Biogeochemistry (2005) 72: 1–34
DOI 10.1007/s10533-004-0076-3


# Estimating the molecular composition of a diverse range of natural organic materials from solid-state $^{13}$C NMR and elemental analyses

PAUL N. NELSON[1] and JEFFREY A. BALDOCK[2,*]
*[1]James Cook University, Cairns Qld 4870, Australia; [2]CSIRO Land and Water/CRC for Greenhouse Accounting, Glen Osmond SA 5064, Australia; *Author for correspondence (e-mail: jeff.baldock@csiro.au; phone: +61-8-8303-8537; fax: +61-8-8303-8550)*

Received 10 December 2001; accepted in revised form 7 February 2004

**Key words:** Humus, Molecular composition, Nuclear magnetic resonance, Organic matter chemistry, Sediment, Soil

**Abstract.** Most techniques for determining the chemical nature of natural organic matter in soil, sediment and water require prior extraction or concentration steps that are not quantitative and that create artifacts. $^{13}$C nuclear magnetic resonance (NMR) analysis can avoid these problems, but it gives little information at the scale of molecules. Here we show that the molecular composition of a diverse range of natural organic materials could be inferred from $^{13}$C NMR analysis combined with C and N analysis. Forty-six different organic materials including undecomposed and decomposed plant materials, soil organic matter, phytoplankton, and the organic matter found in freshwater, estuarine and marine sediments were examined. A mixing model simultaneously solved a series of equations to estimate the content of four biomolecule components representing the organic materials produced in greatest abundance by plants and other organisms (carbohydrate, protein, lignin and aliphatic material) and two additional components (char and pure carbonyl). Based on defined molecular structures for each component, signal intensities for $^{13}$C NMR spectra were predicted and compared with measured values. The sum of the absolute differences in signal intensity between the measured and predicted spectral regions was <7% for the terrestrial materials. For aquatic materials the fit of the predicted to measured signal intensities was not as good. Predicted molecular compositions correlated well with independent analyses of cellulose, protein and lignin contents of plant samples and char contents of soil samples. Across all samples, carbohydrates accounted for 10–76% of the sample C (40–76% in plants and 10–42% in soils, sediments and phytoplankton), protein for 2–80% (21–80% in phytoplankton and marine water column samples and 2–36% in plants, soils and sediments), lignin for 0–36%, aliphatic materials for 2–44%, char for 0–38% and carbonyl for 0–22%. For the soils, sediments and decomposed plant materials, the close correspondence between actual signal intensities and those predicted using known biomolecular components, suggested that either 'humic' structures can be approximated by mixtures of common biologically derived molecules or that humic structures did not exist in significant amounts.

**Abbreviations:** CP – cross-polarisation; DP – Direct polarisation; MAS – magic-angle-spinning; NMR – nuclear magnetic resonance; UV NMR – ultra-violet photo-oxidation of the <53$\mu$m fraction followed by $^{13}$C NMR analysis

## Introduction

Natural organic matter in soil, water and sediments plays an important role in many ecosystem functions. It provides energy and nutrients for organisms; acts as an important pool of elements in global cycles; buffers against chemical changes such as

acidification; alters the mobility and reactivity of pollutants such as heavy metals and synthetic organic chemicals; and influences the physical properties of soils and sediments (Baldock and Skjemstad 1999). Due to its complex chemical nature, and its ability to form strong associations with mineral materials, natural organic matter has proved difficult to characterise chemically, and despite its importance, its chemical nature is not yet fully understood.

Traditional approaches to characterising the molecular composition of natural organic materials have relied on a variety of prior extraction, concentration and chemical degradation techniques. These techniques include both general procedures that extract a range of different organic molecules and more specific procedures designed to quantify individual classes of biomolecules. The most common general extraction procedure uses strong alkaline and acidic solutions to extract and fractionate organic material into 'humic acid', 'fulvic acid' and 'humin'. Application of a variety of analytical procedures to the isolated materials has identified a large range of known biochemicals and left substantial portions of the materials unidentified. Characterisation of the unidentified materials has lead to the development of generic chemical 'humic' structures thought to be representative of decomposed organic materials in natural systems (Schulten and Schnitzer 1993) and of various pathways of humic substance formation (Waksman 1936; Hatcher and Spiker 1988; Hedges 1988; Baldock and Nelson 2000).

Specific extraction techniques have been developed for protein, carbohydrate, lignin and lipid fractions. Protein content has been estimated by determining the quantity of amino acids released by hydrolysis in hydrochloric acid (Beavis and Mott 1996, 1999). Carbohydrate content has been estimated by quantification of the products of sulphuric acid hydrolysis (Baldock et al. 1987). Lignin has been estimated by quantification of the products of CuO oxidation (Hedges and Ertel 1982; Reeves 1995) and tetramethylammonium hydroxide thermochemolysis (Hatcher et al. 1995; Baldock et al. 1997a). A variety of organic solvents and solvent mixes have been used to extract and quantify various lipid fractions (Heng and Goh 1981; Ma'shum et al. 1988; Wu et al. 1995). Several of these specific extraction techniques can be combined, as in the proximate-analysis approach, in an attempt to completely define the molecular composition of organic materials (Preston et al. 1997).

A problem with using specific and general extractions prior to analysis of the molecular composition of natural organic materials is that the procedures are not quantitative or selective and that artifacts can form. For example, Preston et al. (1997) showed that the Klason lignin fraction isolated from various forest litter samples by a proximate-analysis approach contained significant amounts of lipid carbon. Typically, selective chemical degradation techniques identify only 20–80% of organic materials present in natural samples e.g., litter and soil organic horizons examined by Ogner (1985), Kögel et al. (1988), Kögel-Knabner et al. (1988) and marine organic materials examined by Wakeham et al. (1997).

There are two possible reasons why identification of known biochemicals does not account for all the organic matter in samples of soil, sediment and water. The first is that much of the organic matter may be in the form of complex 'humic'

molecules of the type proposed by Schulten and Schnitzer (1993). The second is that the organic matter may consist of potentially identifiable molecules, but they are associated with each other and with mineral particles so intimately that they cannot be fully and non-destructively isolated and identified. In order to determine which of the two possibilities is correct, we need to examine the nature of the organic materials quantitatively and *in-situ*, without prior extraction or degradation.

Recent advances in the use of solid-state $^{13}C$ nuclear magnetic resonance (NMR) as a tool for examining the chemical structure of organic materials in natural environments has allowed non-destructive and potentially quantitative examinations of organic carbon chemistry *in-situ*, without prior extraction and/or degradation procedures (Wilson 1987; Kinchesh et al. 1995a, b; Smernik and Oades 2000a, b). Although the results of solid-state $^{13}C$ NMR can be used to quantitatively assess the structural environment of each C atom, provided correct NMR parameters are first defined, no direct information pertaining to how the various C atoms are joined together into molecular structures is provided. Therefore to date, the results of NMR studies have usually been reported in terms of C types, rather than molecular structures.

A method of inferring molecular structures of natural organic matter from $^{13}C$ NMR spectra was formulated by Nelson et al. (1999) and subsequently also applied to marine plankton and particles sinking through the marine water column (Hedges et al. 2001, 2002). The method begins by assuming that the organic matter in natural ecosystems exists in the form of molecules produced by organisms (biomolecules). More than 90% of all biomolecules can be classified into a few distinct types (e.g., proteins, carbohydrates, lipids, etc.), with each type having a characteristic $^{13}C$ NMR spectrum and N:C ratio. To predict the molecular composition of a sample (the content of each type of biomolecule), the distributions of $^{13}C$ NMR signal intensity defined for each biomolecule can be mathematically mixed together so that the N:C ratio, and the signal intensity in some regions of the NMR spectrum are equated to that measured for the sample. This mixing process estimates (models) the proportion of organic C found in a sample of organic matter that can be attributed to each distinct type of biomolecule. The deviation between predicted and actual signal intensities for the remaining regions can be used to give an indication of the goodness of fit between the distribution of $^{13}C$ NMR signal intensity measured for the sample and that predicted for the modelled mixture of biomolecules. A high level of agreement between the measured and modelled distributions of $^{13}C$ NMR signal intensity suggests that the organic matter in the samples might consist purely of biomolecules, rather than 'humic' structures formed by other abiotic means.

In this study, we set out to extend our previous use of this biomolecular mixing model (Nelson et al. 1999; Hedges et al. 2001, 2002) to determine whether the chemistry of natural organic matter in samples taken from a more diverse range of environments and exhibiting various extents of decomposition could be explained by a mixture of common biomolecules. It is proposed that, if the biomolecular mixing model can successfully account for the distribution of $^{13}C$ NMR signal intensity in samples of natural organic material, then the modelling process will

provide a useful extension of current NMR methodologies and allow the overall gross molecular composition of natural organic materials to be estimated.

## Material and methods

### *Samples*

The 46 samples of natural organic matter examined in this study included fresh and decomposed plant materials, soils, soil particle size fractions, phytoplankton, aquatic sediments (freshwater, estuarine and marine), marine plankton, and particles in the marine water column. For soil and sediment samples, low concentrations of carbon and the presence of paramagnetic minerals can make the acquisition of solid-state $^{13}C$ NMR spectra difficult. Some samples, as indicated in the subsequent descriptions of sample origins, were demineralised prior to application of the elemental and solid-state $^{13}C$ NMR analyses.

Several pasture grasses and legumes covering a range of digestibilities were analysed by $^{13}C$ NMR and for molecular components using conventional methods typical of those used to assess ruminant feed quality. The samples included an oaten hay (*Avena sativa*) from southern Australia, and Seca stylo (*Stylosanthes scabra*) stems, wet season Buffel grass (*Cenchrus ciliaris*) and dry season native grasses (mixed species) from northern Australia. The conventional feed quality measurements included crude protein (total Kjeldahl nitrogen multiplied by 6.25) and crude cellulose and lignin by the van Soest method (Goering and van Soest 1972). Crude cellulose, which includes hemicellulose, was determined by subtracting 'acid detergent lignin' from 'neutral detergent fibre'. Crude lignin, which includes aliphatic material such as cutin, was the 'acid detergent lignin' fraction. Additional samples of plant material and decomposed plant material included: (1) corn (*Zea maize* L.) and alfalfa (*Medicago sativa*) residues collected at the end of the growing season from a research site near Winchester, Ont. Canada (45°04′ N, 75°21′ W) and (2) maple (*Acer saccharum*) leaves collected on litter traps placed in a woodlot located approximately 3 km from the agricultural research site. Samples of the corn litter that had been placed in 15 cm×10 cm litter bags with 1 mm openings and buried at 15 cm depth in a corn and forest soil (equivalent to 7 Mg $ha^{-1}$) for 0, 306, 387 and 582 days were also examined.

The soils examined originated from Canada and Australia. The Canadian soils were collected from the 0–10 cm layer of a Brookston clay loam (280 g $kg^{-1}$ sand, 350 g $kg^{-1}$ silt, and 370 g $kg^{-1}$ clay) near Woodslee, Ont. Canada (42°13′ N, 82°44′ W). The corn and grass pasture soils were collected from the 0–10 cm layer of research plots supporting corn in monoculture and bluegrass (*Poa pratensis* L.) in monoculture from 1959 until sample collection in 1995. The maple forest soil was collected from the 0–10 cm layer of mineral soil under a remnant mixed deciduous woodlot adjacent to the corn and grass pasture research plotsfs, All Canadian soils were examined in their whole state (no size separation or demineralisation procedure was applied). The eight Australian soils were collected from the 0–5 or 0–10 cm depth layer from different environments in Australia. $^{13}C$ NMR

spectra had been obtained following pre-treatment with 2% HF according to Skjemstad et al. (1994) to remove paramagnetic materials. Details of the classification, location, properties and preparation of the eight Australian soils were described by Skjemstad et al. (1999b). Soil particle size fractions of a Mollisol from Millicent, South Australia (37°36′ S, 140°22′ E) were examined using the data of Baldock et al. (1992).

The phytoplankton *Botryococcus braunii*, *Chorelia pyrenosidosa*, *Dunaliella tertiolecta* and *Scenedesmus obliquus* were examined using the data of Zelibor et al. (1988). The Lake Washington sediment (0–30 cm layer) was collected in 1981 from the northern part of Lake Washington (47°39.0′ N, 122°15.3′ W, Washington State, USA) under 59 m of water using a box core. The Chesapeake Bay sediment sample (37°11′ N, 76°11′ W, Virginia, USA) and Buffalo River sediment sample (Buffalo, New York, USA) were standard reference materials of the US National Institute of Standards and Technology (reference numbers 1646a and 8704, respectively). All three sediment samples may have been anthropogenically influenced, and the Chesapeake Bay and Buffalo River samples were demineralised according to Gélinas et al. (2001) prior to analysis.

The marine surface plankton, sinking particles and sea floor sediment samples were depth sequences collected from the Arabian Sea (Lee et al. 1998) and the Equatorial Pacific Ocean (Wakeham et al. 1997). These depth sequences are typical of open ocean regions in terms of organic composition and disappearance of organic matter with depth (Hedges et al. 2001). The water column samples were analysed intact, but the sediment samples were demineralised according to Gélinas et al. (2001) prior to analysis.

### *$^{13}$C NMR spectroscopy*

We obtained solid-state $^{13}$C NMR spectra of the plant materials (fresh and decomposed), Canadian soils, freshwater and estuarine sediments, and marine samples in the following way. Spectra were acquired at a $^{13}$C frequency of 50.3 MHz on a Varian Unity 200 spectrometer. Samples were ground to a fine powder, packed into a 7 mm diameter cylindrical zirconia rotor with Kel-F end-caps and spun at 5000±100 Hz in a Doty Scientific MAS probe. A conventional cross-polarisation/magic-angle-spinning (CP/MAS) pulse sequence (Wilson 1987) was used with a 1.0 ms contact time. Variable contact time experiments (Wilson 1987) were used to establish the optimum contact time. An inversion recovery pulse sequence (Wilson 1987) was used to estimate relaxation times (T1H values) for each sample analysed and thereby define the duration of the recycle delay required between pulses (seven times the longest T1H), Spectral distributions (the distribution of total signal intensity amongst various chemical shift ranges) for each sample were calculated as documented by Baldock and Smernik (2002), by integrating the signal intensity in seven chemical shift regions based on previous work (Levy et al. 1980; Wilson 1987), The chemical shift regions used and their corresponding labels were: carbonyl (210–165 ppm), O-aromatic (165–145 ppm), aromatic

(145–110 ppm), $O_2$-alkyl (110–95 ppm), O-alkyl (95–60 ppm), *N*-alkyl/methoxy (60–45 ppm), and alkyl (45 to −10 ppm). The labels only indicate major types of C found in each region. It is acknowledged that some overlap of C types in adjacent regions may exist and that the use of discrete chemical shift values may not handle shoulders on larger peaks very well. However, the spectral distributions used for the model components included the overlap that occurs for those materials. It was critical that the same discrete set of chemical shift limits was used to describe the distribution of signal intensity for each of the model components and the samples examined in this study. This was necessary to ensure consistency in the mathematical process used to estimate molecular composition of the samples based on the distribution of signal intensity in the model components.

Solid-state $^{13}C$ NMR spectra of the remaining samples were obtained from the following sources: (1) Mollisol particle size fractions (Baldock et al. 1992), (2) Australian soils with known char contents (Skjemstad et al. 1999b), and (3) phytoplankton (Zelibor et al. 1988). Spectral distributions for these samples were determined by integrating the intensity found in the seven chemical shift regions previously identified, using the original data or digitised versions of published spectra. The Australian soils had been examined using conventional CP/MAS NMR and also using Bloch decay NMR. In this paper we use the term *direct polarisation* (DP) rather than *Bloch decay.*

### *Mixing model for inferring molecular structures from $^{13}C$ NMR spectra*

The approach used was the 'model components and simultaneous equations' method of Nelson et al. (1999), which is fully described below, including some minor modifications to the previous version. The biomolecules chosen for the mixing model were those produced in greatest abundance by plants (i.e., cellulose, hemi-cellulose, protein, lignin, cutin and suberin) and microorganisms (i.e., protein [structural and enzymes]; chitin; structural and muco-polysaccharides; and aliphatic membrane components). These categories of biomolecules were reduced to four model components having characteristic and distinctive $^{13}C$ NMR spectra and N:C ratios: (A) carbohydrate (representing cellulose, hemi-cellulose, muco-polysaccharides and smaller molecular weight saccharides), (B) protein (representing proteins, peptides and amino acids), (C) lignin, and (D) an aliphatic component (representing cutin, suberin and aliphatic membrane components). In addition to protein and carbohydrate, chitin (an amino-sugar polymer produced by a variety of terrestrial and aquatic organisms) is a biomolecule produced in abundance in natural ecosystems (Muzzarelli and Muzzarelli 1998). However, neither chitin nor glycoproteins were included as model components, because they have $^{13}C$ NMR spectra and N:C ratios intermediate between, and indistinguishable from, a mixture of protein and carbohydrate. Likewise, protein-phenols are represented by the protein and lignin model components. Two additional model components were included in the analysis. A pure carbonyl component (component E) allowed for the presence of additional carboxyl C in the other components (A, B, C and D), for example uronic

acids in the carbohydrate component and fatty acids in the aliphatic component. Char was included as component F, as it is known to exist in some soils and sediments (Gustafsson and Gschwend 1998; Skjemstad et al. 1999b).

For each model component, the spectral distribution of signal intensity in the seven spectral regions was determined in one of two ways, For the protein, aliphatic and carbonyl components, spectral distributions were calculated using the known chemical shifts of all the C atoms in chemical structures of pure materials. For the protein and aliphatic material components, the pure materials used are described below. For the carbohydrate, lignin and char components, spectral distributions were obtained from spectra of representative materials. Two sets of model components were used; a 'terrestrial' set, applied to all samples, and an 'aquatic' set, applied to the aquatic samples only. The two sets differed only in the protein and aliphatic components (Table 1).

The spectral distribution of the terrestrial protein component was calculated using the weighted average of the amino acid composition of HCl hydrolysates isolated from sub-tropical soils (Stevenson 1994), which was similar to that of temperate soils (Stevenson 1994). The spectral distribution of the aquatic protein component was calculated using the weighted average amino acid composition of phytoplankton (Cowie and Hedges 1992). The amino acid composition of phytoplankton was similar to that of vascular plant tissues, macrophytes, zooplankton, and bacteria and fungi (Cowie and Hedges 1992).

The spectral distribution of the terrestrial aliphatic component was calculated from the chemical structure of cutin (Kolattukudy 1980). $^{13}C$ NMR and pyrolysis gas chromatography mass spectrometry analyses have shown cutin to be an important aliphatic constituent of soil organic matter (Köbgel-Knabner et al. 1992). For the aquatic aliphatic component, stearic acid was used (Hedges et al. 2001).

For the carbohydrate, lignin and char model components, spectral distributions were derived by integrating spectra of pure examples. For the carbohydrate component, cellulose was used. The spectrum of cellulose is identical to that of glucose. Using a hexose to represent carbohydrates underestimated the signal in the 110–95 ppm region relative to that in the 95–60 ppm region in many samples. This was probably due to the presence of pentoses, which are produced by plants and microorganisms (Oades 1984). The spectral distribution of the lignin component was derived from the mean of spectra for Norway spruce lignin and red alder lignin (Wilson 1987). That of the char component was derived from a spectrum of charcoal fragments picked from a sandy soil by hand.

To predict the molecular composition of each sample, we mathematical-mixed the model components together so that the N:C ratio and the signal intensity in some regions of the NMR spectrum for the model mixture equalled those for the sample. The mixing process was accomplished by solving Equation (1)–(6) Solution of these equations defined a proportion of each model component for that sample (*a–f*, shown as percentage in Figures 2,4–6,8,9). The proportions of the various model components were used to calculate a predicted spectral distribution for that sample by summing, for each chemical shift region, the product of (a) the proportion of each model component and (b) the present C in that region for that model component (from Table 1). The spectral regions used in the equations (45 to −10, 95–60,

*Table 1.* Molar N:C ratio, and distribution of C (as a % of total) in spectral regions, for model components (see text for sources).

| Chemical shift region (ppm) | (A) Carbohydrate | (B)Protein | | (C) Lignin | (D) Aliphatic | | (E) Carbonyl | (F) Char |
|---|---|---|---|---|---|---|---|---|
| | | Terrestrial | Aquatic | | Terrestrial | Aquatic | | |
| 210–165 | 0.0 | 26.4 | 30.4 | 4.6 | 6.6 | 5.6 | 100.0 | 5.6 |
| 165–145 | 0.0 | 2.5 | 1.0 | 19.5 | 0.7 | 0.0 | 0.0 | 16.1 |
| 145–110 | 1.0 | 7.5 | 4.5 | 30.6 | 3.6 | 0.0 | 0.0 | 73.9 |
| 110–95 | 15.7 | 0.0 | 0.0 | 8.6 | 0.0 | 0.0 | 0.0 | 4.3 |
| 95–60 | 79.0 | 2.1 | 2.9 | 12.5 | 9.0 | 0.0 | 0.0 | 0.0 |
| 60–45 | 4.3 | 21.9 | 24.7 | 13.8 | 4.5 | 0.0 | 0.0 | 0.0 |
| 45 to −10 | 0.0 | 39.6 | 36.6 | 10.5 | 75.6 | 94.4 | 0.0 | 0.0 |
| Molar N:C | 0.0 | 0.320 | 0.266 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

210–160 and 145–110 ppm) were chosen because they differed most widely between model components (Table 1).

$$a + b + c + d + e + f = 1 \tag{1}$$

$$a(\alpha_A) + b(n_B) + c(n_C) + d(n_D) + e(n_E) + f(n_F) = n_{sample} \tag{2}$$

$$a(\alpha_A) + b(\alpha_B) + c(\alpha_C) + d(\alpha_D) + e(\alpha_E) + f(\alpha_F) = \alpha_{sample} \tag{3}$$

$$a(\beta_A) + b(\beta_B + c(\beta_C) + d(\beta_D) + e(\beta_E) + f(\beta_F) = \beta_{sample} \tag{4}$$

$$a(\chi_A) + b(\chi_B) + c(\chi_C) + d(\chi_D) + e(\chi_E) + f(\chi_F) = \chi_{sample} \tag{5}$$

$$a(\delta_A) + b(\delta_B) + c(\delta_C) + d(\delta_D) + e(\delta_E) + f(\delta_F) = \delta_{sample} \tag{6}$$

In Equations (1)–(6), $a$, $b$, $c$, $d$, $e$ and $f$ equal the proportions of components A (carbohydrate), B (protein), C (lignin), D (aliphatic material), E (carbonyl) and F (char) in the model. Equation (1) ensures that the sum of all component proportions equals 1. In Equation (2), n equals the N:C ratio of the component (or sample) specified (e.g., $n_A$ equals the N:C ratio of component A, the carbohydrate component). Molar N:C ratios of the model components are shown in Table 1. In Equations (3)–(6), $\alpha$, $\beta$, $\chi$ and $\delta$ equal the proportions of carbon in the specified components (or sample) that resonate in the 45 to −10, 95–60, 210–165 and 145–110 ppm chemical shift regions, respectively (e.g., $\alpha_A$ equals the proportion of total signal in the 45 to −10 ppm region in the carbohydrate component).

In samples where solution of the six equations generated a proportion for a given model component that was more negative than −1%, Equation (6) was eliminated from the model, and all instances of that component in Equations (1)–(5) were removed. The model was then solved again to generate a new solution. If there was still a component having a proportion more negative than −1%, Equation (5) was removed and all instances of that component in Equations (1)–(4) were removed; and so on until all component proportions were zero or positive. If the proportion of a given component was between −1 and 0%, it was assumed to be 0, Reducing the number of equations reduced the number of constrained spectral regions and therefore generally increased the difference between actual and predicted spectral distributions, For phytoplankton samples, lignin and char were omitted from the initial set of equations, and for the plant samples char was omitted from the initial set of equations, because those components are known not to exist in those samples.

### *Assessment of the agreement between measured and predicted spectral distributions*

The ability of the model to accurately define molecular composition can be assessed through a comparison of the predicted and measured distributions of spectral

intensity. However, in the modelling process, the simultaneous solving of some combination of the Equations (1)–(6)forces predicted signal intensities to be equated to the measured intensities for some spectral regions. If the number of equations simultaneously solved is defined as n, the number of spectral regions with predicted signal intensities equal to the measured signal intensities is n−2, and the number not equated to measured signal intensities is 7− ($n$−2). For example, when all six equations are used, the predicted signal intensities for the four regions $\alpha$, $\beta$, $\chi$ and $\delta$ are all equal to their measured signal intensities and any lack of fit between predicted and measured data can only be associated with the three remaining spectral regions.

An indication of goodness of fit between predicted and measured data for an individual sample was derived by calculating the sum of the absolute value of the differences between the measured and predicted signal intensities obtained for all seven spectral regions, hereafter referred to as the 'error', Since the total NMR signal intensity of all samples analysed was normalised to a value of 100, the magnitude of the error value indicates the percentage of the total measured signal intensity not accounted for by the predicted distribution of signal intensity derived from the predicted molecular composition.

## Results and discussion

### *Correspondence between measured and predicted spectral distributions*

Good agreement between the measured and predicted distributions of $^{13}$C NMR signal intensity was obtained for all the terrestrial samples (Tables 2–4). The error was always $\leq$ 9%, with a mean of 3.6% and median of 3.2%, indicating that the predicted molecular composition accounted for ≥91% and often ≥95% of the total measured $^{13}$C NMR signal intensity for the terrestrial samples. The difference between measured and predicted intensities in any individual spectral region of the terrestrial samples never accounted for more than 4.5% of the total signal intensity. Variations of this magnitude within individual spectral regions are comparable to maximum variation (3.9% of total signal intensity) associated with replicate analyses of well ground and homogenised heated wood samples determined using the same spectrometer (Baldock and Smernik 2002).

For the eight Australian soils, the fit between measured and predicted distributions of $^{13}$C NMR signal intensity was better when the measured values were derived from the cross polarisation (CP) analysis than from the direct polarisation (DP) analysis. It is perhaps not surprising that less accurate fits were obtained for the direct polarisation data. The spectral distributions used for the molecular components included in the modelling procedure were either calculated or derived from CP analyses. Significant variations in the distribution of signal intensity have been observed between CP and DP $^{13}$C NMR analyses, particularly in the aryl (145–110 ppm) and alkyl (45–10 ppm) regions (Skjemstad et al. 1999b; Smernik et al. 2002a, b). As a result, when estimates of molecular composition are made

*Table 2*. Carbon and nitrogen contents and measured and predicted $^{13}C$ NMR spectral distributions of fresh and decomposed plant samples. The spectral regions are defined by their chemical shift (ppm). Italics indicates spectral regions in which the modeled and actual values of percent C were set to be equal. 'Error' is defined in the text and 'Components' notation is given in Table 1.

| Sample | Organic C content (g $kg^{-1}$) | Total N content (g $kg^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Australian pasture samples | | | | | | | | | | | | | |
| Buffel grass | 378 | 14.7 | 0.0333 | Measured | *5.4* | 2.1 | 5.0 | 13.3 | *59.8* | 5.6 | *8.8* | 2.8 | A, B, C, D, E |
| | | | | Predicted | *5.4* | 2.1 | 4.6 | 12.3 | *59.8* | 7.0 | *8.8* | | |
| Seca Stylo | 424 | 26.8 | 0.0542 | Measured | *12.0* | 3.3 | 7.6 | 9.3 | *43.7* | 7.8 | *16.4* | 1.7 | A, B, C, D, E |
| | | | | Predicted | *12.0* | 3.4 | 6.7 | 9.3 | *43.7* | 8.5 | *16.4* | | |
| Native grasses | 419 | 3.6 | 0.0074 | Measured | *3.1* | 3.5 | 6.4 | 15.1 | *62.5* | 5.0 | *4.4* | 3.3 | A, B, C, D, E |
| | | | | Predicted | *3.1* | 3.6 | 6.6 | 13.5 | *62.5* | 6.4 | *4.4* | | |
| Oaten hay | 374 | 43.6 | 0.0999 | Measured | *7.8* | 1.8 | 5.0 | 9.5 | *45.8* | 9.8 | *20.3* | 3.0 | A, B, C, D |
| | | | | Predicted | *9.0* | 1.5 | 4.3 | 9.0 | *45.8* | 10.1 | *20.3* | | |
| Canadian plant litter | | | | | | | | | | | | | |
| Maple litter | NA | NA | 0.0250 | Measured | *7.1* | 2.7 | 7.3 | 9.1 | *47.0* | 6.6 | *20.1* | 2.5 | |
| | | | | Predicted | *7.1* | 3.0 | 6.1 | 9.8 | *47.0* | 6.9 | *20.1* | | A, B, C, D, E |
| Alfalfa litter | NA | NA | 0.0429 | Measured | *8.4* | 1.5 | 5.2 | 10.3 | *53.0* | 8.1 | *13.5* | 2.5 | |
| | | | | Predicted | *8.4* | 2.2 | 4.7 | 10.9 | *53.0* | 7.4 | *13.5* | | A, B, C, D, E |
| Corn litter | NA | NA | 0.0109 | Measured | *4.2* | 2.2 | 5.5 | 14.1 | *62.4* | 5.7 | *6.0* | 2.1 | |
| | | | | Predicted | *4.2* | 2.8 | 5.4 | 13.1 | *62.4* | 6.1 | *6.0* | | A, B, C, D, E |

*Table 2.* (Continued).

| Sample | Organic C content (g kg$^{-1}$) | Total N content (g kg$^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Corn litter buried in agricultural soil | | | | | | | | | | | | | |
| 193 days | NA | NA | 0.0228 | Measured | *4.8* | 2.7 | 6.3 | 12.8 | *56.1* | 8.0 | *9.3* | 3.1 | |
| | | | | Predicted | *4.8* | 3.7 | 6.9 | 12.1 | *56.1* | 7.2 | *9.3* | | A, B, C, D, E |
| 306 days | NA | NA | 0.0257 | Measured | *6.9* | 4.3 | 8.0 | 10.3 | *45.1* | 11.4 | *14.1* | 6.4 | |
| | | | | Predicted | *6.9* | 5.6 | 9.8 | 10.4 | *45.1* | 8.2 | *14.1* | | A, B, C, D, E |
| 387 days | NA | NA | 0.0317 | Measured | *12.4* | 7.5 | 12.2 | 9.3 | *37.4* | 9.0 | *12.1* | 0.6 | |
| | | | | Predicted | *12.4* | 7.3 | 12.3 | 9.5 | *37.4* | 9.1 | *12.1* | | A, B, C, D, E |
| 582 days | NA | NA | 0.0321 | Measured | *9.4* | 5.4 | 11.2 | 10.3 | *39.1* | 10.9 | *13.8* | 4.7 | |
| | | | | Predicted | *9.4* | 7.0 | 11.9 | 9.7 | *39.1* | 9.1 | *13.8* | | A, B, C, D, E |
| Corn litter buried in forest soil | | | | | | | | | | | | | |
| 193 days | NA | NA | 0.0295 | Measured | *5.9* | 2.8 | 6.6 | 11.7 | *52.7* | 9.3 | *11.0* | 3.9 | |
| | | | | Predicted | *5.9* | 4.0 | 7.4 | 11.5 | *52.7* | 7.6 | *11.0* | | A, B, C, D, E |
| 306 days | NA | NA | 0.0380 | Measured | *5.9* | 3.0 | 7.8 | 11.1 | *48.8* | 10.2 | *13.2* | 4.3 | |
| | | | | Predicted | *5.9* | 4.6 | 8.4 | 10.8 | *48.8* | 8.4 | *13.2* | | |
| 387 days | NA | NA | 0.0374 | Measured | *8.0* | 3.6 | 8.6 | 10.1 | *43.8* | 10.6 | *15.3* | 4.3 | |
| | | | | Predicted | *8.0* | 5.1 | 9.3 | 10.0 | *43.8* | 8.6 | *15.3* | | A, B, C, D, E |
| 582 days | NA | NA | 0.0504 | Measured | *9.9* | 4.5 | 9.8 | 9.1 | *39.1* | 11.3 | *16.2* | 3.7 | |
| | | | | Predicted | *9.9* | 5.8 | 10.3 | 9.2 | *39.1* | 9.5 | *16.2* | | A, B, C, D, E |

NA: data not available.

*Table 3*. Carbon and nitrogen contents and measured and predicted $^{13}C$ NMR spectral distributions of whole soils and soil particle size fractions. The spectral regions are defined by their chemical shift (ppm). Italics indicates spectral regions in which the modelled and actual values of percent C were set to be equal. 'Error' is defined in the text and 'Components' notation is given in Table 1.

| Sample | Organic C content ($g\ kg^{-1}$) | Total N content ($g\ kg^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Canadian soil under various managements | | | | | | | | | | | | | |
| Corn | 20.3 | 1.8 | 0.0760 | Measured | *13.7* | 5.2 | 9.8 | 7.6 | *27.3* | 10.4 | *26.0* | 0.8 | A, B, C, D, E |
| | | | | Predicted | *13.7* | 5.3 | 10.0 | 7.2 | *27.3* | 10.5 | *26.0* | | |
| Grass pasture | 43.3 | 2.9 | 0.0574 | Measured | *11.8* | 3.9 | 8.2 | 7.8 | *31.0* | 10.8 | *26.4* | 3.1 | A, B, C, D, E |
| | | | | Predicted | *11.8* | 4.7 | 8.9 | 7.9 | *31.0* | 9.3 | *26.4* | | |
| Maple forest | 64.8 | 4.9 | 0.0648 | Measured | *13.1* | 5.1 | 10.1 | 9.4 | *28.2* | 9.9 | *24.1* | 3.6 | A, B, C, D, E |
| | | | | Predicted | *13.1* | 5.9 | 10.8 | 7.7 | *28.2* | 10.3 | *24.1* | | |
| Particle size fractions isolated from Australian Mollisol | | | | | | | | | | | | | |
| 250–2000$\mu$m | 300 | 7.6 | 0.0216 | Measured | *7.6* | 5.8 | 10.9 | 10.0 | *38.2* | 9.2 | *18.3* | 2.9 | A, B, C, D, E |
| | | | | Predicted | *7.6* | 6.7 | 11.5 | 9.3 | *38.2* | 8.5 | *18.3* | | |
| 53–250$\mu$m | 211 | 10.3 | 0.0420 | Measured | *11.7* | 5.3 | 11.6 | 7.7 | *27.5* | 10.9 | *25.3* | 4.2 | A, B, C, D, E |
| | | | | Predicted | *11.7* | 6.9 | 12.1 | 7.0 | *27.5* | 9.5 | *25.3* | | |
| 20–53$\mu$m | 139 | 8.1 | 0.0501 | Measured | *15.4* | 6.6 | *14.9* | 7.2 | *24.7* | 8.9 | *22.4* | 1.3 | A, B, C, D, E, F |
| | | | | Predicted | *15.4* | 6.9 | *14.9* | 6.5 | *24.7* | 9.2 | *22.4* | | |
| 2–20$\mu$m | 244 | 21.2 | 0.0745 | Measured | *17.2* | 5.7 | *15.8* | 4.8 | *21.8* | 8.8 | *25.9* | 0.8 | A, B, C, D, E, F |
| | | | | Predicted | *17.2* | 5.3 | *15.8* | 5.2 | *21.8* | 8.8 | *25.9* | | |
| 0.2–2$\mu$m | 55 | 6.8 | 0.1058 | Measured | *18.0* | 4.5 | *14.8* | 5.4 | *22.0* | 8.5 | *26.8* | 3.2 | A, B, C, D, E, F |
| | | | | Predicted | *18.0* | 4.0 | *14.8* | 4.8 | *22.0* | 9.6 | *26.8* | | |
| < 0.2$\mu$m | 26 | 3.5 | 0.1143 | Measured | *12.4* | 2.3 | *7.8* | 5.8 | *16.0* | 8.2 | *47.6* | 6.1 | A, B, C, D, E, F |
| | | | | Predicted | *12.4* | 2.5 | *7.8* | 2.7 | *16.0* | 11.0 | *47.6* | | |

*Table 4.* Carbon and nitrogen contents and measured and predicted $^{13}C$ NMR spectral distributions of Australian soils with known char contents, using CP/MAS (CP) and DP/MAS (DP) NMR analyses. The spectral regions are defined by their chemical shift (ppm). Italics indicates spectral regions in which the modelled and actual values of percent C were set to be equal. 'Error' is defined in the text and 'Components' notation is given in Table 1.

| Sample | Organic C Content (g $kg^{-1}$) | Total N content (g $kg^{-1}$) | Molar N:C ratio | NMR technique | Percentage of C resonating in spectral regions | | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| SS6 | 70.5 | 6.1 | 0.0804 | CP | Measured | *12.0* | 4.3 | *10.2* | 8.3 | *28.9* | 10.1 | *26.1* | 3.3 | A, B, C, D, E, F |
| | | | | | Predicted | *12.0* | 5.3 | *10.2* | 6.7 | *28.9* | 10.8 | *26.1* | | |
| | | | | DP | Measured | *11.0* | 6.2 | *16.9* | 8.1 | *19.2* | 8.4 | *26.7* | 8.1 | |
| | | | | | Predicted | *11.0* | 8.5 | *16.9* | 5.8 | *19.2* | 11.9 | *26.7* | | |
| SS7 | 143.0 | 7.2 | 0.0546 | CP | Measured | *13.1* | 5.6 | *14.2* | 7.7 | *23.7* | 7.6 | *28.0* | 3.7 | A, B, C, D, E, F |
| | | | | | Predicted | *13.1* | 6.1 | *14.2* | 5.9 | *23.7* | 9.0 | *28.0* | | |
| | | | | DP | Measured | *11.3* | 8.9 | *27.0* | 8.2 | *14.3* | 5.2 | *21.8* | 8.8 | |
| | | | | | Predicted | *11.3* | 10.5 | *27.0* | 5.5 | *14.3* | 9.7 | *21.8* | | |
| SS8 | 27.3 | 1.6 | 0.0680 | CP | Measured | *11.6* | 5.2 | *18.9* | 6.8 | *24.3* | 9.3 | *23.9* | 2.4 | A, B, C, D, E, F |
| | | | | | Predicted | *11.6* | 6.4 | *18.9* | 6.1 | *24.3* | 8.8 | *23.9* | | |
| | | | | DP | Measured | *14.1* | 8.3 | *34.8* | 5.9 | *13.1* | 4.9 | *15.7* | 5.1 | |
| | | | | | Predicted | *14.1* | 9.7 | *34.8* | 5.0 | *13.1* | 7.7 | *15.7* | | |
| ACU1 | 25.6 | 2.3 | 0.0787 | CP | Measured | *12.5* | 3.3 | *14.3* | 6.1 | *26.5* | 8.5 | *28.8* | 1.2 | A, B, C, D, E, F |
| | | | | | Predicted | *12.5* | 3.9 | *14.3* | 5.6 | *26.5* | 8.4 | *28.8* | | |
| | | | | DP | Measured | *11.9* | 6.1 | *27.1* | 6.6 | *17.6* | 6.1 | *22.3* | 5.4 | |
| | | | | | Predicted | *11.9* | 7.6 | *27.1* | 5.1 | *17.6* | 8.5 | *22.3* | | |
| URB-P | 26.7 | 2.7 | 0.0791 | CP | Measured | *11.2* | 3.9 | *13.8* | 6.6 | *27.2* | 9.0 | *28.2* | 1.3 | A, B, C, D, E, F |
| | | | | | Predicted | *11.2* | 4.5 | *13.8* | 6.0 | *27.2* | 9.1 | *28.2* | | |
| | | | | DP | Measured | *12.7* | 7.9 | *27.7* | 5.7 | *17.3* | 5.0 | *20.2* | 4.4 | |
| | | | | | Predicted | *12.7* | 8.1 | *27.7* | 5.2 | *17.3* | 8.7 | *20.2* | | |

*Table 4.* (Continued).

| Sample | Organic C Content (g $kg^{-1}$) | Total N content (g $kg^{-1}$) | Molar N:C ratio | NMR technique | Percentage of C resonating in spectral regions | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| B211 | 34.3 | 3.2 | 0.0786 | CP | Measured | *12.5* | 4.1 | *14.6* | 6.2 | *26.2* | 9.7 | *26.7* | 1.6 | A, B, C, D, E, F |
| | | | | | Predicted | *12.5* | 4.9 | *14.6* | 5.9 | *26.2* | 9.2 | *26.7* | | |
| | | | | DP | Measured | *12.4* | 7.6 | *30.9* | 7.1 | *15.8* | 6.2 | *17.6* | 6.1 | |
| | | | | | Predicted | *12.4* | 9.2 | *30.9* | 5.3 | *15.8* | 8.9 | *17.6* | | |
| Qld | 38.4 | 3.3 | 0.0748 | CP | Measured | *9.7* | 3.6 | *10.2* | 7.0 | *25.8* | 9.2 | *34.5* | 2.8 | A, B, C, D, E, F |
| | | | | | Predicted | *9.7* | 4.4 | *10.2* | 5.6 | *25.8* | 9.8 | *34.5* | | |
| | | | | DP | Measured | *12.1* | 6.8 | *23.4* | 5.8 | *17.2* | 5.6 | *25.4* | 5.4 | |
| | | | | | Predicted | *12.1* | 7.7 | *23.4* | 5.0 | *17.2* | 9.3 | *25.4* | | |
| Buck | 15.8 | 1.6 | 0.0844 | CP | Measured | *11.2* | 3.2 | *11.9* | 6.3 | *28.3* | 10.4 | *28.8* | 2.0 | A, B, C, D, E, F |
| | | | | | Predicted | *11.2* | 4.2 | *11.9* | 6.1 | *28.3* | 9.6 | *28.8* | | |
| | | | | DP | Measured | *10.8* | 6.4 | *24.5* | 7.3 | *18.1* | 6.8 | *23.2* | 6.7 | |
| | | | | | Predicted | *10.8* | 8.1 | *24.5* | 5.4 | *18.1* | 9.9 | *23.2* | | |

using DP analyses, it is suggested that the distribution of signal intensities associated with the model components be derived from DP analyses.

The fits between measured and predicted distributions of signal intensity obtained for the aquatic samples were not as good as those obtained for the terrestrial samples (Tables 5 and 6). Using the 'terrestrial' model components, the error ranged from 0.5 to 22.7% with a mean of 7.8% and a median of 5.8%. Within the aquatic samples, the agreement between measured and predicted was poorest for the single-species phytoplankton samples, intermediate for the multi-species marine plankton samples and best for the sediment samples. The poor performance of the model in the case of the single phytoplankton and plankton species was attributed to the fact that the model used generic, average environmental compositions for the components. Particular species of phytoplankton or plankton may produce relatively large quantities of certain molecules that do not necessarily reflect the broad averages produced by many organisms in a community. For example, some organisms are known to produce large quantities of particular proteins or aliphatic constituents. This problem was not so pronounced in the plant samples because of the predominance of cellulose in those samples and the fact that the carbohydrate component was an accurate representation of the cellulose produced by all plant species examined.

For the aquatic samples, use of the 'aquatic' model components resulted in worse fits (larger errors with maximum 26.7%, mean 11.9%, median 9.8%) than use of 'terrestrial' model components for all samples except the Equatorial Pacific 500 m sediment trap (Tables 5 and 6). The main difference between 'aquatic' and 'terrestrial' model components was in the aliphatic component (Table 1). Stearic acid was used as the 'aquatic' aliphatic component while cutin was used for the 'terrestrial' aliphatic component. It would therefore appear that stearic acid may not be the most appropriate 'aquatic' aliphatic component and perhaps a more complex structure containing a portion of unsaturated carbon and additional functional groups may be more appropriate. Indeed, Hedges et al. (2002) found that oleic acid with its two unsaturated carbons served as a better 'aquatic' aliphatic component than stearic acid. However, a more complex material similar to terrestrial cutin may serve as an even better approximation of the 'aquatic' aliphatic component.

*Correspondence between measured and predicted molecular composition*

The ability of the modelling procedure to accurately predict molecular composition was examined using additional data for the Australian pasture samples and the Australian soils. For the Australian pasture samples, the predicted molecular composition was compared with that derived from conventional wet chemical analyses (Figure 1). The predicted proportions of each molecular component calculated by the model are expressed on a total C basis. To allow these values to be compared with the molecular composition measured by the wet chemical methods, predicted data were converted to mass proportions using C contents of 43.7% for the protein component, 64.0% for the lignin component, 44.4% for the carbohydrate

*Table 5*. Carbon and nitrogen contents and measured and predicted $^{13}C$ NMR spectral distributions of aquatic organic materials, using terrestrial model components. The spectral regions are defined by their chemical shift (ppm). Italics indicates spectral regions in which the modelled and actual values of present C were set to be equal. 'Error' is defined in the text and 'Components' notation is given in Table 1.

| Sample | Organic C content ($g\ kg^{-1}$) | Total N content ($g\ kg^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Phytoplankton | | | | | | | | | | | | | |
| *Botryococcus* | 617 | 49 | 0.0681 | Measured | 18.7 | 3.7 | 8.3 | 5.5 | *21.5* | 8.4 | *33.9* | 22.7 | A, B, D, E |
| | | | | Predicted | 30.1 | 0.8 | 3.0 | 3.6 | *21.5* | 7.2 | *33.9* | | |
| *Chlorella* | 507 | 57 | 0.0971 | Measured | 14.1 | 3.0 | 6.5 | 7.6 | *30.3* | 10.9 | *27.6* | 18.2 | A, B, D, E |
| | | | | Predicted | 23.2 | 0.9 | 3.4 | 5.5 | *30.3* | 9.1 | *27.6* | | |
| *Dunaliella* | 508 | 127 | 0.2128 | Measured | 20.1 | 4.2 | 8.9 | 4.4 | *14.9* | 11.6 | *36.0* | 15.4 | A, B, D, E |
| | | | | Predicted | 23.5 | 1.7 | 5.6 | 2.4 | *14.9* | 15.8 | *36.0* | | |
| *Scenedesmus* | 526 | 59 | 0.0971 | Measured | 14.1 | 2.7 | 5.8 | 6.5 | *26.6* | 9.0 | *35.3* | 11.5 | A, B, D, E |
| | | | | Predicted | 19.6 | 1.0 | 3.7 | 4.6 | *26.6* | 9.3 | *35.3* | | |
| Freshwater and estuarine sediments | | | | | | | | | | | | | |
| Chesapeake Bay | 16 | 1.6 | 0.0877 | Measured | *14.7* | 6.4 | *20.3* | 2.0 | *12.9* | 9.3 | *34.4* | 2.4 | A, B, C, D, E, F |
| | | | | Predicted | *14.7* | 5.6 | *20.3* | 3.2 | *12.9* | 8.9 | *34.4* | | |
| Buffalo River | 159 | 8.8 | 0.0474 | Measured | *11.3* | 8.6 | *28.0* | 3.5 | *14.1* | 7.2 | *27.2* | 1.8 | A, B, C, D, E, F |
| | | | | Predicted | *11.3* | 8.0 | *28.0* | 4.4 | *14.1* | 6.9 | *27.2* | | |
| Lake Washington | 47 | 4.4 | 0.0804 | Measured | *7.8* | 5.7 | *13.2* | 6.3 | *19.0* | 9.8 | *38.2* | 7.7 | A, B, C, D |
| | | | | Predicted | *10.0* | 5.8 | *11.0* | 4.6 | *19.0* | 11.2 | *38.3* | | |

*Table 5.* (Continued).

| Sample | Organic C content (g kg$^{-1}$) | Total N content (g kg$^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Equatorial Pacific Ocean | | | | | | | | | | | | | |
| Plankton | 217 | 36.3 | 0.1434 | Measured | 13.6 | 0.4 | 6.5 | 2.0 | *19.8* | 13.4 | *44.3* | 5.8 | A, B, D, E |
| | | | | Predicted | 14.4 | 1.4 | 4.8 | 3.1 | *19.8* | 12.2 | *44.3* | | |
| 500 m | 99 | 13.1 | 0.1135 | Measured | *14.3* | 1.6 | 7.8 | 3.5 | *21.8* | 11.4 | *39.7* | 3.2 | A, B, C, D, E |
| | | | | Predicted | *14.3* | 2.6 | 6.4 | 4.1 | *21.8* | 11.2 | *39.7* | | |
| 3000 m | 64 | 6.5 | 0.0866 | Measured | *12.4* | 2.5 | 7.0 | 5.0 | *26.9* | 13.1 | *33.2* | 5.3 | A, B, C, D, E |
| | | | | Predicted | *12.4* | 3.7 | 7.8 | 5.6 | *26.9* | 10.4 | *33.2* | | |
| Sediment | 2 | 0.4 | 0.1429 | Measured | *12.2* | 5.7 | 7.7 | 8.2 | *29.7* | 11.5 | *25.0* | 6.8 | A, B, C, D |
| | | | | Predicted | *12.9* | 4.0 | 8.4 | 6.5 | *29.7* | 13.5 | *25.0* | | |
| Arabian Sea | | | | | | | | | | | | | |
| Plankton | 210 | 36.0 | 0.1469 | Measured | 11.2 | 2.8 | 8.1 | 4.6 | *20.9* | 11.6 | *40.8* | 12.0 | A, B, D, E |
| | | | | Predicted | 16.5 | 1.4 | 4.7 | 3.4 | *20.9* | 12.3 | *40.8* | | |
| 500 m | 91 | 1.2 | 0.1108 | Measured | 11.3 | 1.6 | 7.6 | 3.7 | *23.7* | 10.6 | *41.6* | 2.0 | A, B, D, F |
| | | | | Predicted | 11.8 | 1.7 | 6.9 | 4.1 | *23.7* | 10.3 | *41.6* | | |
| 3000 m | 62 | 7.7 | 0.1071 | Measured | *11.7* | 2.2 | *8.8* | 4.4 | *23.8* | 10.5 | *38.7* | 0.5 | A, B, C, D, E, F |
| | | | | Predicted | *11.7* | 2.5 | *8.8* | 4.4 | *23.8* | 10.3 | *38.7* | | |
| Sediment | 10 | 1.0 | 0.0840 | Measured | *11.4* | 5.5 | *11.8* | 6.6 | *24.6* | 11.2 | *28.8* | 1.2 | A, B, C, D, E, F |
| | | | | Predicted | *11.4* | 6.0 | *11.8* | 6.0 | *24.6* | 11.3 | *28.8* | | |

*Table 6.* Carbon and nitrogen contents and measured and predicted $^{13}C$ NMR spectral distributions of aquatic organic materials, using aquatic model components. The spectral regions are defined by their chemical shift (ppm). Italics indicates spectral regions in which the modelled and actual values of percent C were set to be equal. 'Error' is defined in the text abd 'Components' notation is given in Table 1.

| Sample | Organic C content (g $kg^{-1}$) | Total N content (g $kg^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Phytoplankton | | | | | | | | | | | | | |
| *Botryococcus* | 617 | 49 | 0.0681 | Measured | 18.7 | 3.7 | 8.3 | 5.5 | *21.5* | 8.4 | *33.9* | 25.4 | A, B, D, E |
| | | | | Predicted | 31.4 | 0.2 | 1.4 | 4.1 | *21.5* | 7.5 | *33.9* | | |
| *Chlorella* | 507 | 57 | 0.0971 | Measured | 14.1 | 3.0 | 6.5 | 7.6 | *30.3* | 10.9 | *27.6* | 18.4 | A, B, D, E |
| | | | | Predicted | 23.3 | 0.4 | 2.0 | 5.8 | *30.3* | 10.6 | *27.6* | | |
| *Dunaliella* | 508 | 127 | 0.2128 | Measured | 20.1 | 4.2 | 8.9 | 4.4 | 14.9 | 11.6 | *36.0* | 26.7 | A, B, D |
| | | | | Predicted | 24.7 | 0.8 | 3.7 | 2.0 | 12.5 | 20.3 | *36.0* | | |
| *Scenedesmus* | 526 | 59 | 0.0971 | Measured | 14.1 | 2.7 | 5.8 | 6.5 | *26.6* | 9.0 | *35.3* | 15.1 | A, B, D, E |
| | | | | Predicted | 20.3 | 0.4 | 2.0 | 5.1 | *26.6* | 10.4 | *35.3* | | |
| Freshwater and estuarine sediments | | | | | | | | | | | | | |
| Chesapeake Bay | 16 | 1.6 | 0.0877 | Measured | *14.7* | 6.4 | *20.3* | 2.0 | *12.9* | 9.3 | *34.4* | 3.3 | A, B, C, D, E, F |
| | | | | Predicted | *14.7* | 4.8 | *20.2* | 3.6 | *12.9* | 9.3 | *34.4* | | |
| Buffalo River | 159 | 8.8 | 0.0474 | Measured | *11.3* | 8.6 | *28.0* | 3.5 | *14.1* | 7.2 | *27.2* | 2.7 | A, B, C, D, E, F |
| | | | | Predicted | *11.3* | 7.7 | *28.0* | 4.8 | *14.2* | 6.8 | *27.2* | | |
| Lake Washington | 47 | 4.4 | 0.0804 | Measured | 7.8 | 5.7 | *13.2* | 6.3 | *19.0* | 9.8 | *38.2* | 8.6 | A, B, C, D, F |
| | | | | Predicted | 12.1 | 3.5 | *13.2* | 4.5 | *19.0* | 9.5 | *38.2* | | |

*Table 6.* (Continued).

| Sample | Organic C content (g kg$^{-1}$) | Total N content (g kg$^{-1}$) | Molar N:C ratio | Percentage of C resonating in spectral regions | | | | | | | Error (%) | Components |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 210–165 | 165–145 | 145–110 | 110–95 | 95–60 | 60–45 | 45 to −10 | | |
| Equatorial Pacific Ocean | | | | | | | | | | | | | |
| Plankton | 217 | 36.3 | 0.1434 | Measured | 13.6 | 0.4 | 6.5 | 2.0 | 19.8 | 13.4 | *44.3* | 12.6 | A, B, D |
| | | | | Predicted | 17.9 | 0.5 | 2.6 | 3.1 | 17.4 | 14.2 | *44.3* | | |
| 500 m | 99 | 13.1 | 0.1135 | Measured | 14.3 | 1.6 | 7.8 | 3.5 | *21.8* | 11.4 | *39.7* | 3.0 | A, B, D, F |
| | | | | Predicted | 14.7 | 1.4 | 6.5 | 4.3 | *21.8* | 11.7 | *39.7* | | |
| 3000 m | 64 | 6.5 | 0.0866 | Measured | 12.4 | 2.5 | 7.0 | 5.0 | *26.9* | 13.1 | *33.2* | 8.8 | A, B, D, F |
| | | | | Predicted | 11.8 | 2.3 | 10.7 | 5.7 | *26.9* | 9.5 | *33.2* | | |
| Sediment | 2 | 0.4 | 0.1429 | Measured | 12.2 | 5.7 | 7.7 | 8.2 | *29.7* | 11.5 | *25.0* | 16.1 | A, B, D, F |
| | | | | Predicted | 16.9 | 1.3 | 6.5 | 5.8 | *29.7* | 14.8 | *24.9* | | |
| Arabian Sea | | | | | | | | | | | | | |
| Plankton | 210 | 36.0 | 0.1469 | Measured | 11.2 | 2.8 | 8.1 | 4.6 | 20.9 | 11.6 | *40.8* | 19.7 | A, B, D |
| | | | | Predicted | 18.0 | 0.5 | 2.7 | 3.6 | 19.8 | 14.6 | *40.7* | | |
| 500 m | 91 | 1.2 | 0.1108 | Measured | 11.3 | 1.6 | 7.6 | 3.7 | *23.7* | 10.6 | *41.6* | 9.8 | A, B, D, F |
| | | | | Predicted | 14.3 | 0.7 | 3.6 | 4.6 | *23.7* | 11.5 | *41.5* | | |
| 3000 m | 62 | 7.7 | 0.1071 | Measured | 11.7 | 2.2 | 8.8 | 4.4 | *23.8* | 10.5 | *38.7* | 6.7 | A, B, D, F |
| | | | | Predicted | 14.0 | 1.3 | 6.3 | 4.7 | *23.8* | 11.2 | *38.7* | | |
| Sediment | 10 | 1.0 | 0.0840 | Measured | *11.4* | 5.5 | *11.8* | 6.6 | *24.6* | 11.2 | *28.8* | 1.5 | A, B, C, D, E, F |
| | | | | Predicted | *11.4* | 5.3 | *11.8* | 6.2 | *24.6* | 11.9 | *28.8* | | |

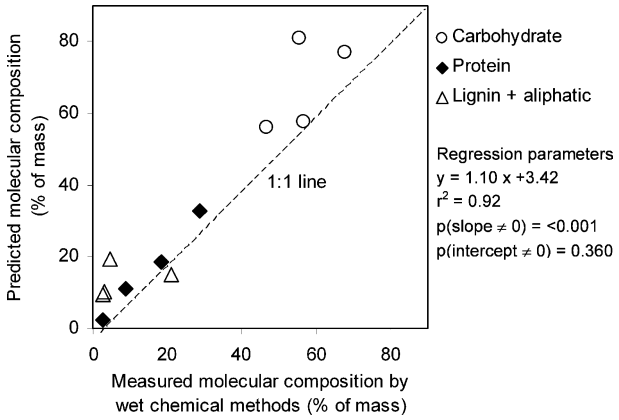


*Figure 1.* Predicted molecular composition of Australian pasture samples compared with their corresponding molecular composition measured using conventional wet chemical analyses. The labels use the model component names; for the wet chemical analyses, *carbohydrate* represents neutral detergent fibre minus acid detergent lignin, *protein* represents crude protein, and *lignin + aliphatic* represents acid detergent Lignin.

component, and 50.0% for all other components. A significant relationship between the predicted and measured data existed, with the variation in the measured data accounting for 92% of the variation in the predicted data. The predicted proportions tended to be higher than conventionally measured proportions, as indicated by the slope (>1) of the regression line. The higher proportions obtained for the predicted composition reflected the fact that the conventionally derived fractions accounted for only 70–80% of the total sample mass whereas the predicted components accounted for 96–100% of the total sample mass (the remainder consisting of the carbonyl component). The 20–30% of sample mass not accounted for by the conventional analysis consists largely of cytoplasm constituents (mostly proteins and saccharides) and pectins (Goering and van Soest 1972; Lowry et al. 1994). In addition, the conventional method tends to underestimate lignin content, especially in tropical grasses (Lowry et al. 1994).

For the eight Australian soils, the ability of the model to estimate the content of the char component was examined. The char content of each sample had been measured previously using an ultra-violet photo-oxidation treatment of the <53$\mu$m fraction followed by $^{13}$C NMR analysis (UV NMR) (Skjemstad et al. 1999b). The predicted char contents were highly correlated with char contents determined by the UV NMR procedure (Figure 2). The value of 0.50 for the slope of the relationship is consistent with the fact that CP/MAS $^{13}$C NMR does not detect all C in highly aromatic structures with low proton contents such as those found in char (Skjemstad et al. 1999b; Smernik et al. 2002a, b). The modelling analysis was repeated using the DP $^{13}$C NMR data rather than CP $^{13}$C NMR data. The DP $^{13}$C NMR analysis theoretically detects all forms of C, including that in non-protonated aromatic structures, with a high efficiency unless a significant amount of para-magnetic

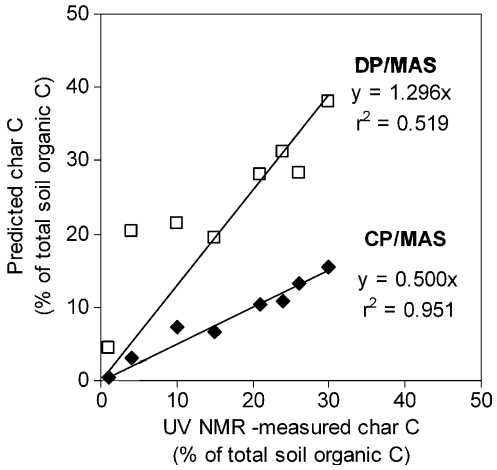


*Figure 2.* The relationship between predicted content of char carbon (using CP and DP $^{13}$C NMR spectra) and previously determined estimates of char carbon content (using UV photo-oxidation of the <53$\mu$m fraction followed by NMR (Skjemstad et al. 1999b), in eight Australian soils.

material is present in the sample. Char contents estimated using the DP $^{13}$C NMR data were higher than those obtained by Skjemstad et al. (1999b) as indicated by the value of 1.30 obtained for the slope of the relationship between predicted and measured char contents (Figure 2). Skjemstad et al. (1999b) pointed out that the UV NMR technique was conservative, and provided a measure of the minimum char content of soil because of the potential existence of char in the > 53 $\mu$m soil fraction and the potential for some oxidation of char during the UV oxidation process. The CP and DP $^{13}$C NMR spectra used in this study were obtained from whole unfractionated soil that was not exposed to the UV oxidation process. Thus, the actual char content of a sample may be estimated better by the modelling approach (using DP data) than by the UV NMR technique. The substantial differences between the molecular compositions determined using CP and DP spectra suggest that in any samples containing char, modelling using CP spectra will underestimate char content. Char content was not specifically targeted by the modelling approach, so the ability of the modelling technique to predict it over a wide range of values means that the prediction of the other molecular components must also have been reasonably accurate.

### *Molecular composition of plant residues*

The predicted molecular composition of C in the three Canadian plant litters differed (Figure 3). The corn litter was relatively rich in carbohydrate, consistent with a high requirement for the structural carbohydrate cellulose in the large sterm and leaf

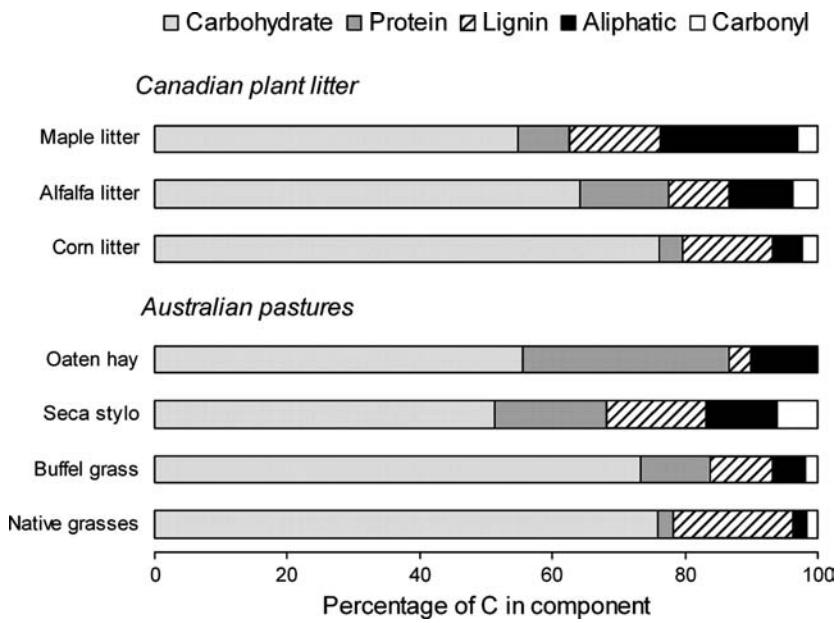


*Figure 3.* Predicted molecular composition of organic carbon in the Canadian and Australian plant samples.

structures. The alfalfa litter contained the largest concentration of protein, consistent with its ability to fix N through the root/rhizobia symbiosis and its high molar N:C ratio. The maple leaf litter contained the lowest concentration of carbohydrate, moderate concentrations of protein and lignin, and the highest concentration of aliphatic material, consistent with the presence of waxy cuticular materials on leaf surfaces.

Of the Australian pasture samples, the native grasses had the highest predicted contents of carbohydrate and lignin and the lowest protein contents. Given the low nitrogen status of soils and the water limited environment conditions under which these native grasses grow, such low protein contents and high contents of structural polymers are understandable. In progressing from the native grasses through Buffel grass and Seca stylo to the oaten hay, the predicted contents of carbohydrates tended to decrease and protein contents increased. The high protein content of the oaten hay may have resulted partly from an overestimation of the N:C ratio of the organic components present in these residues due to the presence of nitrate-N.

Changes in predicted molecular composition of corn shoot residues decomposing in agricultural and forest soil are shown in Figure 4. In both soils, a general trend of decreasing carbohydrate and increasing protein and lignin contents with time was noted with increasing duration of decomposition. Other studies with a wide range of plant residues decomposing in a variety of environments have consistently shown a

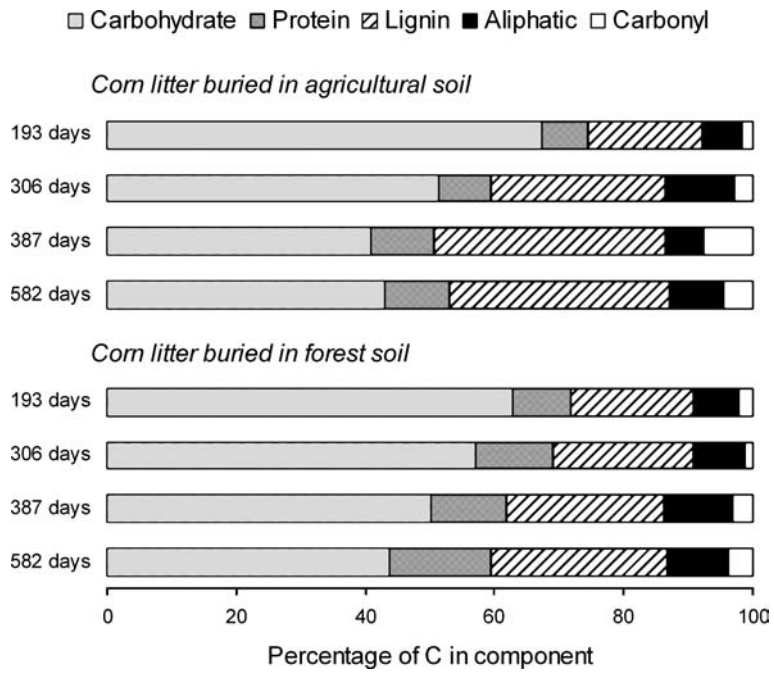


*Figure 4.* Predicted molecular composition of organic carbon in corn litter buried in an agricultural soil and a forest soil for increasing durations.

similar rapid loss of carbohydrate structures and an accumulation of lignin structures, using wet chemical molecular analyses (e.g., Cortez et al. 1996; Hamadi et al. 2000; Tian et al. 2000; Sariyildiz and Anderson 2003) and inference from $^{13}C$ NMR measurements (e.g., Baldock et al. 1997a; Quideau et al. 2001, 2000).

### *Molecular composition of organic matter in several Canadian and Australian soils*

The molecular composition of the organic matter in soils under corn, grass pasture and maple forest was similar (Figure 5), despite differences in the composition of plant litter added to the soils (Figure 3). Carbohydrates accounted for the largest fraction, with protein, lignin and aliphatic materials making similar contributions.

In the Mollisol particle size fractions, there was a general tendency for carbohydrate and lignin content to decrease and protein and aliphatic material contents to increase with decreasing particle size (Figure 5). These changes with decreasing particle size are consistent with a change from dominantly plant-derived materials to microbial products, as noted by Baldock et al. (1992) and Nelson et al. (1999). The high aliphatic content of the finest fraction is consistent with an increasing

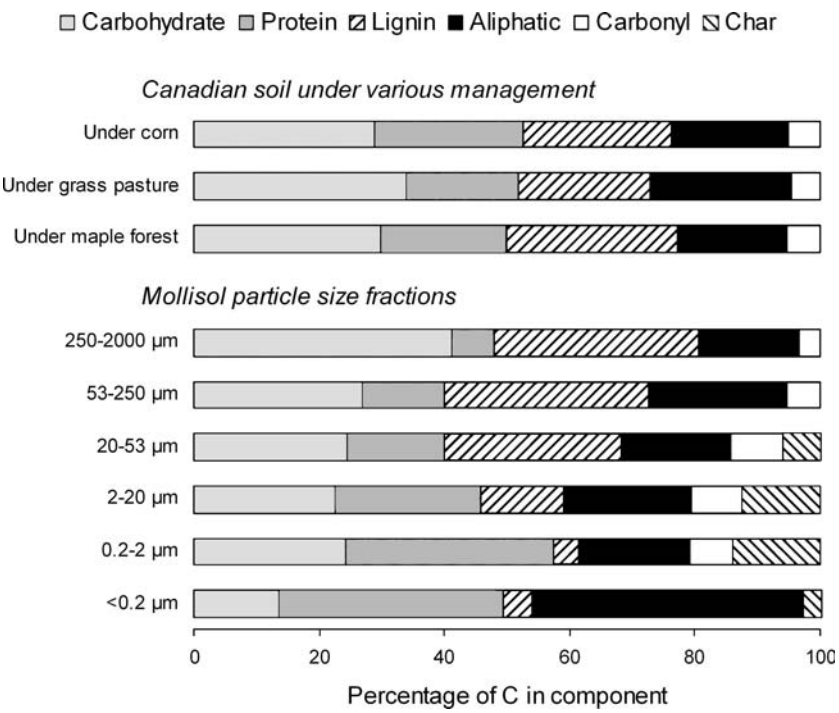


*Figure 5.* Predicted molecular composition of organic carbon in three Canadian soils under different vegetation, and various particle size/density fractions of an Australian Mollisol.

extent of decomposition being associated with decreasing particle size since aliphatic carbon has been shown to be one of the more recalcitrant forms of carbon in soils (Baldock et al. 1997b). $^{14}C$ and $^{15}N$-labelling studies have also shown a concentration of resistant plant remains in sand fractions and microbial biomass and metabolites in clay fractions (Ladd et al. 1996). Char was concentrated in the 0.2–53$\mu$m fractions, consistent with the observations of Skjemstad et al. (1999b, 2002).

The molecular composition of organic C in the Australian soils predicted from the CP analyses differed, consistent with the fact that the soils were collected from ecosystems with different vegetation, climate and landuse. The predicted proportions of soil C contributed by the molecular components ranged from 23–31% for carbohydrate, 17–26% for protein, 5–23% for lignin, 18–31% for aliphatic and 1–15% for char (Figure 6). The predicted molecular compositions derived using the CP and DP analyses differed, with more char and lignin predicted using the DP $^{13}C$ NMR signal intensity data. The higher char and lignin contents were balanced by lower carbohydrate and aliphatic contents. The protein content was the same in results based on CP and DP analyses because it was determined by the amount of N in the samples.

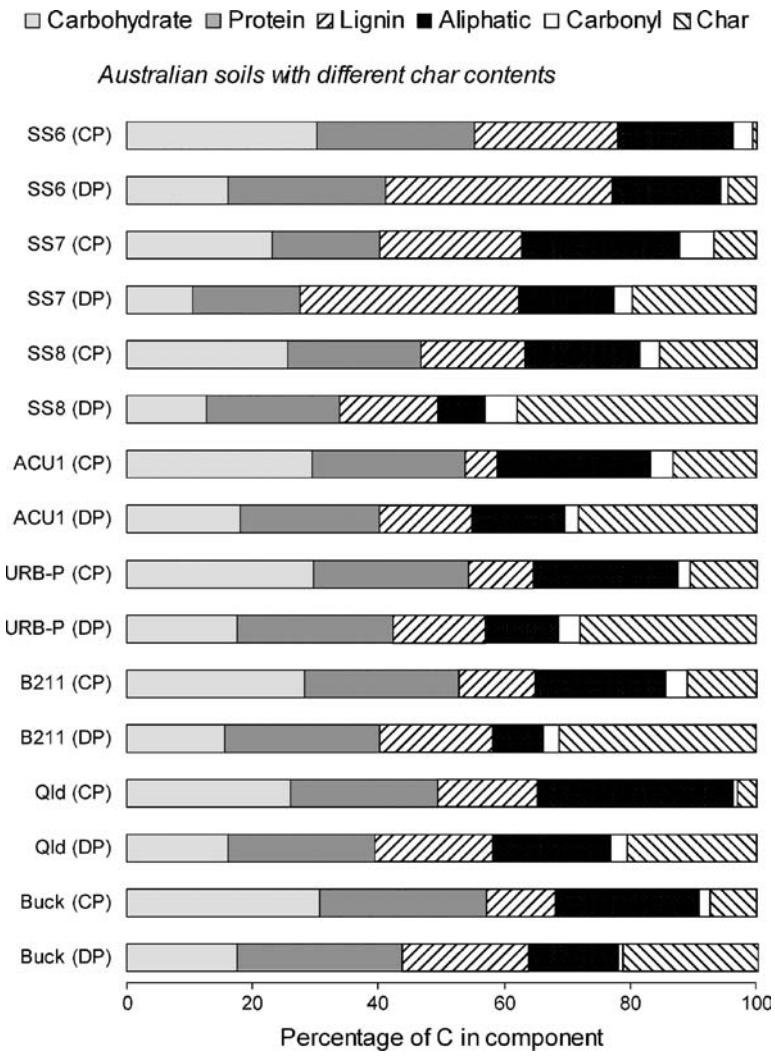


*Figure 6.* Predicted molecular composition, using CP or DP NMR analysis, of organic carbon in Australian soils with different char contents.

### *Molecular composition of phytoplankton, freshwater sediments and marine water columns*

Of the phytoplankton, B. *braunii*, C. *pyrenosidosa* and S. *obliquus* were similar, consisting of similar proportions of carbohydrate, protein and aliphatic material, and with a substantial carbonyl component. D. *tertiolecta* was quite different, with a very

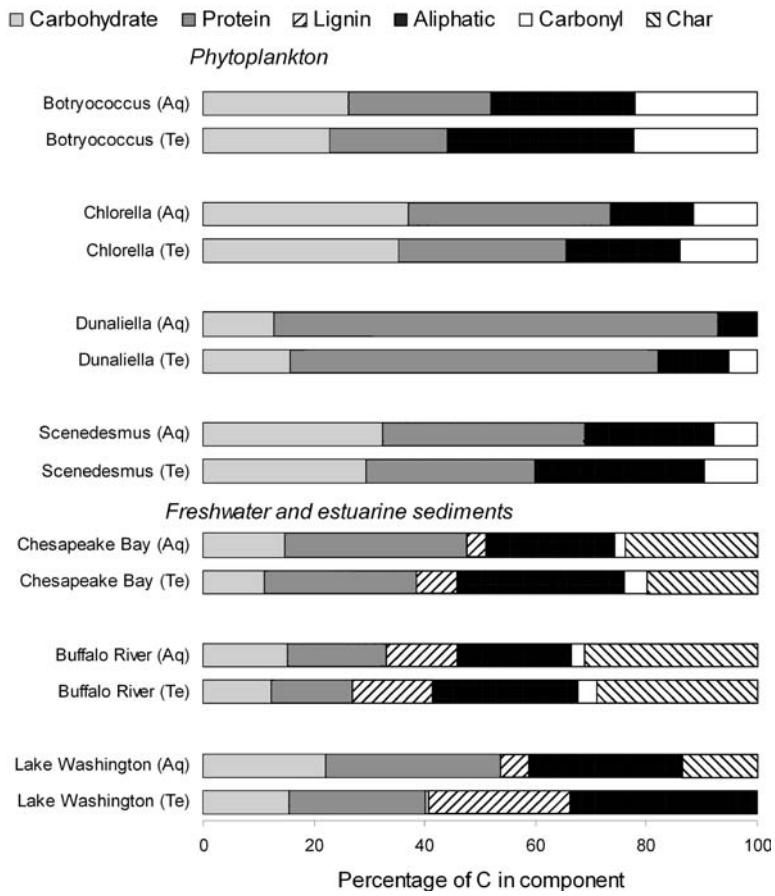


*Figure 7.* Predicted molecular composition of organic carbon in phytoplankton and freshwater and estuarine sediments using the terrestrial (Te) or aquatic (Aq) sets of model components.

high protein content. Unlike the others, D. *tertiolecta* does not produce a cellulose cell wall (Zelibor et al. 1988).

The freshwater and estuarine sediments contained similar proportions of carbohydrate and aliphatic material, and different proportions of protein, lignin and char (Figure 7). They tended to have lower proportions of carbohydrate than the terrestrial soil samples. It appears that there is a higher input of char and/or anthropogenically derived soot into Chesapeake Bay and the Buffalo River than into Lake Washington.

In the marine water column samples, previous studies have shown a marked mineralisation of the sinking organic matter, with only a few percent of C fixed by primary producers reaching the ocean floor (Wakeham et al. 1997; Lee et al. 1998). The fraction of organic matter that was measurable using a comprehensive molecular

analyses decreased steadily with depth, from around 80% near the surface to around 20% in the sea floor sediments (Wakeham et al. 1997). Wakeham et al. (1997) concluded that detailed study was needed of the biologically and analytically recalcitrant material that predominates at depth. For the molecularly well-characterised plankton samples, the NMR results corresponded with known biochemical compositions (Hedges et al. 2001). The composition of plankton was similar at the two sites. However, the changes with depth were different at the two sites. In the Pacific the proportion of carbohydrate increased and aliphatic material decreased consistently with depth. Protein content decreased with depth in the water column, but was high on the sea floor (Figure 8). Our results were consistent with the absolute decreases in measurable amino acids and lipids noted previously (Wakeham et al. 1997). In the Arabian Sea samples the changes with depth were not as marked. Lignin was a significant component at depth, which was unexpected, considering that it is not produced by marine organisms. The aromatic signal in the marine samples may arise from aromatic structures in aliphatic or protein molecules or other unsaturated carbon structures rather than from lignin. However, lignin may actually be present, as biomarkers from vascular land plants have been found in the sediments (Wakeham et al. 1997). Overall, the character of the total organic materials changed much less with depth than did the concentration of particular biomolecules found by Wakeham et al. (1997). Hedges et al. (2001) suggested that the increasing biological and analytical recalcitrance with depth might be due to physical protection by silicates, carbonates, aluminosilicates or other inorganic species, or perhaps by complex glycoproteins.

### *Molecules not accounted for in the model*

In the model, protein was the only N-containing component of organic matter. Other N-containing materials, such as nucleotides and other heterocyclic compounds are known to occur in soil and other materials. However, $^{15}N$ NMR studies have indicated that soil organic N is dominated by amine/amide structures with little contribution from heterocyclic N structures (Knicker et al. 1995; Knicker and Lüdemann 1995; DiCosty et al. 2003). Inclusion of a model component representing nucleotides was examined, but no improvements in model performance were noted.

Carbon in glycoproteins or peptidoglycans was attributed to the carbohydrate and protein components by the molecular mixing model. Similarly, biomolecules containing amino-sugars (e.g., chitin) could not be differentiated from a mixture of carbohydrate and protein. It might be possible to improve the ability of the modelling approach to account for the presence of these types of biomolecules by including additional equations based on independent analyses such as the quantity of amino-sugars and amino acids in extracts or hydrolysates. Alternatively, the composition of the protein or carbohydrate components used in the model could be altered to include estimates of glycoprotein, peptidoglycan and amino-sugars, pro-

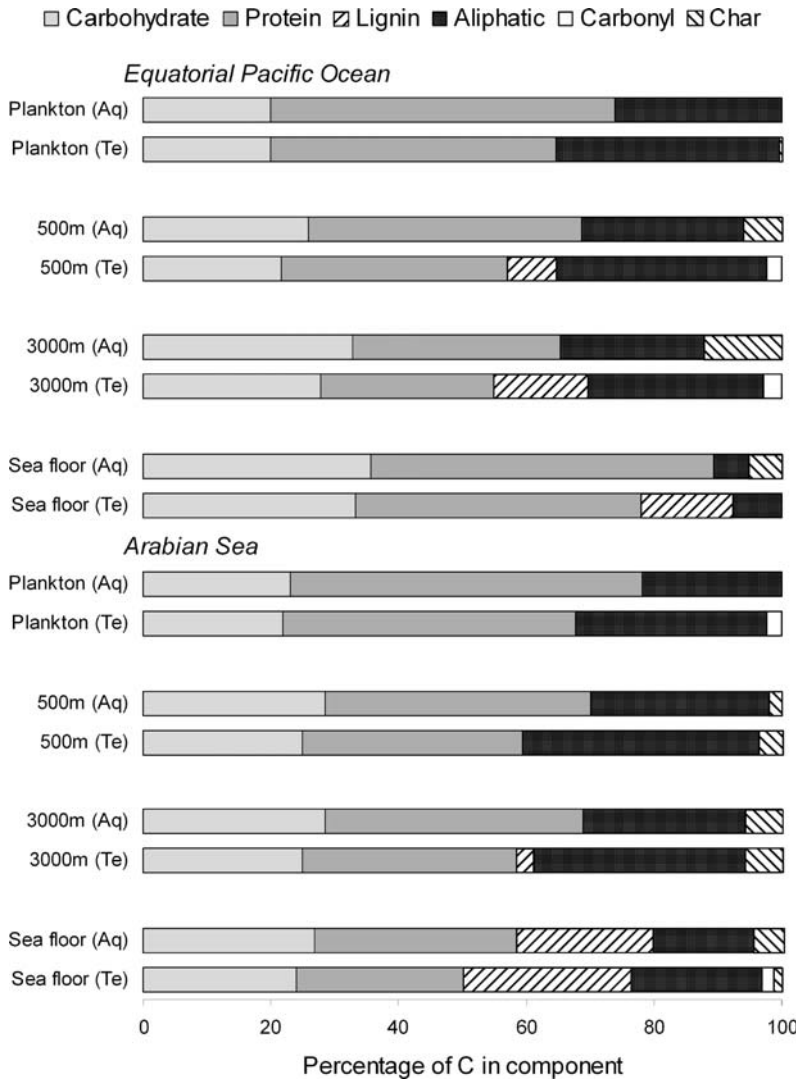


*Figure 8.* Predicted molecular composition of organic carbon in marine water column samples using the terrestrial (Te) or aquatic (Aq) sets of model components.

vided it can be shown that variations in the proportions of these biomolecules within ecosystems are consistent. Such an approach will form the basis of subsequent work with the model aimed at more accurately defining the composition of the biomolecular components to better reflect changes in ecosystem type and dominant forms of organic carbon inputs.

The results presented here do not prove that the organic matter in the samples examined consisted solely of the molecular components described, nor do they disprove the existence of humic structures in the samples. However, the degree to which measured $^{13}C$ NMR signal intensities and N:C ratios could be accounted for by mixtures of known biomolecules was striking. The presence of a 'humic' component would not be expected for the fresh plant residues phytoplankton or plankton samples. However, for the highly decomposed litter, the soils and soil fractions, and the various aquatic sediments, a significant amount of organic C would likely be extracted using the traditional alkaline extraction and acid precipitation methodology, be unidentifiable as known molecules, and thus be classified as 'humic' substance (MacCarthy et al. 1990). Even in these samples, the predicted NMR signal intensities based on known biomolecules corresponded well with the actual signal intensities for the whole sample.

## Conclusions

Our data show that the distribution of signal intensity in $^{13}C$ NMR spectra of samples collected from a diverse range of environments could be explained using a simple mixture of common biomolecules. For several of the plant and soil samples it was possible to compare our results with measurements of molecular components made by other means, and for those samples there was close correspondence. The results and methodology presented allow:

(1) An extension of the information typically derived from $^{13}C$ NMR analyses so that an indication of how the various types of C are joined together into molecules can be obtained.
(2) A biologically informative means of estimating gross chemical characteristics of natural organic materials found across a range of diverse environments.
(3) An indication that the organic matter in highly decomposed organic materials is composed mainly of a mixture of known biomolecules, not 'humic' molecules produced by abiotic means.

Although some uncertainty in the ability of the biomolecular mixing model to accurately represent the true molecular composition of natural organic materials undoubtedly remains, this methodology provides a significant step forward from the more traditional approach of classifying organic matter on the basis of its chemical solubility in acid and alkaline solutions. In addition, given that complete molecular analysis of natural organic material by conventional chemical degradation techniques cannot quantitatively account for all organic material found in a sample, and given the analytical time and expense required to perform such analyses, the combined use of $^{13}C$ NMR, C and N elemental analysis and the biomolecular mixing

model offers an effective alternative for characterising the molecular composition of natural organic materials. Further refinement of the number and composition of the molecular components included in the model and a tuning of these components to different ecosystems should help to reduce the uncertainties associated with the predicted molecular compositions.

## Acknowledgements

Some samples and $^{13}C$ NMR spectra were kindly supplied by John Hedges, Ed Gregorich, Jan Skjemstad, Riaz Ahmed and David Coates. David Coates carried out van Soest analyses of the Australian pasture samples. Phillip Clarke carried out some of the $^{13}C$ NMR analyses of pure materials for the model components.

## References

Baldock J.A. and Nelson P.N. In: Sumner M. (ed), Handbook of Soil Science. CRC Press, Boca Raton, FL, USA, pp. B25–B84.

Baldock J.A. and Skjemstad J.O. In: Peverill K.I.Sparrow L.A. and Reuter D.J. (eds), Soil Analysis: An Interpretation Manual. CSIRO Publishing, Collingswood, Vic, Australia, pp. 159–170.

Baldock J.A. and Smernik R.J. 2002. Chemical composition and bioavailability of thermally altered *Pinus resinosa* (Red pine) wood. Org. Geochem. 33: 1093–1109.

Baldock J.A., Kay B.D. and Schnitzer M. 1987. Influence of cropping treatments on the monosaccharide content of the hydrolysates of a soil and its aggregate fractions. Can. J. Soil Sci. 67: 489–499.

Baldock J.A., Oades J.M., Waters A.G., Peng X., Vassallo A.M. and Wilson M.A. 1992. Aspects of the chemical structure of soil organic materials as revealed by solid-state $^{13}C$ NMR spectroscopy. Bio-geochemistry 16: 1–42.

Baldock J.A., Oades J.M., Nelson P.N., Skene T.M., Golchin A. and Clarke P. 1997a. Assessing the extent of decomposition of natural organic materials using solid-state $^{13}C$ NMR spectroscopy. Aust. J. Soil Res. 35: 1061–1084.

Baldock J.A., Sewell T. and Hatcher P.G. 1997b. Decomposition induced changes in the chemical structure of fallen red pine, white spruce and tamarack logs. In: Cadisch G. and Giller K.E. (eds), Driven by Nature: Plant Litter Quality and Decomposition. CAB International, Wallingford, UK, pp. 75–83.

Beavis J. and Mott C.J.B. 1996. Effects of land use on the amino acid composition of soils: 1. Manured and unmanured soils from the Broadbalk continuous wheat experiment, Rothamsted, England. Geoderma 72: 259–270.

Beavis J. and Mott C.J.B. 1999. Effects of land use on the amino acid composition of soils: 2. Soils from the Park Grass experiment and Broadbalk Wilderness, Rothamsted, England. Geoderma 91: 173–190.

Cortez J., Demard J.M., Bottner P. and Monrozier L.J. 1996. Decomposition of Mediterranean leaf litters: a microcosm experiment investigating relationships between decomposition rates and litter quality. Soil Biol. Biochem. 28: 443–452.

Cowie G.L. and Hedges J.I. 1992. Sources and reactivities of amino acids in a coastal marine environment. Limnol. Oceanogr. 37: 703–714.

DiCosty R.J., Weliky D.P., Anderson S.J. and Paul E.A. 2003. $^{15}N$-CPMAS nuclear magnetic resonance spectroscopy and biological stability of soil organic nitrogen in whole soil and particle-size fractions. Org. Geochem. 34: 1635–1650.

Gélinas Y., Baldock J.A. and Hedges J.I. 2001. Demineralization of marine and freshwater sediments for CP/MAS $^{13}C$ NMR analysis. Org. Geochem. 32: 677–693.

Goering H.K. and van Soest P.J. 1972. Forage fibre analyses (apparatus, reagents, procedures and some applications). Agriculture Handbook No. 379. United States Department of Agriculture.

Gustafsson Ö. and Gschwend P.M. 1998. The flux of black carbon to surface sediments on the New England continental shelf. Geochim. Cosmochim. Acta 62: 465–472.

Hamadi Z., Steinberger Y., Kutiel P., Lavee H. and Barness G. 2000. Decomposition of Avena sterilis litter under arid conditions. J. Arid Environ. 46: 281–293.

Hatcher P.G. and Spiker E.C. 1988. Selective degradation of plant biomolecules. In: Frimmel F.H. and Christman R.F. (eds), Humic Substances and Their Role in the Environment. John Wiley and Sons, New York, NY, USA, pp. 59–74.

Hatcher P.G., Nanny M.A., Minard R.D., Dible S.D. and Carson D.M. 1995. Comparison of two thermochemolytic methods for the analysis of lignin in decomposing gymnosperm wood: the CuO oxidation method and the method of thermochemolysis with tetramethylammonium hydroxide (TMAH). Org. Geochem. 23: 881–888.

Hedges J.I. 1988. Polymerization of humic substances in natural environments. In: Frimmel F.H. and Christman R.F. (eds), Humic Substances and Their Role in the Environment. John Wiley and Sons, New York, NY, USA, pp. 45–58.

Hedges J.I. and Ertel J.R. 1982. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. Anal. Chem. 54: 174–178.

Hedges J.I., Baldock J.A., Gélinas Y., Lee C., Peterson M. and Wakeham S.G. 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. Nature 409: 801–804.

Hedges J.I., Baldock J.A., Gélinas Y., Lee C., Peterson M.L. and Wakeham S.G. 2002. The biochemical and elemental compositions of marine plankton: a NMR perspective. Mar. Chem. 78: 47–63.

Heng S. and Goh K.M. 1981. A rapid method for extracting lipid components from forest litter especially adapted for ecological studies. Commun. Soil Sci. Plant Anal. 12: 1283–1292.

Kinchesh P., Powlson D.S. and Randall E.W. 1995a. $^{13}C$ NMR studies of organic matter in whole soils: I. Quantitation possibilities. Eur. J. Soil Sci. 46: 125–138.

Kinchesh P., Powlson D.S. and Randall E.W. 1995b. $^{13}C$ NMR studies of organic matter in whole soils: II. A case study of some Rothamsted soils. Eur. J. Soil Sci. 46: 139–146.

Knicker H. and Lüdemann H.-D. 1995. N-15 and C-13 CPMAS and solution NMR studies of N-15 enriched plant material during 600 days of microbial degradation. Org. Geochem. 23: 329–341.

Knicker H., Almendros G., González-vila F.J., Lüdemann H.-D. and Martin F. 1995. $^{13}C$ and $^{15}N$ NMR analysis of some fungal melanins in comparison with soil organic matter. Org. Geochem. 23: 1023–1028.

Kögel I., Hempfling R., Zech W., Hatcher P.G. and Schulten H.-R. 1988. Chemical composition of the organic matter in forest soils 1. Forest litter. Soil Sci. 146: 124–136.

Kögel-Knabner I., Zech W. and Hatcher P.G. 1988. Chemical composition of the organic matter in forest soils: The humus layer. Z. Pflanzener. Boden. 151: 331–340.

Kögel-Knabner I., Hatcher P.G., Tegelarr E.W. and de Leeuw J.W. 1992. Aliphatic components of forest soil organic matter as determined by solid-state $^{13}C$ NMR and analytical pyrolysis. Sci. Total Environ. 113: 89–106.

Kolattukudy P.E. 1980. Cutin, suberin and waxes. In: Stumpf P.K. and Conn E.E. (eds), The Biochemistry of Plants. Vol. 4. Lipids. Academic Press, New York, NY, USA, pp. 571–638.

Ladd J.N., van Gestel M., Jocteur Monrozier L. and Amato M. 1996. Distribution of organic $^{14}C$ and $^{15}N$ in particle-size fractions of soils incubated with $^{14}C$, $^{15}N$-labelled glucose/$NH_4$, and legume and wheat straw residues. Soil Biol. Biochem. 28: 893–905.

Lee C., Murray D.W., Barber R.T., Buesseler K.O., Dymond J., Hedges J.I., Honjo S., Manganini S.J., Marra J., Moser C., Perterson M.L., Prell W.L. and Wakeham S.G. 1998. Particulate organic carbon fluxes: compilation of results from the 1995 US JGOFS Arabian Sea Process Study. Deep-Sea Res. Pt. II 45: 2489–2501.

Levy G.C., Litcher R.L. and Nelson G.L. 1980. Carbon-13 Nuclear Magnetic Resonance Spectroscopy. 2nd ed. John Wiley and Sons, New York, NY, USA.

Lowry J.B., Conlan L.L., Schlink A.C. and McSweeny C.S. 1994. Acid detergent dispersible lignin in tropical grasses. J. Sci. Food Agric. 65: 41–49.

MacCarthy P., Malcolm R.L., Clapp C.E. and Bloom P.R. 1990. An introduction to soil humic substances. In: MacCarthy P., Malcolm R.L., Clapp C.E. and Bloom P.R. (eds), Humic Substances in Crop and Soil Sciences: Selected Readings. Soil Science Society of America, Madison, WI, USA, pp. 1–12.

Ma'shum M., Tate M.E., Jones G.P. and Oades J.M. 1988. Extraction and characterisation of water repellent materials from Australian soils. J. Soil Sci. 39: 99–110.

Muzzarelli R.A.A. and Muzzarelli C. 1998. Native and modified chitins in the biosphere. In: Stankiewicz B.A. and van Bergen P.F. (eds), Nitrogen-Containing Macromolecules in the Bio- and Geosphere. American Chemical Society, Washington, DC, USA, pp. 148–162.

Nelson P.N., Baldock J.A., Clarke P., Oades J.M. and Churchman G.J. 1999. Dispersed clay and organic matter in soil: their nature and associations. Aust. J. Soil Res. 37: 289–315.

Oades J.M. 1984. Soil organic matter and structural stability: mechanisms and implications for management. Plant Soil 76: 319–337.

Ogner G. 1985. A comparison of four different raw humus types in Norway using chemical degradations and CPMAS $^{13}C$ NMR spectroscopy. Geoderma 35: 343–353.

Preston C.M., Trofymow J.A., Sayer B.G. and Niu J. 1997. $^{13}C$ nuclear magnetic resonance spectroscopy with cross-polarization and magic-angle spinning investigation of the proximate-analysis fractions used to assess litter quality in decomposition studies. Can. J. Bot. 75: 1601–1613.

Quideau S.A., Anderson M.A., Graham R.C., Chadwick O.A. and Trumbore S.E. 2000. Soil organic matter processes: characterization by $^{13}C$ NMR and $^{14}C$ measurements. For. Ecol. Manag. 138: 19–27.

Quideau S.A., Chadwick O.A., Benesi A., Graham R.C. and Anderson M.A. 2001. A direct link between forest vegetation type and soil organic matter composition. Geoderma 104: 41–60.

Reeves A.D. 1995. The use of organic markers in the differentiation of organic inputs to aquatic systems. Phys. Chem. Earth 20: 133–140.

Sariyildiz T. and Anderson J.M. 2003. Interactions between litter quality, decomposition and soil fertility: a laboratory study. Soil Biol. Biochem. 35: 391–399.

Schulten H.-R. and Schnitzer M. 1993. A state of the art structural concept for humic substances. Naturwissenschaften 80: 29–30.

Skjemstad J.O., Clarke P., Taylor J.A., Oades J.M. and Newman R.H. 1994. The removal of magnetic materials from surface soils–a solid-state C-13 CP/MAS NMR study. Aust. J. Soil Res. 32: 1215–1229.

Skjemstad J.O., Taylor J.A. and Smernik R.J. 1999b. Estimation of charcoal (char) in soils. Commun. Soil Sci. Plant Anal. 30: 2283–2298.

Skjemstad J.O., Reicosky D.C., Wilts A.R. and McGowan J.A. 2002. Charcoal carbon in US agricultural soils. Soil Sci. Soc. Am. J. 66: 1249–1255.

Smernik R.J. and Oades J.M. 2000a. The use of spin counting for determining quantitation in solid state $^{13}C$ NMR spectra of natural organic matter. 1. Model systems and the effects of paramagnetic impurities. Geoderma 96: 101–129.

Smernik R.J. and Oades J.M. 2000b. The use of spin counting for determining quantitation in solid state $^{13}C$ NMR spectra of natural organic matter. 2. HF-treated soil fractions. Geoderma 96: 159–171.

Smernik R., Baldock J.A., Oades J.M. and Whittaker A.K. 2002a. Determination of $T_1\rho$ H relaxation rates in charred and uncharred wood and consequences for NMR quantitation. Solid State Nucl. Mag. Res. 22: 50–70.

Smernik R., Baldock J.A. and Oades J.M. 2002b. Impact of remote protonation on $^{13}C$ CPMAS NMR quantitation of charred and uncharred wood. Solid State Nucl. Mag. Res. 22: 71–82.

Stevenson F.J. 1994. Humus Chemistry. Genesis, Composition and Reactions. John Wiley and Sons, New York, NY, USA.

Tian X., Takeda H. and Azuma J. 2000. Dynamics of organic-chemical components in leaf litters during a 3.5-year decomposition. Eur. J. Soil Biol. 36: 81–89.

Wakeham S.G., Lee C., Hedges J.I., Hernes P.J. and Peterson M.L. 1997. Molecular indicators of diagenetic status in marine organic matter. Geochim. Cosmochim. Acta 61: 5363–5369.

Waksman S.A. 1936. Humus. Origin, Chemical Composition, and Importance in Nature. Balliere, Tindall and Cox, London, UK.

Wilson M.A. 1987. N.M.R. Techniques and Applications in Geochemistry and Soil Chemistry. Pergamon Press, Oxford, UK.

Wu Q.-L., Schleuss U. and Blume H.P. 1995. Investigation on soil lipid extraction with different organic solvents. Z. Pflanzener. Boden. 158: 347–350.

Zelibor J.L., Romankiw L., Hatcher P.G. and Colwell R.R. 1988. Comparative analysis of the chemical composition of mixed and pure cultures of green algae and their decomposition residues by $^{13}C$ nuclear magnetic resonance spectroscopy. Appl. Envion. Microb. 54: 1051–1060.