done, the procedure described for case  $d = p$  is used to estimate the diet from the faecal samples  $F_1, F_2, \dots, F_q$ . The estimated diet consists of the non-negative coefficients of the linear combination of the vectors identified with the subsets of the partition.

However, there is no satisfactory criteria to decide how species should be grouped (in other words, how to define a suitable  $d$  - partition of the set of vectors  $P_1, P_2, \dots, P_p$ ), and how to weight species within each group (in other words, how to select a vector which adequately represents a whole subset of vectors). The choices made with respect to those questions will affect the final result and will ultimately determine the estimated diet.

We suggest an alternative approach to estimate diet composition that avoids these difficulties and allows arbitrary numbers of plant species and n-alkanes.

#### The model

Let  $P_1, P_2, \dots, P_p$  be the  $d$  - vectors representing the  $p$  potentially selected plant species, where  $d$  is the number of n-alkanes used, and let  $C$  be the cone generated by these vectors. If  $P$  is one of the vectors  $P_1, P_2, \dots, P_p$ , let  $C_P$  be the cone generated by the remaining vectors. We denote by  $F_i, i = 1, 2, \dots, q$ , the projection onto  $C$  of the point on the  $d$  - Euclidean space representing each of the  $q$  faecal samples, and define  $\Phi = \{F_1, F_2, \dots, F_q\}$  as the set of 'corrected' faecal samples.

We define feasible diet as every non-negative solution of the linear system  $Ax = F_i$  for any point  $F$  in  $\Phi$ , where  $A$  is the  $d \times p$  matrix with columns  $P_1, P_2, \dots, P_p$ . There may be infinitely many feasible diets. Nevertheless, the set of diets is bounded by specified upper and lower bounds on the values of each component. Linear programming is a suitable tool to address questions about dietary composition such as:

- which are the maximum and minimum proportions of each species in the diet?
- which are the maximum and minimum proportions of each functional group (e.g. grasses, legumes, browse) in the diet?
- which is the minimum proportion of a particular plant species in those diets that satisfy some nutritional requirement (e.g. levels of nitrogen intake or tannin contents)?

In particular, if  $x_p$  denotes the component of  $x$  corresponding to plant  $P$ , the knowledge of  $M(i,P) = \max\{x_p, \text{subject to } Ax = F_i \text{ and } x \geq 0\}$  and  $m(i,P) = \min\{x_p, \text{subject to } Ax = F_i \text{ and } x \geq 0\}$ , for  $i = 1, 2, \dots, q$ , allow us to answer questions regarding presence or absence of plant species in the diet. For instance:

- which species are never eaten (absent species)?
- which species are always eaten (mandatory species)?
- which species are identifiable in some faecal samples but not in all (conditional mandatory species)?

We cannot confirm if species that do not fall in any of the previous categories (optional species) are eaten or not. Contrary to the absent species, whose corresponding concentrations of n-alkanes are not recovered in any faeces, there is evidence of the chemical markers of an optional plant species in faeces. However, the concentration of n-alkanes of an optional species is identical to those derived from certain combinations of other potentially selected plants. To confirm if an optional species is actually eaten may require further fieldwork to collect evidence of the utilisation of that species.

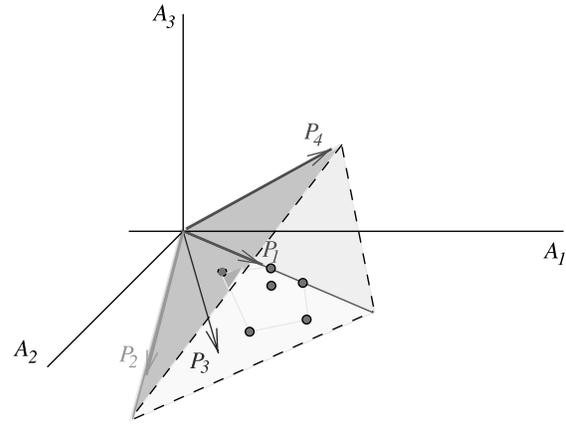
**Absent species.** Absent species are the plants not included in any diet. Species  $P$  is absent if the corresponding coefficient in every non-negative solution of  $Ax = F_i$  for every  $F_i$  in  $\Phi$ , is equal to zero. This can be determined by checking whether  $M(i,P) = 0$ , for  $i = 1, 2, \dots, q$ .

In Figure 4 species  $P_4$  is the unique absent plant.

Once the absent species have been identified they can be removed from further consideration. Figure 5 represents the example in Figure 4 after the absent plant species  $P_4$  have been removed. From now on we assume that there are no absent species.

**Mandatory species.** Mandatory species are the plants that are present in every diet. Species  $P$  is mandatory if the corresponding coefficient in every nonnegative solution of  $Ax = F_i$  for every  $F_i$  in  $\Phi$ , is greater than zero. From a geometrical point of view,  $P$  is mandatory if there is no  $F_i$  in the cone  $C_P$ . To decide if species  $P$  is mandatory corresponds to check if  $m(i,P) > 0$ , for  $i = 1, 2, \dots, q$ .

In Figure 5 species  $P_1$  is the unique plant which is mandatory.

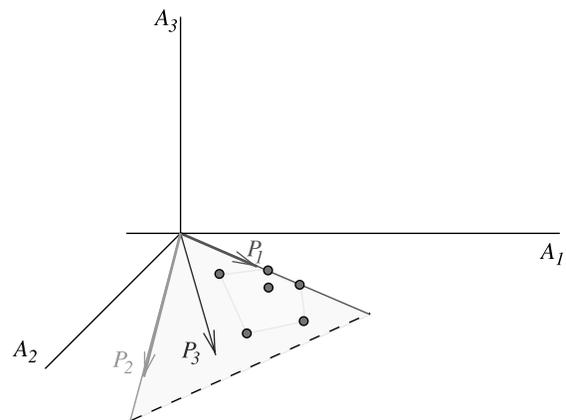


**Figure 4** Example representing three n-alkanes (axes  $A_1$ ,  $A_2$  and  $A_3$ ), four plant species (vectors  $P_1$ ,  $P_2$ ,  $P_3$  and  $P_4$ ), and six faecal samples (the six dots in the figure). The shaded area is the three-dimensional cone generated by vectors  $P_1$ ,  $P_2$ ,  $P_3$  and  $P_4$ .

**Conditional mandatory species.** Conditional mandatory species are plants that are obligatory for some diets, but not for every diet. Species  $P$  is conditional mandatory if for some, but not all,  $F_i$  in  $\Phi$  there is no non-negative solution of  $Ax = F_i$  with the coefficient corresponding to  $P$  equal to zero. Equivalently, species  $P$  is conditional mandatory if there are points of  $\Phi$  which are in  $C_P$  and others that are not. Testing if species  $P$  is conditional mandatory can be achieved by checking whether, for some but not every  $i = 1, 2, \dots, q$ ,  $m(i,P) > 0$ .

If plant  $P_2$  ( $P_3$ ) were not present in Figure 5,  $P_3$  ( $P_2$ ) would be an example of one such species.

**Optional species.** Optional species are the plants that are not in any of the previous cases. Species  $P$  is optional if for every  $F_i$  in  $\Phi$  there are non-negative solutions of  $Ax = F_i$  for which the coefficient  $P$  is equal to zero, but for at least



**Figure 5** Example representing three n-alkanes (axes  $A_1$ ,  $A_2$  and  $A_3$ ), three plant species (vectors  $P_1$ ,  $P_2$ , and  $P_3$ ), and six faecal samples (the six dots in the figure), obtained by removing  $P_4$  from the example in Fig. 4. The shaded area is the two-dimensional cone generated by vectors  $P_1$ ,  $P_2$ , and  $P_3$ .

one right hand side  $F$  a non-negative solution exists for which the coefficient of  $P$  is greater than zero. Species  $P$  is optional if ( $P$  is not absent and)  $F$  is contained in  $C_P$ . If species  $P$  is optional then  $m(i,P) = 0$ , for every  $i = 1, 2, \dots, q$ , and  $M(i,P) > 0$ , for at least one  $i$ .

In Figure 5,  $P_2$  and  $P_3$  are both optional species.

Optional species are the only plants for which we cannot conclude if they were eaten or not. Hence it may be interesting to identify a minimum size subset of plants, with no optional species, that generate a cone that still contains  $\Phi$ . This can be achieved solving a mixed integer linear programming problem (with  $pq$  continuous variables and  $p$  binary variables, and  $p(q + 1)$  constraints). It may also be relevant to examine if a given group of optional species (e.g. legumes, grasses) are eaten. If, excluding from the input every plant of the group, the cone generated by the remaining species no longer contains  $\Phi$ , we can conclude that at least one of the plants is eaten.

We tested this model on two data sets where these questions were highlighted in estimating the proportions of each plant species in the diets.

#### The data sets

The two data sets consist of records of the n-alkane profiles of plant species and faeces of sheep and deer collected in experiments conducted in Australia and Portugal, respectively.

*The Australian sheep data set.* Australian sheep data set came from a grazing experiment conducted at Ginninderra Experimental Station (35°12' S, 149°12' E) in New South Wales, Australia. This experiment aimed to assess the effects of grazing management on the productivity and persistence of Lotus species (Ayres and Blumenthal, 2000) and consists of the N-alkane profiles of four sown cultivars of Lotus (*L. corniculatus* cv. Goldie, *L. corniculatus* cv. Prostrate, *L. pedunculatus* cv. Grasslands Maku and *L. pedunculatus* cv. Sharnae), two sown grasses (*Phalaris aquatica* and *Austrodanthonia richardsonii*), and volunteer grasses (*Vulpia myuros*, *Bromus catharticus* and *Festuca arundinacea*) and legumes (*Trifolium repens*, *T. striatum* and *T. glomeratum*).

The faecal data consists of the n-alkane composition of faecal samples collected from four different sheep that grazed the study area for 5 days. After a preliminary period of 2 days to allow passage of faeces from previously grazed material, four rectal grab samples of faeces were obtained from each sheep over three consecutive days. Faecal samples were bulked within sheep across days, resulting in one bulk faeces sample per sheep. These samples were stored at  $-18^\circ\text{C}$  before sample preparation and analysis as described in Kelman *et al.* (2003).

During the grazing period, samples of whole plants of the sward species were harvested from random sites within the sward, using electric clippers. For each species, the harvested material was bulked until there was sufficient amount for chemical analysis and then stored at  $-18^\circ\text{C}$ .

Faecal and plant samples were freeze dried and ground prior to the extraction of cuticular wax n-alkanes by direct saponification, followed by solvent separation and purification through silica gel. Purified N-alkanes were quantified by gas chromatography relative to docosane (C22) and tetratriacontane (C34) N-alkanes as internal standards. Faecal n-alkane recovery corrections were based on those reported by Dove and Oliván (1998).

Proportions of plant species available at the sward, before the grazing trial, were 0.31 and 0.07 of the legumes *Lotus* spp. and *Trifolium* spp., respectively, and 0.27, 0.17 of the grasses *Phalaris aquatica* and *Vulpia myuros*, respectively. Remaining proportion (0.18) of the sward was composed by the grasses *A. richardsonii*, *B. catharticus* and *Festuca arundinacea* (Kelman *et al.*, 2003).

*The Portuguese deer data set.* This data set was collected in a study area located in Vila Viçosa (38°47' N, 7°25' W) in southern Portugal as part of a general programme on the foraging behaviour of red deer (*Cervus elaphus*) in a Mediterranean environment. Samples of leaf and stem of cork oak (*Quercus suber*), holm oak (*Quercus rotundifolia*), gum cistus (*Cistus ladanifer*) blackberry (*Rubus ulmifolius*), were collected in the beginning of summer (June).

About 30 g dry weight of a sample of leaf and stem (diameter not greater than 0.5 cm) of each browse species was collected from 20 different individuals and pooled within species. Samples were kept in sealed polythene bags, tagged, and stored at  $-15^\circ\text{C}$  in a freezer until they were analysed.

Samples of species of the herbage layer, consisting of about 30 g of dry weight, were also collected in the study area where deer had been seen feeding previously. These samples, consisting of approximately 30 g dry weight of plant leaf and stem, were collected from the following species: *Vulpia bromoides*, *Phalaris brachystachys*, *Holcus lanatus*, *Gaudinia fragilis*, *Agrostis porreti*, *Briza maxima*, *Avena barbata*, *Lolium multiflorum* and *Bromus hordeaceus* (grasses); *Trifolium arvensis*, *Trifolium angustifolium*, *Vicia sativa*, *Trifolium campestre*, *Trifolium subterraneum* and *Ornithopus compressus* (legumes); *Chamameleum mixtum* and *Coleostepus micorrhis* (Asteraceae) *Spergularia purpurea* and *Echium plantagineum* (other species).

Nine samples of fresh faeces of red deer were also collected. Each sample consisted of the full group of faecal pellets found at a particular location. These samples were kept in sealed polythene bags, tagged, and stored at  $-15^\circ\text{C}$  until they were oven dried at  $60^\circ\text{C}$  for 48 h.

After milling through a 1-mm mesh, plant and faecal samples were analysed for n-alkane composition by gas chromatography using the method of Mayes *et al.* (1986) with the modification described by Salt *et al.* (1992). n-alkanes C22 and C34 were used as internal standards.

Proportions of plant species available at the sward were 0.50, 0.25, 0.22 and 0.02 of grasses, legumes, Asteraceae and 'other species', respectively (Bugalho and Milne, 2003).

## Results

The computational results were obtained using Microsoft Excel with the LINDO Systems What'sBest! solver for the optimisation problems (a solver tolerance of  $10^{-6}$  was given).

In both data sets all the faecal samples were slightly out of the cone generated by the plants. This could result from laboratory errors in measuring the concentrations of the markers. Hence projections onto the cone had to be performed. This was achieved, using quadratic programming, by finding the point in the cone which is closest (the Euclidean distance was used) to each faecal sample. The 'projected faeces' obtained this way differed at most 2% on each coordinate from the original ones, which is a value below the usually admitted analytical error of 3 to 4% when measuring n-alkane concentrations (H. Dove, personal communication). For the purpose of classification of the plant species (mandatory, absent, cond.mandatory) we assumed that values per unit less than 0.01 are equal to zero, to cope with round-off errors of the solver. The contributions of each plant species in the diets are expressed as proportions.

### The Australian sheep data set

For this example we used nine n-alkanes (C25 to C33) and 12 plant species. There were two mandatory legumes (*T. repens* and *T. striatus*) and three mandatory grasses (*Vulpia myuros*, *Bromus catharticus* and *Festuca arundinacea*) in the diet of sheep. Among mandatory legumes *T. repens* had the highest proportion in the diet (proportions varying between 0.25 and 0.34) while *T. striatus* (0.05 and 0.10) was consumed at comparatively lower proportions. Mandatory grasses *V. myuros* (0.09 to 0.23 of the diet), *B. catharticus* (0.14 to 0.27) and *F. arundinacea* (0.11 to 0.27) occurred with approximately similar proportions in the diet (Table 1).

There were four conditional mandatory legumes (*T. glomeratum*, *L. pedunculatus* Sharnae, *L. pedunculatus* Maku

and one conditional mandatory grass (*F. arundinacea*) in the diet. When occurring, cluster clover, *L. pedunculatus* Sharnae and *L. pedunculatus* Maku reached 0.03, 0.04 and 0.02, respectively, of the diet, but were not detectable in remaining samples. The grass *P. aquatica* reached 5% in faecal F1 but was absent from remaining samples (Table 1).

Both varieties of *L. corniculatus* (legumes) and *A. richardsonii* (a grass) were absent from the diet. No species were classified as optional (Table 1). The minimum and maximum proportions of each plant species corresponding to each  $F_i$  in Table 1 almost coincide. Actually, the small differences between these values are certainly due to round-off computational errors. Note that since three species are absent from the diet, the nine remaining intervenient plant species equal the number of n-alkanes. Consequently (S) becomes a square linear system with a unique solution.

### The Portuguese deer data set

In this example we used nine n-alkanes (C25 to C33) and 17 plant species. There were three mandatory legumes (*Trifolium arvensis*, *T. subterraneum* and *Vicia sativa*) a mandatory Asteraceae (*Chamameleum mixtu*) and a mandatory browse species (*Cistus ladanifer*) in the diet of red deer. The mandatory legumes *T. arvensis*, *V. sativa* and *T. subterraneum* had proportions of 0.03 to 0.22, 0.05 to 0.30 and 0.02 to 0.09, respectively. The only mandatory browse species, *C. ladanifer*, varied between 0.21 and 0.42 whilst *C. mixtum* (Asteraceae) varied between 0.04 and 0.17 of the diet (Table 2).

There were six conditional mandatory species: *T. angustifolium*, *Ornithopus compressus* (legumes), *Lolium multiflorum* (grass), *Coleostepus mycorris* (Asteraceae), *Quercus suber* (browse) and *Echium plantagineum* (an herb). As compared with other conditional mandatory species *T. angustifolium* occurred in most of the samples and reached relatively high proportions in some of them (e.g. faecal samples  $F_2$  and  $F_9$ , with 0.42 and 0.31, respectively).

**Table 1** Results for the Australian sheep data set: classification of plant species, and minimum and maximum proportions of each plant species in the diet of sheep (values less than 0.01 appear as 0;  $F_i$  represents each faecal sample projection)

	Status <sup>†</sup>	$F_1$		$F_2$		$F_3$		$F_4$	
		Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
<i>L. corniculatus</i> Prostrate	absent	0	0	0	0	0	0	0	0
<i>L. corniculatus</i> Goldie	absent	0	0	0	0	0	0	0	0
<i>L. pedunculatus</i> Sharnae	cond. mand.	0.04 518	0.04 518	0	0	0.01 374	0.01 374	0.01 812	0.01 812
<i>L. pedunculatus</i> Maku	cond. mand.	0	0	0	0	0.01 972	0.01 972	0	0
<i>A. richardsonii</i>	absent	0	0	0	0	0	0	0	0
<i>P. aquatica</i>	cond. mand.	0.04 810	0.04 810	0	0	0	0	0	0
<i>V. myuros</i>	mand.	0.22 702	0.22 702	0.09 375	0.09 375	0.12 632	0.12 632	0.12 293	0.12 293
<i>Bromus catharticus</i>	mand.	0.25 366	0.25 366	0.26 611	0.26 611	0.13 777	0.13 777	0.21 798	0.21 798
<i>F. arundinacea</i>	mand.	0.11 879	0.11 879	0.18 392	0.18 392	0.23 013	0.23 013	0.23 105	0.23 105
<i>T. striatus</i>	mand.	0.05 146	0.05 146	0.08 413	0.08 413	0.09 908	0.09 908	0.04 962	0.04 962
<i>T. glomeratum</i>	cond. mand.	0	0	0	0	0	0	0.03 343	0.03 343
<i>T. repens</i>	mand.	0.24 747	0.24 747	0.36 785	0.36 785	0.37 324	0.37 324	0.32 686	0.32 686

<sup>†</sup> Cond. = conditional; mand = mandatory.

**Table 2** Results for the Portuguese deer data set: classification of plant species, and minimum and maximum proportions of each plant species in the diet of deer (values less than 0.01 appear as 0; F<sub>1</sub> represents each faecal sample projection)

	Status <sup>†</sup>	F <sub>1</sub>		F <sub>2</sub>		F <sub>3</sub>		F <sub>4</sub>		F <sub>5</sub>		F <sub>6</sub>		F <sub>7</sub>		F <sub>8</sub>		F <sub>9</sub>		
		Min.	Max.																	
<i>V. bromoides</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>S. purpurea</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>H. lanatus</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>A. barbata</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>L. multiflorum</i>	cond. mand.	0.03 895	0.03 926	0	0	0	0	0	0	0	0	0	0	0	0	0.02 044	0.02 082	0	0	0
<i>E. plantaginum</i>	cond. mand.	0.06 156	0.06 184	0	0	0	0	0	0	0	0	0	0	0.19 162	0.19 175	0.0928	0.09 314	0	0	0
<i>C. mixtu</i>	mand.	0.0444	0.0445	0.12 092	0.12 097	0.02 143	0.02 148	0.0772	0.07 723	0.11 132	0.11 136	0.17 312	0.17 317	0.07 724	0.07 728	0.12 361	0.12 374	0.05 581	0.05 585	0
<i>C. mycorris</i>	cond. mand.	0.09 946	0.0997	0	0	0.04 612	0.0463	0	0	0	0	0	0	0	0	0.12 212	0.1224	0	0	0
<i>T. arvensis</i>	mand.	0.15 865	0.15 873	0.04 913	0.04 916	0.21 216	0.21 218	0.2183	0.21 835	0.17 349	0.17 354	0.1769	0.17 695	0.13 009	0.13 013	0.1533	0.15 339	0.03 098	0.03 103	0
<i>T. angustifolium</i>	cond. mand.	0	0	0.42 273	0.42 289	0	0	0.19 022	0.19 035	0.2634	0.26 354	0.11 155	0.11 169	0.21 499	0.21 513	0	0	0.31 929	0.31 955	0
<i>V. sativa</i>	mand.	0.30 179	0.30 179	0.05 837	0.05 841	0.14 611	0.14 613	0.04 807	0.0481	0.12 263	0.12 267	0.14 662	0.14 667	0.11 466	0.1147	0.12 501	0.12 504	0.21 039	0.21 049	0
<i>T. subterraneum</i>	mand.	0.08 733	0.08 735	0.06 268	0.06 269	0.03 966	0.03 967	0.04 723	0.04 723	0.08 467	0.08 467	0.04 897	0.04 898	0.02 464	0.02 465	0.07 467	0.07 469	0.01 737	0.01 737	0
<i>O. compressus</i>	cond. mand.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02 615	0.02 618	0
<i>Q. suber</i>	cond. mand.	0	0	0.07 815	0.07 818	0.19 483	0.19 489	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Q. rotundifolia</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>C. ladanifer</i>	mand.	0.20 716	0.20 717	0.20 771	0.20 771	0.33 947	0.33 949	0.41 873	0.41 871	0.24 422	0.24 421	0.34 254	0.34 253	0.24 646	0.24 646	0.28 719	0.2872	0.33 954	0.33 951	0
<i>R. ulmifolius</i>	absent	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

<sup>†</sup> Cond. = conditional; mand = mandatory.

Remaining mandatory species were detectable in a lower number of samples but some still reaching relatively high values.

This is the case of *Q. suber* and *E. plantagineum* that occurred with proportions of 0.19 in faecal samples  $F_3$ , and  $F_8$ , respectively, or *C. mycorris* with 0.12 in faecal samples  $F_9$  (Table 2). *O. compressus* and *L. multiflorum* have always very low proportions and may have been misclassified as conditional mandatory. *Vulpia bromoides*, *Holcus lanatus*, *Avena barbata* (grasses), *R. ulmifolius*, *Quercus rotundifolia* (browse) and *Spergularia purpurea* (herb) were absent from the diet of red deer. There were no species classified as optional (Table 2). As for the case of the Australian sheep data set, the minimum and maximum proportions of each plant species corresponding to each  $F_i$  in Table 2 are similar. From the original set of 17 plant species, two are absent and two other (*O. compressus* and *L. multiflorum*) have very low contributions, suggesting that only nine plant species (as many as the number of n-alkanes) are effective in solving the linear system (S). This explains the observed similarity of the minimum and maximum values.

## Discussion

The approach described in the present paper does not limit the number of dietary components to be identified in the diet of herbivores. This is particularly useful when using n-alkanes to estimate diet composition of animals feeding on complex plant communities. In rangelands or other natural systems, where herbivores have large numbers of plant species available for selection, the method described may provide relevant information to the analyst by identifying plants that are certainly eaten and others which are excluded from the diet. Additionally, by giving information on minimum and maximum proportions of a particular plant species in the faeces, the method provides an approach to quantification of diet composition. Two real data sets allowed testing of the approach described.

### *The Australian sheep data set*

Our model suggested strong selection of clovers by sheep and particularly of *T. repens*. From proportions of 0.07 in the sward (see methods), *T. repens* was found in the diet, for instance is at proportions of 0.37. High consumption of *T. repens* is in agreement with results shown by Kelman *et al.* (2003) for the same data set. Conversely, Lotus species were not mandatory and, when occurring in the diet, were consumed at very low proportions, particularly when compared with proportions available at the sward. A maximum of 0.04 of the diet was reached by *L. pedunculatus* Sharnae. Lotus species are characterised by the presence of tannins in their stem and leaves (Reed, 1995; Van Soest, 2004). Tannins are anti-nutritional components of plant species which may depress diet digestibility (Van Soest, 2004). As a consequence, plants rich in tannins, such as *Lotus* spp. are expected to be avoided by sheep. If ingested

at low levels, however, tannins may form complexes with proteins, protecting them from microbial enzymes in the rumen and increasing protein digestibility in the lower digestive tract (Reed, 1995; Bento, 2004). Other potential positive benefits such as antioxidant, antibacterial and anti-helminthic properties could account for the intake of low doses of tannins (Clauss *et al.*, 2003; Iason, 2005). Low proportions of *Lotus* spp. in the diet of sheep, as found here, were not unlikely. Indeed Verheyden-Tixier *et al.* (2000) found that roe deer (*Capreolus capreolus*), a wild ruminant, actively included a certain amount of tannins in their daily ration. It is known that concentrations of condensed tannins are genetically determined and may vary widely among varieties of the same plant species. The present approach potentially allowed the differentiation of the two varieties of *L. pedunculatus* Sharnae and Maku in the diet of sheep that may be of potential interest. Grasses were included in the diet at proportions similar to their availability in the sward as could also be expected.

### *The Portuguese deer data set*

Our computational results suggest that deer consumed different species of legumes. Although we do not have data on availability of individual species, the group of legumes comprised 0.25 of the sward (Bugalho and Milne, 2003) and for legumes such as *Trifolium angustifolium* or *Vicia sativa* proportions in the diet were in some cases well above 0.25, suggesting that deer were selecting legumes for their diets. A grazing exclusion experiment in the same study site has also shown that the proportion of legumes were significantly higher in areas where deer grazing had been excluded which again suggests selection of legumes by red deer (Bugalho *et al.*, 2006). Legumes such as *V. sativa* and *T. subterraneum*, two cultivated and ameliorated species, were mandatory in red deer diet. This could be expected as cultivated species generally have higher nutritive value and are preferentially selected than wild varieties. With respect to browse, only a shrub (*C. ladanifer*) was mandatory in the diet, as estimated by our model, with *Q. suber* (conditional mandatory) found only in one faecal sample and *Q. rotundifolia* absent from the diet. In Mediterranean environments, such as that occurring in the Portuguese deer study area, browse species are predominantly consumed by ruminants between mid-summer to the end of summer, when most of the herbage layer is senescent and of low nutritive value (Seligman, 1996; Bugalho and Milne, 2003). Indeed present data was collected in the beginning of June, when a proportion of green plant material was still available in the herbage layer and thus a very high consumption of browse should not be expected.

## Conclusions

Previous approaches for estimating diet composition using n-alkanes require that the number of markers is greater or

equal than the number of dietary components. However, this constraint is not realistic in natural or semi-natural systems where a high number of plant species is generally available for selection by the herbivores and usually exceeds the number of plant markers. While the method described in this paper needs further validation with additional data sets, it gives relevant information to the analyst by identifying plants which are certainly eaten and others which are excluded from the diet. Furthermore, since it provides minimum and maximum proportions of each particular plant species in the faeces, the method also allows us to quantify the diet composition.

## Acknowledgements

We thank Hugh Dove and Walter Kelman for permission to use the Australian sheep data set. We also thank Kathleen Merrill for corrections on an early version of the manuscript. This work was partially supported by Fundação para a Ciência e Tecnologia (FCT), EU FEDER and POCTI Program, namely by Post-Doctoral grant SFRH 727/2000 and Project POCTI/AGR/63 322/2004 (MNB), Center for Logic and Computation (MLC), and Matemática Aplicada IISA (JOC).

## References

- Ayres JF and Blumenthal MJ 2000. Lotus grazing management for weaner production. Final report, project DAN082. Meat and Livestock Australia, New South Wales Agriculture, Australia.
- Bento H 2004 The effects of condensed tannins in rumen and post-rumen digestion of nutrients. PhD thesis, University of Aberdeen.
- Bugalho MN, Lecomte X, Pile S and Caldeira MC 2006. Effects of deer on oak regeneration and species composition of a Mediterranean grassland. *Revista de Ciências Agrárias* In press.
- Bugalho MN and Milne JA 2003. The composition of the diet of red deer (*Cervus elaphus*) in a Mediterranean environment: a case of summer nutritional constraint? *Forest Ecology and Management*. 181, 23-29.
- Bugalho MN, Milne JA and Racey PA 2001. The foraging ecology of red deer (*Cervus elaphus*) in a Mediterranean environment: is a larger body size advantageous? *Journal of Zoology (London)* 255, 285-289.
- Clauss M, Lason K, Gehrke J, Lechner-Doll M, Fickel J, Grune T and Streich WJ 2003. Captive roe deer (*Capreolus capreolus*) select for low amounts of tannic acid but not quebracho: fluctuation of preferences and potential benefits. *Comparative Biochemistry and Physiology Part B* 136, 369-382.
- Dove H and Mayes RW 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 and Mayes RW 1996. Plant wax components: a new approach to estimating intake and diet composition in herbivores. *Journal of Nutrition* 126, 13-26.
- Dove H, Mayes RW and Freer M 1996. Effect of species, plant part and plant age on the n-alkane concentrations in the cuticular wax of pasture plants. *Australian Journal of Agricultural Research* 47, 1333-1347.
- Dove H and Moore AD 1995. Using a least-squares optimization procedure to estimate botanical composition based on the n-alkanes of plant cuticular wax. *Australian Journal of Agricultural Research* 46, 1535-1544.
- Dove H and Oliván M 1998. Using synthetic or bees-wax alkanes for estimating supplement intake in sheep. *Animal Production in Australia* 22, 189-192.
- Duncan AJ, Mayes RW, Lamb CS, Young SA and Castillo I 1999. The use of naturally occurring and artificially applied n-alkanes as markers for estimation of short-term diet composition and intake in sheep. *Journal of Agricultural Science, Cambridge* 132, 233-246.
- Hulbert IA, Iason GR and Mayes RW 2001. The exibility of an intermediate feeder: dietary selection by mountain hares measured using n-alkane analysis. *Oecologia* 129, 197-205.
- Hutchings MR, Gordon IJ, Robertson E, Kyriazakis I and Jackson F 2000. Effects of parasitic status and level of feeding motivation on the diet selected by sheep grazing grass/clover swards. *Journal of Agricultural Science* 135, 65-75.
- Iason GR 2005. The role of plant secondary metabolites in mammalian herbivory: ecological perspectives. *Proceedings of the Nutrition Society* 64, 124-131.
- Kelman WM, Bugalho MN and Dove H 2003. Cuticular wax alkanes and alcohols used as markers to estimate diet composition of sheep (*Ovis aries*). *Biochemistry Systematics and Ecology* 31, 919-927.
- Mayes RW, Beresford NA, Lamb CS, Barnett CL, Howard BJ, Jones B-E V, Eriksson O, Hove K, Pedersen O and Staines BW 1994. Novel approaches to the estimation of intake and bioavailability of radiocaesium in ruminants grazing forested areas. *Science of the Total Environment* 157, 289-300.
- Mayes RW and Dove H 2000. Measurement of dietary nutrient intake in free-ranging mammalian herbivores. *Nutrition Research Reviews* 13, 107-138.
- Mayes RW, Lamb CS and Colgrove PM 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.
- Rao SJ, Iason GR, Hulbert IAR, Mayes RW and Racey PA 2003. Estimating diet composition for mountain hares in newly established native woodland: development and application of plant-wax faecal markers. *Canadian Journal of Zoology* 81, 1047-1056.
- Reed JD 1995. Nutritional toxicology of tannins and related polyphenols in forage legumes. *Journal of Animal Science* 73, 1516-1528.
- Salt CA, Mayes RW, Colgrove PM and Lamb CS 1994. The effects of season and diet composition on the radiocaesium intake by sheep grazing heather moorland. *Journal of Applied Ecology* 31, 125-136.
- Salt CA, Mayes RW and Elston DA 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.
- Seligman NG 1996. Management of Mediterranean grasslands. In *The ecology and management of grazing systems* (ed. J. Hodgson and AW Illius), pp. 359-392. CAB International, Wallingford.
- Van Soest PJ 1994. *Nutritional ecology of the ruminant*, fourth edition. O&B Books, Corvallis, OR.
- Verheyden-Tixier H, Duncan P, Scehovic J, Yani A, Gleizes M and Lila M 1997. Food selection by European roe deer (*Capreolus capreolus*): effects of plant chemistry, and consequences for the nutritional value of their diets. *Journal of Zoology (London)* 242, 229-245.