Global Change Biology (2001) **7**, 883–902

Quantifying uncertainty in estimates of C emissions from above-ground biomass due to historic land-use change to cropping in Australia

DAMIAN J. BARRETT*†, IAN E. GALBALLY‡ and R. DEAN GRAETZ§
*CSIRO Plant Industry, GPO Box 1600 Canberra, ACT, 2601, †Cooperative Research Centre for Greenhouse Accounting, GPO Box 475 Canberra, ACT, 2601, ‡CSIRO Atmospheric Research, Private Bag No.1, Aspendale Victoria, 3195, §CSIRO Earth Observation Centre, GPO Box 3023, Canberra, ACT 2601

Abstract

Quantifying continental scale carbon emissions from the oxidation of above-ground plant biomass following land-use change (LUC) is made difficult by the lack of information on how much biomass was present prior to vegetation clearing and on the timing and location of historical LUC. The considerable spatial variability of vegetation and the uncertainty of this variability leads to difficulties in predicting biomass C density (t_C ha^{-1}) prior to LUC. The issue of quantifying uncertainties in the estimation of land based sources and sinks of CO_2, and the feasibility of reducing these uncertainties by further sampling, is critical information required by governments world-wide for public policy development on climate change issues. A quantitative statistical approach is required to calculate confidence intervals (the level of certainty) of estimated cleared above-ground biomass. In this study, a set of high-quality observations of steady state above-ground biomass from relatively undisturbed ecological sites across the Australian continent was combined with vegetation, topographic, climatic and edaphic data sets within a Geographical Information System. A statistical model was developed from the data set of observations to predict potential biomass and the standard error of potential biomass for all 0.05° (approximately 5 × 5 km) land grid cells of the continent. In addition, the spatial autocorrelation of observations and residuals from the statistical model was examined. Finally, total C emissions due to historic LUC to cultivation and cropping were estimated by combining the statistical model with a data set of fractional cropland area per land grid cell, f_{Ac} (Ramankutty & Foley 1998). Total C emissions from loss of above-ground biomass due to cropping since European colonization of Australia was estimated to be 757 Mt_C. These estimates are an upper limit because the predicted steady state biomass may be less than the above-ground biomass immediately prior to LUC because of disturbance. The estimated standard error of total C emissions was calculated from the standard error of predicted biomass, the standard error of f_{Ac} and the spatial autocorrelation of biomass. However, quantitative estimates of the standard error of f_{Ac} were unavailable. Thus, two scenarios were developed to examine the effect of error in f_{Ac} on the error in total C emissions. In the first scenario, in which f_{Ac} was regarded as accurate (i.e. a coefficient of variation, CV, of f_{Ac} = 0.0), the 95% confidence interval of the continental C emissions was 379–1135 Mt_C. In the second scenario, a 50% error in estimated cropland area was assumed (a CV of f_{Ac} = 0.50) and the estimated confidence interval increased to between 350 and 1294 Mt_C. The CV of C emissions for these two scenarios was 25% and 29%. Thus, while accurate maps of land-use change contribute to

Correspondence: D. J. Barrett, CSIRO Plant Industry, GPO Box 1600, Canberra, ACT, 2601, Australia, tel +61 26246 5558, +61 26246 5560, e-mail d.barrrett@pi.csiro.au

decreasing uncertainty in C emissions from LUC, the major source of this uncertainty arises from the prediction accuracy of biomass C density. It is argued that, even with large sample numbers, the high cost of sampling biomass carbon may limit the uncertainty of above-ground biomass to about a CV of 25%.

Keywords: Carbon, greenhouse gas emissions, uncertainty analysis, vegetation clearing, geographical information systems

Received 4 January 2001; revised version received and accepted 12 May 2001

Introduction

Emissions to the atmosphere of carbon from the oxidation of above- and below-ground plant biomass and soil organic matter following land-use change (LUC) to agriculture are important sources of CO_2 contributing to the enhanced greenhouse effect globally. It is estimated that these emissions contribute historically about half as much as fossil fuel emissions to the overall global greenhouse effect and currently about 25% of annual emissions (Houghton 1999; Houghton & Hackler 1999; Houghton *et al*. 1999). To estimate the amount of carbon (C) lost to the atmosphere by clearing of above-ground biomass requires information on above-ground biomass C density (t_C ha^{-1}) of the landscape prior to clearing, the area which has undergone LUC and the above-ground biomass in disturbed areas after clearing. Most of the work to date has focused on reducing the uncertainty of the land area cleared of vegetation using satellite mapping both in Australia (e.g. Woodgate & Black 1988; Queensland Department of Natural Resources 1999) and world-wide (e.g. DeFries *et al*. 1999; Ramankutty & Foley 1998, 1999). However, there may be an equivalent or greater uncertainty in the assignment of above-ground biomass prior to vegetation clearing. The problem of determining the amount of biomass cleared and the uncertainty of this estimate occurs irrespective of the subsequent fate of the biomass after clearing.

Emissions from the oxidation of above-ground biomass carbon are estimated as the product of the above-ground biomass C density (q; t_C ha^{-1}) in its pre-cleared state and the area of land cleared of vegetation (A_L). Estimating the above-ground biomass loss from past vegetation clearing over large regions (of the order of 10^5 km^2), such as continental Australia, is difficult because (i) there are few prior measures of biomass on lands where clearing has occurred, (ii) above-ground biomass is highly variable, (iii) the task of measuring biomass at many locations is not possible given financial constraints, (iv) the location of past (and future) clearing is uncertain, and (v) it is virtually impossible to validate estimates of emissions from past vegetation clearing except globally by examining the C balance of the atmosphere. To quantify reliably the impact of clearing of above-ground biomass on atmospheric CO_2 concentrations using ground-based measurements of biomass, it is necessary to address adequately these five difficulties and to quantify uncertainties in emissions.

In the present context, we define *uncertainty* as inadequate knowledge regarding the statistical parameters (mean, variance and skew) of the frequency distribution of a variable of interest; it is measured here by the standard error of the mean. This is distinct from the definition of *variability*, being the quantitative value of statistical parameters of the frequency distribution; it is measured here by the standard error of the population. The combination of uncertainty and variability of model input variables determines the range and likelihood of model output. Here, modelled output is the predicted C emissions per grid-cell for a continental raster. When calculating C emissions as the product of fraction of land area cleared per grid-cell and mean potential above-ground biomass C density prior to clearing, the appropriate uncertainty is the standard error of the population of residuals of the statistical model. This is because, at the grid-cell scale, we have no information on the exact biomass at sites prior to clearing. The 'true' biomass cleared could have occurred at any point on the frequency distribution function for that grid cell. In contrast, when calculating total above-ground biomass of many grid cells, the standard error of the mean is the appropriate uncertainty because the variability within a grid-cell does not contribute to the grid-cell total if an unbiased mean value is known. Thus, uncertainty in estimated C emissions arises from variability in biomass (i.e. natural heterogeneity) across a grid cell, uncertainty in this variability and uncertainty in the estimate of the fraction of grid-cell area cleared of vegetation. Variability in the fraction of area cleared does not exist because there are only two possible states for this term: uncleared land and land cleared for cropping.

Qualitative sources of uncertainty may also impact on the potential range of emissions. These sources arise from a lack of process understanding or confusion or ambiguity of definitions (see Morgan & Henrion 1990 for a

comprehensive treatment). For the case of above-ground biomass, qualitative errors are likely to be insignificant but may be considerable for other terrestrial sources and sinks. Accurately quantified estimates of uncertainties of CO_2 emissions from land clearing are needed to improve understanding of the global carbon cycle, to report in national inventories prepared under the Intergovernmental Panel on Climate Change national greenhouse gas inventory uncertainty guidelines (Penman *et al*. 2000), and may subsequently be useful for evaluating commitments under the Kyoto Protocol to the United Nations Framework Convention on Climate Change (UN FCCC 1997).

Remotely sensed data archives provide an opportunity to classify land cover and, by tracking changes through time, quantify the location and area undergoing LUC (e.g. Graetz *et al*. 1995; Tucker & Townsend 2000). However, there are difficulties in interpreting the reflected radiation signal as a land cover type (e.g. Kimes *et al*. 1998) and this uncertainty translates to uncertainty in the detection of areas of LUC. Due to the sparseness of data on above-ground biomass at large (continental) scales, and the high resolution and improvement of current satellite imagery, the majority of the uncertainty in estimates of C emissions from above-ground biomass due to LUC will derive from the estimate of above-ground biomass prior to the changed land use.

In Australia, total emissions from the Land Use Change and Forestry (LUCF) Sector of the National Greenhouse Gas Inventory are substantial, being about 25% of the total anthropogenic emissions (NGGIC, 1998). In addition, the uncertainty in these emissions estimates contributes a greater proportion to the total uncertainty of greenhouse gas emissions than other sectors such as fossil fuel, gas flaring and cement manufacture. This situation has arisen because many LUCF inventory terms are estimated on the basis of few observations and are accompanied by qualitative or 'expert judgement' measures of uncertainty. Current estimates of carbon released from land clearing at national or continental scales (e.g. NGGIC 1998; Houghton & Hackler 1999) have at best relatively simple quantitative assessments of uncertainty (e.g. Brown & Gaston 1995; Alexeyev *et al*. 1995; Schuft *et al*. 1998), although robust error analyses at the forest to regional scale have been developed by Gertner & Kohl (1992), Ciezewski *et al*. (1996), Phillips *et al*. (2000) and Schroeder *et al*. (1997).

The small sample numbers used in the LUCF sector characteristically have large standard errors and may poorly represent the population from which they are drawn. Normally more sampling would be recommended but there are no estimates available as to the likely sampling effort required to reduce the uncertainty in land based C emissions at the continental scale to any particular pre-defined level. Explicit determination and incorporation of uncertainty *via* quantitative statistical analysis will improve the level of reliability of emissions estimates from vegetation clearing in a transparent and verifiable manner.

In this paper, we calculate the predicted value and the standard errors of above-ground biomass C density prior to LUC based on observations from minimally disturbed sites. The standard error of the mean was used to examine the relationship between uncertainty in 'steady-state' biomass C density and sample number. The standard error of the residuals was used to estimate the magnitude and uncertainty of C emissions from the clearing of above-ground biomass arising for cultivation and cropping. First, we examine the spatial distribution of the standard error of mean above-ground biomass at the scale of continental Australia and identify regions where major reductions in uncertainty would occur if further sampling was conducted. Second, we explore the relationship between standard error of mean above-ground biomass and sample number to examine the propensity for small sample numbers to yield biased estimates and quantify the rate of convergence of the standard error of the mean towards zero as sample number increases. Finally, we quantify the impact of error in mapping LUC on uncertainties of estimated C emissions to the atmosphere from above-ground biomass by combining the statistical model of above-ground biomass, information on the spatial correlation of biomass and a spatial data set of land area under cultivation (Ramankutty & Foley 1998). The methodology used here is generally applicable to other locations and other biosphere C stocks where emissions estimates are made on the basis of few observations. Emissions from below-ground biomass, littermass C and soil C decay will be addressed in future papers; although some of the issues related to the estimation of the release of carbon from land clearing and regrowth for national greenhouse gas inventories have been addressed previously (Wang *et al*. 1997).

Materials and methods

Linear model formulation

This analysis develops a statistical model of steady state biomass for a continent, in this case Australia, as a function of climatic, topographic, vegetation and edaphic variables. The term 'steady state' (i.e. equilibrium) is defined as used in compartmental analysis (Jacquez 1985) to describe a system in which the rate of change of compartment mass (a 'reservoir' or 'stock') with respect to time is zero. Thus, for above-ground biomass, the

vegetation stock has attained the maximum value when at steady state, given climatic and edaphic constraints. In reality, this condition may never exist as all systems are disturbed at varying frequencies and at a range of scales. Nonetheless, the trajectories of natural systems in state space may oscillate around steady state. Where disturbance frequencies of natural systems are low, such as where no recent stand-replacing fire events have occurred, stocks may be close enough to the equilibrium point – given the level of natural heterogeneity – to assume steady state. Higher disturbance frequency may reduce biomass density considerably and because this information is not available in the present work, the estimates of C emissions based on steady state values are an upper limit which may, in the future, be revised downward when further information on disturbance and pre-clearing vegetation stocks becomes available.

The biomass model is calibrated using a geographically referenced high-quality biomass data set (see section *Application to the Australian Continent*, below, for selection criteria of observations) that includes all acceptable published biomass measures plus associated rainfall, temperature, vegetation, topographic and soil attributes. Published biomass data from the scientific literature, being peer reviewed are the best available information on the variability of above-ground biomass and were used to develop the observation data set in a GIS (ArcView, Environmental Systems Research Institute, Redlands, CA, USA). From the calibrated model, the spatial distribution of above-ground biomass and its standard error are predicted.

A multiple linear regression (MLR) statistical model (Sokal & Rohlf 1995) was developed to predict biomass prior to LUC. Data for auxiliary variables were extracted from coverages within the GIS by overlaying the geographical locations of study sites on the existing data sets. Model development continued by iterative fitting of various combinations of the auxiliary variables (vegetation, topography, climate and soil variables) until the most parsimonious model was produced, i.e. the model that explained the greatest proportion of variation in the dependent variable (biomass observations) using the least number of auxiliary variables. The MLR model partitions variance in the observation data set into systematic and random terms and it is the measure of variance 'explained' by the systematic term that constitutes the model's ability to make predictions (statistical inferences) in parameter space.

Spatial autocorrelation

Examination of the spatial correlation of observations and residuals is necessary for three reasons: (i) the least squares method used to parameterize the regression coefficients of the MLR model assumes independence of residuals and so the presence of autocorrelation would lead to bias in the prediction variance; (ii) the presence of autocorrelation in the residuals implies that the regression model is incomplete and that further auxiliary variables are required to explain variation in the dependent variable; and (iii) variances of totals obtained by summing predicted values and their variances over many grid cells will be underestimated if the autocorrelation is not considered.

Moran's statistic of spatial autocorrelation, I_d, is a measure of the dependency of ecological variables on neighbouring observations (Cliff & Ord 1981; Legendre 1993). The spatial correlation of observed above-ground biomass was calculated and is presented as a correlogram of I_d against distance class, d, where d is a measure of the distance separating observations of above-ground biomass. Moran's statistic is the ratio of covariance to variance of observations grouped by distance class:

$$I_d = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(q_i - \overline{q})(q_j - \overline{q})}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \right) \sum_{i=1}^{n} (q_i - \overline{q})^2} \quad (1)$$

where q is an observation of above-ground biomass for the i, $j = 1$ to n observations (subject to $i \neq j$), the overbar represents the sample mean of observations and $w_{i\,j}$ is the weighting applied to each observation dependent on the distance between observations. In the present work, Euclidean distances between observations were calculated by Pythagoras' theorem between all pairs of observations using the latitude and longitude of each study in a square matrix (Czaplewski *et al*. 1994) and then grouping observations into the $i = 1$ to n distance classes. The form of the weighting coefficients was taken from Legendre & Fortin (1989) where observations within a distance class were assigned a weight of 1.0 and 0.0 otherwise.

Relationship between sample size and standard error of the mean

The data set of above-ground biomass observations was used to explore the decrease in uncertainty in predicted biomass with increase in sample number used to generate the MLR model. The decrease in uncertainty arises from the central tendency of large samples to converge on the 'true' mean value. Thus, not only does the standard error of the mean decrease with larger sample numbers but the accuracy of the estimated standard error increases. The approach outlined below explores how

chance selection of observations from the literature may bias statistical inferences made about biomass particularly when sample numbers are small. By re-sampling the data set and re-calculating the predicted value and standard error of mean biomass for different sample sizes, the relationship between the uncertainty of model prediction and sample number was developed. Specifically, the observation data set of biomass was re-sampled 40 times (with replacement) for 15 sample sizes (n = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 and 75). For each of these 600 realizations, the regression coefficients of the MLR model were re-calculated and used to predict biomass and standard error of mean biomass at a nominal value of each of the auxiliary variables (values at about the mid-range of each auxiliary variable were used). Details of the calculation of standard errors of the mean (and of the residuals) from the MLR are given in Appendix 1.

Estimated C emissions from land use change and their uncertainty

The MLR model was combined with a data set of fractional cropland area per grid-cell to estimate C emissions from historic land-use change to cultivation and cropping on the assumption that all carbon in plant matter is oxidized to the atmosphere after LUC. The fractional cropland area data were developed by Ramankutty & Foley (1998) using the DISCover data set of Loveland & Belward (1997). The DISCover data set was developed using an unsupervised classification of monthly maximum value composites of the 1-km NOAA-AVHRR Normalized Difference Vegetation Index data between April 1992 and March 1993. Vegetation classification is based on the spectral characteristics and the temporal variation of the radiation signal received by the satellite. The preliminary clustering of the unsupervised classification was followed by a post-classification assignment of vegetation classes using existing national vegetation, land use, land cover and soil data sets (Loveland & Belward 1997). A subsequent validation study of the DISCover data set showed that, for the cropland vegetation class world-wide, 64% of the validation samples (i.e. 18 out of 28) were correctly assigned (Scepan 1999).

Ramankutty & Foley (1998) reclassified the 'Seasonal Land Cover Regions' scheme of the DISCover-1.2 data set into six categories of increasing proportion of crop area. They defined cropland as arable land used for cropping or in fallow including land used for temporary pasture and cocoa, coffee and rubber tree crops (of which there are none in Australia). Permanent pastures were omitted. In the Australian context, this is interpreted here as native pastures, sown perennial pastures and sown annual pastures. To spatially distribute statistics of cropland area for geo-political units, the fraction of cropland area of each of the six cropland categories was varied iteratively until the errors between the aggregated cropland area based on the DISCover data set and the area reported by government agencies for each political unit were minimized. The constraint used in their analysis was that the rank-order of crop area of the six vegetation classes remained constant during the minimization process. For Australia, the political units were the eight states and territories of the Australian mainland and Tasmania. The 1990 cropland area data of the Australian Bureau of Statistics (ABS) were used to calculate the proportion of total cropland area for each state and territory in that year. These proportions were then applied to the United Nations Food and Agriculture Organization statistics for 1992 to arrive at the aggregated cropland area values.

The estimate of C emissions from the $(i,j)^{\text{th}}$ grid cell of the continental raster was calculated as

$$\hat{F}_{ij} = \hat{q}_{ij}\hat{A}_{\text{L}ij} \tag{2}$$

where $\hat{q}$ is the predicted above-ground biomass of the grid-cell and $\hat{A}_{\text{L}ij}$ is the predicted cropland area per grid cell,

$$\hat{A}_{\text{L}ij} = \hat{f}_{\text{AC}ij}\hat{A}_{\text{C}ij} \tag{3}$$

where $\hat{f}_{\text{AC}ij}$ is the predicted fractional cropland area and $A_{\text{C}ij}$ is the grid-cell total area. For convenience we used a log-transformation of equation 2 to simplify error propagation because the residuals of above-ground biomass were log normally distributed (see Results):

$$\ln(\hat{F}_{ij}) = \ln(\hat{q}_{ijh}) + \ln(\hat{A}_{\text{L}ij}). \tag{4}$$

Exponentiation of Equation 4 gives the estimated modal (geometric mean) C emissions. However, for a log-normal model, the estimated arithmetic mean C emissions per grid cell are skewed relative to the mode. The unbiased mean C emissions are calculated as (Harris & Stocker 1998)

$$\hat{F}_{ij} = \exp[\ln(\hat{F}_{ij}) + \frac{1}{2}\sigma^2_{\ln(\hat{F}_{ij})}].$$

The latitudinal total C emissions (i.e. of the i^{th} row) is

$$\hat{F}_i = \sum_{j=1}^{M} \hat{F}_{ij} \tag{5}$$

and the continental total C emissions were calculated as

$$\hat{F} = \sum_{i=1}^{N} \sum_{j=1}^{M} \hat{F}_{ij}. \quad (6)$$

The estimated standard error of C emissions of the (i^{th}, j^{th}) grid cell after exponentiation, $\sigma_{\hat{F}_{ij}}$, was calculated from the first-order Taylor expansion of equation 4 for the random variables $\hat{q}_{ij}$ and $\hat{A}_{Lij}$ assuming that the area of each grid cell is known with zero error (see Appendix 2). The spatial correlation of $\hat{q}_{ij}$ was incorporated into the estimated variance of the total latitudinal and total continental emissions as described in Appendix 3.

Application to the Australian continent

A geographically referenced data set of above-ground biomass measurements was established based on opportune, but robust, estimates of plant mass for known locations across Australia. To maintain data quality, the following criteria determined acceptance of reported biomass measurements into the observation data set:

- Only data from the published scientific literature were used. This ensured measurements met a standard of peer review.
- Studies were included where latitude and longitude of each site was noted by the authors or, where absent, was identifiable from a described feature using Australian Surveying and Land Information Group (AUSLIG) 1 : 1000 000 Maps to 0.1 degree accuracy.
- Where multiple sites were measured by authors, but were described by a single latitude and longitude reference, the average biomass was used for all sites with the consequential loss of a small amount of information on short-scale spatially correlated variance.
- Where an age sequence of vegetation was measured at a single site, the oldest age in the sequence was used, on the assumption that the oldest site would be closest to steady state conditions.
- The description of vegetation species composition, growth form and structure from each study was checked against the following criterion of steady state. Published studies were considered to be at or near steady state when the author's description of the growth form of the overstorey vegetation matched the description of the tallest stratum in the Australian Surveying & Land Information Group 1788 Australian Vegetation Classification (AUSLIG 1990). Thus, these observations of above-ground biomass are interpreted as the C stock under minimally disturbed conditions where the return period of stand replacing fires is longer than the time to maximum biomass. This approach ensured that overstorey vegetation composition was the same in 1788 as the present but may ignore land use leading to thinning or thickening of biomass in the period 1788 to present unless remnant vegetation was historically reserved from wood harvesting or grazing activity. Such effects, if they occur, will add to the uncertainty of the estimate *via* the residual mean square of the statistical model.

The statistical models were developed using four pre-existing data sets to supply values for the auxiliary variables:

1 The AUSLIG Digital Vegetation Atlas (AUSLIG 1990) from which vegetative growth form of the tallest stratum (T, tall trees > 30 m height; M, medium trees 10 –30 m; L, low trees < 10 m; S, tall shrubs > 2 m; Z, low shrubs < 2 m; H, hummock grasses; G, tussock or tufted grasses; F, other herbaceous plants and percentage foliage projected cover attributes (1, < 10% cover; 2, 10–30% cover; 3, 30–70% cover; 4, > 70% cover) were extracted.

2 The digital 1 : 2000 000 Atlas of Australian Soils Northcote *et al*. (1968) combined with look-up tables developed by McKenzie & Hook 1992) to convert soil classification to soil attributes. The soil attribute used in this study was gross nutrient status of the soil (1 = 'low; vegetation likely to have a major response to N, P and K along with most micronutrients', 2 = 'moderate; responses by vegetation to N and P with occasional response to some micronutrients', and 3 = 'high; responses by vegetation to N and P uncommon except after intensive farming').

3 The 0.25° Australian Bureau of Meteorology Research Centre and the National Climate Centre interpolated climate surfaces for monthly maximum and minimum temperature and rainfall across Australia (Jones 1999). These data sets were re-processed to yield mean annual rainfall and temperature from the monthly data.

4 The Australian National University, Centre for Resource and Environmental Science, 0.05° Digital Elevation Model (ANUDEM) from Hutchinson (1998) was used to obtain elevation and delineate the land/sea mask for the continent in conjunction with the Bureau of Meteorology climate surfaces.

Results and discussion

Observation data set of above-ground biomass C density

A total of 76 study sites were obtained from 49 published studies of above-ground biomass (Table 1 and Fig. 1) that met the quality control criteria outlined in the Methods, above. Fourteen of these observations were subsequently omitted from parameterization of the MLR

Table 1 The geographically referenced biomass data set of the Australian continent based on a sample of sites (n = 76) from the published scientific literature

Latitude (°)	Longitude (°)	State	Vegetation type[1]	Growth form and foliage[2]	Temperature[3] (°C)	Rainfall[3] (mm)	Factual key[4]	GNS[5]	Biomass[6] (t_C ha^{-1})	Source[7]
12.67	132.92	NT	1	M2	27.55	1458	Va73	1	1.4	1
12.67	132.92	NT	1	M2	27.55	1458	Va73	1	1.4	1
24.50	149.80	Qld	2	M2	21.27	695	CC23	2	53.6	2
28.50	153.00	NSW	3	M3	19.08	1277	Ml2	3	172.5	3
28.50	153.00	NSW	3	M3	19.08	1277	Ml2	3	167.5	3
28.50	153.00	NSW	3	M3	19.08	1277	Ml2	3	176.8	3
37.27	149.40	NSW	4	M3	13.41	982	Pb5	2	135.0	4
38.33	146.25	Vic	5	T3	12.36	1037	Tb14	1	147.6	5
37.43	145.13	Vic	5	T3	13.57	753	Mk3	2	166.5	6
33.55	150.37	NSW	6	M3	13.40	1058	Mb2	1	188.2	7
30.20	153.13	NSW	7	T3	17.10	1671	Pb17	2	196.2	8
37.00	149.50	NSW	8	T3	12.57	906	Pb5	2	196.0	9
27.28	149.75	Qld	9	M3	19.50	616	My3	1	53.5	10
35.87	140.45	SA	10	Z3	15.66	474	Ca2	1	11.9	11
35.87	140.45	SA	11	Z3	15.66	474	Ca2	1	4.1	12
26.42	146.22	Qld	12	L2	20.81	452	B12	1	1.1	13
31.90	141.45	NSW	13	L2	17.83	278	F5	1	1.0	14
27.70	153.45	Qld	14	M3	20.08	1320	B9	1	46.6	15
32.67	116.08	WA	15	M3	16.61	989	JZ1	1	119.8	16
37.42	145.13	Vic	16	T3	13.57	753	Mk3	2	374.1	17
37.42	145.13	Vic	17	T3	13.57	753	Mk3	2	417.9	17
37.43	145.58	Vic	18	T3	12.50	1092	Tb1	1	294.5	18
37.43	145.58	Vic	19	T3	12.50	1092	Tb1	1	168.0	18
37.59	145.60	Vic	20	T3	13.52	1124	Mh5	2	214.2	19
30.42	153.00	NSW	21	T3	18.53	1662	Tb43	1	214.7	20
38.33	146.25	Vic	22	T3	12.36	1037	Tb14	1	9.6	21
26.68	146.17	Qld	23	L2	20.87	439	My4	1	21.8	22
26.33	146.27	Qld	23	L2	20.70	474	Nb3	2	21.0	22
26.38	146.25	Qld	23	L2	20.70	474	CC19	2	18.9	22
32.12	139.37	SA	24	L1	17.33	303	BB1	1	0.6	23
37.42	149.55	Vic	25	M3	13.82	963	Pb5	2	147.3	24
32.85	145.98	NSW	26	M2	17.73	476	My1	1	0.0	25
32.85	145.98	NSW	27	M2	17.73	476	My1	1	0.0	25
32.85	145.98	NSW	28	M2	17.73	476	My1	1	4.5	26
25.47	153.07	Qld	29	M3	21.33	1190	B18	1	207.0	27
34.45	116.03	WA	30	T3	15.49	1075	Uc1	1	112.1	28
35.03	144.97	NSW	31	L2	16.50	389	CC3	2	1.6	29
30.92	146.50	NSW	32	M2	19.15	465	Mx1	1	24.6	30
29.87	115.25	WA	33	S2	19.27	526	Ca27	1	6.3	31
23.70	133.88	NT	34	L1	20.83	329	AB71	1	1.8	32
23.70	133.88	NT	35	L1	20.83	329	AB71	1	1.2	32
23.70	133.88	NT	36	L1	20.83	329	AB71	1	0.6	32
38.13	145.13	Vic	37	M2	14.16	787	Fu28	1	4.3	33
38.13	145.13	Vic	37	M2	14.16	787	Fu28	1	4.6	33
26.50	153.00	Qld	38	M3	20.28	1646	Wa18	1	2.6	34
35.60	141.10	NSW	39	S2	15.44	458	Ca2	1	60.1	35
32.88	137.38	SA	40	L1	17.25	264	DD2	1	0.2	36
32.88	137.38	SA	40	L1	17.25	264	DD2	1	0.2	36
23.55	133.60	NT	41	S1	20.06	325	My118	1	6.7	37
33.82	151.15	NSW	42	M3	17.23	1087	Tb35	1	57.8	38
14.47	132.32	NT	43	L1	27.25	1009	My70	1	0.9	39
32.42	142.42	NSW	44	L1	18.63	315	II1	2	0.2	40
35.42	148.80	ACT	45	M3	9.56	910	Fa3	1	141.0	41
35.36	148.80	ACT	46	M3	9.56	910	Fa3	1	117.7	42
23.57	133.57	NT	47	S1	20.06	325	My118	1	2.7	43

Table 1 *(continued)*

Latitude (°)	Longitude (°)	State	Vegetation type[1]	Growth form and foliage[2]	Temperature[3] (°C)	Rainfall[3] (mm)	Factual key[4]	GNS[5]	Biomass[6] (t_C ha^{-1})	Source[7]
23.57	133.60	NT	48	S1	20.06	325	My118	1	0.2	43
30.28	145.90	NSW	49	L2	19.63	383	Mx1	1	1.4	43
26.48	146.47	Qld	49	L2	20.70	474	My4	1	2.1	43
20.35	142.63	NSW	49	G3	25.38	492	MM44	2	4.5	43
19.67	145.75	Qld	50	M1	23.60	586	Kb24	2	13.1	43
20.18	146.72	Qld	50	M1	23.57	693	Qa14	2	8.7	43
20.23	145.87	Qld	50	M1	23.40	642	MS8	1	11.9	43
20.23	145.87	Qld	50	M1	23.40	642	MS8	1	11.8	43
14.52	132.27	NT	50	L1	27.17	977	My70	1	14.2	43
14.80	131.80	NT	50	M2	27.12	939	My70	1	20.0	43
12.67	132.42	NT	50	M3	27.60	1428	JY2	1	25.9	43
33.96	151.22	NSW	51	M3	17.23	1087	Cb27	1	20.3	44
34.13	143.30	NSW	52	S2	17.23	385	DD3	1	0.1	45
35.03	144.97	NSW	53	L2	16.50	389	CC3	2	1.3	46
35.03	144.97	NSW	53	L2	16.50	389	CC3	2	1.2	46
35.03	144.97	NSW	53	L2	16.50	389	CC3	2	1.2	46
31.90	141.50	NSW	53	L2	17.85	293	F5	1	0.5	47
38.87	146.40	Vic	54	M3	13.01	974	Cb12	1	3.4	48
36.03	140.50	SA	54	S1	15.45	487	Ya11	1	12.6	49
39.03	146.32	Vic	55	M3	13.53	936	A5	1	8.1	49
38.87	146.40	Vic	56	M3	13.01	974	Cb12	1	6.1	49

[1]Author's description of growth form of overstorey vegetation: 1, eucalypt woodland; 2, brigalow woodland (*Acacia harpophylla*); 3, sub-tropical rain forest; 4, *E. seiberi* forest; 5, *E.obliqua* forest; 6, *E. maculata* forest; 7, *E. grandis* forest; 8, *E. seiberi/E. muellerana/E. obliqua* forest; 9, Brigalow woodland; 10, *Banksia ornata/Xanthoria australis* coastal heath; 11, *E. incrasata/Melaleuca uncinata*; 12, *Eremophila gilesii* shrubland; 13, *Atriplex vesicaria* shrubland; 14, *E. stignata/E.umbra* forest; 15, *E. marginata/E.calophylla* forest; 16, *E. regnans* forest; 17, *E. sieberi* forest; 18, *E. regnans* forest; 19, *E. obliqua* forest; 20, *E. regnans* forest; 21, *E. grandis* forest; 22, *E. globulus* forest; 23, brigalow woodland; 24, *Maireana sedifolia* shrubland; 25, *E. agglomerata/E. muellerana/E.sieberi* forest; 26, *E. oleosa/E.dumosa* forest; 27, *E. incrsata* forest; 28, *E.oleosa/E.dumosa* forest; 29, *E. pilularis* forest; 30, *E.diversicolor* forest; 31, *Atriplex vesicaria* shrubland; 32, *E.populnea* woodland; 33, *Banksia* shrubland; 34, *Astrebla* grassland; 35, open woodland; 36, *Acacia* woodland; 37, *Leptospermum myrsinoides/Banksia marginata* shrubland; 38, *E.micrantha* forest; 39, *E. radiata/E. dalrypleana* forest; 40, *Mareiana sedifolia/Atriplex* shrubland; 41, *Eragrostis eriopoda/Acacia aneura* shrubland; 42, *Avicenia mariana* forest; 43, grassland; 44, *E. largiflorens/Casuraina stricta*; 45, *E. delegatensis* forest; 46, *E. pauciflora/E dives* forest; 47, *Acacia aneura* wooland; 48, native grassland; 49, *Acacia aneura* woodland; 50, Eucalypt savanna; 51, *Banksia/Leptospermum/Acacia* woodaland; 52, *Atriplex numularia/Maireana pyrimidata* shrubland; 53, *Atriplex vesicaria/Kocia aphylla* shrubland; 54, *Banksia* coastal heath; 55, *Leptospermum myrsinoides/Casuarina pusilla*; 56, *Leptospermum myrsinoides/Xanthorea australis* shrubland.

[2]Growth form and foliage cover of tallest stratum from 1788 AUSLIG vegetation classification at location of each study. The AUSLIG data set uses an alphanumeric code to describe growth form (T, tall trees >30 m; M, medium trees 10–20 m; L, low trees <10 m; S, shrubs >2 m; Z, shrubs <2 m; G, tussocky or tufted grasses and graminoids) and projected foliage cover (1, foliage cover <10%; 2, foliage cover 10–30%; 3, foliage cover 30–70%; 4, foliage cover >70%).

[3]25 years mean annual temperature and mean annual rainfall of the 0.25° grid cell within which the study resides based on Bureau of Meteorology interpolated climate surfaces.

[4]'Factual Key' soil description of Northcote *et al.* (1968) from the Atlas of Australian Soils at the study site location. The Factual Key describes the characteristic soils comprising each polygon. An abbreviated description is provided here: A5; Coastal dunes of deep calcareous sand, some red-brown sandy soils, AB71; Red earthy sands on plains, B12; Siliceous sands, other sandy soils and some clay pans, B18; Dunes of silicous sands and leached sands, B9; Siliceous sands of dunes and estuaries, BB1; Hills and plains of loamy soils, brown sands and brown calcareous earths, Ca2; leached sands, sandy alkaline yellow mottled soils. Clay substrata, Ca27; leached sands, occasional sand dunes. Clay substrata, Cb12; Coastal plains of leached sands and undescribed soils, Cb27; Coastal plains of chiefly leached sands, dunes of siliceous or calcareous sands, CC19; Plains of chiefly grey and brown clays, CC23; alluvial plains of grey and brown clays, CC3; riverine plains of cracking clay with grey and brown clays, DD2; Plains of chiefly brown calcareous earths, DD3; Plains of chiefly brown calcareous earths having distinct micro-relief, F5; Hills and plains of shallow loamy soils, Fa3; Mountainous soils of loamy soils with areas of yellow brown and red earths, Fu28; Hilly lands having shallow gravely loams or duplex soils, II1; Level plains of cracking clay soils, JY2; Dissected rolling lowlands of gravels, shallow sands and occasional red or yellow earths, JZ1; Dissected plateau of sandy loams particular to the western part of the Darling Range in W.A., Kb24; Dark clays of occasional deep soils, Mb2; Dissected sandstone plateau of chiefly leached yellow earths and siliceous sands, Mh5; Mountainous soils of red or yellow earths or shallow grey brown sandy soils, Mk3; Dissected tableland and escarpment of red or yellow earths, Ml2; Steep rounded or hilly terrain of chiefly friable earths, MM44; undulating grey or brown clay plains, MS8;

Undulating plains with drainage depressions of loamy yellow earths, Mx1; Undulating plains of low dunes and occasional stony ridges. Soils: sandy and loamy red earths, My1; Undulating tablelands of generally neutral red earths, My118; Plains of neutral red earths, My3; Undulating plains of red earths with some yellow earths, My4; Undulating plains with some dune or stony ridges. Soils of sandy or loamy red earths, My70; flat plains of neutral loamy red earths, Nb3; Plains and drainage lines of crusty loamy soils, Pb17; Hilly of acidic red soils, Pb5; Mountainous of acidic red soils, Qa14; Hilly to steep terrain of loamy red duplex soils, Tb1; Hills and plains of acidic yellow mottled soils and neutral red soils, Tb14; Undulating of acidic yellow mottled soils, Tb35; Dissected plateau of acidic yellow and yellow mottled soils, Tb43; Undulating to Hilly terrain of acidic yellow mottled soils, Uc1; Steep terrain of neutral to acidic yellow mottled soils and some red earths, Va73; Flat to gently sloping plains and flood-plains of yellow mottled and yellow soils. Some gravels, Wa18; Hilly terrain of acidic yellow mottled soils, Ya11; Hills and plains of sandy alkaline yellow mottled soils with some leached sands.

[5]Gross Nutrient Status (GNS) of soil based on McKenzie & Hook (1992) interpretation of the *Atlas of Australian Soils*. Levels are: 1, low; 2, moderate; 3, high. See text for further description.

[6]Biomass reported within each study as t_{DM} ha^{-1} converted to t_C ha^{-1} assuming a dry weight to C conversion coefficient of 0.45.

[7]Sources of published studies used to construct data set:

1 Cook GD, Andrew MH (1991) *Australian Journal of Ecology* **16**, 375–384. 2 Scanlan JC (1991) *Australian Journal of Ecology* **16**, 521–529. 3 Turner J, Lambert MJ, Kelly J (1989) *Annals of Botany* **63**, 635–642. 4 Turner J, Lambert MJ, Holmes G (1992) *Forest Ecology and Management* **55**, 135–148. 5 Baker TG, Attiwill PM (1985) *Forest Ecology and Management* **13**, 41–52. 6 Attiwill P (1979) *Australian Journal of Botany* **27**, 439–458. 7 Ash J, Helman, C (1990) *Cunninghamia* **2**, 167–182. 8 Turner J, Lambert MJ (1983) *Forest Ecology and Management* **6**, 155–168. 9 Turner J, Lambert MJ (1986) *Oecologia* **70**, 140–148. 10 Moore AW, Russell JS, Coaldrake JE (1967) *Australian Journal of Botany* **15**, 11–24. 11 Specht RL, Rayson P, Jackman ME (1958) *Australian Journal of Botany* **6**, 59–88. 12 Specht R (1966) *Australian Journal of Botany* **14**, 361–371. 13 Burrows WH (1972) *Australian Journal of Botany* **20**, 317–329. 14 Charley JL, Cowling SW (1968) *Proceedings of the Ecological Society of Australia* **3**, 28–38. 15 Westman W, Rogers R (1977) *Australian Journal of Botany* **25**, 171–191. 16 Hingston F, Dimmock G, Turton A (1980) *Forest Ecology and Management* **3**, 183–207. 17 Ashton DH (1976) *Journal of Ecology* **64**, 171–86. 18 Feller M (1980) *Australian Journal of Ecology* **5**, 309–333. 19 Tajchman S, Benyon R, Bren L, Kochenderfer J, Plan C *et al.* (1996) *Biomass and Bioenergy* **11**, 383–386. 20 Bradstock R (1981) *Australian Forest Research* **11**, 111–127. 21 Cromer R, Williams E (1982) *Australian Journal of Botany* **30**, 265–278. 22 Pressland AJ (1975) *Australian Journal of Botany* **23**, 965–976. 23 Noble IR (1977) *Australian Journal of Botany* 25, 639–653. 24 Stewart H, Flinn D, Aeberli B (1979) *Australian Journal of Botany* **27**, 725–740. 25 Holland P (1968) *Australian Journal of Botany* **16**, 615–622. 26 Holland P (1969) *Ecology* **50**, 212–218. 27 Applegate GB (1989) *Australian Forestry* **52**, 195–196. 28 Grove T, Malajczuk N (1985) *Forest Ecology and Management* **11**, 59–74. 29 Rixon (1969) In *The Biology of* Atriplex (ed. R Jones) 128 p. 30 Harrington G (1979) *Australian Journal of Botany* **27**, 135–143. 31 Low A, Lamont B (1990) *Australian Journal of Botany* **38**, 351–359. 32 Friedel M (1981) *Australian Journal of Botany* **29**, 219–231. 33 Jones R (1968) *Australian Journal of Botany* **16**, 589–602. 34 Connor DJ, Wilson GL (1968) *Australian Journal of Botany* **16**, 117–123. 35 Turner J (1980) *Australian Forest Research* **10**, 289–294. 36 Andrew MH, Lange RT (1986) *Australian Journal of Ecology*. **11**, 395–409. 37 Ross MA (1977) *Australian Journal of Ecology* **2**, 257–268. 38 Briggs SV (1977) *Australian Journal of Ecology* **2**, 369–373. 39 Ive JR (1976) *Australian Journal of Ecology* **1**, 185–196. 40 Robertson G (1988) *Australian Journal of Ecology* **13**, 519–528. 41 Crane WJB Raison RJ (1980) *Australian Forestry* **43**, 253–260. 42 Keith H Raison RJ Jacobsen KL (1997) *Plant and Soil* **196**, 81–99. 43 Walker BH, Langridge JL (1997) *Journal of Biogeography* **24**, 813–825. 44 Maggs J, Pearson C (1977) *Oecologia* **31**, 239–250. 45 Marshall JK (1979) In *Studies of the Australian Arid Zone* (eds RD Graetz and KMW Howes). 196 p. 46 Leigh JH Mulham WE (1966) *Australian Journal of Experimental Agriculture and Animal Husbandry* **6**, 460–467. 47 Roshier DA Nicol HI (1998) *Rangelands Journal* **20**, 3–25. 48 Groves RH (1965) *Australian Journal of Botany* **13**, 281–289. 49 Groves RH Specht RL (1965) *Australian Journal of Botany* **13**, 261–280.

model because they were regarded as biased. In five cases, measurements consisted of only the grass component beneath a woody overstorey, six sites were of heath vegetation on sandy soils of exceptionally low nutrient status in otherwise high rainfall zones, one site was exposed to salt spray in a coastal location, another site was a young fertilized plantation forest and the methods in another study were incomplete such that a confident estimate of total above-ground biomass could not be made.

Figure 1 shows the observation data used in development of the MLR model plotted as a function of mean annual rainfall, along with data omitted from model parameterization and a set of independent measures of above-ground biomass of central Queensland woodlands (Burrows *et al*. 2000). These independent observations overlay and have a similar dispersion about the regression function to that of the continental data set and so partly validate the data used to construct the MLR model. It was assumed that the biomass measurements of the observation data set represented the maximum biomass possible for each grid cell given biophysical constraints of water availability, temperature, nutrient supply and minimal disturbance. Furthermore, it was assumed that the whole data set of biomass observations was representative of the range of variation in above-ground biomass at the continental scale. Given that ecological studies of vegetation biomass are likely to be biased towards the highest biomass and least disturbed sites within the landscape, particularly in developed

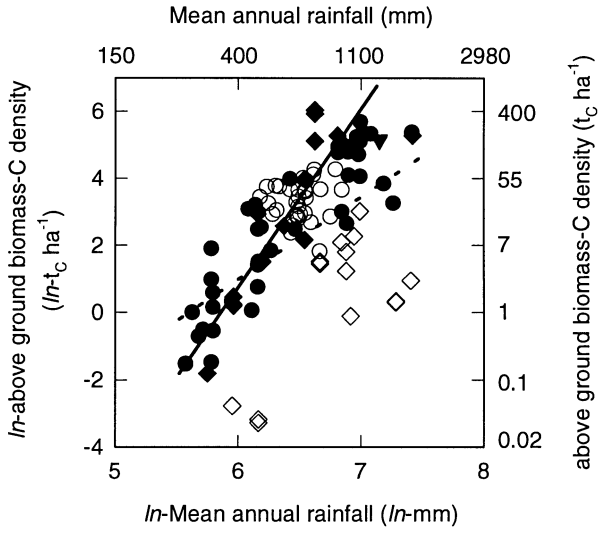

Fig. 1 Scatterplot of observations of above-ground biomass density as a function of mean annual rainfall. ●, ◆ and ▲ depict observations used in the development of the multiple linear regression (MLR) model for Gross Nutrient Status classes 1, 2 and 3, respectively (see Methods for details of nutrient classes). ◇, observations not included in the MLR model due to bias. ○, independent observations of above-ground biomass by Burrows *et al*. (2001) for sites in central Queensland not used in the MLR. Solid and dotted lines depict the MLR model for mean annual temperature of 10 °C and 25 °C, respectively.

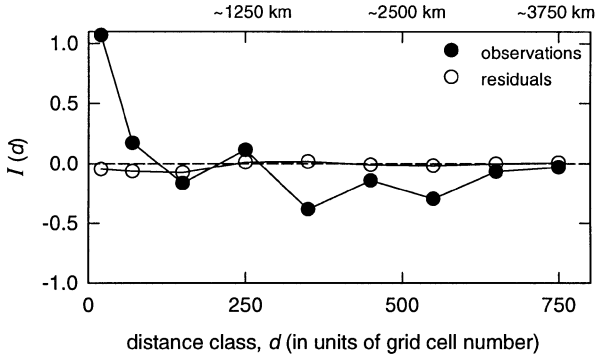

Fig. 2 Correlogram of spatial autocorrelation of above-ground biomass observations and residuals of the Multiple Linear Regression model against distance class, *d*. Spatial correlation was calculated as Moran's *I* statistic, I_d (equation 1). Distance classes are in units of grid-cell number having 0.05° resolution (i.e. *c.* 5 × 5 km).

countries (Iverson *et al*. 1994; Brown & Gaston 1995; Schroeder *et al*. 1997), the assumption of steady state in the present work is reasonable.

The values of above-ground biomass reported in the data set varied from 0.04 t_C ha^{-1} in the lowest rainfall regions of the continent to approximately 420 t_C ha^{-1} for the most productive forests of Victoria (Table 1), which is in broad agreement with the maximum value of 400 t_C ha^{-1} by Grierson *et al*. (1992) for Victorian upland forests where heights of *Eucalyptus regnans* trees on undisturbed sites can attain nearly 80 m (Ashton 1975). These biomass maxima are comparable to the upper range of biomass reported for forests globally. For example, the maximum biomass of lowland tropical forests of Malaysia and central Africa are of the order of 400–500 t_C ha^{-1} (Brown *et al*. 1993; Brown & Gaston 1995).

Statistical inferences of above-ground biomass C density

It was necessary to transform the biomass data by natural logarithm (Fig. 1) to maintain homogeneity of the residuals as required under the assumptions of the least-squares regression model. In the most parsimonious MLR model, 81% of the variability of natural log-transformed biomass was explained by the auxiliary variables of mean annual rainfall, mean annual temperature, GNS and the interaction between mean annual rainfall and mean annual temperature. Other possible models that included vegetation structure or canopy cover as significant terms were rejected in favour of the rainfall/temperature/soil nutrient model (despite a greater 'explained' variance) because of circularity arising from inclusion of vegetation characteristics as auxiliary variables in a model used to explain another vegetation characteristic (i.e. above-ground biomass). The final MLR model was

$$\ln(\hat{q}) = -37.80 + \beta_i + 6.52\ln(x_R) + 0.933\,(x_T) - 0.1578\,(\ln(x_R)\cdot x_T) \quad (7)$$

where $\beta_i = 0$ for GNS class 1, 0.26 for class 2 and 0.10 for class 3, ln-x_R is log-transformed mean annual rainfall and x_T is mean annual temperature. Surprisingly, the inclusion of other soil terms such as soil hydraulic conductivity, soil pH and soil depth did not contribute significantly to the explained variance of the final model. Thus, any variation in biomass induced by these other soil properties is small and not systematic at the continental scale.

The interaction between the effects of mean annual rainfall and mean annual temperature on C stock arises from the effect of environmental variables on C influx by photosynthesis and efflux by respiration and mortality. At mean annual rainfall < 550 mm an increase in mean annual temperature is likely to decrease soil water availability by increased evaporation from the soil leading to a reduction in vegetation productivity and hence biomass density per unit rainfall. Biomass turnover is faster more in tropical than in subtropical and

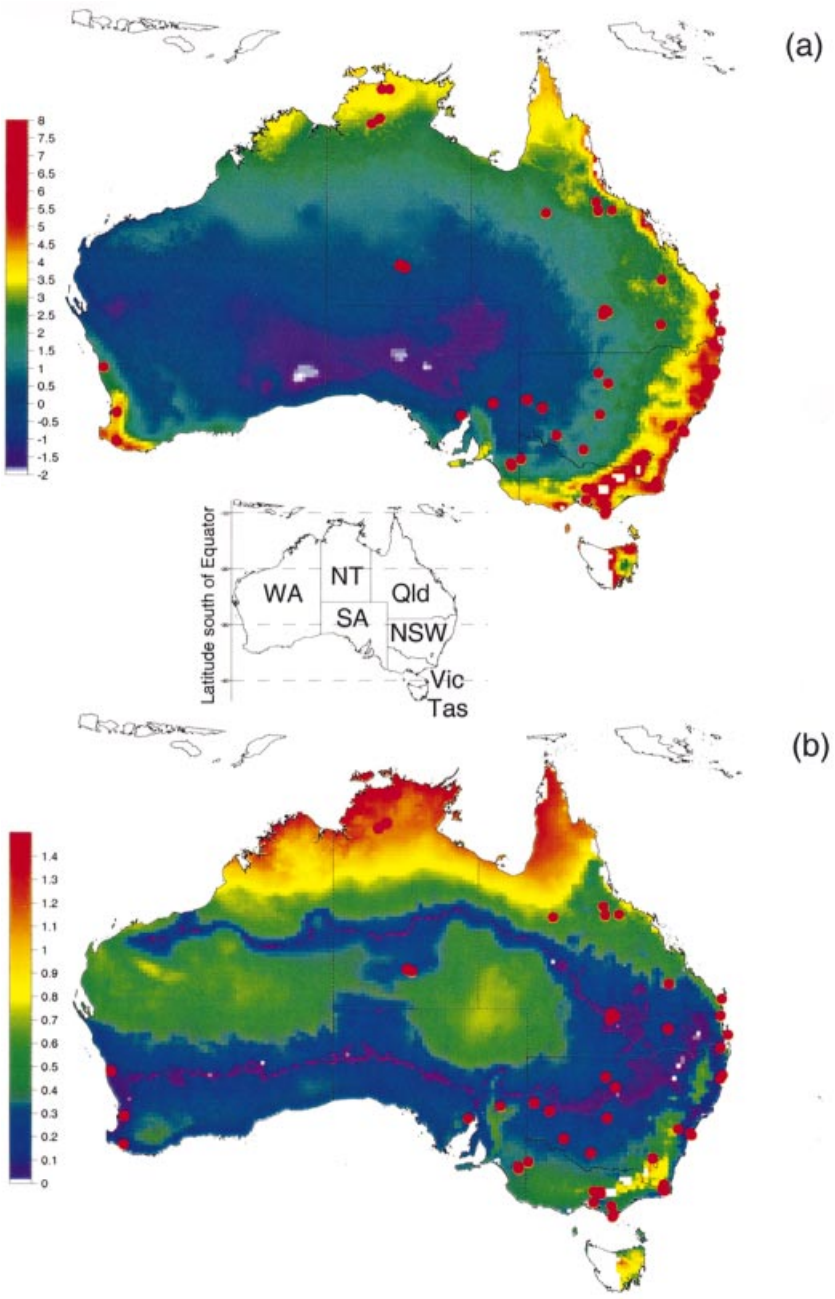

Fig. 3 Spatial distribution of (a) the estimated above-ground biomass C density ($\ln(t_C\ ha^{-1})$) and (b) the log-transformed standard error of mean above-ground biomass C density ($\ln(t_C\ ha^{-1})$) from the MLR model developed from the data in Fig. 1. Red circles depict the location of point observations of above-ground biomass from Table 1. The location map depicts the Australian states: WA, Western Australia; NT, Northern Territory; SA, South Australia; Qld, Queensland; NSW, New South Wales; Vic, Victoria; Tas, Tasmania.

temperate regions of Australia (Barrett 1999), so for a given net primary production a lower biomass results. For a given rainfall and temperature, locations with low nutrient status (GNS class = 1) had lower above-ground biomass than those of medium or high nutrient status (GNS classes 2 and 3).

Observations of above-ground biomass were spatially autocorrelated with values of I_d = *c.* 1.0 at distance scales of up to 20 grid cells (i.e. approximately 100 km; Fig. 2 closed symbols) and decreasing approximately as the inverse of distance up to 120 grid cells (*c.* 600 km). While this result is largely self-evident (in that above-ground biomass is likely to be more similar for nearby points than for locations further away), it is necessary to quantify the extent of the spatial correlation in order to accurately calculate the variance of total C emissions (Appendix 3). The significant negative correlation (I_d = *c.* –0.4) at distances of 360 cell units (i.e. *c.* 1800 km) depicts a characteristic scale separating the higher biomass of the fringes of the continent from the arid shrublands of the interior (Fig. 3a).

The absence of spatial autocorrelation in the residuals (Fig. 2 open symbols) indicates that the systematic variation in observations of above-ground biomass is

well accounted for by variation in the auxiliary variables of the MLR model. This result increases confidence that both predicted above-ground biomass and predicted variance of above-ground biomass of each grid cell are unbiased.

The unexplained variance, of approximately 19%, remaining after fitting the most parsimonious model was likely due to phenomena at distance scales shorter than a grid-cell that cause deviations of observations from the model function (Fig. 1). These phenomena include: (i) the effect of localized variation in microclimate or topography on soil moisture and soil nutrient availability which controls biomass at scales not discernible at the resolution of the continental data sets; (ii) differences between actual rainfall, temperature and soil nutrient content at the study site and the values of these variables for each respective grid cell derived from continental data sets; (iii) the likely differences between species in their ecophysiological responses to the availability of soil water, nutrients and temperature; and (iv) the presence of past disturbance causing deviation of measured biomass from the true maximum steady state value.

Spatial distribution of $\hat{q}_{ij}$ and standard error of mean above-ground biomass

The continental surface of predicted above-ground biomass is shown in Fig. 3(a) and the one standard error surface of mean biomass in Fig. 3(b). In central-southern Queensland, central-west New South Wales and western Victoria a combination of low biomass, relatively small climatic and edaphic variation and high sample density results in relatively small standard errors of mean above-ground biomass. However, where variation in climate and soil is high, such as throughout the high relief mountain regions of south-eastern Australia (south-east NSW and eastern Victoria), the range of the standard error of mean biomass is large. For these regions, sub-grid cell scale variation in biomass contributes to the large standard errors *via* the residual mean square (see equation A3) resulting in large uncertainties. Throughout the tropical regions of Australia, for example north of about 18° S, the standard error of mean biomass is also relatively large, but in this case the uncertainty is due primarily to low sample density (n) rather than necessarily from large sub-grid cell natural variability.

Total above-ground biomass C stock for the Australian continent based on predicted modal biomass C density from the MLR (Fig. 3a) is 10.7 Gt_C which is about 77% of the total above- and below-ground biomass stock of 13.9 Gt_C estimated by Barrett (1999). Given a continental land area of 769 Mha, the biomass C stock of Australia is relatively small compared to the 51 Gt_C of above- and below-ground biomass C stock of Africa (Gaston *et al*. 1998), the 176 Gt_C of above-ground biomass of tropical south and south-east Asia (Iverson *et al*. 1994) or the 58–81 Gt_C of above- and below-ground biomass of Brazil (Schroeder & Winjum 1995). The extensive contiguous tracts of lowland tropical forests in Africa, Asia and South America result in large vegetation C stocks for these countries. As a comparison, the combined above- and below-ground biomass stock of the forest sector alone in China, Russia, Canada and the conterminous United States of America is 4.3, 27.9, 14.0 and 12.1 Gt_C, respectively (Kurz & Apps 1994; Alexeyev *et al*. 1995; Turner *et al*. 1995; Fang *et al*. 1998). In Australia, large areas of 'low woodland' and 'open woodland' formations have biomass C densities in the range of 5–30 t_C ha^{-1}. These biomass C densities are relatively low compared with 170–250 t_C ha^{-1} for tropical lowland forests of Africa (Gaston *et al*. 1998; Houghton & Hackler 1999), 228–342 t_C ha^{-1} of tropical Amazonian forests of Brazil (Brown *et al*. 1995) and up to 600 t_C ha^{-1} in south-east Asia (Iverson *et al*. 1994).

The standard error mean biomass surface (Fig. 3b) provides important information on the spatial distribution of the uncertainty in biomass derived from sample density and random variation in biomass observations (the residual mean square). The random variation in biomass observations (residuals) was spatially uncorrelated (Fig. 2), although it is possible that at finer scales than the present work, some random variation could actually be spatially correlated. For example, recent vegetation mapping studies in tropical rain forest have demonstrated autocorrelation of some vegetation properties (e.g. tree density and tree diameter which are both related to biomass) at distances from 0 to 35 km (Atkinson 2000). To accurately quantify autocorrelation at patch, regional and continental scales requires a hierarchical sampling strategy involving nested plots at a range of distance scales. As it is not financially feasible to hierarchically sample a region as large as a continent, information on the spatial distribution of uncertainty along with other information sources (e.g. GIS coverage of remnant vegetation, road access and areas of strategic interest) are needed to identify priority locations where further sampling would maximize the reduction in uncertainty particularly where financial constraints limit sample size.

Sampling and uncertainty of above-ground biomass density

The risk of using small sample sizes to represent large regions is biased sampling of the biomass distribution of that region. Figure 4 quantifies the rate at which uncertainty in estimated biomass decreases with increase in

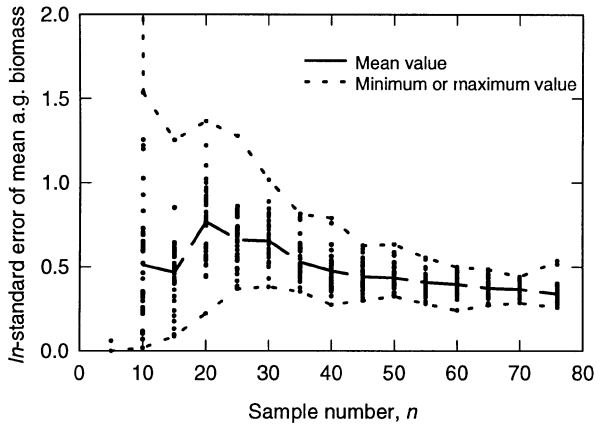

Fig. 4 Uncertainty of above-ground biomass as a function of sample number expressed as the ln-transformed standard error of mean biomass for a nominal site. The plot was generated by randomly re-sampling 40 times the data set of observations of above-ground biomass for each sample number (n = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 and 75). For each of the 600 realizations the coefficients of the MLR model were re-calculated. The nominal site had a mean annual rainfall of 1000 mm, a mean annual temperature of 12 °C and a predicted biomass C density of 123 t_C ha^{-1}. Dots depict individual realizations, dotted lines the maximum and minimum values of all realizations, and the dashed line the mean values of all realizations.

sample number at a nominal site. It shows that, not surprisingly, low sample numbers result in large standard errors of the mean (high uncertainty) for estimated biomass. The confidence interval of predicted biomass at this site based on less than 10 observations of above-ground biomass taken randomly from the data set are so wide that they encompass a range greater than the observed range of biomass across the continent, i.e. 0–400 t_C ha^{-1}. Thus, it is not possible to predict the biomass at any location within a large region (of the order of 10^5 km^2) to any certainty with less than 10 samples of biomass taken from within that region.

The maximum and minimum extremes of the standard errors in Fig. 4 define the range within which the estimated uncertainty could reside depending on which particular selection of samples happened to be chosen at any time. It is entirely possible that for small sample sizes further sampling may apparently *increase* uncertainty in cases where the original sample underestimated the standard error of the mean. The narrowing of the range between maximum and minimum curves with increase in sample number indicates increasing representativeness (i.e. decreasing bias) of the sample and a more confident estimate of the standard error.

Uncertainty in the estimate of biomass decreases rapidly with an increase in sample number. For example, a sample number of n = 40 yields a standard error of mean log-biomass of 0.48 which, when back transformed, corresponds to a confidence interval of 77–199 t_C ha^{-1} about the predicted value of 123 t_C ha^{-1}. By increasing the sample number to n = 75, the confidence interval narrowed to between 88 and 173 t_C ha^{-1}. The narrower confidence interval, based on more samples, is between 29% and 40% either side of the predicted value. This range is of a similar magnitude to the confidence intervals of predictions based on more intensively sampled forest inventory measurements (Table 2). The 95% confidence intervals of these studies ranged from approximately 9% of mean biomass for a *Eucalyptus sieberi* forest site to 44% for a *Eucalyptus obliqua/E. dives* forest in Victoria. Similar 95% confidence intervals of biomass have been reported for tropical Amazonian rain forest (± 50% of the mean value of 300 t_C ha^{-1}; Brown *et al*. 1995) and African biomes (coefficients of variation from 15 to 123% for biomass from 22 to 412 t_C ha^{-1}; Brown & Gaston 1995).

Thus, even with intensive sampling of forests (of scales 10^0 to 10^2 km^2), considerable random variance remains (e.g. Atkinson 2000) and the confidence intervals of these studies cannot be reduced to less than 20% – 30% of the mean value. If these results are generally applicable to woody vegetation world-wide, it sets a limit to the maximum accuracy possible for predicted biomass prior to LUC and, hence, the estimated C emissions based on biomass estimates. In other words, an uncertainty of the order of 20% to 30% of the estimated biomass C density of forests may be irreducible given natural heterogeneity and the logistical constraints of sampling.

C emissions from historic LUC to cropping in Australia and their uncertainty

Figure 5 shows the spatial distribution of C loss to the atmosphere estimated by combining the surface in Fig. 3 with a data set of fractional cropland area per grid-cell (Ramankutty & Foley 1998). The extensive cropping regions of south-eastern and south-western Australia are visible as well as the relatively smaller, more recently established cropping areas of central Queensland and the lower rainfall areas of central western New South Wales; although there are some apparent outliers, such as in parts of mid-north Queensland, the Northern Territory, and south-east Western Australia, indicating erroneous classification.

The total cropland area of continental Australia for 1992 including land under pasture phase of a crop-pasture rotation was estimated from the product of area per grid cell and fraction of cropland area per grid cell to

Table 2 Uncertainty in local scale biomass estimates from the published scientific literature. The 95% confidence intervals are expressed as a percentage of the mean biomass for each study

Vegetation[1]	Method[2]	Area (ha)[3]	95% CI[4]	Source[5]
E. regnans forest	Allometric	2.4	11	1
E. oliqua/E. dives forest	Allometric	0.16*	44	2
E. seiberi forest	Allometric	4.3	9	3
E.populnea woodland	Allometric	17.0	30	4
Coastal heath	Harvest	0.0012	18	5
E. pilularis forest	Harvest/allometric	?	26	6

[1]Author's description of dominant overstorey vegetation.
[2]Two methods are used to estimate biomass: (i) Allometric methods require the destructive harvesting of a representative sample of individual plants to calibrate regression equations. These equations are then used to predict biomass of a plot area based on non-destructive measurements usually of trunk girth. Allometric methods are generally used in forests. (ii) Harvesting methods are generally reserved for shrub, heath and grass ecosystems where all above-ground material is removed from sample quadrats and weighed.
[3]Total area sampled reported by authors or calculated as plot number ×area per plot. *= difficulty in interpreting plot area information. ? = no data.
[4]The coefficients of variation and standard errors cited by authors represent a measure of the confidence in the value of biomass based on predictions using allometric relationships. The 95% confidence interval of the predicted mean was approximated as 2 times the standard error reported by the authors and expressed as a percentage of the mean value. Due to possible errors in interpretation, these values should be regarded as approximate.
[5]Literature sources are:
1 Tajchman S *et al*. (1996) *Biomass and Bioenergy* **11,** 383–386. 2 Feller M (1980) *Australian Journal of Ecology* **5,** 309–333. 3 Turner J *et al*. (1992) *Forest Ecology and Management* **55,** 135–148. 4 Harrington G (1979) *Australian Journal of Botany* **27,** 135–143. 5 Jones R (1968) *Australian Journal of Botany* **16,** 589–602. 6 Applegate GB (1989) *Australian Forestry* **52,** 195–196.

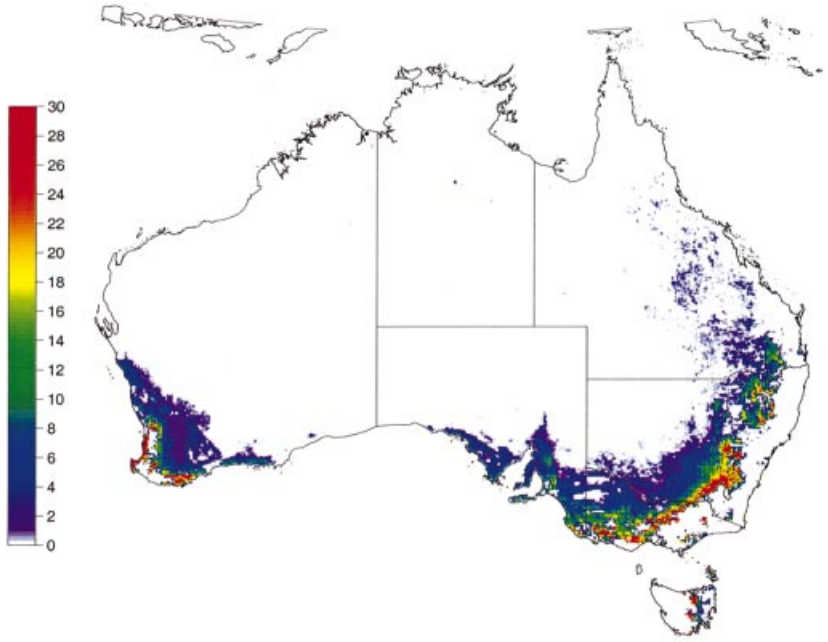

Fig. 5 The spatial distribution of above-ground C biomass lost to the atmosphere by historic land-use change to cultivation and cropping in Australia (t_C ha^{-1}). The map was generated by combining the fractional cropland area data set of Ramankutty & Foley (1998) with the MLR model of above-ground biomass (Fig. 3a).

be 34.5 Mha. The Australian Bureau of Statistics estimates 17.3 Mha of 'crops' in 1993, primarily of cereals, and a further 29.0 Mha of 'sown pasture and grasses' that includes both sown perennial, sown annual pastures and pasture phase of a crop rotation cycle. Cropping predominates on agricultural land in regions with <600 mm rainfall in south-eastern and south-western Australia and with <700 mm rainfall in northern NSW and southern Queensland, whereas meat and wool production is the primary land use in areas with higher rainfall. The

Ramankutty & Foley (1998) data set excludes 'permanent pastures' but does include temporary pastures under pasture-crop rotation. Therefore, the land area under sown perennial and sown annual pasture can be estimated by difference (29.0 + 17.3–34.5) as 11.8 Mha. This estimate agrees reasonably well with the area of total annual (2.97 Mha) and perennial (8.0 Mha) sown pastures in 1993 of 10.9 Mha estimated from NOAA-AVHRR and Landsat TM data by Hill *et al*. (1999). This increases confidence that the aggregate cropland area as estimated by Ramankutty & Foley (1998) (and interpreted here as excluding the sum of sown perennial and annual pastures of Australia) is reasonably accurate at the continental scale. However, for any given grid cell, error in the estimate of fractional cropland area is likely and this error is unknown.

The effect of error in the estimation of fractional cropland area per grid cell, $\hat{f}_{A_{Cij}}$, on calculated C emissions from clearing of above-ground biomass for cultivation and cropping is shown in Fig. 6. If the cropland areas are mapped accurately such that the coefficient of variation (CV) of $\hat{f}_{A_{Cij}}$ is zero, the 95% confidence intervals comprise error from the biomass estimate only (Fig. 6a). This error increases with increase in latitude from the equator because the log-normal standard error of the residuals increases with the predicted value and so larger C emissions estimates have larger absolute errors. The total mean C emissions for all land grid cells of continental Australia is 757 Mt_C with a 95% confidence interval of 379–1135 Mt_C (i.e. a standard error for total C emissions of 193 Mt_C yielding a CV for C emissions of 25%). These emissions represent about 0.7% of the estimated 105 Gt_C total contribution to global C emissions by historical LUC of forested land to agriculture world-wide (Houghton 1999).

If cropland area was mapped inaccurately, estimating C emissions from LUC becomes much more difficult. In Fig. 6b, the CV of $\hat{f}_{A_{Cij}}$ was set to 0.50. Both the confidence intervals and the predicted continental total mean C emissions are increased by the increase in error of mapping the cropland area. In this case, the 95% confidence interval for continentally aggregated C emissions ranged from 350 to 1294 Mt_C around a total of 822 Mt_C (standard error = 241 Mt_C), corresponding to a CV for C emissions of 29%. The increase in the estimate of the total occurs because grid cell means are sensitive to an increase in the standard error of the residuals when the residual distribution is positively skewed. The continental total based on the modal value, however, is stable at 496 Mt_C with respect to the size of the error term. Thus, while accurate maps of land-use change decrease uncertainty in C emissions, the lower limit of the standard error of total C emissions is governed by error in biomass which, as stated above, may be irreducible beyond a certain point given financial and logistical constraints of biomass sampling. The presence of a skewed distribution of above-ground biomass increases emphasis on the need for accurate estimation of the residual distribution function because of the sensitivity of the mean to the presence of outliers in skewed distributions.

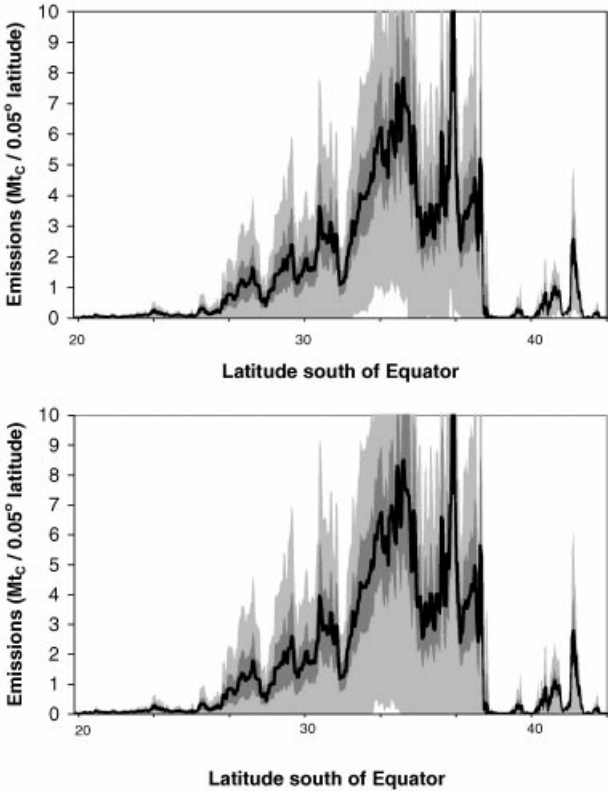

Fig. 6 The distribution of latitudinal total mean C emissions to the atmosphere from historical LUC to cultivation and cropping in Australia (black line) in Mt_C per 0.05° latitude and the cumulative probability distribution function of C emissions (shaded regions either side of black line) calculated from the standard error of the mean (equation A8). The light shaded region encompasses the 5th to 95th percentiles of the prediction interval of latitudinal total C emissions. Dark shaded region defines the 25th to 75th percentiles of the prediction interval. In (a)the coefficient of variation (CV) of the fractional cropland area data set is 0.0; in (b)the CV of the fractional cropland area is 0.50.

Carbon emissions based on steady state estimates of above-ground biomass are usually thought to overrepresent C loss to the atmosphere from LUC. This is because actual biomass of the landscape prior to LUC may be less than the bioclimatic potential (Iverson *et al*. 1994) due to frequent disturbance such as high fire frequency in Australia's northern tropical savannas (e.g. Smith *et al*. 1998). However, it is unlikely given the relative infrequency of stand replacing fires at higher

latitudes in Australia and the independent nature of the observations, that the predictions from the statistical model are significantly lower than pre-LUC values. While evidence does exist for recent vegetation thickening due to relaxation of disturbance by fire in south-eastern coastal Australia (e.g. Lunt 1997) as well as in the savannas of central and north Queensland (Burrows *et al*. 1990) and rangelands of central west New South Wales, there is little information on disturbance frequencies in areas now under intense agriculture. In the Darling Downs of south-east Queensland, Fensham & Fairfax (1999) found little change in tree/understorey density and composition of remnant vegetation after the onset of agricultural activity in the 1840s and 1850s, but we cannot assume this is the case throughout the cropping region of Australia. Thus, the estimated C emissions from LUC to cropping reported here should be regarded as an upper limit based on potential above-ground biomass until new information becomes available.

Implications for estimates of C emissions from LUC in Australia and elsewhere

Biomass is burned quickly following vegetation clearing and so it is impossible to retrospectively measure and validate estimated C emissions from this source at the land surface. Therefore, *post hoc* sampling and measurement of remnant vegetation must be performed and models developed to infer biomass prior to LUC. A major difficulty in intensively cropped regions of the world is that few areas of representative remnant vegetation remain with which to calibrate these models. Another difficulty is establishing whether the original vegetation was at a maximal steady state value or whether prior conditions held the landscape at some reduced biomass.

When prescribing the accuracy desirable for estimates of C emissions from LUC in greenhouse gas inventories, one course of action is to accept that the uncertainty of this estimate will be large and to develop policies accordingly. Another possibility is to attempt to reduce the uncertainties by further sampling of biomass of remnant vegetation in areas currently undergoing clearing. This course of action, of reducing uncertainties by further sampling, needs to be weighed against the potential cost incurred by undertaking extensive biomass sampling over large regions. Due to the high cost of making biomass measurements to high accuracy (Catchpole & Wheeler 1992), it is likely that only relatively small sample sizes are possible at continental scales, particularly for forest biomes where biomass measurements are most difficult. Hence, strategies for ensuring maximal confidence in the estimated biomass per unit sampling effort are necessary.

The standard error of mean above-ground biomass for any grid cell (Fig. 3b) provides information on the uncertainty of the predicted value from the statistical model. However, it does not provide any information on the biomass distribution within the 25–30 km^2 area of a grid cell. For example, if all observations in Table 1 happened by chance to be exactly equivalent to the true potential biomass for each grid cell, the residual mean square, $\sigma^2_{\ln(q)}$, in equation A2 would be zero and the standard error of the mean would also be zero; in other words, we would be entirely confident that the predicted value of each grid cell was accurate but we would still know nothing about the distribution function (and the degree of skew) of biomass within a grid cell. Some information on grid-cell scale variability (i.e. natural heterogeneity) does exist via the residual mean square due to deviations between observations and the model function (Fig. 1), but the small sample sizes ensure that large uncertainties remain in the estimated parameters of the biomass distribution contributing to large uncertainties to the estimate of C emissions. Given enough samples, the predicted biomass of any landscape can be known to a high level of confidence; however, the logistical constraints of sampling inevitably yield small sample sizes that, when combined with large natural variability in biomass C density, results in relatively large and possibly irreducible uncertainties in continental scale forest inventories.

In the present work, we have shown that by using a statistical approach a quantitative determination of uncertainty in estimates of carbon emissions from clearing of undisturbed biomass is possible for large data sparse regions of the world. Objective critera for the selection of samples of above-ground biomass from the literature, of using biomass data to generate spatially explicit distributions of the predicted value and standard errors of above-ground biomass, and techniques for reducing uncertainties in above-ground biomass through further sampling by examining the spatial distribution of the error term of a statistical model were presented. The methodologies presented here ensure that the confidence interval of estimated C emissions is quantified by a transparent methodology. The minimum uncertainty attainable for estimates of biomass of regions that have undergone intensive LUC is limited by the cost and logistics of sampling biomass carbon rather than by the accuracy with which LUC can be mapped given current satellite technology. This is not to say that continued effort to accurately map LUC by satellite is unimportant. Rather, it is necessary to recognize that some uncertainties are perhaps irreducible given logistical constraints. In that case, identification and communication of the relevant reducible and irreducible uncertainties, as well as the quantitative level of these uncertainties, are critical

for the development of appropriately informed understanding of carbon release from land-use change.

Acknowledgements

We gratefully acknowledge Jenny Langridge, Andrew Ash and Trevor Hobbs for discussions on site suitability in the data set used in this study, and Bill Burrows for making site data available in central Queensland. Mark Howden, Roger Gifford, Chris Mitchell and Andrew Moore provided valuable comments on earlier drafts. This work was supported by funding from the Australian Government (Australian Greenhouse Office) to both the CSIRO Climate and Atmosphere Sector as part of Biosphere Working Group research activities and to the Australian Cooperative Research Centre for Greenhouse Accounting.

Appendix 1 Variance of predicted above-ground biomass density per grid cell

The regression coefficients, β_0, β_1,…β_n, of the MLR model were estimated using least-squares by solving for the column vector, $\boldsymbol{\beta}$, in (Sokal & Rohlf 1995)

$$\boldsymbol{Y} = \boldsymbol{\beta S} \tag{A1}$$

where S is Sums of Squares-Sums of Products matrix of the auxiliary variables (i.e. $\Sigma(X_i - \bar{X}_i)^2$ and $\Sigma(X_i - \bar{X}_i)(X_j - \bar{X}_j)$) and $\boldsymbol{Y}$ is the column vector of sums of products of the auxiliary variables and the independent observations of above-ground biomass (i.e. $\Sigma(X_i - \bar{X}_i)(Y_i - \bar{Y})$). Thus,

$$\boldsymbol{\beta} = \boldsymbol{GY} \tag{A2}$$

where $\boldsymbol{G}$ is the 'Gaussian' matrix; $\boldsymbol{G} = \boldsymbol{S}^{-1}$.

As discussed in the text, observations of above-ground biomass C density were log-normally distributed and so the standard error of the mean (equation A3a) and standard error of the residuals (equation A3b) of log-transformed above-ground biomass, $\sigma_{\ln(\hat{q}_{ij})}$, for any grid cell were calculated from the residual mean square of the MLR model, $\sigma^2_{\ln(q)}$, sample number, n, and the auxiliary variables,

$$\sigma^2_{\ln(\hat{q}_{ij})} = \sigma^2_{\ln(q)}\left(\frac{1}{n} + c\right) \tag{A3a}$$

$$\sigma^2_{\ln(\hat{q}_{ij})} = \sigma^2_{\ln(q)}\left(1 + \frac{1}{n} + c\right) \tag{A3b}$$

where $c = \boldsymbol{X}^{\mathrm{T}}\ \boldsymbol{G}\ \boldsymbol{X}$, and where $\boldsymbol{X}$ is the column vector of deviations of auxiliary variables (i.e. $(X_i - \bar{X}_i)$). In equation A3a, the RHS accounts for the uncertainty of the regression model. The extra residual mean square term in equation A3b accounts for the unexplained variance of the residuals, which in the present circumstance is largely due to natural heterogeneity of above-ground biomass within a grid cell.

Appendix 2 Variance of estimated C emissions form LUC per grid cell

The variance of estimated C emissions of the $(i, j)^{\mathrm{th}}$ grid cell was calculated from the logarithm of the standard error of the residuals of above-ground biomass from equation A3b, $\sigma_{\ln(\hat{q}_{ij})}$, and the standard error of cropland area per grid cell, $\sigma_{\ln(\hat{A}_{\mathrm{L}ij})}$. The $\sigma_{\ln(\hat{A}_{\mathrm{L}ij})}$ is unknown and so to proceed it was assumed that the error in fractional cropland area data of Ramankutty & Foley 1998) data set was log-normally distributed. This assumption simplifies calculations but implies that the error in the estimate of cropland area increases proportionally with the extent of cropland per grid cell. While there is no evidence for this assumption, the alternative, that errors are independent of the estimate, would yield a more conservative error in C emissions. Thus, the present estimate of uncertainty of C emissions per grid cell should be regarded as 'pessimistic' in this respect. The effect of uncertainty in $\sigma_{\ln(\hat{A}_{\mathrm{L}ij})}$ on $\sigma_{\ln(\hat{F}_{ij})}$ was investigated by setting the coefficient of variation (CV) of $\sigma_{\ln(\hat{A}_{\mathrm{L}ij})}$ to two values, 0 and 0.50, which correspond to a standard error of either 0 or 50% in the estimated cropland area per grid cell.

The estimated variance of log-transformed C emissions from clearing of above-ground biomass for the $(i, j)^{\mathrm{th}}$ grid cell, $\sigma^2_{\ln(\hat{F}_{ij})}$, was calculated from the first-order Taylor expansion of equation 4 evaluated at the predictions of $\ln(\hat{q}_{ij})$ and $\ln(\hat{A}_{\mathrm{L}ij})$,

$$\sigma^2_{\ln(\hat{F}_{ij})} = \sigma^2_{\ln(\hat{q}_{ij})} + \sigma^2_{\ln(\hat{A}_{\mathrm{L}ij})} + 2\sigma_{\ln(\hat{q}_{ij})}\,\sigma_{\ln(\hat{A}_{\mathrm{L}ij})}\,\rho_{\hat{q}_{ij}\ \hat{A}_{ij}}. \tag{A4}$$

The use of first-order methods to estimate the variance is justified because the second partial derivatives of equation 4 are zero and so the linearity assumption of the method is met. The correlation between cropland area and above-ground biomass density, $\rho_{\hat{q}_{ij}\ \hat{A}_{ij}}$, was set to zero because we expect independence between the estimate of cropland area and biomass density at the grid-cell scale.

For the log-normal model, the error in cropland area arises from the fractional cropland data (assuming zero error for grid-cell area, A_{C}) and was calculated from the coefficient of variation (CV) of $f_{A_{\mathrm{C}}}$;

$$\sigma_{\ln(A_{\mathrm{L}})} = \sigma_{\ln(f_{A\mathrm{C}})} = \ln(\mathrm{CV} + 1). \tag{A5}$$

An unbiased estimate of the standard error of C emissions from the $(i, j)^{\mathrm{th}}$ grid cell was calculated from the log-normal predicted value and standard error terms (Harris & Stocker 1998);

$$\sigma^2_{\hat{F}_{ij}} = \exp[2\ln(\hat{F}_{ij}) + \sigma^2_{\ln(\hat{F}_{ij})}](\exp[\sigma^2_{\ln(\hat{F}_{ij})}] - 1). \tag{A6}$$

The 95% confidence intervals of total (back-transformed) C emissions per grid cell are

$$\mathrm{UCI}(i, j) = \hat{F}_{ij}\, e^{(1.96\ \sigma_{\ln(\hat{F}ij)})} \tag{A7a}$$

and

$$\mathrm{LCI}(i,j) = \frac{\hat{F}_{ij}}{e^{(1.96\ \sigma_{\ln(\hat{F}ij)})}}. \qquad \text{(A7b)}$$

Appendix 3 Variance of latitudinal and continental total C emissions

Calculation of the standard error of total C emissions from LUC is problematic because accurate estimation of the correlation terms of predicted above-ground biomass between grid cells is difficult with such a small sample size. Here the standard error of total C emissions from LUC for each 0.05°-latitude band and for the continent were estimated by summing the variances of estimated emissions for each grid cell and the covariances of spatially correlated grid cells.

The variance of latitudinal total C emissions was calculated from the standard error in equation A6 as

$$\sigma^2_{\hat{F}_i} = \sum_{j=1}^{M}\sum_{k=1}^{M}(\sigma_{\hat{F}_{ij}}\sigma_{\hat{F}_{ik}}\rho_{q_{ij}\,q_{ik}}) \qquad \text{(A8)}$$

and the variance of continental total C emissions as

$$\sigma^2_{\hat{F}} = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{l=1}^{M}(\sigma_{\hat{F}_{ij}}\sigma_{\hat{F}_{kl}}q_{q_{ij}\,q_{kl}}) \qquad \text{(A9)}$$

where *i, k and j, l* refer to row and column numbers of the continental raster of standard errors of C emissions from equation A6. In the special case of independence of biomass between grid cells, $\rho_{q_{ij}\,q_{kl}} = 1.0$ when $i = k$ and $j = l$ or $\rho_{q_{ij}\,q_{kl}} = 0.0$ otherwise. Equation A9 then reduces to

$$\sigma^2_{\hat{F}} = \sum_{i=1}^{N}\sum_{j=1}^{M}\sigma^2_{\hat{F}_{ij}}$$

(i.e. the sum of the variances). However, the correlogram (see Fig. 2) indicated that above-ground biomass measurements were highly correlated when close together with correlation decreasing as distance increased to *c.* 600 km. Therefore, the correlation coefficient in equation A9 and A10 were calculated as $\rho_{q_{ij}\,q_{kl}} = 1.0$ at zero distance, then weighted by an exponential decrease in the correlation coefficient with distance up to 120 grid cells (*c.* 600 km) and, finally, $\rho_{q_{ij}\,q_{kl}} = 0$ otherwise. The 5th, 25th, 75th and 95th percentiles were calculated from $\sigma_{\hat{F}_i}$ and $\sigma_{\hat{F}}$ and reported in Fig. 6.

References

Alexeyev V, Birdsey R, Stakanov V, Korotkov I (1995) Carbon in vegetation of Russian forests: methods to estimate storage and geographical distribution. *Water Air and Soil Pollution*, **82**, 271–282.

Ashton DH (1975) Studies of litter in *Eucalyptus regnans* forests. *Australian Journal of Botany*, **23**, 413–433.

Atkinson PM (2000) Mapping the forests of Cameroon. In: *Vegetation Mapping* (eds Alexander R, Millington AC), pp. 283–303. John Wiley and Sons, Ltd, Chichester, UK.

AUSLIG (1990) *Atlas of Australian Resources: Vegetation*. Commonwealth Government Printer, Canberra, Australia.

Barrett DJ (1999) Steady State Carbon Mean Residence Time in the Australian Terrestrial Biosphere. *EOS* (Suppl.), Transactions, American Geophysical Union, 80 (46), F51.

Brown S, Gaston G (1995) Use of forest inventories and geographic information systems to estimate biomass density of tropical forests – application to tropical Africa. *Environmental Monitoring and Assessment*, **38**, 157–168.

Brown S, Hall CAS, Knabe W, Raich J, Trexler MC, Woomer P (1993) Tropical forests – their past, present, and potential future role in the terrestrial carbon budget. *Water Air and Soil Pollution*, **70**, 71–94.

Brown IF, Martinellis LA, Thomas WW, Moreira MZ, Ferreira CAC, Victoria RA (1995) Uncertainty in the biomass of Amazonian forests: an example from Rondomia, Brazil. *Forest, Ecology and Management*, **75**, 175–189.

Burrows WH, Cater JO, Scanlan JC, Anderson ER (1990) Management of savannas for livestock production in north-east Australia: contrasts across a tree-grass continuum. *Journal of Biogeography*, **17**, 503–512.

Burrows WH, Hoffman MB, Compton JF, Back PV, Tait LJ (2001) Allometric relationships and community biomass estimates for some dominant eucalypts in central Queensland woodlands. *Australian Journal of Botany*, **48**, 707–714.

Catchpole WR, Wheeler CJ (1992) Estimating plant biomass: a review of techniques. *Australian Journal of Ecology*, **17**, 121–131.

Ciezewski CJ, Turner DP, Phillips DL (1996) Statistical analysis of error propagation in national level carbon budgets. In: *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium. May 21–23, 1996, Fort Collins, Colorado* (eds Mowrer HT, Czaplewski RL, Hamre RH), pp. 649–658. General Technical Report RM-GTR-277. USDA Forest Service.

Cliff AD, Ord JK (1981) *Spatial Processes: Models and Applications*. Pion Ltd, London, UK.

Czaplewski RL, Reich RM, Bechtold WA (1994) Spatial autocorrelation in growth of undisturbed natural pine stands across Georgia. *Forest Science*, **40**, 314–328.

DeFries RS, Townshend JRG, Hansen MC (1999) Continuous fields of vegetation characteristics at the global scale at 1 km resolution. *Journal of Geophysical Research*, **104**, 16911–16925.

Fang JY, Wang GG, Liu GH, Xu SL (1998) Forest biomass of China: an estimate based on the biomass–volume relationship. *Ecological Applications*, **8**, 1084–1091.

Fensham RJ, Fairfax RJ (1997) The use of the land survey record to reconstruct pre-European vegetation patterns in the darling

downs, Queensland, Australia. *Journal of Biogeography*, **24**, 827–836.

Gaston G, Brown S, Lorenzini M, Singh KD (1998) State and change in carbon pools in the forests of tropical Africa. *Global Change Biology*, **4**, 97–114.

Gertner G, Kohl M (1992) An assessment of some nonsampling errors in national survey using an error budget. *Forest Science*, **38**, 525–538.

Graetz RD, Wilson MA, Campbell SK (1995) *Landcover Disturbance Over the Australian Continent: A Contemporary Assessment*. Biodiversity Series Paper no. 7. Biodiversity Units, Department of Environment, Sport and Territories, Canberra.

Grierson PF, Adams MA, Attiwill PM (1992) Estimates of carbon storage in the above-ground biomass of Victoria's forests. *Australian Journal of Botany*, **40**, 631–640.

Harris JW, Stocker H (1998) *Handbook of Mathematics and Computational Science*. Springer, Inc., New York.

Hill MJ, Vickery PJ, Furnival EP, Donald GE (1999) Pasture land cover in eastern Australia from NOAA-AVHRR NDVI and classified landsat TM. *Remote Sensing of Environment*, **67**, 32–50.

Houghton RA (1999) The annual net flux of carbon to the atmosphere from changes in land use 1850–1990. *Tellus*, **51B**, 298–313.

Houghton RA, Hackler JL (1999) Emissions of carbon from forestry and land-use change in tropical Asia. *Global Change Biology*, **5**, 481–492.

Houghton RA, Hackler JL, Lawrence KT (1999) The US carbon budget: Contributions from land-use change. *Science*, **285**, 574–578.

Hutchinson MF (1988) Calculation of hydrologically sound digital elevation models. In: *Proceedings of the Third International Symposium on Spatial Data Handling, August 17–19, 1988, Sydney Australia*, pp. 117–133. International Geographical Union, Columbus, Ohio.

Iverson LR, Brown S, Prasad A, Mitasova H, Gillespies AJR, Lugo AE (1994) Use of GIS for estimating potential and actual forest biomass for continental South and Southeast Asia. In: *Effects of Land-Use Change on Atmospheric CO_2 Concentrations: South and Southeast Asia as a Case Study* (ed. Dale VH), pp. 67–116. Springer Verlag, New York.

Jacquez JA (1985) *Compartmental Analysis in Biology and Medicine*. 2nd edn. The University of Michigan Press. Ann Arbor, Michigan, USA.

Jones DA (1999) Characteristics of Australian land surface temperature variability. *Theoretical and Applied Climatology*, **63**, 11–31.

Kimes DS, Nelson RF, Skole DL, Salas WA (1998) Accuracies in mapping secondary tropical forest age from sequential satellite imagery. *Remote Sensing of Environment*, **65**, 112–120.

Kurz WA, Apps MJ (1994) The carbon budget of Canadian forests: a sensitivity analysis of changes in disturbance regimes, growth rates and decomposition rates. *Environmental Pollution*, **83**, 55–61.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.

Legendre P, Fortin MJ (1989) Spatial pattern in ecological analysis. *Vegetatio*, **80**, 107–138.

Loveland TR, Belward AS (1997) The IGBP-DIS global 1 km land cover data set, discover – first results. *International Journal of Remote Sensing*, **18**, 3291–3295.

Lunt ID (1997) Tree densities last century on the Lowland Gippsland Plain, Victoria. *Australian Geographical Studies*, **35**, 342–348.

McKenzie N, Hook J (1992) *Interpretations of the Atlas of Australian Soils*. Consulting Report to the Environmental Resources Information Network. Technical Report 94/1992. CSIRO, Division of Soils, Canberra.

Morgan MG, Henrion M (1990) *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge, UK.

NGGIC (1998) *Workbook for CO_2 from the biosphere.* National Greenhouse Gas Inventory Committee. http://www.greenhouse.gov.au/

Northcote KH, Beckmann GG, Bettenay E *et al.* (1968) *Atlas of Australian Soils.* Sheets 1–10 with explanatory data. CSIRO & Melbourne University Press, Melbourne.

Penman J, Kruger D, Galbally IE *et al.* (2000) *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories.* Intergovernmental Panel on Climate Change, 1 v., Kanagawa, Japan. http://www.ipcc-nggip.iges.or.jp/gp/report.htm [electronic publication].

Phillips DL, Brown SL, Schroeder PE, Birdsey RA (2000) Toward error analysis of large-scale forest carbon budgets. *Global Ecology and Biogeography*, **9**, 305–313.

Queensland Department of Natural Resources (1999) *Land cover change in Queensland 1995 – 1997: A Statewide Landcover and Trees Study Report*. Queensland Government, Australia.

Ramankutty N, Foley JA (1998) Characterizing patterns of global land use: an analysis of global croplands data. *Global Biogeochemical Cycles*, **12**, 667–685.

Ramankutty N, Foley JA (1999) Estimating historical changes in global land cover: croplands from 1700 to 1992. *Global Biogeochemical Cycles*, **13**, 997–1027.

Running SW, Loveland TR, Peirce LL, Nemani RR, Hunt ER (1995) A remote sensing based vegetation classification logic for global land cover analysis. *Remote Sensing of Environment*, **51**, 39–48.

Scepan J (1999) Thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering and Remote Sensing*, **65**, 1051–1060.

Schroeder P, Brown S, Mo JM, Birdsey R, Cieszewski C (1997) Biomass estimation for temperate broadleaf forests of the United States using inventory data. *Forest Science*, **42**, 424–434.

Schroeder PE, Winjum JK (1995) Assessing Brazil's carbon budget. 2. Biotic fluxes and net carbon balance. *Forest Ecology and Management*, **75**, 87–99.

Schuft MJ, Barker JR, Cairns MA (1998) Spatial distribution of carbon stocks in southeast Mexican forests. *Geocarto International*, **13**, 77–86.

Smith RCG, Craig RL, Steber MT, Marsden AJ, McMillan CE, Adams J (1998) Use of NOAA-AVHRR for fire management in Australia's tropical savannas. In: *Proceedings of the Land AVHRR Workshop: 9th Australasian Remote Sensing Photogrammetry Conference, 24 July, 1998, Sydney* (ed. McVicar TR), pp. 1–10.

Sokal RR, Rohlf FJ (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. 3rd edn. W.H. Freeman, New York.

Tucker CJ, Townsend JRG (2000) Strategies for monitoring

tropical deforestation using satellite data. *International Journal of Remote Sensing*, **21**, 1461–1471.

Turner DP, Koerper GJ, Harmon ME, Lee JJ (1995) A carbon budget for forests of the conterminous United States. *Ecological Applications*, **5**, 421–436.

UN FCCC (1997) *Kyoto Protocol to the United Nations Framework convention on climate change.* Report to the conference of the parties, Third session, Kyoto, 1–10 December 1997. **FCCC/CP/1997/CRP**.6. 24pp.

Wang YP, Galbally IE, Meyer CP, Smith CJ (1997) A comparison of two algorithms for estimating carbon dioxide emissions after forest clearing. *Environmental Modelling and Software*, **12**, 187–195.

Woodgate PW, Black PJ (1988) *Forest cover changes in Victoria 1869–1987.* Department of Conservation, Forest and Lands, Melbourne.