



A model of potential carbon dioxide efflux from surface water across England and Wales using headwater stream survey data and landscape predictors

B. G. Rawlins¹, B. Palumbo-Roe¹, D. C. Goody², F. Worrall³, and H. Smith¹

¹British Geological Survey, Keyworth, Nottingham, UK

²British Geological Survey, Wallingford, UK

³Department of Earth Sciences, Durham University, Durham, UK

Correspondence to: B. G. Rawlins (bgr@bgs.ac.uk)

Received: 9 October 2013 – Published in Biogeosciences Discuss.: 24 October 2013

Revised: 24 February 2014 – Accepted: 26 February 2014 – Published: 8 April 2014

Abstract. Measurements of CO₂ partial pressures ($p\text{CO}_2$) in small headwater streams are useful for predicting potential CO₂ efflux because they provide a single concentration representing a mixture from different hydrological pathways and sources in the catchment. We developed a model to predict $p\text{CO}_2$ in headwater streams from measurements undertaken on snapshot samples collected from more than 3000 channels across the landscape of England and Wales. We used a subset of streams with upstream catchment areas (CA) of less than 8 km² because below this scale threshold $p\text{CO}_2$ was independent of CA. A series of catchment characteristics were found to be statistically significant predictors of $p\text{CO}_2$, including three geomorphic variables (mean altitude, mean catchment slope and relief) and four groups of dominant catchment land cover classes (arable, improved grassland, suburban and all other classes). We accounted for year-round, temporal variation in our model of headwater $p\text{CO}_2$ by including weekly or monthly analyses of samples from three headwater catchments with different land use and geomorphic features. Our model accounted for 24 % of the spatial and temporal variation in $p\text{CO}_2$.

We combined predictions from the $p\text{CO}_2$ model (on a 1 km grid) and monthly runoff volumes (litres) on 0.5° resolution grid across England and Wales to compute potential C fluxes to the atmosphere. Our model predicts an annual average potential C flux of 65.4 kt C across England and Wales (based on free C concentrations), with lower and upper 95 % confidence values of 56.1 and 77.2 kt C, respectively.

1 Introduction

There is increasing interest in approaches to compute fluxes in the global carbon cycle, including the role of freshwater channels (Raymond et al., 2013; Benstead and Leigh, 2012) which are active conduits for the transfer of greenhouse gases to the atmosphere from the terrestrial biosphere (Battin et al., 2009). Evasion of carbon dioxide (CO₂) from surface waters may account for between 10 % (Sun et al., 2011) of net ecosystem exchange at the continental scale to as much as 70 % in peatland catchments (Hope et al., 2001). In a recent study, Butman and Raymond (2011) used stream water $p\text{CO}_2$ concentrations from samples collected at flow gauging stations (channels ranging from 1st to 10th order) in combination with estimates of stream surface area and gas transfer velocities to compute CO₂ efflux across the USA. For many continents, although some geochemical data sets exist to estimate CO₂ efflux (Regnier et al., 2013), it may be possible to develop alternative approaches to compute stream water $p\text{CO}_2$ concentrations (and fluxes) from more readily available, landscape data.

Measurements of CO₂ partial pressures ($p\text{CO}_2$) in small headwater streams may be particularly useful for predicting CO₂ efflux because (i) it has been suggested that the provenance of 77 % of the CO₂ evasion from channels in large river basins such as the Amazon is from soil respiration (Johnson et al., 2008), and headwater streams are closely connected to the soil; (ii) they provide a single concentration representing a mixture from different hydrological pathways and sources within the catchment including soil water, both shallow and

deeper groundwater contributions (Jones and Mulholland, 1998) and in-stream (e.g. through the mineralisation of organic carbon in stream water and bed sediments) sources of CO_2 ; and (iii) $p\text{CO}_2$ concentrations typically decline with increasing stream order/catchment size (Butman and Raymond, 2011; Li et al., 2013) and, accounting for lengthy upstream evasive losses (based on measurements from larger channels), may be prone to substantial error because the gas transfer coefficient (K_{CO_2}) exhibits a considerable degree of spatial and temporal variation (Wallin et al., 2011).

In our experience, measurements of headwater stream $p\text{CO}_2$ from large, landscape-scale surveys are rare. In this study we investigated those catchment characteristics which account for variations in $p\text{CO}_2$ using data from a headwater stream survey across England and Wales. We hypothesised that a range of upstream catchment characteristics – defined by the delineated watershed of a particular sampling location – could account for variations in headwater $p\text{CO}_2$. A similar approach was applied to channel water chemistry data from the channels of 814 substantially larger catchments (catchment area (CA) from < 1000 to $> 10\,000\text{ km}^2$) across North America and Canada. In this study, multiple analyses were available at each site from which a mean $p\text{CO}_2$ was computed (Lauerwald et al., 2013). In their study, Lauerwald et al. (2013) found that mean air temperature, mean catchment slope gradient and mean annual precipitation explained 43 % of the variation in the negative logarithm of mean $p\text{CO}_2$ in rivers from which these catchment characteristics were calculated; the proportion of agricultural land was not found to be a statistically significant predictor of mean $p\text{CO}_2$. Other studies have also shown that catchment characteristics including CA, dominant land cover class (including urban/suburban) (Butman and Raymond, 2011; Prasad et al., 2013; Li et al., 2013) and geomorphic features such as slope and elevation (Jones et al., 1998) account for variations in channel $p\text{CO}_2$. However, no previous studies have used landscape-scale data on *headwater* $p\text{CO}_2$ values from which evasive loss of CO_2 would be substantially smaller than for larger channels with larger (Strahler) stream orders (e.g. > 3). Lauerwald et al. (2013) reported there was no observable effect of including a temporal component into their multiple regression approach based on inclusion of average sampling year.

Both stream CA and (Strahler) stream order account for variations in headwater $p\text{CO}_2$ (Butman and Raymond, 2011) because in small headwater streams (lower stream orders) there is a greater net loss of CO_2 from the stream surface, due to the effect of flow turbulence on the gas transfer coefficient. By delineating stream catchment boundaries it is possible to establish landscape predictors of stream $p\text{CO}_2$ based on geomorphic features and land cover for entire landscapes from widely available digital terrain and remotely sensed data, respectively. It may be possible to develop a statistical model to predict $p\text{CO}_2$ for those parts of similar landscapes where stream chemistry data are unavailable.

The vast majority of the sampling sites in our snapshot (one visit per site) survey data set were headwater streams (CA 1–10 km²). We wished to investigate whether it is possible – by undertaking a statistical analysis for catchments across a range of scales and considering other predictors – to determine a threshold scale below which stream $p\text{CO}_2$ are independent of CA. In large-scale studies undertaken to date, headwater streams generally have larger $p\text{CO}_2$ values than channels of greater stream order (Butman and Raymond, 2011) and a large proportion of the CO_2 evades from the headwater; in the case of the Amazon Basin this was estimated to be as large as 90 % of the total CO_2 flux from channels (Johnson et al., 2008). By restricting our analysis to those data below this CA threshold we will have confidence that our model does not underestimate $p\text{CO}_2$ due to downstream net loss. Even though headwater streams have greater potential for CO_2 evasion (compared to the higher order channels) due to their larger gas transfer coefficients associated with larger velocities and turbulent flow, the *net* loss is restricted, as shallow groundwater water entering the headwaters will have similar $p\text{CO}_2$ values to the stream they enter because they connect the local soil environment to the channel. In other words, evasion of CO_2 in the headwater streams is not substantially greater than inputs from shallow groundwater or in-stream sources, such as the turnover of fluvial carbon.

Where model predictions are available for headwater $p\text{CO}_2$ across a landscape it may be possible to use a simplifying approach to compute potential CO_2 evasion fluxes. The approximate timescale for dissolved $p\text{CO}_2$ to reach equilibrium with the atmosphere is around 100 hours in a large river system (Morel and Hering, 1993), similar to the water residence time of many large waterbodies of England and Wales. In this study we compute CO_2 evasion fluxes, assuming that channel residence times are sufficient for all free CO_2 to evade and also that there are limited downstream changes in water chemistry because mixing of flow from headwater channels over a range of bedrock types will tend towards a mean chemical composition in higher order channels, and changes in pH associated with CO_2 evasion are also smaller in larger channels. This is a different approach to efforts to model downstream losses based on gas transfer velocities and stream hydraulics (Raymond et al., 2012).

Headwater $p\text{CO}_2$ exhibit seasonal variations (Jones et al., 1998), so to compute potential CO_2 annual evasion fluxes it is necessary to account for this. In this paper, the large-scale survey data set of headwater stream samples was undertaken between June and September. To encompass year-round variation we included data sets in our predictive model for a series of small ($< 10\text{ km}^2$) headwater catchments which were monitored for a full calendar year or more, one of which is described in detail by Dinsmore et al. (2010).

In this paper we develop a statistical model to predict monthly headwater stream $p\text{CO}_2$ using large-scale survey data from around 3000 locations across England and Wales,

and a series of associated catchment characteristics. We determine an optimum set of land cover class groupings to include in our model by undertaking a series of significance tests. We also investigate the incorporation of the temporal variation of headwater $p\text{CO}_2$ into our predictive model. We apply the resulting model using landscape characteristics to predict $p\text{CO}_2$ values in flow from cells on a 1 km grid across England and Wales. After converting $p\text{CO}_2$ to free C concentrations in water, we compute potential monthly C fluxes (using runoff volumes on a 0.5° resolution grid) across England and Wales. We also calculate the 95 % confidence intervals for potential C fluxes based on the model predictions.

2 Methods

2.1 Headwater stream $p\text{CO}_2$

2.1.1 Large-scale headwater survey

The methods used in the large-scale headwater survey are described in detail in Johnson et al. (2005). Headwater stream samples were collected during the summer months (June–September) of the years between 1998 and 2002 (inclusive). Sampling was undertaken between 9.00 and 17.00 LT each day but the precise time was not recorded. In our study we wished to avoid potential bias introduced by including stream water sampling and analyses undertaken during large channel flows (greater than mean flow), so we selected only those samples from sites where there had been no substantial rainfall for more than 7 days. We used information from local flow gauging stations to determine whether local flow conditions were below mean flow (Sect. 2.3). Details of each sampling location were recorded on a field card and collated in a database with a unique sample identifier. The sampling locations are shown in Fig. 1.

A total of four separate water samples were collected from each site so that a range of measurements could be made. These include (analyses in parentheses):

1. a 250 mL Nalgene polyethylene bottle (alkalinity titration);
2. a 30 mL polyethylene bottle (conductivity and pH);
3. a 60 mL Nalgene bottle which is filtered ($0.45\ \mu\text{m}$) and subsequently acidified (see below; prior to analysis by ICP-AES);
4. a 60 mL Nalgene bottle which is filtered ($0.45\ \mu\text{m}$) not acidified (ion chromatography/TOC analyser).

All samples were collected from the middle of the stream. In the case of samples 1 and 2 (above) the sample containers were thoroughly rinsed three times. The containers were then submerged in the streams and sealed underwater to ensure that all the air had been expelled. On return to the field base

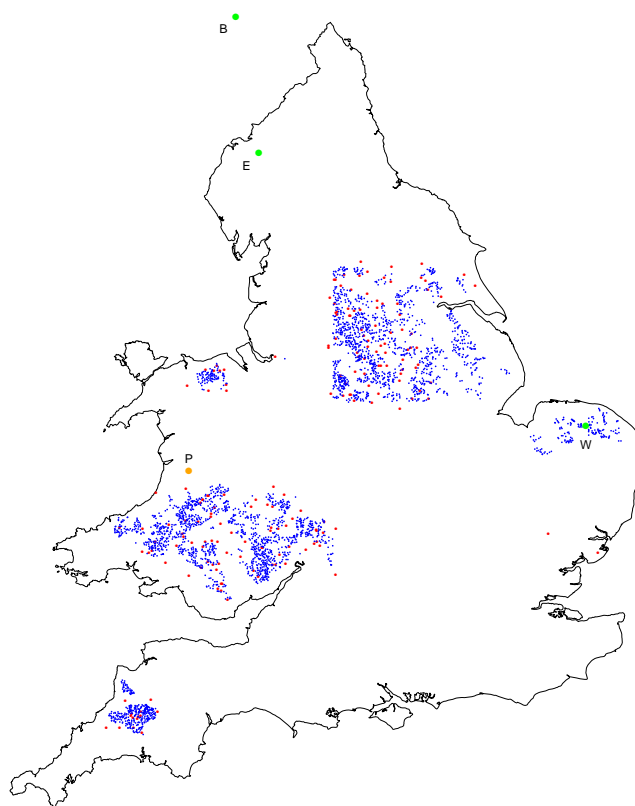


Fig. 1. Locations of sites referred to in this study across England and Wales (coastline shown). Blue symbols show the headwater stream survey sampling locations. Red discs show the nearest national river flow archive gauging stations to the survey sites. Green symbols are headwater sites from which temporal (weekly or monthly) data were available for $p\text{CO}_2$ and stream temperature: E = Eden (Pow), W = Wensum, B = Black burn in Scotland. Orange disc shows the location of the Upper Hafren at Plynlimon (P) from which stream water temperature data was used.

all samples were refrigerated at around 4°C prior to further analysis.

Conductivity and pH were measured on the evening of sample collection and the containers were removed from refrigeration around 1 hour prior to undertaking the analyses; this allowed the water sample temperature to reach between approximately 10 and 15°C at the time of measurement. Conductivity was measured using a Hannah HI9033 portable conductivity meter calibrated with a conductivity standard and thermometer. pH was determined using a radiometer PHM80 meter with combination electrode using buffer solutions (pH 4, 7 and 9). Alkalinity measurements were made the day after sample collection based on a simple laboratory titration method using a bromocresol green indicator. Samples 3 and 4 were used to determine the concentrations of a range of cations (including Ca^{2+} , Mg^{2+} , Na^+ , K^+ ; ICP-AES Fisons Instruments ARL 3580) and anions (NO_3^- , SO_4^{2-} , Cl^-). For a subset of samples dissolved organic car-

Table 1. Features of three headwater catchments where weekly or monthly stream water measurements or speciation-based estimates of $p\text{CO}_2$ were included in the predictive model (cf: Fig. 2).

	Regional catchment (headwater catchment)		
	Esk (Black burn)	Eden (Pow)	Wensum (Blackwater)
Catchment area (km^2)	3.4	10	8
Dominant land cover	Bog	Impr. grass	Arable
Mean elevation (m)	280	99	48
Mean slope ($^\circ$)	1.4	1.4	0.73
Relief (m)	60	96	26
Monitoring frequency	weekly	monthly	weekly
$p\text{CO}_2$ measurements ^a	D	S	S
Reference	Dinsmore et al. (2010) ^b	Owen et al. (2012)	Wensum Alliance (2013)

^a D = direct, S = estimated by PHREEQC speciation using measurements of alkalinity, pH plus major anions and cation concentrations for the Wensum; ^b these data can be accessed via the DOI in Dinsmore et al. (2013a).

bon (DOC) was determined using Shimadzu TOC 5000 analyser, purging all inorganic carbon with hydrochloric acid. Quality control was undertaken on these data using blank waters and duplicates collected in the field.

For each sample we calculated the theoretical partial pressure of CO_2 that would be in equilibrium with the dissolved inorganic carbon, using the measurements of pH, alkalinity, the seven major cations and anions (listed above) and the DOC data where available, applied to the aqueous speciation model PHREEQC (Parkhurst and Appelo, 1999) with the phreeqc.dat database. Temperature measurements were not available for the water samples and we assumed a stream water temperature of 12°C in all these speciation calculations. We associated a unique site identifier with the estimated $p\text{CO}_2$ value at each headwater survey site.

2.1.2 Seasonal variations in stream water $p\text{CO}_2$

To account for seasonal variations in stream water $p\text{CO}_2$ values we included data from three headwater catchments (with differing dominant land use types and mean elevations) where monitoring was undertaken on either a weekly or monthly basis through complete calendar years. The features of these three catchments and the associated data are summarised in Table 1. In our subsequent analysis, we restrict our model to include only stream water data for catchments with a total area of $< 8 \text{ km}^2$; note the Eden (Pow) headwater catchment has an area of 10 km^2 , and we include it despite the larger catchment area because it meets our other criteria in terms data availability, land cover (grassland) and with an intermediate elevation. In the case of the Pow and Wensum catchments, we used measurements of pH, alkalinity and dissolved cations and anions to predict $p\text{CO}_2$ values using the same approach as the headwater survey (based on the speciation model PHREEQC Parkhurst and Appelo, 1999). For these catchments stream water temperature data at the time of sampling were also available for the speciation calculations. We converted the sampling date to a numeric

value representing day between 1 and 365 (1 January has a value of 1).

2.2 Catchment characteristics

2.2.1 Catchment area

We used the ArcHydro extension and a 5 m resolution digital surface model (NEXTMap Britain elevation data from Intermap Technologies, Intermap, 2009) to create drainage catchments upstream of all the stream sampling sites ($n = 3274$). We created a series of catchment polygons with a unique sample identifier so we could estimate catchment properties from other landscape data. We calculated the total area of each catchment polygon and associated it with the unique identifier from each headwater survey site.

2.2.2 Geomorphic variables

Previous studies (Butman and Raymond, 2011) have shown that over prolonged periods total rainfall is positively correlated with CO_2 evasion due to larger stream surface area and increased flushing of CO_2 from the soil. In temperate climates there is typically a strong correlation between altitude and total rainfall, so the former may be a useful proxy for the latter where accurate rainfall data is unavailable at fine scales (e.g. $< 10 \text{ km}^2$). Other geomorphic features such as catchment slope may also account for variations in stream water $p\text{CO}_2$ because it influences contact time between soil and percolating water. We used the digital surface model to compute mean elevation (m), mean slope ($^\circ$), and relief (difference between minimum and maximum elevation (m)) for each of the headwater catchments in our study based on the catchment polygons. We also calculated the same measures for a 1 km grid across all of England and Wales.

2.2.3 Land cover class

The dominant land cover class in each catchment was determined by intersecting the catchment polygons with a 25 m pixel Land Cover map of Great Britain 2000 (Fuller et al., 1994) and identifying for each catchment the class with the largest number of pixels. The dominant class is that which accounts for the largest proportion of land cover in the catchment; in some cases this could be < 50 %. We also calculated the dominant land cover class for the 1 km grid across England and Wales.

2.2.4 Soil and geology

We also wished to establish whether spatial data on soils and hydrogeology could account for variations in headwater $p\text{CO}_2$. We used a simplified classification to determine the dominant parent material (PM) group for each catchment polygon and associated the PM code with each catchment (Lawley and Smith, 2008). We also determined the mean BFIHOST value for each catchment. The BFIHOST value is an index value (between 0 and 1) relating to hydrological source of river flow. It is a data set that was derived for the UK from a combination of information on catchment baseflow index (BFI) and associated maps classified by the hydrology of their soil types and substrates (HOST). A BFIHOST value of 1 implies that river flow is entirely related to groundwater sources (no runoff contributions), whilst a value of 0 implies all flow is from shallow runoff. The 1 km grid BFIHOST data for England and Wales were derived from a combination of information on catchment baseflow index (Gustard et al., 1992) and associated maps classified by the hydrology of their soil types and substrates (HOST) (Boorman et al., 1995).

2.3 Removal of survey sites with the largest flows

We considered that bias might be introduced into our predictive model if we included stream sites which had large flows when they were sampled (Dinsmore et al., 2010). Time series of headwater stream flow exhibit strong positive skewness; we chose to exclude sites where data from local gauging stations showed that flow on the stream sampling date was larger than mean daily flow. We used the *ann* function in the R package *yaImpute* (Crookston and Finley, 2007) to determine the nearest neighbouring National River Flow Archive (<http://www.ceh.ac.uk/data/nrfa/index.html>) flow gauging station (based on their coordinates). This was the nearest gauging station regardless of whether it was sited on the same channel. We identified a total of 93 local gauging stations (Fig. 1) which were the nearest neighbouring stations to the selected sampling sites and which also had flow data spanning the full period of sampling. We then extracted the mean daily flow data for each of these gauging stations for the full-year period spanning the dates over each full year for

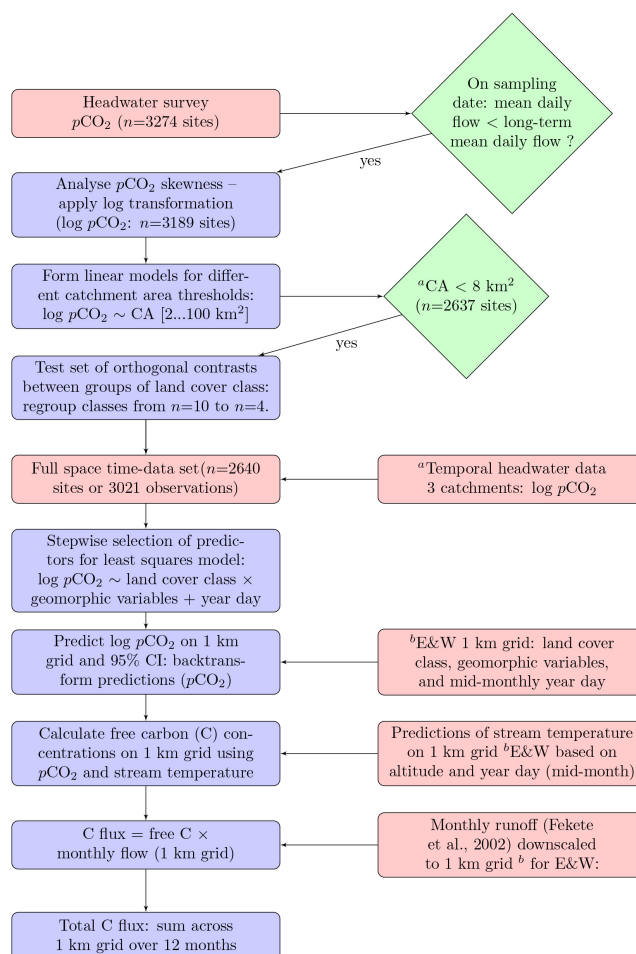


Fig. 2. Flow diagram showing stages in the development and application of a model to predict potential C (CO_2) fluxes from surface water across England and Wales. Red boxes = data sets, green boxes = data selection, blue boxes = computations. ^a see supplementary file for exploratory analysis; ^b England and Wales.

which the samples were collected (1 January 1988 to 31 December 2003). We computed the mean of the mean daily flow values for this period for each of the 93 gauging stations and compared these to mean daily flow on the sampling date recorded for each nearest neighbouring, large-scale survey site. By doing so, we identified 85 sites (from a total of 3274 sites) where mean flow was exceeded. We removed the data for these sites from our survey data set prior to further statistical analysis (Fig. 2).

2.4 Model of stream water $p\text{CO}_2$

We undertook preliminary exploratory statistical analyses using the headwater survey data. The steps we undertook in developing and applying a model to predict stream water $p\text{CO}_2$ across England and Wales using the available data are summarised in Fig. 2. These include selecting subsets of the data, applying a transformation to the $p\text{CO}_2$ values, predicting

$p\text{CO}_2$ on the transformed values, including the calculation of confidence intervals for these predictions, and computing potential C fluxes based on flow from a 1 km grid across England and Wales. We describe below the rationale for developing the model and its implementation.

The calculated $p\text{CO}_2$ values at the stream survey sites were strongly positively skewed (skewness coefficient = 2.66). We found that a logarithmic transformation of the $p\text{CO}_2$ applied after the addition of a small positive value (linear shift) of $3e^{-05}$ produced a variate with an normal approximately normal distribution (skewness coefficient = -0.06). We undertook all subsequent statistical analyses using this transformed data, and back-transformed the values to original units for the calculation of C fluxes (Fig. 2).

2.4.1 Catchment area threshold

The stream sites which were sampled as part of the large-scale headwater survey exhibited a range of catchment areas from 0.01 to 254 km² with a mean of 5.49 km². We needed to ensure that estimates of $p\text{CO}_2$ from all these sites were representative of headwaters; there is strong evidence (Butman and Raymond, 2011) that as CA (or stream order) increases from the smallest catchments, $p\text{CO}_2$ decreases because evasion exceeds inputs of CO₂ from a combination of groundwater and in-stream sources. In their study, Lauerwald et al. (2013) did not observe any correlation between CA and negative log $p\text{CO}_2$ because their analyses were based on substantially larger catchments (generally greater than 1000 km²; two orders of magnitude larger than in our study) from which a substantial proportion of CO₂ will have evaded to the atmosphere. To investigate this further, we formed a scatterplot of $p\text{CO}_2$ versus CA (Fig. 3). This suggested that those sample sites with the larger CA had – on average – lower $p\text{CO}_2$ values. To determine if there was a threshold CA below which this effect was not statistically significant, we fit ordinary least squares models (using the *lm* function in the R computing environment R Development Core Team, 2012). In each case the predictors were CA values for subsets of headwater sites truncated at thresholds of < 2, < 4, ..., < 20, < 50, < 100 and < 250 km² and the predictand was the series of transformed $p\text{CO}_2$ values. We investigated (i) whether CA is a statistically significant predictor of $p\text{CO}_2$ ($P < 0.05$) for this range of CA values, and (ii) if there is a threshold at which CA was no longer statistically significant.

2.4.2 Land cover: orthogonal contrasts

In their study, Butman and Raymond (2011) showed that land use was strongly related to area-normalised C fluxes from surface water for a range of drainage basin regions across the USA. Preliminary exploratory analyses (Fig. 4) suggested that catchment-dominant land cover class would likely be a statistically significant predictor of headwater stream $p\text{CO}_2$

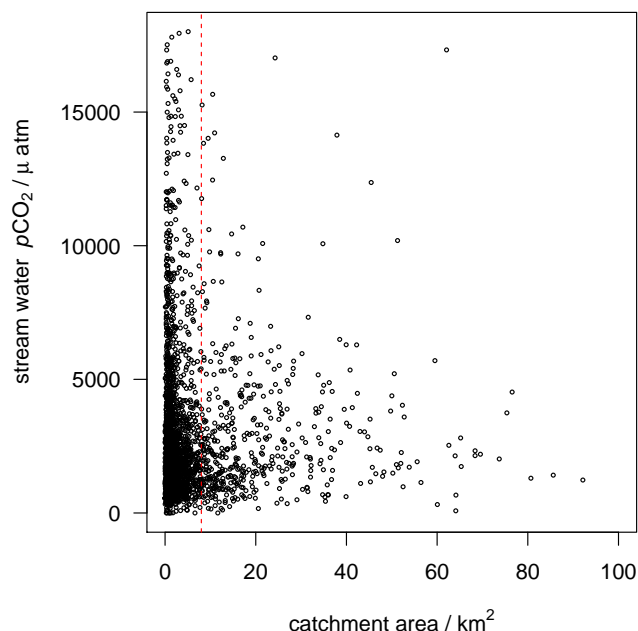


Fig. 3. Scatterplot of headwater catchment area (CA) versus stream water $p\text{CO}_2$ (μatm). The threshold chosen to select a subset of the data corresponding to the smallest headwater catchments (< 8 km²) is shown by the red dashed line.

in our survey data from England and Wales. There are reasons why we might expect the magnitude of ecosystem soil respiration (a major factor in controlling streamwater $p\text{CO}_2$) to reflect land cover type. For example, the addition of nutrients (fertilisers) to maximise agricultural production enhances both net primary production and also soil heterotrophic respiration, leading to larger soil gas $p\text{CO}_2$ values (Smith, 2005). Lowland areas with agricultural land cover generally have smaller mean annual rainfall than upland regions, so water:soil contact time in the former regions is longer and there is also less dilution of gaseous CO₂. Therefore we might expect mean $p\text{CO}_2$ values in streams draining agricultural catchments (arable or improved grassland) to be greater than those draining less managed or semi-natural habitats. Catchments dominated by urban land use may also have larger $p\text{CO}_2$ values because of nutrient inputs to managed gardens and increased heterotrophic respiration associated with nutrient loads in urban waste water. We wished to determine the most appropriate set of land cover classes for inclusion in our statistical model. The 10 dominant land cover classes (from the Land Cover Map) were: Arable, Bog, Broadleaf, Coniferous (Forest), Fen, Heather, Heather Grass, Improved Grassland, Rough Grassland and Suburban. We formed a hierarchical classification (Fig. 5) of these groups and undertook a statistical analysis using five orthogonal contrasts based on it:

1. managed land versus less managed land + urban;
2. urban versus less managed;

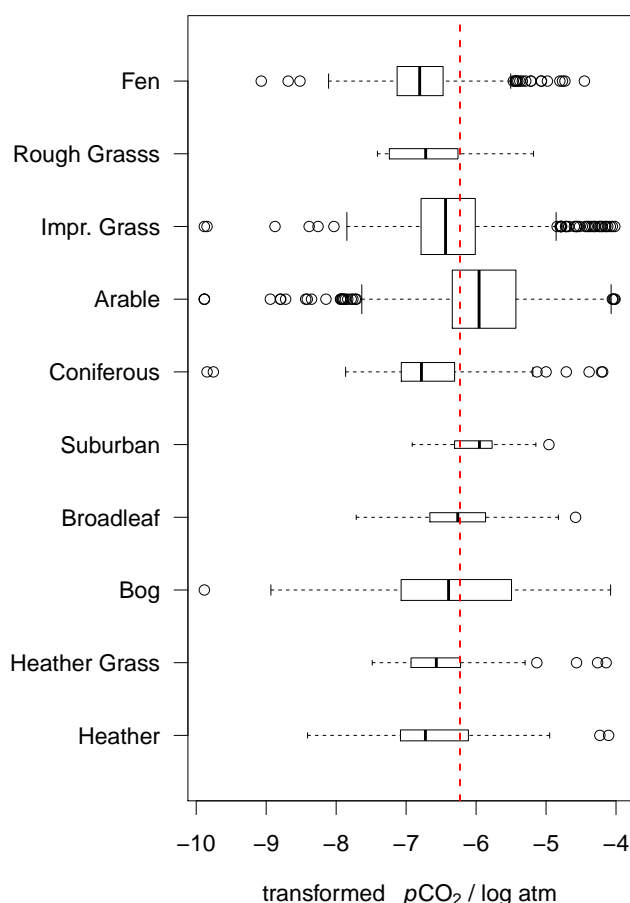


Fig. 4. Boxplot of headwater stream $p\text{CO}_2$ concentrations by dominant land cover class for each catchment ($n = 2637$) based on broad habitat types. Box widths are proportional to number of samples in each class. The red dashed line is the overall median of the data.

3. within less managed (forested versus non-forested);
4. within less managed: non-forested (wetter versus drier);
5. within managed (arable versus improved grassland).

The orthogonal treatment contrasts were entered using the contrasts function in the R environment (R Development Core Team, 2012). Each of the contrasts was tested using the aov command to determine whether they were statistically significant ($P < 0.05$).

2.4.3 Temporal variation

To account for seasonal variations in $p\text{CO}_2$ we converted sampling dates from all the headwater surveys into a year day value (numeric values between 1 and 365). We expressed the year day values on a radian scale (range from zero to $2 \times \pi$). To account for any seasonal trend we included terms for both sine and cosine of the radian-based units for year day in the ordinary least squares regression model.

2.4.4 Refining the model

We formed separate linear models for the prediction of $p\text{CO}_2$ from both the large headwater stream survey data set for those catchments with a total area of less than 8 km^2 , and also for the temporal data for the three headwater catchments. We investigated which of the various predictors (landscape-related and seasonal) accounted for statistically significant variations in the $p\text{CO}_2$ data. The models formed in these analyses and their interpretation are presented in a supplementary file to this paper.

Using a subset of the initial headwater survey data (based on the CA threshold) and including the temporal data from three headwater catchments (Fig. 2) we performed stepwise selection of the model predictors using the *stepAIC* function from the MASS library (Venables and Ripley, 2002). This tests the inclusion of predictors based on the Akaike information criterion; the k value (multiple of the degrees of freedom for penalty) was 2 and the mode of stepwise search was forwards and backwards. In our model specification we included an interaction term between elevation and the land cover classification (based on the findings from the orthogonal contrasts) as we considered this may be significant.

2.5 Monthly flow volumes

Our measurements of $p\text{CO}_2$ from headwater streams represent pathways combining shallow and deeper flow routes. Their relative magnitude depends on both geomorphology and the physical properties of local bedrock, any Quaternary deposits and the soils overlying them. We extracted data for E&W for composite monthly runoff in millimetres (mm) published by Fekete et al. (2002) on a 30 min (0.5°) resolution grid. The discharge data are from the World Meteorological Organization Global Runoff Data Centre and the water-balance model was applied using long-term mean monthly climate forcings (air temperature and precipitation). We selected the monthly runoff (mm) from the 0.5° resolution grid and extracted the values for the 1 km grid cells across E&W.

2.6 Model of headwater stream temperature

The conversion of $p\text{CO}_2$ to free CO_2 concentration in freshwater relies on the computation of Henry's constant which is temperature dependent. There is a substantial degree of variation in annual stream temperature in temperate regions such as the UK due to a combination of seasonal air temperature and variations in altitude. We used year-round data on stream water temperature from three catchments with widely differing mean elevations in England and Wales (Upper Hafren; Neal et al., 2012, 550 m, Pow; 99 m, Wensum; 48 m) to establish an ordinary least squares model of stream temperature. The predictors of stream water temperature were altitude (alt) and a combination of sine and cosine coefficients of the transformed (zero to $2 \times \pi$) year day (yr day) values

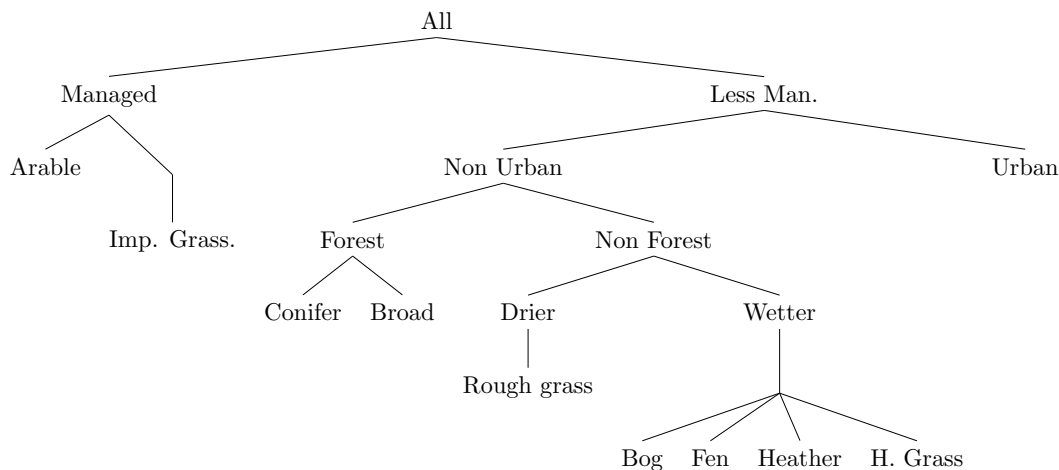


Fig. 5. Hierarchical classification of the ten land cover classes used to define groups for statistical analysis using orthogonal contrasts. Less Man. refers to less managed land cover types in contrast to more intensively managed (agricultural) land.

for each observation. The sum of two sinusoidal terms is a phase-shifted sine curve with a period of 365 days. The regression model took the form

$$T_i = \alpha + \beta_1 \text{alt}_i + \beta_2 \sin \text{yr day}_i + \beta_3 \cos \text{yr day}_i + \epsilon_i, \quad (1)$$

where α and β are coefficients of the ordinary least squares model and ϵ is the random component of the linear relationship. We included interaction terms between the predictors and used the same stepwise selection procedure described above.

2.7 Potential monthly carbon fluxes and their uncertainty

We used the linear model (*lm* function in the R package) selected from the stepwise procedure (Sect. 2.4.4) to predict $\log p\text{CO}_2$ in flowing water on 1 km grid across England and Wales. We made these predictions for each of the 12 months based on (i) the geomorphic and land cover predictors, and (ii) the year day value for the midpoint of each calendar month (Fig. 6). We used the model to predict the values for the 95 % confidence intervals (using the interval argument in the predict function) for each 1 km grid cell. We then back-transformed the predictions and the confidence interval values onto the original scale; the back-transformed values are the median values in the original units. We then used the model of stream water temperatures to convert $p\text{CO}_2$ (atm) to a dissolved gas concentration using Henry's Law and so estimated free C concentrations in water (mg L^{-1}). This concentration can be converted to a quantity of potential C evasion when it is multiplied by flow volume (Fig. 2).

3 Results

3.1 Predictive model

3.1.1 Catchment area threshold

The formation of linear models for the prediction of $\log p\text{CO}_2$ based on subsets of a range of CA thresholds (< 2 , < 4 , ..., < 20 , < 50 , < 100 and $< 250 \text{ km}^2$) showed that CA was a statistically significant predictor above 8 km^2 , but not at smaller CA thresholds. This threshold is highlighted in Fig. 3. We infer that at positions on streams which drain catchments finer than this threshold area, the rate of CO_2 evasion is balanced by combined inputs from groundwater and in-stream sources. To ensure our final model of stream water $p\text{CO}_2$ was not biased by inclusion of observations from coarser catchments, we undertook all subsequent analysis on those catchments with areas less than 8 km^2 .

3.1.2 Land cover: orthogonal contrasts

The results from significance tests for five orthogonal, land cover contrasts are shown in Table 2. It shows that the first (managed land versus less managed land + urban), second (urban versus less managed) and fifth (arable versus improved grassland) orthogonal contrasts were all statistically significant ($P < 0.05$). By contrast, the third (forested versus non-forested) and fourth (non-forested: wetter versus drier) contrasts were not statistically significant. It is noteworthy that there was no evidence for statistically significant larger $p\text{CO}_2$ values in wetter areas of land cover (that includes wetlands) where one might expect greater concentrations of both dissolved organic carbon and CO_2 to influence CO_2 export from soils. These findings suggest that the most effective reclassification of the land cover classes would be to group all the less managed, non-urban classes to form a

Table 2. Results from orthogonal contrasts between groups of selected land cover classes based on the hierarchical classification (Fig. 5).

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Contrast1	−0.021	0.003	−6.82	1.1×10^{-11}
Contrast2	−0.024	0.008	−3.13	0.002
Contrast3	−0.007	0.009	−0.74	0.47
Contrast4	0.012	0.01	1.23	0.22
Contrast5	0.088	0.006	15.4	$< 2 \times 10^{-16}$

Contrast 1 to 5 were: (1) managed land vs. less managed land+urban, (2) urban vs. less managed, (3) within less managed (forested vs. non-forested), (4) within less managed: non-forested wetter vs. drier, (5) within managed (arable vs. improved grassland).

Table 3. Summary statistics for geomorphological and land cover predictors (for catchments derived from 2637 sample sites) used to form the model of stream $p\text{CO}_2$, and for a 1 km grid ($n = 147829$) across England and Wales.

	Model data ^a	1 km grid
Geomorphologic data		
Minimum of mean elevation (m)	0	0
Mean of mean elevation (m)	215	124
Maximum of mean elevation (m)	722	945
Minimum of mean slope (°)	0.09	0
Mean of mean slope (°)	4.76	3.78
Maximum of mean slope (°)	34.8	33.2
Minimum relief (m)	0	0
Mean relief (m)	99.5	58
Maximum relief (m)	626	669
Land cover class ^b (%)		
Arable	34.7	41
Improved grassland	34.1	28.8
Suburban	0.007	0.09
Less managed	30.5	21.4

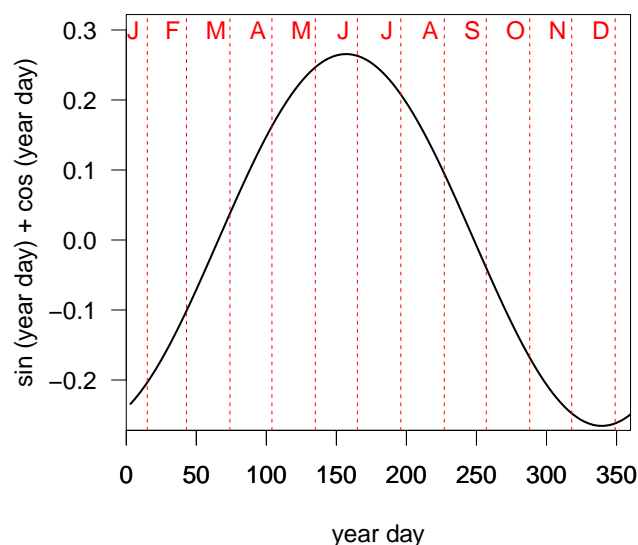
^a Data from catchments of headwater streams with CA < 8 km². ^b Dominant land cover class in catchment or grid cell.

single class, and retain three other separate classes: arable, improved grassland and urban. We undertook this reclassification before undertaking the stepwise model selection procedure, reported in the next section.

3.1.3 Final model data

Summary statistics for the final data set used to form the model for prediction of headwater stream $p\text{CO}_2$ are summarised in Table 3. It is based on data from 2637 locations from the large headwater survey where $p\text{CO}_2$ measurements were restricted to 4 months (June to September), and three small headwater catchments where measurements were made through full calendar years.

The predictors (based on stepwise selection) which were included in the final linear model of log $p\text{CO}_2$ in streams across England and Wales are shown in Table 4. The sum of

**Fig. 6.** The sum of sine and cosine model coefficients multiplied by year day highlighting the seasonal effect of temperature on headwater stream $p\text{CO}_2$. Differences in the magnitude of the coefficients used to apply the model at the midpoint of each calendar month (J=January, F=February, etc.) are highlighted by the dashed red lines.

the model coefficients for sine and cosine functions multiplied by year day are presented in Fig. (6) highlighting the effect of temperature on ecosystem respiration and stream $p\text{CO}_2$ values throughout the year. The residuals were close to a normal distribution (histogram not shown: skewness coefficient = −0.31). Although the land use interaction terms LM and Urban have p values < 0.05 they are included in the final model because land use has statistically significant interactions (with elevation) and so its terms must be included to give a simple interpretation to the interaction. This model accounted for 24% of the variance (adjusted $R^2 = 0.24$) in $p\text{CO}_2$ values in the combined spatial and temporal data set. Although there is a considerable amount of variation which the model cannot account for, we consider its performance is reasonable given that (i) $p\text{CO}_2$ in stream water exhibits a

Table 4. Summary of the model from stepwise selection of predictors of $\log p\text{CO}_2$ in stream water based on geomorphic variables, land cover classification and year day. The coefficient (Estimate) for the three land cover classes shown are expressed as differences from the Arable land cover class (not shown).

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	-5.53	0.04	-131	$< 2 \times 10^{-16}$
Sin(yr day)	0.11	0.03	3.903	9.70×10^{-05}
Cos(yr day)	-0.24	0.03	-7.904	3.76×10^{-15}
Mean Slope	-0.06	0.006	-10.5	$< 2 \times 10^{-16}$
Relief	-0.0009	0.0002	-3.802	0.0001
Mean Elev.	-0.003	0.0004	-7.61	3.75×10^{-14}
IG ^a	-0.39	0.07	-5.44	5.75×10^{-08}
Urban	-0.09	0.36	-0.260	0.79
LM ^b	-0.04	0.087	-0.47	0.64
Elev. ^c :IG ^b	0.002	0.0004	4.48	7.67×10^{-06}
Elev. ^c :Urban	0.001	0.002	0.47	0.63
Elev. ^c :LM ^b	0.0015	0.0004	3.72	0.0002

^a Improved grassland, ^b less managed (non-urban), ^c mean elevation.

substantial degree of spatial (Jones and Mulholland, 1998) and temporal (Dinsmore et al., 2010) heterogeneity, and (ii) the landscape predictors can be obtained for most terrestrial landscapes.

Those predictors which accounted for larger proportions of the variation in headwater $p\text{CO}_2$ were the temporal coefficients (sin and cosine of year day), plus mean slope and mean elevation, although in combination the land cover classes also accounted for a reasonable proportion of the variance. Mean catchment elevation has a strong positive correlation with rainfall. Greater rainfall leads to dilution of stream $p\text{CO}_2$ which likely explains why the variation of the latter is accounted for by difference in mean catchment elevation.

It is noteworthy that neither PM class or BFIHOST values were statistically significant predictors of $p\text{CO}_2$. We infer that variations in headwater stream $p\text{CO}_2$ are more closely associated with those factors closely related to the generation of soil gas CO_2 (such as land cover type) and less related to the transport pathways determined by variations in soil parent material and hydrogeology. However, mean catchment slope is a significant predictor of headwater $p\text{CO}_2$. Steeper average slopes likely lead to shorter contact times between soil and water before it reaches a channel, and thus smaller average $p\text{CO}_2$ values.

3.2 Model of headwater stream temperature

The range of headwater stream temperature values for the three sites was 0.89 to 18.1 °C, with a median of 12.25 °C. In combination, the use of sine and cosine functions based on year day and elevation (in metres) accounted for 78 % of the variance in headwater stream temperature. A summary of the model coefficients are presented in Table 5. We considered

the model provided a reasonable basis for predicting stream temperatures on a 1 km grid across England and Wales for the midpoint of each calendar month which we used to convert $p\text{CO}_2$ to its free $\text{CO}_2\text{-C}$ concentration in water.

3.3 Model predictions: England and Wales

To highlight differences in free $\text{CO}_2\text{-C}$ concentrations in flow throughout the year across England and Wales, Fig. 7 presents model predictions for the months of July and January, highlighting the differences in temperature controlling ecosystem respiration. Free $\text{CO}_2\text{-C}$ concentrations in flow are larger in July, reflecting the greater concentrations derived from soil and in-stream respiration, but also greater dilution in January associated with larger quantities of rainfall (Table 6). The maps also reflect the differences in land cover type; the south and east of England is dominated by arable agriculture over neutral soils with fertiliser inputs which enhances ecosystem respiration. By contrast, soil pH is more acidic in the north and west (of England and also in Wales) where land cover types (Improved grassland or less managed habitats) are subject to smaller, or no, nutrient and lime additions. In more intensive agricultural regions, where the pH is maintained around neutral and fertilisers are applied to maximise productivity, soils have larger rates of soil respiration leading to greater production of CO_2 . This may lead to large fluxes of dissolved CO_2 in percolating water, but this depends on the balance between losses in percolation (dissolved CO_2) and gaseous CO_2 emissions from the soil surface. Unfortunately we do not have quantitative data on soil pH at a sufficiently fine resolution to examine this relationship in greater detail.

The predicted median free $\text{CO}_2\text{-C}$ concentrations in May and November across England and Wales were 1.78 and 1.11 mg L^{-1} , respectively. The median free $\text{CO}_2\text{-C}$ concentrations for the lower and upper 95 % confidence intervals were 1.97 and 1.47 mg L^{-1} for May and 0.91 and 1.24 mg L^{-1} for November.

Based on the quantities of flow from each 1 km grid square (Fig. 2) we present the potential monthly evasion fluxes for July and January using data on mean monthly annual rainfall (Fig. 8). For years in which monthly rainfall is near the long-term average quantities, the largest potential C fluxes ($> 70 \text{ kg C km}^{-2}\text{month}^{-1}$) occur generally in the winter months (December–March) in upland ($> 300 \text{ m}$ elevation) areas subject to the largest monthly rainfall (north-west England, Wales and south-west England). The lowest potential fluxes occur in lowland settings with the smallest monthly runoff quantities. Potential C efflux is dominated by flow volumes rather than free C concentrations in surface water; Table 6 shows there is a 14-fold difference in maximum (January) and minimum (July) monthly flow volume, whilst there is less than a two-fold difference in mean free $\text{CO}_2\text{-C}$ concentrations (1.18 and 1.78 mg L^{-1} , December and June respectively). So despite larger free $\text{CO}_2\text{-C}$ concentrations

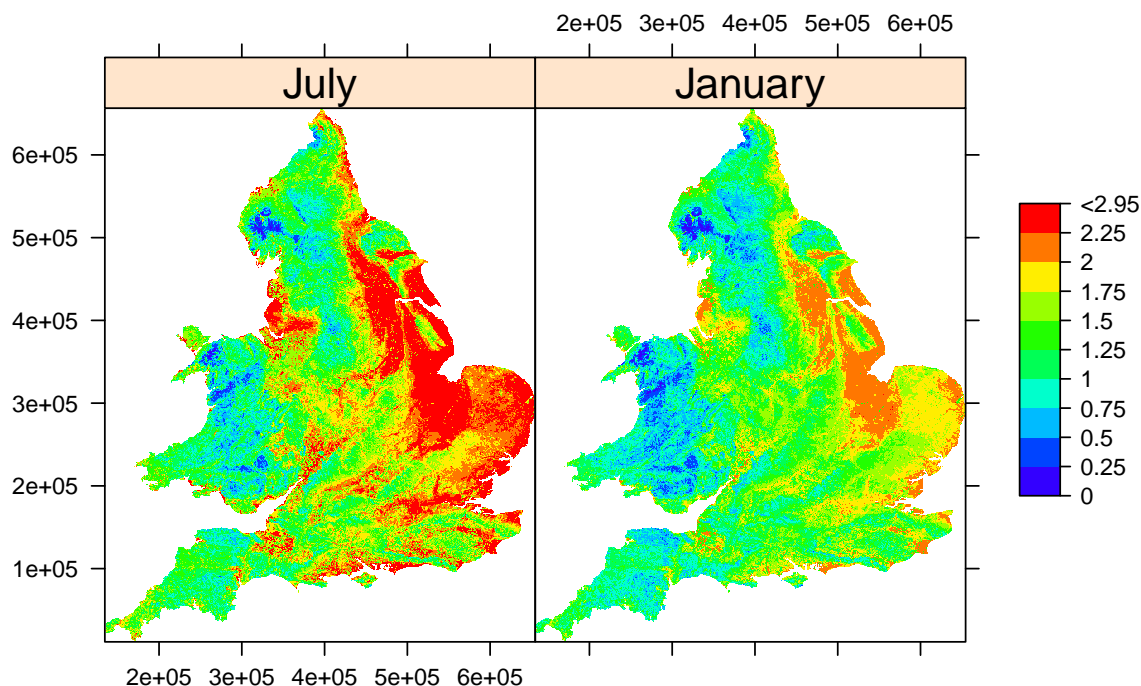


Fig. 7. Predicted free $\text{CO}_2\text{-C}$ concentration (mg L^{-1}) in flow for a 1 km grid across England and Wales: May and November. Coordinates are metres on the British National Grid. Note the maximum class limit is greater than the arithmetic scale used in the other classes.

Table 5. Summary of the ordinary least squares model used for the prediction of daily headwater stream temperature for the 1 km grid across England and Wales. Year day (yr day) is the transformed numeric value of day in the calendar year. A colon (:) denotes inclusion of an interaction (a product term) of predictors.

	Estimate	Std. Error	<i>t</i> value	<i>p</i> value
Intercept	10.2	0.024	434	$< 2 \times 10^{-16}$
sin(yr day)	-2.37	0.05	-47.6	$< 2 \times 10^{-16}$
cos(yr day)	-3.23	0.031	-103	$< 2 \times 10^{-16}$
Elev.	-0.004	0.0001	-38.1	$< 2 \times 10^{-16}$
sin(yr day):cos(yr day)	-2.64	0.088	-29.8	$< 2 \times 10^{-16}$
sin(yr day):Elev.	0.001	0.0002	8.17	3.38×10^{-16}
cos(yr day):Elev.	0.0003	0.0001	2.147	0.031
sin(yr day):cos(yr day):Elev	0.005	0.0003	17.0	$< 2 \times 10^{-16}$

in surface water during summer months (assuming average runoff), potential C fluxes are substantially smaller than for average winter months.

Our model predicts a mean annual flux of $0.44 \text{ t C km}^{-2} \text{ yr}^{-1}$ or $1.6 \text{ g CO}_2 \text{ m}^{-2} \text{ yr}^{-1}$ (expressed on the basis of total land area). This latter value is 25 times smaller than the figure of $40.4 \text{ g CO}_2 \text{ m}^{-2} \text{ yr}^{-1}$ for this region published by Raymond et al. (2013). In a study from a peatland stream and catchment in Scotland (within the region of the value published by Raymond et al., 2013), Hope et al. (2001) gave a figure of $51.7 \text{ g CO}_2 \text{ m}^2 \text{ yr}^{-1}$ ($14.1 \text{ g C m}^{-2} \text{ yr}^{-1}$) for evasion per unit area of watershed. Peatland or bog environments, which have substantially larger channel evasion fluxes than most land cover types, account for

a relatively small proportion ($< 5\%$) of land cover across England and Wales according to the Land Cover Map of England and Wales. We assume that the evasion fluxes per unit land area presented by Raymond et al. (2013) are based on studies from these peatland environments and this would account for the substantially larger fluxes reported in their study compared to our figure. Our estimate should be towards the upper end of actual evasion rates because our model assumes that all free CO_2 evades.

The total potential annual C efflux from flow across England and Wales (summing across 12 months) is 65.4 kt C (kilotonnes carbon), with the lower and upper 95% confidence intervals of 56.1 and 77.2 kt C , respectively. The confidence intervals we present only reflect the uncertainty

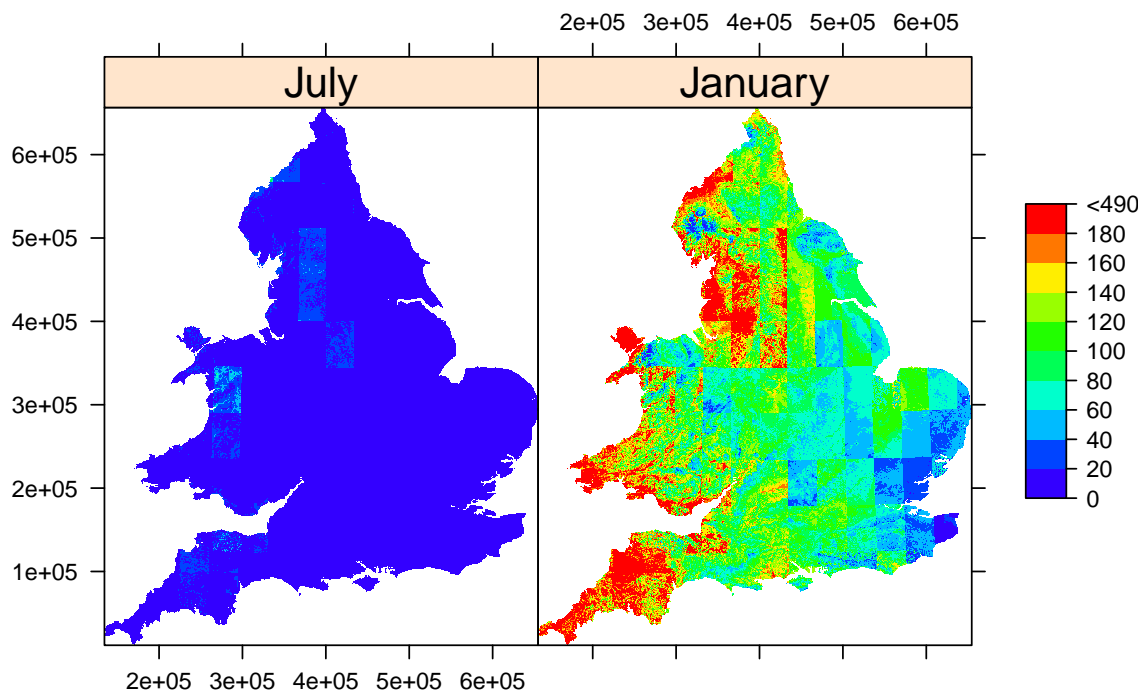


Fig. 8. Predicted potential monthly source of carbon (kg) efflux based on flow for a 1 km grid across England and Wales (May and November). Coordinates are metres on the British National Grid. Note the maximum class limit is greater than the arithmetic scale used in the other classes.

Table 6. Summary of monthly predictions across all of England and Wales for total flow volume (gigalitres; GL) based on composite runoff (Fekete et al., 2002) on a 0.5° resolution grid, mean concentration of free $\text{CO}_2\text{-C}$ (mg L^{-1}) and potential carbon efflux (kilotonnes carbon; kt C).

	Flow (Gl)	Mean free C (mg L^{-1})	C efflux ^a (kt C)
Jan	14003	1.34	16314
Feb	6651	1.52	9221
Mar	5601	1.69	8527
Apr	3885	1.77	6210
May	2491	1.78	3872
June	1460	1.74	2178
July	978	1.64	1325
Aug	1005	1.47	1127
Sept	1264	1.30	1198
Oct	2503	1.16	2170
Nov	5473	1.11	4830
Dec	8491	1.18	8496
Total	53806	–	65379

^a Potential C efflux – assumes all free C evades to the atmosphere.

associated with prediction of $p\text{CO}_2$ concentrations and do not account for the uncertainties in the computation of flow or with stream temperature estimation (used to compute Henry's constant). For the year 2002, Worrall et al. (2007)

estimated the total fluvial flux of C from the terrestrial biosphere to surface water across English and Welsh rivers to be $1530 \text{ kt C yr}^{-1}$ (megatonnes carbon per year; equivalent to $10.3 \text{ t C km}^2 \text{ yr}^{-1}$). They estimated around 32% (620 kt C) of the total (equivalent to $4.2 \text{ t C km}^2 \text{ yr}^{-1}$) was lost to the atmosphere from surface water. The largest proportion of this flux was DOC (42%), but their flux of excess dissolved CO_2 was $0.37 \text{ Mt C yr}^{-1}$. This estimate was based on measurements of $p\text{CO}_2$ in groundwater from those regions with major aquifers, having made assumptions concerning mixing with surface waters. This estimate by Worrall et al. (2007) for the year 2002 is around 6 times greater than our model prediction of potential CO_2 flux from surface water to the atmosphere ($0.06 \text{ Mt C yr}^{-1}$) based on mean annual flow. In their study, Worrall et al. (2007) computed an average free C concentration in surface water of England and Wales of 5.2 mg C L^{-1} which is between 3 and 4 times larger than the overall mean concentration of 1.48 mg C L^{-1} predicted by our model. It may be helpful to compare the methodology applied by Worrall et al. (2007) and that used in this study to determine why the former has much larger free C concentrations, based on mixing of ground- and surface water.

Our model could be applied using future land use change scenarios to estimate the magnitude of their impact on potential C evasion fluxes from surface water, an approach similar to that used to assess changes in greenhouse gas emissions from soil (Smith et al., 2010). However, certain soil

properties that exert an influence on $p\text{CO}_2$ values in drainage water may be independent of local changes in land use. The model could also be modified to assess differences in the magnitude of potential annual C efflux based on yearly variations in monthly rainfall or changes in mean monthly temperature.

We recognise that our model cannot currently account for all the processes influencing the magnitude of potential CO_2 evasion from surface water at the landscape scale. Our approach does not account for changes in headwater stream CO_2 flux due to variations in discharge; although dissolved CO_2 concentrations in headwater catchments are lower at larger discharges (Dinsmore et al., 2013b), total fluxes increase during storm events. To ensure our model was not biased by the inclusion of $p\text{CO}_2$ measurements from streams at the largest flow conditions, we excluded sites where daily flow was likely greater than long-term mean daily flow (based on local gauging stations). To account for variations in flow would require a much shorter time step (e.g. daily) and parameters to predict concentration–discharge relationships. Our model does not account for in-stream sources of CO_2 in the reaches of larger channels downstream of their headwaters. Our model accounted for a smaller proportion (24 %) of the variation in channel $p\text{CO}_2$ by comparison to that presented by Lauerwald et al. (2013) (43 %). There are two main factors which could account for this difference: (i) Lauerwald et al. (2013) used mean $p\text{CO}_2$ values from a series of water samples compared to the more variable, single snapshot observations used in our study, (ii) the observations reported by Lauerwald et al. (2013) were for substantially larger catchments providing more opportunity for smoothing of the variation, in contrast to our smaller catchments which likely exhibit greater variation.

A further improvement on our current approach would be to compute actual rather than potential CO_2 evasion fluxes using functions which predict variations in downstream gas transfer velocities using parameters of stream hydraulics (Raymond et al., 2012) and channel surface area. A recent study showed that channel-wetted widths across England and Wales are strongly influenced by catchment area and hydrological source of flow (Rawlins et al., 2013).

4 Conclusions

Using analyses of more than 3000 analyses of snapshot headwater samples across different landscape settings of England and Wales, we showed that below a CA threshold of around 8 km^2 there was no statistically significant difference in stream $p\text{CO}_2$ – evasion losses of CO_2 from stream channels below this scale are likely balanced by inputs from groundwater and in-stream sources. We used estimates of $p\text{CO}_2$ from catchments with areas less than 8 km^2 to assess other landscape predictors of stream $p\text{CO}_2$, including data from three catchments in which $p\text{CO}_2$ had been mea-

sured throughout a full calendar year. Based on a series of orthogonal contrasts we found that grouping dominant catchment land cover types into four classes provided the optimum classification for $p\text{CO}_2$. A seasonal variable (expressed as a year day number), mean catchment elevation (interacting with land cover class), mean catchment slope and catchment relief (maximum minus minimum altitude) were also statistically significant predictors of $p\text{CO}_2$. We formed a model from these factors and stream water temperature which accounted for 24 % of the combined spatial and temporal variation in $p\text{CO}_2$ across England and Wales. We predicted free CO_2 -C concentrations in water for a 1 km grid across England and Wales using their catchment characteristics and a model of stream water temperature as predictors. We also used average monthly runoff on a 1 km grid across England and Wales. By combining the predicted free CO_2 -C concentrations and runoff for each 1 km grid cell, we computed monthly and total annual potential C fluxes (65.4 kt C with 95 % confidence intervals 56.1–7.2 kt C) from surface water to the atmosphere, assuming that all CO_2 entering surface water evades.

Supplementary material related to this article is available online at <http://www.biogeosciences.net/11/1911/2014/bg-11-1911-2014-supplement.pdf>.

Acknowledgements. The Land Cover Map and BFIHOST data were provided under licence by the Centre for Ecology and Hydrology (NERC). We would like to thank (i) all the staff and students at the British Geological Survey who were involved in the collection and analysis of samples from headwater streams across England and Wales as part of the G-BASE (Geochemical Baseline Survey of the Environment) project, (ii) staff from the Wensum and Eden Demonstration test catchments for providing their data, (iii) Kerry Dinsmore at CEH for providing the data on stream water $p\text{CO}_2$ concentrations for Black burn, (iv) Ben Marchant and Murray Lark for statistical advice on the application of linear regression models and (v) two anonymous reviewers for their helpful comments. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

Edited by: S. Bouillon

References

- Battin, T. J., Kaplan, L. A., Findlay, S., Hopkinson, C. S., Marti, E., Packman, A. I., Newbold, J. D., and Sabater, F.: Biophysical controls on organic carbon fluxes in fluvial networks (vol 1, pg 95, 2008), *Nature Geosci.*, 2, 595–595, 2009.
- Benstead, J. P. and Leigh, D. S.: An expanded role for river networks, *Nature Geosci.*, 5, 678–679, 2012.
- Boorman, D., Hollis, J., and Lilly, A.: Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom, Tech. rep., Wallingford, <http://nora.nerc.ac.uk/7369/>, 1995.

- Butman, D. and Raymond, P. A.: Significant efflux of carbon dioxide from streams and rivers in the United States, *Nature Geosci.*, 4, 839–842, 2011.
- Crookston, N. L. and Finley, A. O.: yaImpute: An R Package for kNN Imputation, *J. Statist. Software*, 23, <http://www.jstatsoft.org/v23/i10>, 2007.
- Dinsmore, K. J., Billett, M. F., Skiba, U. M., Rees, R. M., Drewer, J., and Helfter, C.: Role of the aquatic pathway in the carbon and greenhouse gas budgets of a peatland catchment, *Glob. Change Biol.*, 16, 2750–2762, 2010.
- Dinsmore, K. J., Billett, M., and Dyson, K. E.: Five year record of aquatic carbon and greenhouse gas concentrations from Auchencorth Moss, doi:10.5285/3f0820a7-a8c8-4dd7-a058-8db79ba9c7fe, 2013a.
- Dinsmore, K. J., Billett, M. F., and Dyson, K. E.: Temperature and precipitation drive temporal variability in aquatic carbon and GHG concentrations and fluxes in a peatland catchment, *Glob. Change Biol.*, 19, 2133–2148, 2013b.
- Fekete, B. M., Vorosmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cy.*, 16, 1042, doi:10.1029/1999gb001254, 2002.
- Fuller, R. M., Groom, G. B., and Jones, A. R.: The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper data, *Photogram. Engin. Remote Sens.*, 60, 553–562, 1994.
- Gustard, A., Bullock, A., and Dixon, J. M.: Low flow estimation in the United Kingdom, *Tech. rep.*, Wallingford, <http://nora.nerc.ac.uk/6050/>, 1992.
- Hope, D., Palmer, S. M., Billett, M. F., and Dawson, J. J. C.: Carbon dioxide and methane evasion from a temperate peatland stream, *Limnol. Oceanogr.*, 46, 847–857, 441CZ, http://aslo.org/lo/toc/vol_46/issue_4/0847.pdf, 2001.
- Johnson, C., Breward, N., Ander, E., and Ault, L.: G-BASE: Baseline geochemical mapping of Great Britain and Northern Ireland., *Geochem.*, 5, 1–13, 2005.
- Johnson, M. S., Lehmann, J., Riha, S. J., Krusche, A. V., Richey, J. E., Ometto, J. P. H. B., and Couto, E. G.: CO₂ efflux from Amazonian headwater streams represents a significant fate for deep soil respiration, *Geophys. Res. Lett.*, 35, L17401, doi:10.1029/2008GL034619, 2008.
- Jones, J., Partick, J., and Mulholland, J.: Influence of drainage basin topography and elevation on carbon dioxide and methane supersaturation of stream water, *Biogeochemistry*, 40, 52–72, 1998.
- Jones, J. B. and Mulholland, P. J.: Carbon dioxide variation in a hardwood forest stream: An integrative measure of whole catchment soil respiration, *Ecosystems*, 1, 183–196, 1998.
- Lauerwald, R., Hartmann, J., Moosdorf, N., Kempe, S., and Raymond, P. A.: What controls the spatial patterns of the riverine carbonate system? — A case study for North America, *Chem. Geol.*, 337–338, 114–127, 2013.
- Lawley, R. and Smith, B.: Digital soil mapping at a national scale: a knowledge and GIS based approach to improving parent material and property information, in: *Digital soil mapping with limited data*, edited by: Hartemink, A., McBratney, A., and de Lourdes Mendonca-Santos, M., Springer, <http://nora.nerc.ac.uk/4926/>, 173–182, 2008.
- Li, S., Lu, X. X., and Bush, R. T.: CO₂ partial pressure and CO₂ emission in the Lower Mekong River, *J. Hydrology*, 504, 40–56, 2013.
- Morel, F. M. M. and Hering, J.: Principles and applications of aquatic chemistry, Wiley-Interscience, 1993.
- Neal, C., Reynolds, B., Rowland, P., Norris, D., Kirchner, J. W., Neal, M., Sleep, D., Lawlor, A., Woods, C., Thacker, S., Guyatt, H., Vincent, C., Hockenhull, K., Wickham, H., Harman, S., and Armstrong, L.: High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge, *Sci. Total Environ.*, 434, 3–12, 2012.
- Owen, G. J., Perks, M. T., Benskin, C. M. H., Wilkinson, M. E., Jonczyk, J., and Quinn, P. F.: Monitoring agricultural diffuse pollution through a dense monitoring network in the River Eden Demonstration Test Catchment, Cumbria, UK, *Area*, 44, 443–453, 2012.
- Parkhurst, D. L. and Appelo, C.: User's Guide to PHREEQC – (Version 2) a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations, *Tech. Rep. 99-4259*, US Geological Survey Water Resources Investigations Report, 1999.
- Prasad, M. B. K., Kaushal, S. S., and Murtugudde, R.: Long-term pCO₂ dynamics in rivers in the Chesapeake Bay watershed, *Applied Geochemistry*, 31, 209–215, 2013.
- R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, ISBN 3-900051-07-0, 2012.
- Rawlins, B. G., Clark, L., and Boyd, D. S.: Using air photos to parameterise landscape predictors of channel wetted width, *Earth Surf. Proc. Land.*, <http://dx.doi.org/10.1002/esp.3469>, 2013.
- Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., Laursen, A. E., McDowell, W. H., and Newbold, D.: Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers, *Limnol. Oceanogr.*, 2, 41–53, 2012.
- Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Dürr, H., Meybeck, M., Ciais, P., and Guth, P.: Global carbon dioxide emissions from inland waters, *Nature*, 503, 355–359, 2013.
- Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., Laruelle, G. G., Lauerwald, R., Luysaert, S., Andersson, A. J., Arndt, S., Arnosti, C., Borges, A. V., Dale, A. W., Gallego-Sala, A., Godderis, Y., Goossens, N., Hartmann, J., Heinze, C., Ilyina, T., Joos, F., LaRowe, D. E., Leifeld, J., Meysman, F. J. R., Munhoven, G., Raymond, P. A., Spahni, R., Suntharalingam, P., and Thullner, M.: Anthropogenic perturbation of the carbon fluxes from land to ocean, *Nature Geosci.*, available online at: doi:10.1038/ngeo1830, 2013.
- Smith, P., Bhogal, A., Edgington, P., Black, H., Lilly, A., Barraclough, D., Worrall, F., Hillier, J., and Merrington, G.: Consequences of feasible future agricultural land-use change on soil organic carbon stocks and greenhouse gas emissions in Great Britain, *Soil Use Manag.*, 26, 381–398, 2010.
- Smith, V. R.: Moisture, carbon and inorganic nutrient controls of soil respiration at a sub-Antarctic island, *Soil Biol. Biochem.*, 37, 81–91, 2005.
- Sun, G., Caldwell, P., Noormets, A., McNulty, S. G., Cohen, E., Moore Myers, J., Domec, J.-C., Treasure, E., Mu, Q., Xiao, J.,

- John, R., and Chen, J.: Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model, *J. Geophys. Res.*, 116, 1–16, 2011.
- Venables, W. N. and Ripley, B. D.: *Modern Applied Statistics with S*, Springer, New York, 4th Edn., <http://www.stats.ox.ac.uk/pub/MASS4>, ISBN 0-387-95457-0, 2002.
- Wallin, M. B., Quist, M. G., Buffam, I., Billett, M. F., Nisell, J., and Bishop, K. H.: Spatiotemporal variability of the gas transfer coefficient (K_{CO_2}) in boreal streams: Implications for large scale estimates of CO_2 evasion, *Global Biogeochem. Cy.*, 25, GB3025, doi:10.1029/2010GB003975, 2011.
- Wensum Alliance: Manual sample water quality data Wensum Demonstration Test Catchment, Tech. rep., University of Norwich (East Anglia: UK), 2013.
- Worrall, F., Guilbert, T., and Besien, T.: The flux of carbon from rivers: the case for flux from England and Wales, *Biogeochemistry*, 86, 63–75, 2007.