Supplementary file associated with *A model of potential carbon dioxide efflux from surface water across England and Wales using headwater stream survey data and landscape predictors* by Rawlins et al., published in Biogeosciences

This supplement presents three items. First (parts A+B) it presents and provides an interpretation of two linear regression models from analysis of datasets described in the paper of Rawlins et al. (2014). The first dataset is for pCO_2 values from the spatial headwater survey of catchments with a total area of less than 8 km^2, limited to collection during the summer months where streams were below mean flow. The second dataset was for pCO_2 values measured over a full calendar year in three headwater catchments, with details in the full paper. Second, (part B) it presents the temporal data for pCO_2, free-C (mg L^{-1}) and stream temperature for the three headwater catchments and an interpretation of their variation. The data from Part A and part B were combined and a spatio-temporal model of streamwater pCO_2 was developed; this is presented in the full paper. Third, Part C provides a meta data description of the full spatial and temporal dataset which is provided as a supplementary file that accompanies the paper by Rawlins et al. (2014).

PART A

Linear model of spatial headwater survey data

These pCO_2 values are from 2637 individual headwater stream sites across England and Wales each with a total catchment area of less than 8 km^2. The samples were all collected during the summer months (between June and September). Based on records made during sampling and also from measurements from local gauging stations, flow at the time of sampling was likely below long-term mean daily flow.

For each catchment we derived various features of the catchment including its mean slope(°), relief (maximum minus minimum elevation), mean elevation (m) and its dominant Land Cover class (IG=improved grassland, LM=less managed, Urban, Arable - coefficient not presented in the Table below) and yday (sine and cosine coefficients of the transformed (zero to $2\times\pi$) year day values for each observation). We then formed an ordinary least squares linear model (using stepwise selection) where the transformed pCO_2 values was the predictand and the various catchment features and the year-day (yday) values were the predictors (Table 1).

The significant predictors in Table 1 include a range catchment features, but significantly also the temporal components (Sine and Cosine of year day) relating to seasonal changes based on day of the year, limited to the summer months. Hence it is clear that a seasonal effect is apparent even within the warmest months (June to September). The model presented in Table 1 accounted for 24% of the variance in transformed pCO_2 values, the same proportion as the full model (24%) combining the spatial and temporal data (see Rawlins et al. 2014). Land cover class (also included in the combined spatial survey data and temporal model) is a statistically significant predictor, included here as a series of land cover groups, as a main effect or as interactions with other predictors.

Table 1 Summary of multiple linear regression model for prediction of log-linear transformed pCO_2 values for headwater survey sites (n=2637) - does not include the data for the three headwater catchments for which there was year round data.

	Estimate	Std. Error	t-value	P-value
Intercept	-6.27	0.145	-43.2	$< 2 \times 10^{-16}$
Sin(yday)	-0.18	0.079	-2.24	0.025
Cos(yday)	-0.87	0.12	-7.2	8.06×10^{-13}
Mean slope	-0.04	0.006	-6.44	1.37×10^{-10}
Relief	-0.001	0.0002	-3.931	8.67×10^{-05}
Mean Elev.	-0.0024	0.0004	-6.23	5.41×10^{-10}
IG	-0.22	0.079	-2.72	0.0067
Urban	-0.077	0.37	-0.21	0.83
LM	-0.49	0.11	-4.62	4.10×10^{-06}
Elev:IG	0.001	0.0005	2.43	0.015
Elev:Urban	0.0009	0.003	0.35	0.73
Elev:LM	0.002	0.0004	4.46	8.68×10^{-06}

Linear model of temporal headwater measurements from three catchments

We combined the data for $p\mathrm{CO_2}$ values from the three headwater catchments (Black water, Pow and Black burn) into a single dataset including the same set of landscape features (Mean elevation, Relief and Mean slope) used in the analysis of the spatial stream survey. In this case the year day values extended across all months of the year (see Figure 1). The $p\mathrm{CO_2}$ values were less strongly (positively) skewed than those in the headwater survey and we chose to use the raw (untransformed) data for fitting linear models by ordinary least squares using stepwise selection (see main text for further details).

Table 2 Summary of multiple linear regression model for prediction of $p\mathrm{CO_2}$ values (milli atmospheres) for three headwater sites.

	Estimate	Std. Error	t-value	P-value
Intercept	6.69	0.47	14.3	$< 2 \times 10^{-16}$
Sin (yday)	-0.22	0.098	-2.21	0.0278
Cos (yday)	-0.96	0.098	-9.84	$< 2 \times 10^{-16}$
Mean slope	-5.07	0.71	-7.12	$< 5.28 \times 10^{-12}$
Mean elevation	0.014	0.002	6.764	$< 5.1 \times 10^{-11}$

We do not consider it justified to draw wider conclusions from the temporal data in relation to the other catchment properties (e.g. mean elevation, land use) because we only have a single elevation value for each of the three individual catchments in each case (i.e. lowland arable (Black water 48 m), moderate elevation grassland (Pow 99 m) and upland bog (Black burn 280 m)). The very small P-value (for the Sin (yday) predictor ($< 2 \times 10^{-16}$) highlights the importance of the seasonal variation on stream $p\mathrm{CO_2}$ values. The model presented in Table 2 accounted for 29% of the variation in $p\mathrm{CO_2}$ values based on all the data (n=384 observations) from the three sites.

PART B - Temporal stream data

We present the temporal variation in $p\mathrm{CO_2}$ values, free C and stream temperature for the three headwater catchments (Figure 1). This shows a seasonal signal in $p\mathrm{CO_2}$ across the four seasons; the largest $p\mathrm{CO_2}$ values typically occur during the summer months, but there is a considerable degree of within season variation. In the case of Black burn, the pCO2-discharge relationships have already been explored in a previous study (Dinsmore and Billett, 2008) where the authors showed 'an overall decrease in stream water $p\mathrm{CO_2}$ concentration with increasing discharge' that was exhibited by a weak negative correlation. By contrast, they showed that excess stream water CO2 was positively correlated with the height of peak discharge. As discussed in Rawlins et al. (2014), one of the limitations of our modelling approach, based on sampling streams below mean flow, is that we cannot account for flow-related variations in $p\mathrm{CO_2}$ values and evasion of free $\mathrm{CO_2}$.

Note, the temporal data for the Pow catchment (plot b in Figure 1) is limited to monthly samples, whilst there is much more frequent data for the other two sites. Figure 1 also highlights the change in the relationship of pCO2 to free C; there is a greater proportion of free C relative to PCO2 at colder by contrast to warmer temperatures due to the change of $\mathrm{CO_2}$ solubility with temperature.

Despite the differences in elevation (range 48 to 280 m), the maximum temperatures for the three catchments are quite similar (15° C), whilst the catchment at the greatest elevation (Black burn; Figure 1 c) has longer periods at the lower temperatures ($< 5°$ C).

PART C - Meta data description of the full dataset provided as a downloadable .csv file

We have provided the full dataset used in our analysis so other researchers can download and use it in their own analyses. Please cite the final version of the paper published in Biogeoscinces as a means of referring to the dataset. The data are in the supplementary file supp.csv with 3021 rows and a series of headers with – data description:

1. Date – calendar date
2. uatm_pco2 – pCO2 value in μ atmospheres
3. Alt – mean catchment altitude (m)
4. slope – mean catchment slope (°)
5. rel – catchment relief (m)
6. newLC – land cover class (aggregated classes - see text)
7. yday – date converted to a calendar day from 1 January
8. log_lin – linear shift and natural logarithmic transformation of field uatm_pco2
9. yday_trig – transformed yday: (yday $\times(2 \times \pi/365)$)
10. ST – classifier either S (spatial survey) or T (temporal site data)

If there are queries regarding these data it may help to write to Barry Rawlins (bgr@bgs.ac.uk) with a query.

References:

Dinsmore, K. J. and M. F. Billett 2008. Continuous measurement and modelling of CO2 losses from a peatland stream during stormflow events. *Water Resources Research* 44: W12417.

Rawlins, B. G., Palumbo-Roe, B., Gooddy, D. C., Worrall, F. and Smith, H. 2014. A model of potential carbon dioxide efflux from surface water across England and Wales using headwater stream survey data and landscape predictors. *Biogeosciences*.
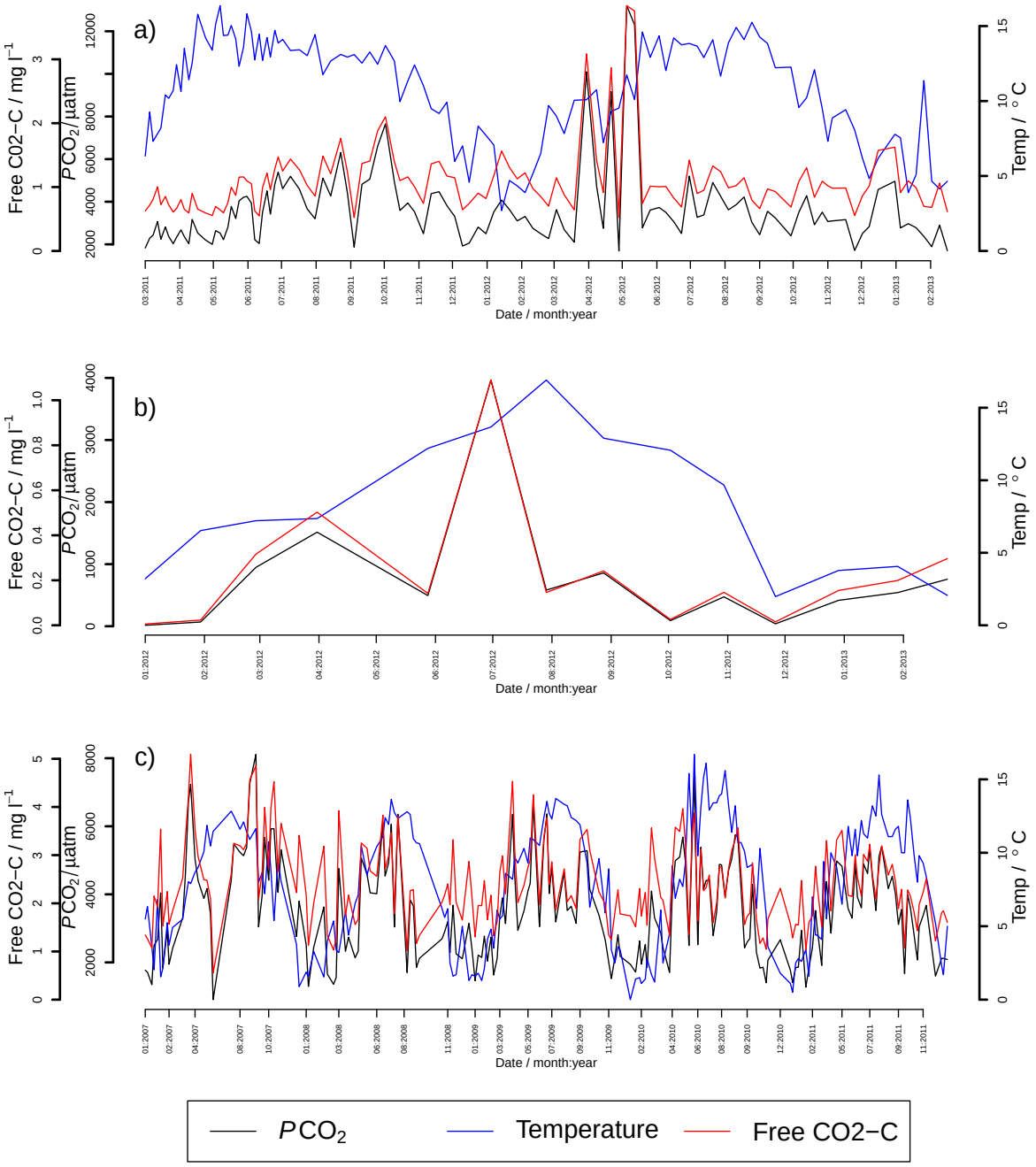
Figure 1: Seasonal variations in stream water PCO_2 (µ atm) , temperature (° C) and free CO_2-C concentrations (mg L^{-1}) for three headwater catchments: a) Blackwater, b) Pow and c) Black burn