Challenges in scaling up greenhouse gas fluxes: experience from the UK Greenhouse Gas Emissions and Feedbacks Programme

Peter Levy[1], Robert Clement[3], Nick Cowan[1], Ben Keane[2], Vasilis Myrgiotis[3], Marcel van Oijen[1], T. Luke Smallman[3], Sylvia Toet[2], and Mathew Williams[3]

[1]Centre of Ecology and Hydrology, Bush Estate, Penicuik, EH26 0QB, UK
[2]University of York, York
[3]University of Edinburgh, School of GeoSciences and NCEO, Mayfield Road, Edinburgh, UK

Key Points:

- We reviewed some of the challenges in accurately estimating fluxes of GHGs at national scale.
- Uncertainty arises from imperfectly-known models, parameters and inputs used in the extrapolation.
- Bayesian principles allow us to quantify this whilst combining information sources in a coherent way.

Corresponding author: Peter Levy, plevy@ceh.ac.uk

Abstract

The role of greenhouse gases (GHGs) in global climate change is now well recognised and there is a clear need to measure emissions and verify the efficacy of mitigation measures. To this end, reliable estimates are needed of the GHG balance at national scale and over long time periods, but these estimates are difficult to make accurately. Because measurement techniques are generally restricted to relatively small spatial and temporal scales, there is a fundamental problem in translating these into long-term estimates on a regional scale. The key challenge lies in spatial and temporal upscaling of short-term, point observations to estimate large-scale annual totals, and quantifying the uncertainty associated with this upscaling. Here, we review some approaches to this problem, and synthesise the work in the recent UK Greenhouse Gas Emissions and Feedbacks Programme, which was designed to identify and address these challenges.

Approaches to the scaling problem included: instrumentation developments which mean that near-continuous data sets can be produced with larger spatial coverage; geostatistical methods which address the problem of extrapolating to larger domains, using spatial information in the data; more rigorous statistical methods which characterise the uncertainty in extrapolating to longer time scales; analytical approaches to estimating model aggregation error; enhanced estimates of C flux measurement error; and novel uses of remote sensing data to calibrate process models for generating probabilistic regional C flux estimates.

Plain Language Summary

Greenhouse gases cause climate change, and we need to know how much is emitted each year across the globe. As well as coming from burning fossil fuels, plants and soil also take up and emit these gases, and we need to be able to quantify this in order to understand how best to tackle climate change. However, we can only measure these emissions over very small areas, at only a few locations, and for relatively short periods of time. Extrapolating from these measurements to a whole country introduces several uncertainties which are often largely ignored. Here, we examine progress in tackling this problem, and focus on better statistical methods to properly identify and account for the errors that are introduced by the large change in scale. Another is the development of instrumentation which can measure the gas emissions over larger scales and run continuously. Earth observation from satellites provides a promising source of data for the future, but cannot yet provide direct measurements of gas emissions. The Bayesian approach to modelling provides us with a coherent method for combining data from different sources, accounting for their uncertainties, and propagating this through to the uncertainties associated with predictions of national scale fluxes.

1 Introduction

The role of greenhouse gases (GHGs) in causing global climate change is now well recognised (IPCC, 2013). Emissions of GHGs from terrestrial ecosystems play an important part in this, and the potential for feedbacks within the climate system which amplify the emissions of GHGs from natural ecosystems is substantial. Accurate estimates are therefore needed of the GHG balance of the land surface at regional and national scales, and over long time periods, if we are to understand the key driver of global change. Because of the large scale involved, in relation to the scale at which we can make observations, this presents a major challenge which spans the domains of biogeochemistry, ecology, remote sensing, and atmospheric science.

For directly measuring GHG fluxes, we have two approaches available, based on either enclosing a small area within a chamber and monitoring the change in GHG concentration, or based on micrometeorological measurements of GHG concentration and

turbulence in the near the surface (see section below). However, both of these operate at scales much smaller than the spatial scale of interest - that of a region, nation or the whole globe. This means that we need to use a model to predict the large-scale flux. The fundamental upscaling issue is that we are forced to rely on predictions from a model which cannot be parameterised or tested at the true scale of interest. If we introduce some generic notation, we can consider this as three inter-related problems[1]. We need to predict y, the large-scale GHG flux, based on parameters θ derived at a small scale and input variables x estimated over the large-scale domain:

$$y = f(\theta, x). \tag{1}$$

Firstly the parameters θ are only inferred from a very small subset of the conditions prevailing over the whole domain. Because of basic sampling error, there is uncertainty in the parameter estimates. Secondly, there is also usually considerable uncertainty in the values of the inputs x over the large-scale domain. GHG fluxes from an ecosystem depend upon such things as incident radiation, leaf area index, soil aerobic status, soil microbial populations, and time elapsed since disturbance events. None of these is easily measured over a wide region, and we inevitably rely on some proxy or modelled estimate, and this introduces uncertainty in the true value of x over the whole domain. Thirdly, the model f is commonly non-linear, which complicates the upscaling procedure. The goal for science in this field is to quantify and reduce the uncertainty associated with parameters θ, input variables x, and model f in making the jump between the small scale of measurement and the large scale of prediction, which we illustrate in Figure 1. Progress here is necessary if we are to estimate the large-scale GHG balance accurately (Leip et al., 2018), and to demonstrate the efficacy of mitigation policies (Gifford, 1994; Smith & Smith, 2004; Smith et al., 2008).

In this paper, we review the challenges in upscaling small-scale GHG flux measurements to produce national-scale estimates. The UK recently developed a novel, multidisciplinary programme to identify and tackle some of these challenges, and our examples come from this programme. The overall aim was to improve the quantification of uncertainty where it arises in the upscaling process, and to reduce this uncertainty by improvements to instrumentation, measurement methods or modelling procedures. Specifically we focused on five challenges which focus on components of the problem, illustrated in Figure 1:

1. **Quantifying uncertainty in spatial upscaling of chamber fluxes to field scale.** Chamber measurements sample only a very small area, even in relation to a single agricultural field. The challenge is to quantify the mean and uncertainty in the estimate of the field-scale mean flux.
2. **Quantifying uncertainty in temporal upscaling of chamber fluxes to annual scale.** Similarly, chamber measurements typically sample only during a very few hours, in relation to the total flux over a year. The challenge here is to quantify the annual cumulative emission and its uncertainty, based on a sparse and spatially variable sample set.
3. **Reducing uncertainty in spatial and temporal upscaling of chamber fluxes via improved instrumentation.** An alternative approach to both of the above

[1] In addition to these, new phenomena may arise at the larger scale because of feedbacks in the system, which are not apparent at the small scale. For example, evapotranspiration from an individual leaf is strongly controlled by the stomatal conductance. However, because of the effect of regional-scale evapotranspiration on the vapour pressure deficit of the air in the boundary layer, regional-scale evapotranspiration is more strongly controlled by radiation input. This is a serious issue with water vapour fluxes, but less so for GHG fluxes themselves, because the magnitude of such feedbacks is much smaller.

is to developed new measurement systems which can provide better spatial and temporal coverage.

4. **Quantifying uncertainty in eddy covariance measurements of field scale fluxes.** Eddy covariance systems are expensive and complex to run, so are rarely operated with any replication. It is therefore usually very difficulty to estimate the systematic and random errors associated with these measurements. Here we evaluate five co-located eddy flux systems to determine the measurement error on net exchanges of CO_2 at both instantaneous times and daily scales.
5. **Quantifying aggregation error in spatial upscaling.** When non-linear models are parameterised at a small scale, but applied at a larger scale, the results will generally be in error wherever small-scale heterogeneity is not accounted for. A further challenge is to estimate and account for this kind of error. We evaluated this in the context of national-scale GHG flux estimates in the UK.

After outlining the basic approach to measuring GHG fluxes, we address each of these challenges in turn, with reference to specific analyses for a range of GHGs. For clarity, each section provides its own methods, results and discussion. We then conclude with a synthesis of the findings from the individual studies. Our synthesis allows us to assess advantages and limitations of current research, and to make suggestions for the development of new studies and approaches necessary to make better inferences about GHG fluxes at regional, national and global scales.

2 Measurement Methods for GHG fluxes

For directly observing GHG fluxes, we have two broad techniques available: chamber-based and micrometeorological. In the former, part of the plant or soil surface is enclosed in a gas-tight chamber, and the flux is inferred from measurements of the mixing ratio. In the case of static (non-steady-state) chambers the mixing ratio is measured on a sequence of gas samples extracted from the chamber over a short time period. From mass balance, the mixing ratio within the chamber is predicted to follow:

$$\chi = \chi_0 + \frac{F}{h\rho}\mathrm{d}t \tag{2}$$

where χ_0 is the initial mixing ratio of a GHG, h is the height of the chamber, ρ is the molar density of dry air, and $\mathrm{d}t$ is the time increment since enclosure. We thus have an inverse problem, which can be rearranged to estimate the flux as:

$$F = \frac{\mathrm{d}\chi}{\mathrm{d}t_0}h\rho \tag{3}$$

where $\mathrm{d}\chi/\mathrm{d}t_0$ is the initial rate of change in the mixing ratio. As an approximation, we can assume linearity in $\mathrm{d}\chi/\mathrm{d}t$, and solve for $\mathrm{d}\chi/\mathrm{d}t_0$ using linear regression. If we account for the non-linearity of diffusion into the chamber, we have to apply non-linear regression, optimisation methods, and potentially, complex 2-D diffusion models (Livingston et al., 2006; Pedersen et al., 2010; Sahoo & Mayya, 2010; Levy et al., 2011). Because part of the ecosystem has to be physically enclosed, the spatial scale of these measurements is necessarily restricted, typically to 0.1 m^2 and rarely more than 1 m^2. Similarly, the temporal scale of measurements is restricted because the physical enclosure changes the environment within - the emitted gas concentrations build up, and the effect of wind and rain is removed.

Micrometeorological techniques make use of measurements in the atmosphere near the surface. Historically, these were based on measuring the gradients in wind and GHG mixing ratios and making some assumptions about the turbulent transport. With the advent of fast-response infra-red analysers for CO_2, and more recently for CH_4 and N_2O based on QCL or CRD laser absorption spectroscopy, the eddy covariance method has

become the default approach (Kroon et al., 2010; Mammarella et al., 2010; Haszpra et al., 2018). If we can assume stationarity and horizontal homogeneity, it follows from mass balance that we can equate the surface flux to the eddy covariance term i.e.:

$$F = \overline{w'\chi'}\overline{\rho} \tag{4}$$

where w' and χ' represent the instantaneous deviations from the means. To measure this term accurately, we need high frequency (10-20 Hz) measurements of the vertical windspeed w and χ at an appropriate height above the surface. Various corrections are required to account for the frequency response of the measurement system, non-zero vertical windspeed, deviations from stationarity, and density fluctuations (Lee et al., 2006; Aubinet et al., 2012). The advantage of the approach is that it measures the integrated surface flux over an area much larger than a chamber, typically several hundred square metres (of the order of a small agricultural field), and can run near-continuously.

3 Spatial upscaling of chamber fluxes to field scale

As described above, chambers used to measure gas fluxes typically have small dimensions (< 1 m^2), several orders of magnitude smaller than the domains, such as agricultural fields, that we want to make inferences about, so the potential for sampling error is large. That is, the naïve sample mean of the chambers may deviate substantially from the true mean for the field. We want to improve this estimate, and quantify the associated uncertainty in extrapolating the field mean. This is a common problem in the area of geostatistics, where observations are only available at point locations, but predictions are required over a larger spatial domain. The classical geostatistical approach to this problem is to represent the spatial domain as a grid of discrete cells (a "raster"), and to use kriging to predict the values at all the unobserved locations in this grid. Kriging and its terminology originated in the mining industry, but is now a widespread and generally applicable technique for extrapolation problems. In essence, it is a form of weighted local averaging, where the estimates of values at unrecorded places are weighted averages of the observations. The kriging weights are calculated on the basis of the semivariogram, which quantifies the form of the increasing variance between pairs of points as the distance between them increases. Graphically, this shows the scale at which values are highly correlated, and how this changes with spatial scale. Prediction at a new location is based on all the observations, each weighted according to the degree of correlation at that distance predicted by the semivariogram. Kriging has been shown to be optimal in the sense that it provides estimates with minimum variance and without bias (in the long-term statistical sense). It is also often described as providing estimates of known variance, but this is only true if the form of the semivariogram is known with certainty; in real-world applications, this is never the case. Here, we use kriging to extrapolate chamber fluxes to the field scale, but for the purposes of characterising the uncertainty correctly, we apply it in a Bayesian framework. In brief, this means we account for the uncertainty in the variogram model, and represent each of the parameters as a probability distribution. Rather than assuming the variance is known, we calculate the posterior distribution of the parameters, given the observed data, and sample many realisations of these to represent the uncertainty.

We can attempt to test the success of this upscaling method because we can also measure at the field scale using eddy covariance. However, eddy covariance also does not directly give the field-scale domain mean, as its spatial sampling characteristics are affected by wind speed, wind direction, sensible heat flux and friction velocity: the so-called "flux footprint" (Schuepp et al., 1990, Schmid and Oke (1990), Leclerc and Foken (2014)). The footprint defines the relative contribution of each element of the surface area to the measured vertical flux, according to the advection-diffusion equation. This acts as a weighting function, such that some areas contribute strongly to the measured flux, and others not at all. If the mean flux F of a scalar over a landscape represented by a discretised gridded domain with dimensions n_x by n_y at time t is given by:

$$\bar{F}_t = \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} F_{xyt} \frac{1}{n_x n_y} \tag{5}$$

where the overbar denotes spatial averaging, eddy covariance effectively measures a weighted mean, where the footprint provides the set of weights, ϕ, to give:

$$\widehat{\bar{F}}_t = \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} F_{xyt} \phi_{xyt} \tag{6}$$

where we use the $\widehat{\text{hat}}$-symbol to indicate that this is an estimator, not $\bar{F}_t$ itself.

So the appropriate way to upscale chamber measurements is to use Bayesian kriging to estimate F_{xyt} over the whole grid, and thereby the domain mean $\bar{F}_t$. However, to compared with eddy covariance, we need to apply the footprint weighting ϕ to estimate the flux that we would expect eddy covariance to measure, $\widehat{\bar{F}}_t$.

Here, we applied the method to data from an arable field of oilseed rape in Lincolnshire, U.K., where fluxes of N_2O were measured by chambers and eddy covariance in the same field (Keane et al., 2017). The Bayesian kriging method provides a way of scaling chamber flux to the field scale, incorporating spatial pattern and its associated uncertainty. This prediction is weighted by the flux footprint to produce our expectation of the flux measured by eddy covariance. This allowed us to compare one method with another, accounting for the difference in spatial sampling characteristics inherent in the two methods as best we can, and account for the associated uncertainty properly.

Figure 2 shows the chambers fluxes of N_2O upscaled by Bayesian kriging and weighted by the footprint probabilities ϕ, to give a valid comparison against the eddy covariance data. The results show that the upscaled values can deviate substantially from the naïve sample mean of the chambers. More importantly, the uncertainty in the mean estimated by Bayesian kriging is substantially larger than the conventional 95 % CI in most cases. This is a source of uncertainty that is typically ignored, and this demonstrates the importance of representing spatial upscaling effects explicitly. This also indicates that we need to be careful in drawing conclusions from such measurements when interpreting the difference in means among experimental treatment plots. The magnitude of error will depend on the spatial pattern in the surface flux; in extreme situations, the sample mean could be quite poorly representative of the large-scale mean. In these cases, spatial upscaling clearly needs to be considered explicitly, and the Bayesian kriging described here provides a rigorous method to do this.

When comparing chambers and eddy covariance measurements, the difference in spatial sampling is usually ignored, and the arithmetic mean of chamber flux samples is compared with the value from eddy covariance for the corresponding time period. This ignore the fact that the flux footprint acts as moving spatial filter, whereby the location and extent of the area that influence the measured flux changes each half-hour, according to wind speed, direction etc. The method described here also provides a rigorous way to compare chamber measurements so that they represent the same area that is sampled by eddy covariance.

4 Temporal upscaling of chamber fluxes to annual scale

There are considerable challenges in interpolating and extrapolating cumulative N_2O fluxes, based on relatively sparse and variable measurements from only a few time points. This leads to substantial uncertainty in the "emission factor" (EF or Ω, the total N_2O released as a percentage of the fertiliser nitrogen added) which is used in the national

inventory. This is analogous to the problem of estimating the spatial mean from measurements at a limited number of locations, but in the time domain. The method most commonly used in the literature to calculate cumulative N_2O fluxes is to interpolate and integrate using trapezoidal rule integration. However, this method is very sensitive to noise in the data, there is no straightforward way to quantify the uncertainty introduced or to extrapolate beyond the sample data, and it does not account for the typically log-normal spatial distribution of fluxes. Here, we used two approaches to examine spatial and temporal upscaling: firstly using a process-based model directly, and secondly using Bayesian emulation of this model.

4.1 DNDC model

DNDC is a process-based biogeochemical model, widely used to estimate agricultural soil N_2O emissions. We parameterised the model using chamber measurements of N_2O fluxes, along with data on harvest crop grain N content, soil mineral N and soil moisture, from four experimental sites between 2010 and 2012, accounting for time lags between measured and simulated time-series (Myrgiotis et al., 2016).

The model was applied across a 3800 km^2 area of Scotland where >90 of its croplands are located. Spatial data on soil properties, crop coverage and weather, and UK-specific crop calendars and fertiliser-use recommendations were used to create model inputs for 2011-2013 at a 1 km^2 resolution (Myrgiotis et al., 2018). The distribution of Ω estimated from the regional simulations was compared against the site data of measurements-based Ω to evaluate the effect of the upscaling process.

Measured and simulated distributions of Ω have similar range and shape (Figure 3), but there were important differences in their distributions, e.g. their inter-quartile range (upper/lower dotted lines in Figure 3). The wider inter-quartile range of the measured Ω is a result of greater variability in weather across the UK field sites and measurement period compared to the conditions over the Scottish arable region for the upscaling period. There were higher temperatures and precipitation at some of the field sites responsible for peaks in Ω during experimental measurements. The mean of the simulated Ω is 0.47%, 14 % less the mean measured Ω of 0.55%. The inter-quartile range of the simulated Ω was around half that found in the measurements. We conclude that upscaling causes substantive changes in Ω linked to the different range of conditions encountered across the wider region.

4.2 Meta-model of N_2O fluxes

To address this issue, we developed an emulator of the DNDC model with which we could apply Bayesian calibration to characterise the uncertainties in the spatial and temporal distribution of emissions. Following a fertilisation event, the time course of N_2O flux is expected to rise to a peak, then decay exponentially. This pattern in time is reproduced by DNDC and similar models, and is well described very simply by the log-normal equation:

$$\mu_t = \frac{1}{\sqrt{2\pi}kt} e^{-(\log(t)-\Delta)^2/2k^2} N_{\text{in}}\Omega \tag{7}$$

where μ_t is the spatial mean of the N_2O flux at time t, Δ and k are analogues for the location and scale parameters, N_{in} is the nitrogen input, and Ω is the fraction of this nitrogen which is released as N_2O. Because the lognormal function integrates to unity at $t = \infty$, Ω is implicitly based on the total cumulative emission, rather than at an arbitrarily defined time. The symbol Δ can be interpreted as the natural logarithm of the delay between fertiliser application and peak flux; k is a decay rate term. This equation

provides a simple meta-model which can be used to emulate the behaviour of DNDC (and similar models).

Because μ_t typically has a very skewed spatial distribution, there is a high probability of the sample means underestimating the true value; the problem increases as variance increases and sample size decreases. Several approaches have been proposed as more efficient estimators of the location and scale of lognormal distributions, but none of these entirely solve the problem when σ is large and n is small, as is generally the case with flux measurements. In this study, we applied a Bayesian approach, using the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling (Gelman et al., 2013). In this way, we estimated the parameters of the underlying distribution.

So, at time t following fertilisation, the mean flux is given by Equation 7, at which time the N_2O flux has a distribution

$$\begin{aligned} F &\sim \ln\mathcal{N}(\mu_{\log,t}, \sigma^2_{\log}) \\ \mu_{\log,t} &= \log(\mu_t) - 0.5\sigma^2_{\log} \end{aligned} \tag{8}$$

To obtain the cumulative flux at time t, we use the standard lognormal cumulative distribution function

$$F_{\mathrm{cum},t} = \Phi\left(\frac{\ln t - \Delta}{k}\right) N_{\mathrm{in}}\Omega \tag{9}$$

where Φ is the cumulative distribution function of the standard Normal distribution. The model was encoded in the JAGS language, and fitted to a number of data sets from across the UK.

Priors for Δ and k were specified as Normal distributions based on the temporal patterns produced by the DNDC model (see below). A Normal distribution was also assigned to $\sigma_{\log}$, based on earlier data from various sites in the UK, mainly from Cowan et al. (2014, 2016). The prior distribution for $\sigma_{\log}$ was truncated at zero to exclude negative values. Ω was given a lognormal distribution, fitted to the data collation of Stehfest and Bouwman (2006) which included data on emission factors from all over the world.

Figure 4 shows the posterior distribution of cumulative fluxes calculated from the UK data sets, expressed as the emission factor, Ω. This distribution is very narrowly defined in some cases (e.g. Dum 2012-10-16, EBS 2009-03-17), and very wide in other cases (e.g. EBS 2007-05-16, EBS 2008-06-18), meaning that uncertainty in the emission factor can be very small or very large. Ω is generally in the range zero to 5 %, but some events have substantially higher emission factors. The value estimated by the trapezoidal method is generally within the posterior distribution of the lognormal model, but the values are usually lower. The emission factor is generally rather poorly constrained by flux chamber measurements, because of the difficulties of accurately estimating the mean of a lognormal distribution with large variance when n is small. The standard approach fails to capture this uncertainty. Our new approach performs well in that it appropriately quantifies the uncertainty, and removes some of the bias by accounting explicitly for the lognormal distribution.

5 Reducing uncertainty in large-scale fluxes via improved instrumentation

The direct measurement of GHG fluxes has developed in tandem with the instrument technologies that allow GHG mixing ratios to be measured. Compared with CO_2, N_2O is less amenable to measurement by infra-red absorption: it is present at lower background concentrations; the typical fluxes are smaller relative to the background concentrations; and the infra-red absorption bands are narrower, making the technicalities of

measurement more difficult. Until recently, N_2O was only accurately measurable by gas chromatography. Hence, for the most part, observations of N_2O fluxes are only available using static chamber methods (Hutchinson & Mosier, 1981; Matson & Harriss, 2009), which necessarily sample small areas (typically < 0.1 m2) over short time periods, based on few (2-4) points. However, N_2O fluxes show a wide variability, ranging over orders of magnitudes on small spatial scales, which is not predictable. This is attributed to two main causes: the variety of unobserved microbial controls on gas production, including their physiological activity and population dynamics; and the sensitivity of physical transport in the soil, which interacts with the measurement process (enclosure) in a complex way (Xu et al., 2006; Sahoo & Mayya, 2010). Fast-response sensors for N_2O have recently become available, making high-precision chamber measurements (Cowan et al., 2014) and the micrometeorological eddy covariance method (Kroon et al., 2010) feasible. The challenge here is to use these new sensors to develop continuous measurement systems for fluxes of N_2O. With continuous measurements, we overcome the need to interpolate and extrapolate in space and time, and thereby remove the large uncertainties this introduces. In the GHGEF programme, we developed and applied two systems which provide near-continuous measurements of N_2O flux: a robotic auto-chamber system using a cavity ring-down spectroscopic instrument ("SkyLine"), and an eddy covariance system based on quantum cascade laser (QCL) spectroscopy.

5.0.1 SkyLine

A detailed description of the system is available in Keane et al. (2018). Briefly, the SkyLine2D automated chamber system used a single, cylindrical chamber (internal diameter 40 cm, height 62 cm), suspended from a motorized trolley mounted on parallel horizontal ropes held above the crop by 2.5-m tall aluminium trellis arches (Figure 5). The trolley repeatedly traversed a transect across the crop of up to 40 m, enabling measurements at high spatial resolution (< 1 m) across replicated manipulations or underlying variation in the landscape. At designated locations where collars were placed in the soil, the chamber automatically lowered and sealed on the collar to conduct a flux measurement. The base of the chamber was fitted with a rubber gasket which formed a gas-tight seal when dropped on the flange of the landing base. Guides around the chamber bases ensured the chamber landed accurately. A vent was fitted into the chamber to minimize pressure differences between the chamber and the external atmosphere (after Xu et al. 2006). The chamber operated as a non-steady state dynamic system, with headspace gas being circulated between the chamber and a cavity ring-down spectroscopic (CRDS) analyser for N_2O (LGR isotopic N_2O analyser, Los Gatos Research, CA, USA) housed in an enclosed shed at one end of the SkyLine2D apparatus (Figure 5). The CRDS analyser operated at 1 Hz, giving a precise measurement of the rise in mixing ratio within the chamber, allowing a relatively short chamber closure (ca. 5 min), thus minimising the time during which the underlying plants and soil are isolated from ambient conditions.

5.0.2 QCL eddy covariance system

The system used a continuous wave quantum cascade laser (QCL) absorption spectrometer (CW-QC-TILDAS-76-CS, Aerodyne Research Inc., Billerica, MA, USA), with an ultra-sonic anemometer (WindMaster Pro 3-axis, Gill, Lymington, UK) to measure fluctuations in 3-D wind components at a frequency of 20 Hz. The QCL was fitted with a laser capable of measuring N_2O with a precision of 0.3 ppb, together with H2O and either CO_2 or CO, using absorption features at 12 and 22 m^{-1}. Internal software fits the observed spectra to a template of known spectral line profiles from the HITRAN (HIgh-resolution TRANsmission) molecular spectroscopic database. Absolute gas concentrations can then be calculated from the strength of the absorption line measured, the temperature, pressure and path length. A vacuum pump (Triscroll 600, Agilent Technolo-

gies, US) was used to draw air through the inlet and instrument with a flow rate of approximately 14 l min^{-1}. Data from the sonic anemometer and QCL was logged in tandem using a custom program written in LabView (National Instruments, TX, USA).

Fluxes were calculated over 30 minute intervals using the EddyPro software (Version 6.2.1, Li-COR, Lincoln, NE, U.S.A.), based on the covariance between gas concentration and vertical wind speed. For flux data taken with a low signal-to-noise ratio, timelag identification by maximisation of the cross covariance introduces systematic biases (Langford et al., 2015). Here, we investigated methods for estimating the timelag, based on either maximising the timelag over a longer time window, or using the timelag established for CO_2. Both CO_2 and N_2O share the same path in the sample line to the measurement cell, and would be expected to travel at the same rate. Fluxes of CO_2 are an order of magnitude larger, so the timelag which gives the maximum covariance is usually clearly defined within each half-hour, and should be equally applicable to N_2O, and not subject to the systematic error described by Langford et al. (2015). Depending which laser was installed in the QCL, measurements of CO_2 were not always available, so the method using N_2O only was also used. After investigating options, we decided on a six-hour window in which we found the timelag which maximised covariance. This timelag was then fixed for all data within the six-hour window, and fluxes were calculated on a 30-minute basis. Standard corrections were made in the flux calculation, following Moncrieff et al. (1997), Ibrom et al. (2007) and Burba (2013). Random uncertainty was estimated by the method of Finkelstein and Sims (2001).

Figure 6 shows a near-continuous time series of measurements of N_2O flux over a whole month by both automated SkyLine chamber and eddy covariance methods. Conventionally, only infrequent static chamber measurements (blue symbols) would be available. Given the variability in the data - irregular peaks in time and spatial variability between chambers - the near-continuous measurements provide a much more accurate estimate of the cumulative emission over the period following fertiliser application, and thereby the emission factor. Measurement techniques such as these, with better data coverage, are clearly needed to enable appropriate temporal upscaling, so that longer-term means and cumulative emissions can be estimated accurately from the observations.

6 Quantifying uncertainty in eddy covariance measurements of field scale fluxes

As described above, eddy covariance (EC) is a one of the key techniques for measuring GHG fluxes, but subject to instrument noise, uncertainties in the processing steps and micrometeorological conditions required to meet the underlying assumptions. The measurement error for eddy covariance is thus challenging to quantify. Due to their expense, EC systems are usually deployed singly, so there is a lack of replication – the most obvious means by which this can be estimated. By using two EC systems 800 m apart (Hollinger & Richardson, 2005) were able to quantify the measurement difference (error) in flux calculations. However, the tower separation was large enough so that footprint regions did not overlap, so the comparison is confounded. Hollinger and Richardson proposed a time–for–space substitution to allow error calculations from a single EC system, the so-called successive days method, and showed it has utility.

Here for the first time we present a comparison of NEE flux estimation from five co-located flux systems, all sampling the same pasture in southwest Scotland (Crichton Research Farm, Dumfries). The low stature of the vegetation allowed the EC systems to be set up within 10 m of each other, each sampling at 5 m above ground level. By having five systems the variance between these can be directly determined for the same conditions to assess instrumental and processing error. All systems used a Gill R3 sonic anemometer; 4 used a Vaisala GMP343 CO_2 sensor and a Honeywell HIH-4000 relative humidity sensor, while one used a LI-COR LI-7500 (Hill et al., 2017). Fluxes were calculated for 30-min periods using EdiRe (version 1.5.0.50). There were 13 days during

June 2015 with almost continuous and simultaneous measurements by all sensor systems. The pasture was cut on 13 May, and LAI was tracked using an LAI-2000 (LiCor Inc), growing from 3 to 4 during the study period from 5-17 June. The deviations of the 30 minute time series, and the daily time series from aggregated data were calculated, and their distributions tested for normality using a Shapiro-Wilk test (R software).

There was broadly a good agreement between all the flux systems in their estimates of NEE at 30 minute and daily time steps (7). The median standard deviation of the 30 minute data was 2 μmol $m^{-2}s^{-1}$. The 30-minute error distribution deviated significantly from normal (P<0.001), being clearly log-normal with a high tails (7). For the daily aggregate NEE estimates, the median value was 0.45 gC $m^{-2}d^{-1}$, and the error distribution was not significantly different from normal (Shapiro-Wilk test p=0.93). These results are close to the results of space-for-time error estimates at Harvard Forest, which reported a daily standard deviation of 0.58 gC $m^{-2}d^{-1}$ (Hill et al., 2012) for NEE.

The value of EC error estimates like these are clear. Errors determine the capacity of EC to detect C sources and sinks, which are driven by small differences between large input and output fluxes. Error data are also important to set the weighting of EC observations in the calibration of C process models. Bayesian calibration methods weigh the importance of observational data constraints in model fitting on the basis of measurement error. Bayesian calibrations propagate measurement uncertainty into model predictions, for example by finding an ensemble of model parameters that produce estimates of NEE consistent with observations and their uncertainty. Eddy covariance data have been used successfully across the UK to produce more robust calibration of C cycle models. The calibrated models then propagate the flux uncertainty from localised and incomplete data set at a few sites into complete and regional assessments of C cycling (Myrgiotis et al., 2020; Smallman et al., 2017; Revill et al., 2016).

7 Quantifying aggregation error in spatial upscaling

7.1 Analytical approach

Typically, we derive a model of the GHG flux based on small-scale measurements as a function of input variables x, measured at the corresponding scale. We want to apply this to a larger scale, usually using the mean value of x, averaged over a grid cell, region, or longer time period. However, the output of a model f at the larger scale, with spatially varying inputs x should be calculated as $\int_{-\infty}^{+\infty} f(x)\,p(x)\,dx$, where $p(x)$ is the probability distribution of input values in the region. We can denote this integral more briefly as $E[f(x)]$, using the expectation operator $E[]$. Whenever models are run with the averaged inputs, what is being calculated is $f(E[x])$ rather than $E[f(x)]$. If the models are nonlinear, this results in an upscaling error δ defined as:

$$\Delta = f(E[x]) - E[f(x)] \tag{10}$$

When models are applied to large regions, the domain is typically subdivided using a spatial grid, with each grid cell covering an area of tens or hundreds of square kilometres. In contrast, the models themselves tend to be based on observations made at much finer scales such as flux chambers, eddy covariance footprints, individual crop fields, or forest stands. The challenge here is in using a small-scale model to estimate GHG fluxes in a large grid cell, accounting for within-cell spatial heterogeneity, and the error this produces.

In the GHGEF programme, we derived a method for characterising this error, based on a multivariate Taylor-expansion approach, a development of earlier work in crop and forest modelling (Bresler & Dagan, 1988; Band et al., 1991; Rastetter et al., 1992). The approach is to estimate Δ and correct for it. For functions of one variable, we can use the formula derived by applying the expectation operator to the second-order Taylor expansion of $f(x)$:

$$E[f(x)] \approx f(E[x]) + \frac{1}{2} Var[x] f^{(2)}(E[x]), \tag{11}$$

where $f^{(2)}(E[x])$ is the second derivative of f evaluated at the mean of x, and Var$[x]$ is the variance of x within the region. Combining this formula with Eq. 10 gives us an approximate formula for the upscaling error of models with one input variable:

$$\widehat{\Delta} = -\frac{1}{2} Var[x] f^{(2)}(E[x]), \tag{12}$$

where we use the $\widehat{\ }$-symbol to indicate that the formula provides an estimator for Δ, not Δ itself. For functions of multiple input variables where x is a vector, multivariate Taylor expansion yields:

$$\widehat{\Delta} = -\frac{1}{2} \mathrm{tr}(S\,H), \tag{13}$$

where S is the variance-covariance matrix of x, H is the Hessian matrix of second order partial derivatives of $f(x)$, and tr denotes the trace (sum of diagonal elements) of the matrix product $S\,H$. For simple models, analysis shows that the formula is exact (i.e. $\widehat{\Delta} = \Delta$), allowing full correction of model upscaling errors. In other cases, the formula provides an approximation.

We demonstrated the application of this approach to example models (Van Oijen et al., 2017) of methane flux (Levy et al., 2012), ammonia and nitrous oxide flux (Flechard et al., 2007). Using high-resolution data, we could calculate the true model output $E[f(x)]$. After averaging the input data in 32-km grid cells, we re-evaluated the model using this, corresponding to $f(E[x])$, and by difference, we calculated the aggregation error Δ. Using the Taylor-expansion approach described above, we then calculated our approximation of this error $\widehat{\Delta}$, for verification against the known values of Δ. Depending on the model, aggregation error could be substantial, ranging from -3 to - 48 %. The error varied spatially (Figure 8), depending on the range of the input variables, their sub-grid variance, and the non-linearity of the model. The $\widehat{\Delta}$ formula gave reasonable approximations to the true error (Figure 8, right-hand panel), correcting model output to within 2 to -9 % of the true values.

For nonlinear models, the effects of spatial upscaling need to be accounted for, and the $\widehat{\Delta}$ formula described here is a generally applicable means to do this. The approach can be applied to more complex, high-dimensional process-based models, but exploration of the accuracy of the approximation is needed.

7.2 An approach using earth observation and model-data fusion

An alternative approach to quantify aggregation error is to use earth observation data (and other relevant mapped products) to provide inputs at increasingly fine scales. We conducted four model-data fusion analyses at differing spatial resolutions of Great Britain's (GB) terrestrial carbon cycle at a monthly time step for an 18 year period (2001-2018) using the CARbon DAta Model fraMework (CARDAMOM, Bloom et al., 2016). CARDAMOM uses a Bayesian approach within an Adaptive Proposal - Markov Chain Monte Carlo (AP-MHMCMC, Haario & Tamminen, 2001; Roberts & Rosenthal, 2009) to estimate location (i.e. pixel) specific ensembles of parameters for an intermediate complexity model of the terrestrial C-cycle (DALEC Bloom & Williams, 2015). The parameter ensembles are consistent with observational constraints, their uncertainties, model structure, meteorology and disturbance (fire & forest loss). From these parameter ensembles we are able directly estimate at pixel level the uncertainty of DALECs C-cycle simulation of terrestrial fluxes and stocks. CARDAMOM analyses were conducted at 4

spatial resolutions, 111 x 111 km, 56 x 56 km, 28 x 28 km and 5 x 5 km, all at monthly temporal resolution. Observational constraints assimilated by CARDAMOM are monthly time series information on leaf area index (LAI), a single estimate of above ground biomass (AGB) and soil C stocks. LAI is extracted from the 1 x 1 km, 8 day product from the Copernicus Service Information (2020). A single per-pixel estimate of AGB for 2017 and its uncertainty is drawn from ESA's 1 x 1 km CCI Biomass product (Santoro, 2021). From this AGB estimate we derive the total woody biomass (which corresponds with DALECs model structure) following Saatchi et al. (2011). A single per-pixel estimate of soil carbon stock is extracted from the SoilGrids database (Hengl et al., 2017). Meteorological drivers are drawn from the ERA5 reanalysis (Hersbach et al., 2020).

Our analyses all estimate that GB's terrestrial ecosystems were a net sink of carbon with a net biome exchange (NBE) of -6.7 to -10 TgC yr^{-1} (-0.32 to -0.41 MgC ha yr^{-1}) between 2001 and 2017 (Table 1). Spatial resolution has a substantial impact on the magnitude and spatial patterning estimated by the CARDAMOM analyses (Table 1, Figure 9). For example, at the coarsest resolution of the model grid (111 x 111 km) NBE is near neutral across much of GB except the far north east and south west. In contrast the highest resolution (5 x 5 km) shows fine scale variation across the whole of GB. Moreover, the range of C flux magnitudes estimates is smaller in the coarser resolution analyses due to aggregation of sub-grid variability. The distributions of pixel-level mean annual fluxes progressively converge towards that estimated by our finest spatial resolution analysis (Figures 9, 10).

The estimates of the gross biological fluxes (i.e. GPP and Reco) are relatively insensitive to spatial resolution between the 5 km, 28 km and 56 km analyses. The mean annual flux estimate of GPP and Reco in the 5 km, 28 km and 56 km resolution analyses vary by less than 0.2 MgC ha yr^{-1} ($< 2\%$; Table 1). Estimation of emissions due to disturbance, both fire and forest cover loss, are progressively underestimated at coarser spatial resolutions (Table 1) and show varied time series dynamics (Figure 10). The mean pixel-level forest loss estimates plateaus between 5 km and 28 km (Table 1). However, there remains disagreement in the temporal interannual variability and magnitude of forest loss estimates particularly after 2010 (Figure 10). Emissions due to fire have not converged in terms of mean annual emissions (Table 1), or temporal dynamics and magnitude (Figure 10). Overall, these results suggest that there remains substantial sub-grid scale disturbance information missing from these analyses, even at 5 km spatial resolution.

8 Conclusions

We reviewed some of the major challenges in accurately estimating net fluxes of GHGs at national scale. These revolve around the difficulties of extrapolating small-scale, short-term observations and models. Uncertainty necessarily arises if these are extrapolated to larger scales because of the imperfectly known models, parameters and inputs used in the extrapolation. Where spatial pattern exists in the small-scale observations, this can be accounted for using a geostatistical approach such as kriging; applying this in the Bayesian framework allows the uncertainty to be propagated correctly. To tackle this same problem in the time dimension, process-based models can be used, which bring in prior knowledge of how we expect the flux to change over time. Temporal upscaling of N_2O fluxes has been difficult because of the lack of high frequency measurements over whole growing seasons. Substantial random error and systematic bias can be introduced when extrapolating in time to estimate cumulative fluxes. Again, the Bayesian approach can be used to characterise the uncertainty in their application. Modelling the spatial and temporal pattern distribution in the observations in a Bayesian framework allows this uncertainty to be accounted for appropriately. Furthermore, new measurement techniques are now allowing near-continuous, long-term observations to be made of these GHG fluxes, which reduce the need for extrapolation. Models applied at large scales need to

account for sub-grid-scale heterogeneity; significant upscaling error can occur otherwise. This aggregation error can be approximated closely by a Taylor-expansion formulation, based only on the model Hessian matrix and the estimated variance-covariance matrix of the inputs.

With advances in computing power, running models of GHG fluxes over large domains at relatively high resolution is increasingly feasible. For example, we can now routinely run models over the UK at 1-km resolution (Robinson et al., 2016), requiring 240,000 grid cells. In moving from the 1 km^2 scale to national scale, the issues of model upscaling do not apply in the same way, so long as the model has been correctly scaled from the original observations to the 1-km^2 scale. This is because the variance over the whole domain is represented, albeit as 1-km^2 means, i.e. we evaluate E[f(x)], so $\Delta = 0$, so the total flux is simply the sum of all the grid cells. In moving between these scales, the main problem encountered is in estimating the input variables accurately over this domain. Because of the large size of the domain, our information is often incomplete or based on proxies (e.g. NDVI) Disney et al. (2016). This is an area where data availability is rapidly improving, with new and improved satellite products becoming available, as well as other new technologies - improved laser-based spectroscopic methods, and unmanned airborne vehicles.

A key point is simply to recognise that the scaling problem exists, and that some attempt to explicitly deal with it, however inadequate, is better than none, and gives a basis for improvement upon. The field of geostatistics has dealt with similar problems over several decades, but the techniques are not yet widely applied in biogeochemistry or process-based modelling. A common theme throughout is the adoption of Bayesian principles for combining multiple information sources in a coherent way, which allows us to make better use of the available data. We expect this trend to continue and expand into the machine learning domain. We can foresee further application of data assimilation methods which combine bottom-up models with atmospheric observations, that are thereby better constrained, potentially more powerful, and easier to interpret in terms of the national-scale inventories.

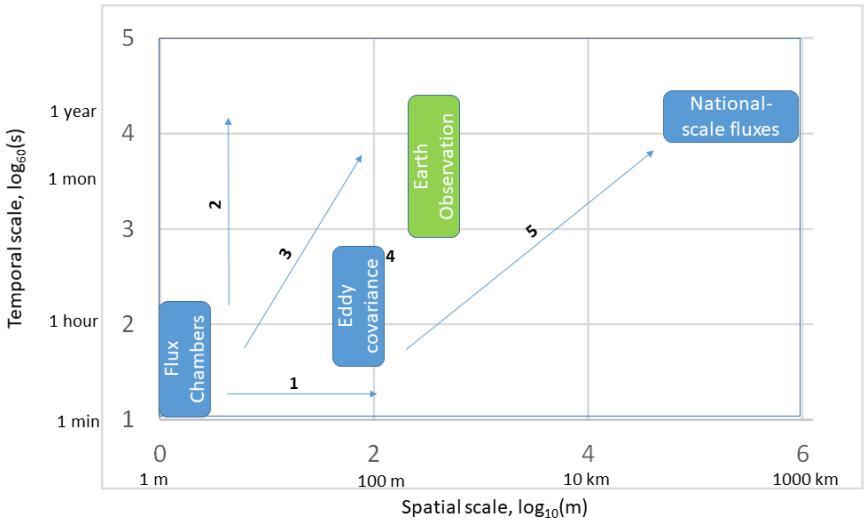

Figure 1. Diagram illustrating the spatial and temporal scales of the upscaling problem, from measurements using chambers and eddy covariance which cover small scales, to national-scale annual fluxes. The numbers respond to the challenges which we identify in tackling this problem. Earth observation represents a rapidly expanding source of data with wide coverage and increasingly fine resolution.

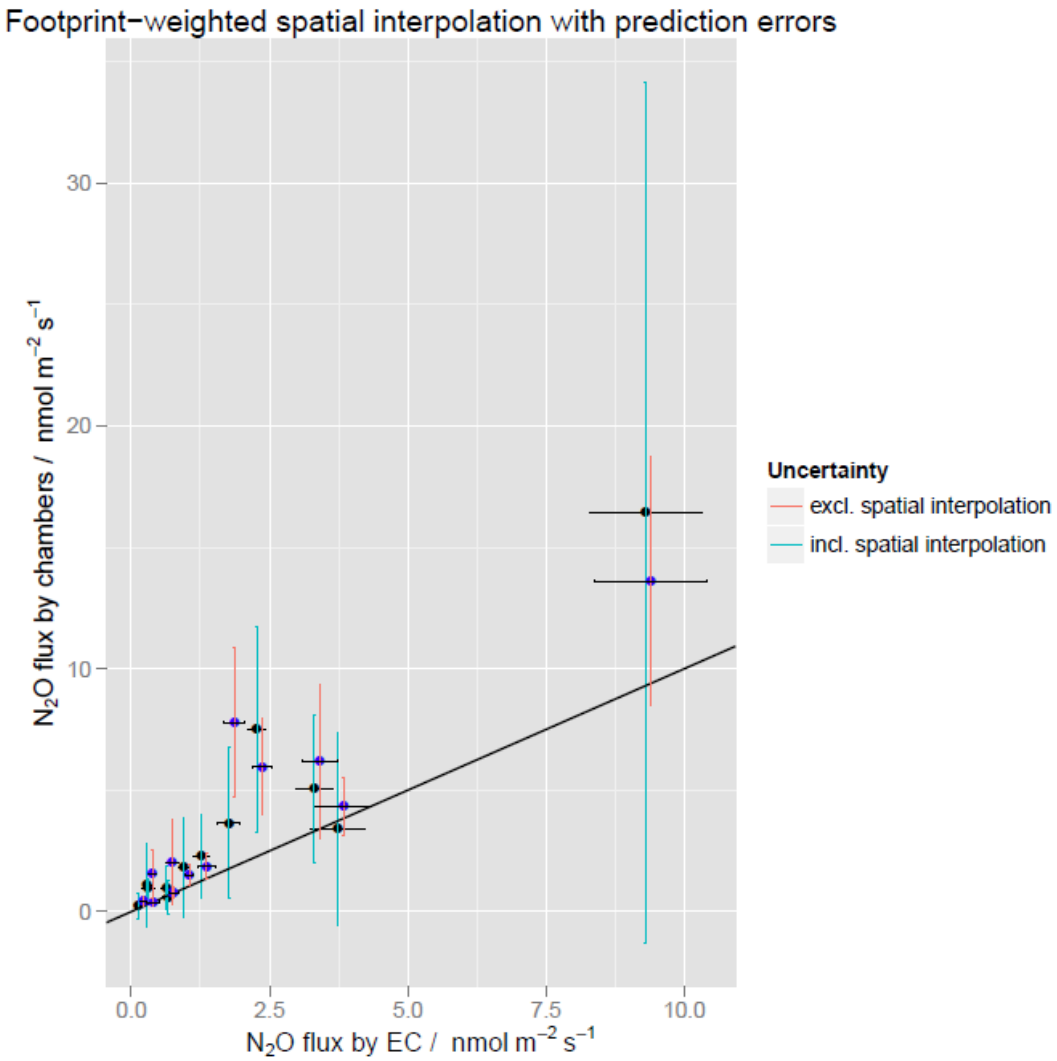

Figure 2. Comparison of fluxes of N_2O measured by chambers and eddy covariance for 12 sampling occasions. For each occasion, symbols show (i) the naïve sample mean of the chamber N_2O fluxes, with the conventional 95 % confidence intervals representing uncertainty (blue symbols, red vertical bars), and (ii) the maximum a posteriori mean flux of the chambers N_2O fluxes produced by Bayesian kriging, with the 2.5 and 97.5 percentiles of the posterior distribution representing uncertainty (black symbols, blue vertical bars). The latter are offset fractionally for visibility. The solid black line shows the 1:1 relationship.

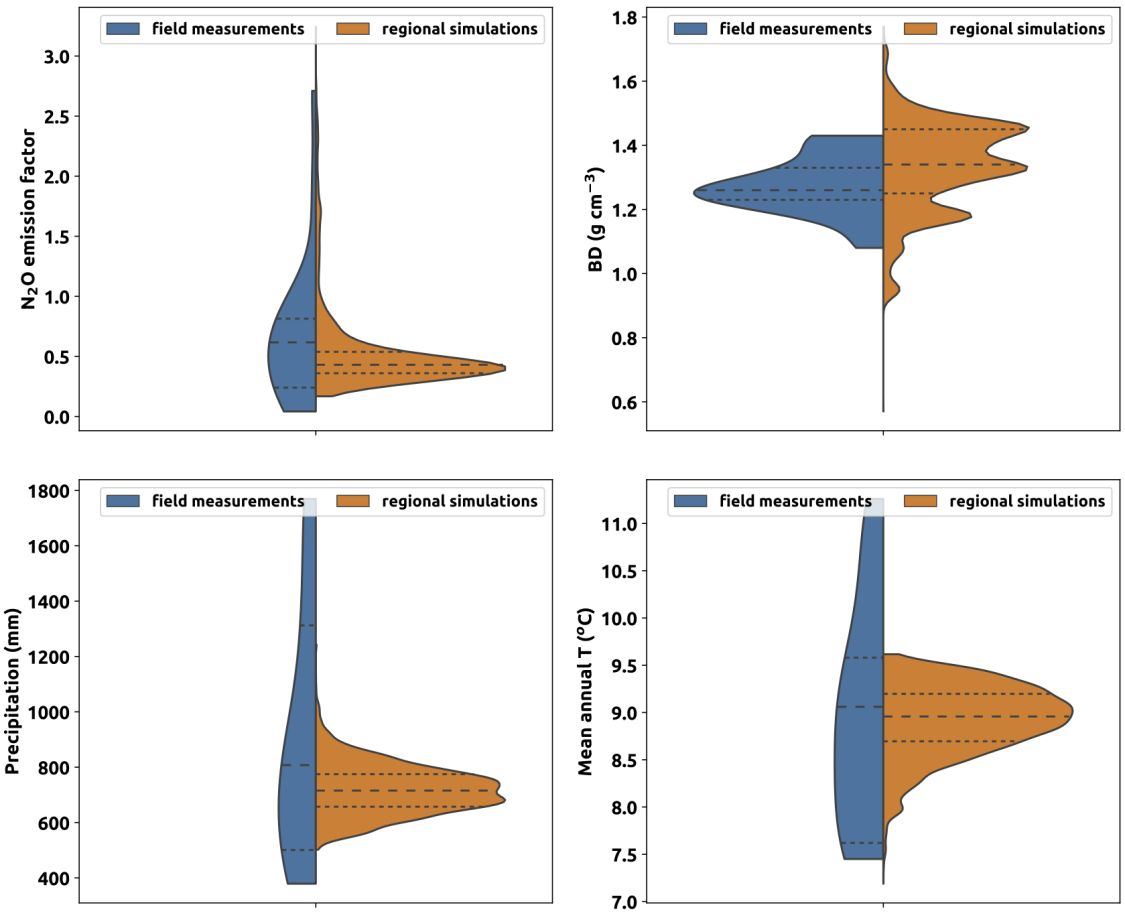

Figure 3. A comparison of distributions of emission factor Ω for N_2O generated from (i) field measurements collected across range of different management approaches, sites and seasons at locations in the UK and (ii) from upscaling estimates based on a model calibrated using the site level data, applied across eastern Scotland over several years. Also shown are the distributions of soil conditions (bulk density, BD) and weather conditions during measurement periods for all sites, and the conditions relevant to the modelled upscaling. The crops grown in the measured and simulated sites were winter wheat, winter barley, spring barley and winter oilseed rape.

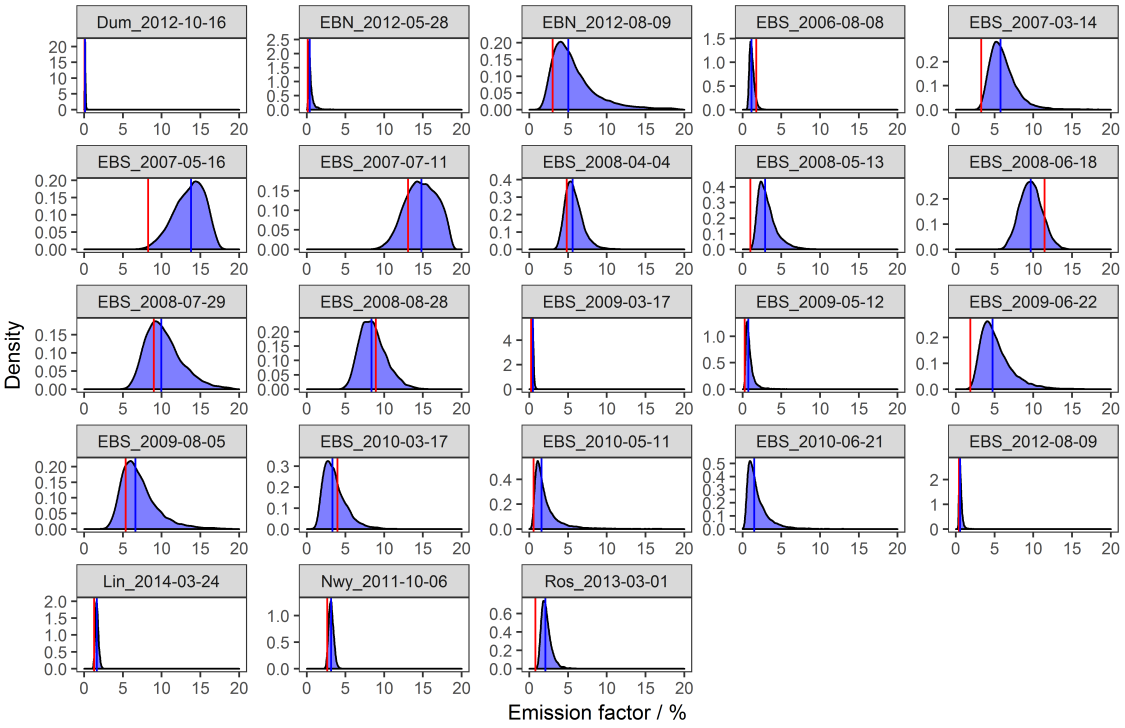

Figure 4. Posterior distribution of emission factor for each fertilizer application obtained by Bayesian estimation of the lognormal model. The blue vertical line shows the median of this distribution and the red vertical line shows the value calculated by the trapezoidal method. A lognormal distribution was used to provide the prior distribution of the emission factor.

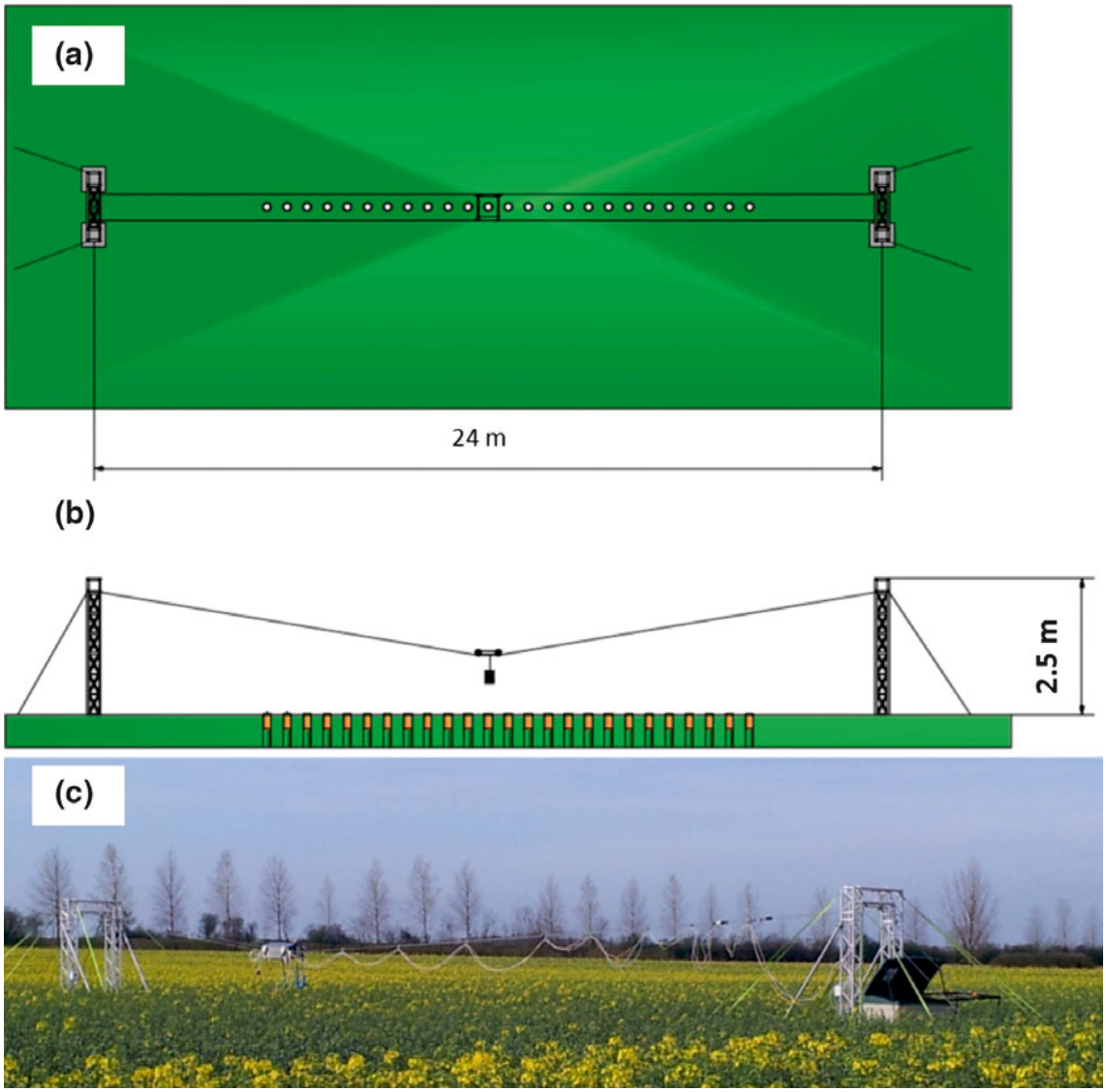

Figure 5. Aerial and side profile schematics of the SkyLine2D system showing (a), the trellis arch supports at either end, supporting the Kevlar ropes between. The motorized trolley is depicted at the mid-point of the two supports (b). Cross section of the in situ system at the OSR field site and (c) the N_2O and CH_4 Los Gatos cavity ring-down analysers were housed in the green garden box by the right-hand trellis support.

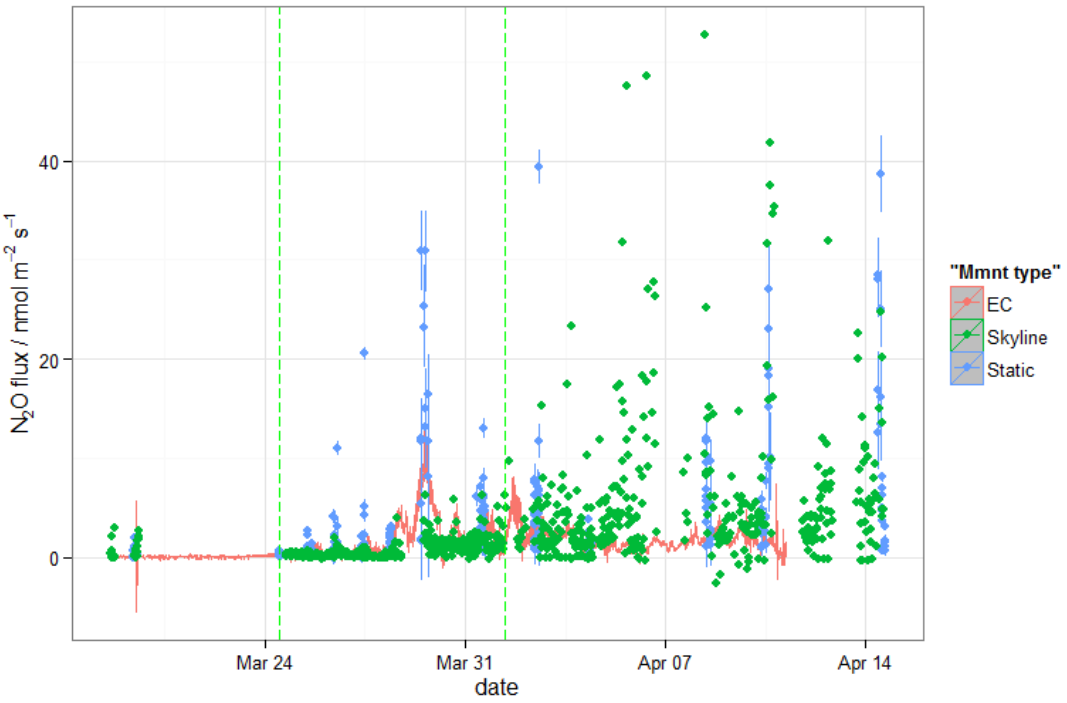

Figure 6. Time course of N_2O fluxes at the Lincolnshire field site, following fertilisation events (dotted vertical lines) as measured by eddy covariance (EC), static chambers, or the SkyLine2D system.

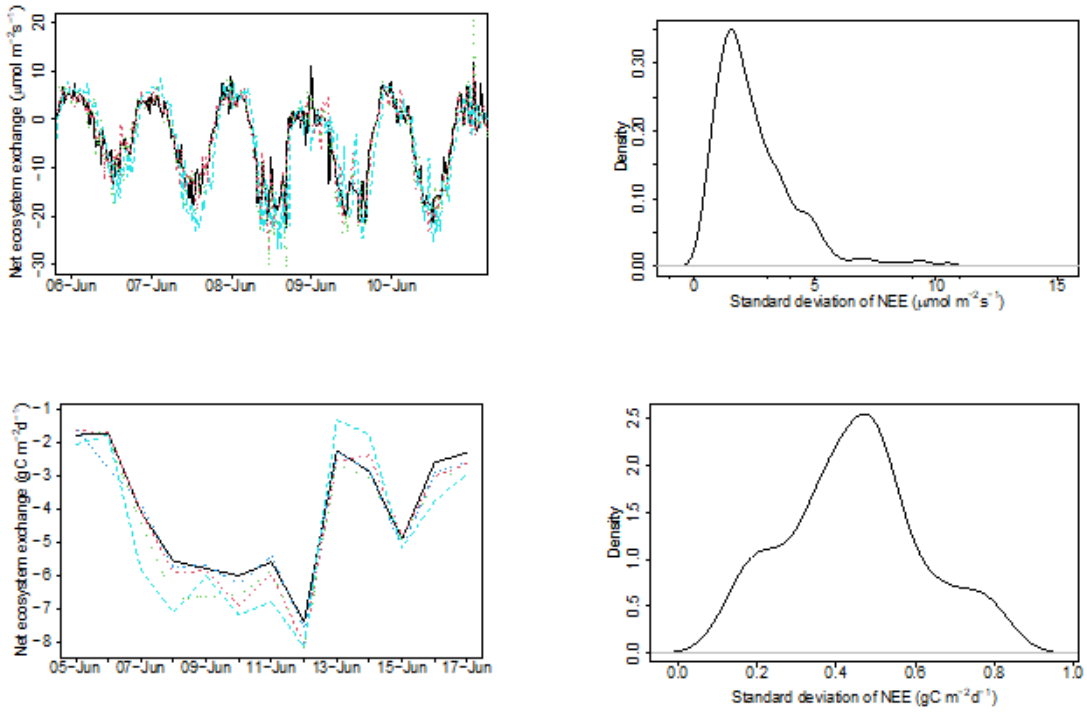

Figure 7. Comparison of five simultaneous eddy flux measurement systems (multiple time series during 2015 in left panels) and their variability (density of standard deviation across all systems for each time interval, right panels) over 30 minute (top panels) and daily (bottom panels) time integrals. The five eddy systems were all installed in the same field, sown with grass, and all within 10 m of each other. For 13 days there were continuous 30-minute time samples available for comparison (except 9-June, only 43 samples available for comparison) or for summing to produce a daily estimate. Only five days are shown in the top left panel for clarity.

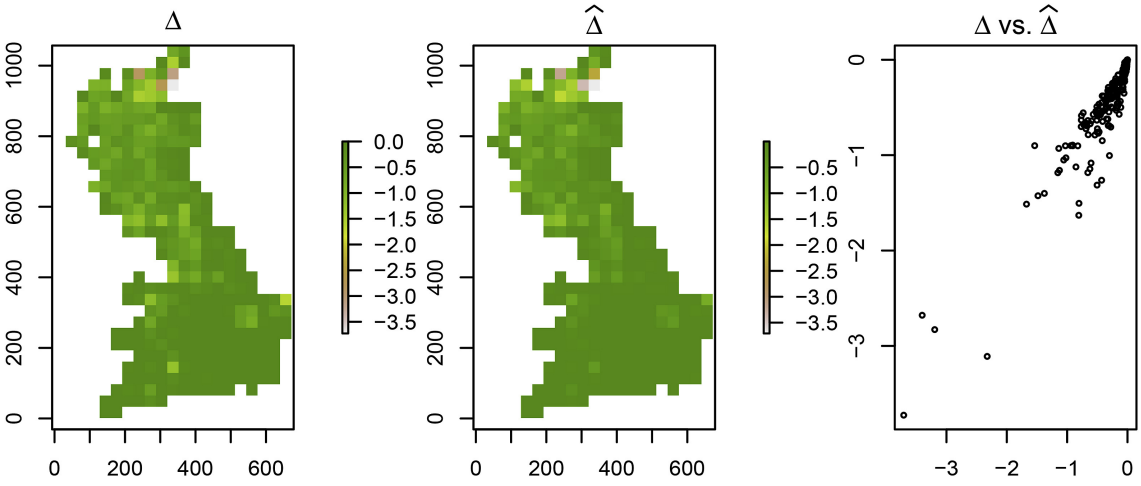

Figure 8. Left panel: The upscaling error Δ for the methane model is calculated by subtracting the correctly upscaled model results (using high-resolution input data) from the incorrect upscaling (using aggregated (mean) input data), and is mapped. Middle panel: The upscaling errors that were predicted by the $\widehat{\Delta}$-formula applied to the inputs and their (co)variances. Right panel: The quality of the error-prediction can be evaluated from the scatterplot of $\widehat{\Delta}$ vs. Δ.

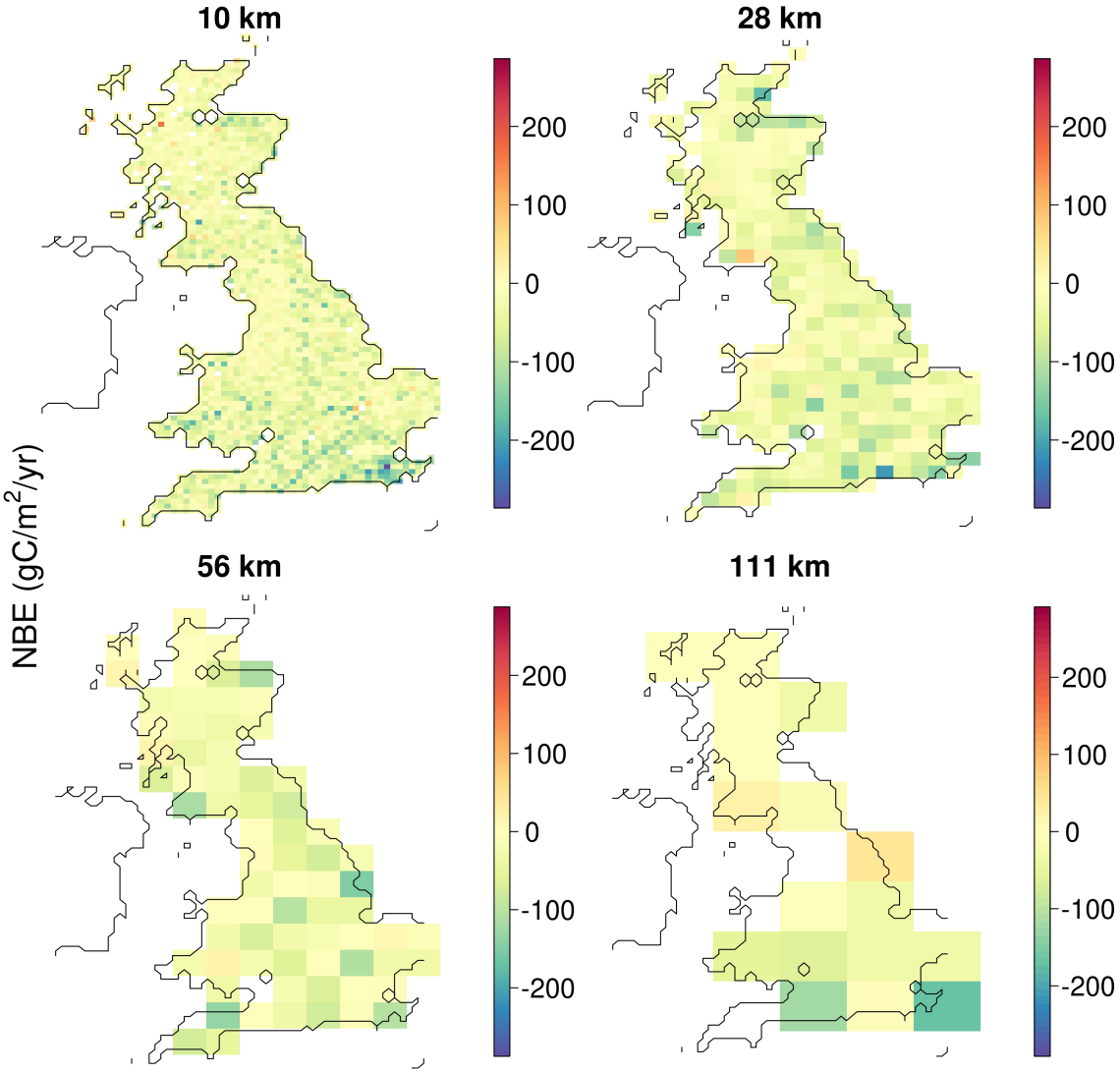

Figure 9. Net Biome Exchange (NBE = -GPP + Reco + Fire) estimated by 4 CARDAMOM analyses at a range of spatial resolutions. A negative value indicates a net uptake of carbon.

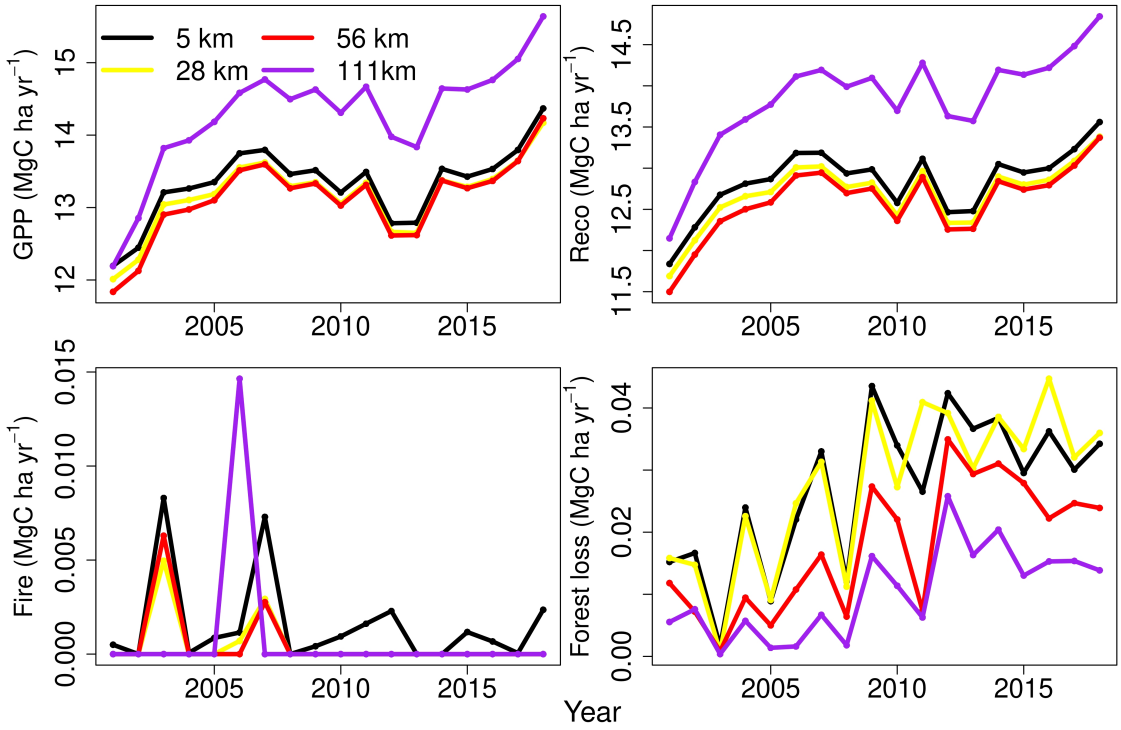

Figure 10. Time series of GB wide mean Gross Primary Productivity (GPP), ecosystem respiration (Reco), fire and forest loss estimated by 4 CARDAMOM analyses at a range of spatial resolutions.

Table 1. Mean annual C-budgets for Great Britain for each target resolution. Budget terms are presented as the mean annual flux in MgC ha^{-1} yr^{-1}. Budget terms presented are gross primary productivity (GPP), ecosystem respiration (Reco), net ecosystem exchange of carbon (NEE = Reco - GPP) C loss due to forest removal (Forest loss) and net biome exchange of carbon (NBE = NEE + Fire loss). The 95 % confidence interval of each term is presented in parenthesis. Fire emissions were low and so are not included in the summary.

Resolution	GPP	Reco	NEE	Forest loss	NBE
5 km	13.3 (10.1 / 16.0)	12.8 (8.6 / 18.3)	-0.39 (-4.1 / 4.4)	0.03 (0.01 / 0.06)	-0.39 (-4.1 / 4.4)
28 km	13.2 (10.0 / 15.9)	12.7 (8.5 / 18.2)	-0.37 (-4.1 / 4.5)	0.03 (0.01 / 0.06)	-0.37 (-4.1 / 4.5)
56 km	13.1 (9.9 / 16.1)	12.6 (8.5 / 18.2)	-0.41 (-4.1 / 4.3)	0.02 (0.01 / 0.04)	-0.41 (-4.1 / 4.3)
111 km	14.3 (11.4 / 16.6)	13.8 (9.9 / 19)	-0.32 (-3.7 / 4.2)	0.01 (0.005 / 0.02)	-0.32 (-3.7 / 4.2)

Acknowledgments

The work was funded by the UK Natural Environment Research Council, grant references: NE/K002481/1, NE/K002619/1 and NE/K002538/1. Data for the programme are being archived at the UK Environmental Information Data Centre (https://eidc.ac.uk/) and the British Atmospheric Data Centre (https://catalogue.ceda.ac.uk/uuid/c117a2e6f451405393cac1c6fbf8f7a3).

References

Aubinet, M., Vesala, T., & Papale, D. (Eds.). (2012). *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer Netherlands.

Band, L. E., Peterson, D. L., Running, S. W., Coughlan, J., Lammers, R., Dungan, J., & Nemani, R. (1991). Forest ecosystem processes at the watershed scale: Basis for distributed simulation. *Ecological Modelling*, *56*, 171–196. doi: 10.1016/0304-3800(91)90199-B

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (2016, February). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, *113*(5), 1285–1290. doi: 10.1073/pnas.1515160113

Bloom, A. A., & Williams, M. (2015, March). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological "common sense" in a model–data fusion framework. *Biogeosciences*, *12*(5), 1299–1315. doi: 10.5194/bg-12-1299-2015

Bresler, E., & Dagan, G. (1988). Variability of yield of an irrigated crop and its causes: 1. Statement of the problem and methodology. *Water Resources Research*, *24*(3), 381–387. doi: 10.1029/WR024i003p00381

Burba, G. (2013). Eddy Covariance Method. *Li-COR Biogeosciences, Lincoln, NE*.

Cowan, N. J., Famulari, D., Levy, P. E., Anderson, M., Bell, M. J., Rees, R. M., ... Skiba, U. M. (2014, September). An improved method for measuring soil N2O fluxes using a quantum cascade laser with a dynamic chamber: Dynamic chamber method. *European Journal of Soil Science*, *65*(5), 643–652. doi: 10.1111/ejss.12168

Cowan, N. J., Levy, P. E., Famulari, D., Anderson, M., Drewer, J., Carozzi, M., ... Skiba, U. M. (2016, August). The influence of tillage on N2O fluxes from an intensively managed grazed grassland in Scotland. *Biogeosciences*, *13*(16), 4811–4821. doi: 10.5194/bg-13-4811-2016

Disney, M., Muller, J.-P., Kharbouche, S., Kaminski, T., Voßbeck, M., Lewis, P., & Pinty, B. (2016, April). A New Global fAPAR and LAI Dataset Derived from Optimal Albedo Estimates: Comparison with MODIS Products. *Remote Sensing*, *8*(4), 275. doi: 10.3390/rs8040275

Finkelstein, P. L., & Sims, P. F. (2001, February). Sampling error in eddy correlation flux measurements. *Journal of Geophysical Research: Atmospheres*, *106*(D4), 3503–3509. doi: 10.1029/2000JD900731

Flechard, C. R., Ambus, P., Skiba, U., Rees, R. M., Hensen, A., Van Amstel, A., ... others (2007). Effects of climate and management intensity on nitrous oxide emissions in grassland systems across Europe. *Agriculture, Ecosystems & Environment*, *121*(1), 135–152.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis, Third Edition* (3edition ed.). Boca Raton: Chapman and Hall/CRC.

Gifford, R. (1994). The global carbon cycle: A viewpoint on the missing sink. *Australian Journal of Plant Physiology*, *21*(1), 1–15. doi: 10.1071/PP9940001

Haario, E. S., H., & Tamminen, J. (2001). An adaptive metropolis algorithm [Journal Article]. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, *7*(2), 223–242.

Haszpra, L., Hidy, D., Taligás, T., & Barcza, Z. (2018, March). First results of tall tower based nitrous oxide flux monitoring over an agricultural region in Central Europe. *Atmospheric Environment*, *176*, 240–251. doi: 10.1016/j.atmosenv.2017.12.035

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... Kempen, B. (2017, February). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. doi: 10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi: 10.1002/qj.3803

Hill, T., Chocholek, M., & Clement, R. (2017). The case for increasing the statistical power of eddy covariance ecosystem studies: Why, where and how? [Journal Article]. *Global Change Biology*, *23*(6), 2154–2165. doi: 10.1111/gcb.13547

Hill, T., Ryan, E., & Williams, M. (2012). The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research [Journal Article]. *Global Change Biology*, *18*, 179–193.

Hollinger, D. Y., & Richardson, A. (2005). Uncertainty in eddy covariance measurements and its application to physiological models, [Journal Article]. *Tree Physiology*, *25*, 873–885.

Hutchinson, G. L., & Mosier, A. R. (1981). Improved Soil Cover Method for Field Measurement of Nitrous Oxide Fluxes1. *Soil Science Society of America Journal*, *45*(2), 311. doi: 10.2136/sssaj1981.03615995004500020017x

Ibrom, A., Dellwik, E., Flyvbjerg, H., Jensen, N. O., & Pilegaard, K. (2007, December). Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. *Agricultural and Forest Meteorology*, *147*(3–4), 140–156. doi: 10.1016/j.agrformet.2007.07.007

IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Keane, B. J., Ineson, P., Vallack, H. W., Blei, E., Bentley, M., Howarth, S., ... Toet, S. (2017). Greenhouse gas emissions from the energy crop oilseed rape (Brassica napus); the role of photosynthetically active radiation in diurnal N2O flux variation. *GCB Bioenergy*, n/a-n/a. doi: 10.1111/gcbb.12491

Kroon, P. S., Hensen, A., Jonker, H. J. J., Ouwersloot, H. G., Vermeulen, A. T., & Bosveld, F. C. (2010, June). Uncertainties in eddy covariance flux measurements assessed from CH4 and N2O observations. *Agricultural and Forest Meteorology*, *150*(6), 806–816. doi: 10.1016/j.agrformet.2009.08.008

Langford, B., Acton, W., Ammann, C., Valach, A., & Nemitz, E. (2015, October). Eddy-covariance data with low signal-to-noise ratio: Time-lag determination, uncertainties and limit of detection. *Atmos. Meas. Tech.*, *8*(10), 4197–4213. doi: 10.5194/amt-8-4197-2015

Lee, X., Massman, W., & Law, B. (2006). *Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis*. Springer Science & Business Media.

Leip, A., Skiba, U., Vermeulen, A., & Thompson, R. L. (2018, February). A complete rethink is needed on how greenhouse gas emissions are quantified for national reporting. *Atmospheric Environment*, *174*, 237–240. doi: 10.1016/j.atmosenv.2017.12.006

Levy, P. E., Burden, A., Cooper, M. D. A., Dinsmore, K. J., Drewer, J., Evans, C., ... Zieliński, P. (2012, May). Methane emissions from soils: Synthesis and analysis of a large UK data set. *Global Change Biology*, *18*(5), 1657–1669. doi: 10.1111/j.1365-2486.2011.02616.x

Levy, P. E., Gray, A., Leeson, S. R., Gaiawyn, J., Kelly, M. P. C., Cooper, M. D. A.,

... Sheppard, L. J. (2011, December). Quantification of uncertainty in trace gas fluxes measured by the static chamber method. *European Journal of Soil Science*, *62*(6), 811–821. doi: 10.1111/j.1365-2389.2011.01403.x

Livingston, G. P., Hutchinson, G. L., & Spartalian, K. (2006, September). Trace Gas Emission in Chambers. *Soil Science Society of America Journal*, *70*(5), 1459–1469. doi: 10.2136/sssaj2005.0322

Mammarella, I., Werle, P., Pihlatie, M., Eugster, W., Haapanala, S., Kiese, R., ... Vesala, T. (2010, February). A case study of eddy covariance flux of N2O measured within forest ecosystems: Quality control and flux error analysis. *Biogeosciences*, *7*(2), 427–440. doi: 10.5194/bg-7-427-2010

Matson, P. A., & Harriss, R. C. (2009). *Biogenic Trace Gases: Measuring Emissions from Soil and Water*. John Wiley & Sons.

Moncrieff, J. B., Massheder, J. M., DeBruin, H., Elbers, J., Friborg, T., Heusinkveld, B., ... Verhoef, A. (1997). A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *Journal of Hydrology*, *189*(1-4), 589–611.

Myrgiotis, V., Blei, E., Clement, R., Jones, S. K., Keane, B., Lee, M. A., ... Williams, M. (2020). A model-data fusion approach to analyse carbon dynamics in managed grasslands [Journal Article]. *Agricultural Systems*, *184*, 102907. doi: 10.1016/j.agsy.2020.102907

Myrgiotis, V., Williams, M., Rees, R. M., Smith, K. E., Thorman, R. E., & Topp, C. F. E. (2016, October). Model evaluation in relation to soil N2O emissions: An algorithmic method which accounts for variability in measurements and possible time lags. *Environmental Modelling & Software*, *84*(Supplement C), 251–262. doi: 10.1016/j.envsoft.2016.07.002

Myrgiotis, V., Williams, M., Topp, C. F. E., & Rees, R. M. (2018, May). Improving model prediction of soil N2O emissions through Bayesian calibration. *Science of The Total Environment*, *624*, 1467–1477. doi: 10.1016/j.scitotenv.2017.12.202

Pedersen, A. R., Petersen, S. O., & Schelde, K. (2010, December). A Comprehensive Approach to Soil-Atmosphere Trace-Gas Flux Estimation with Static Chambers. *European Journal of Soil Science*, *61*(6), 888–902. doi: 10.1111/j.1365-2389.2010.01291.x

Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V., & Hobbie, J. E. (1992). Aggregating Fine-Scale Ecological Knowledge to Model Coarser-Scale Attributes of Ecosystems. *Ecological Applications*, *2*(1), 55. doi: 10.2307/1941889

Revill, A., Bloom, A. A., & Williams, M. (2016). Impacts of reduced model complexity and driver resolution on cropland ecosystem photosynthesis estimates [Journal Article]. *Field Crops Research*.

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive MCMC [Journal Article]. , *18*, 349–367.

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., Finch, J., & Rudd, A. (2016). Climate Hydrology and Ecology Research Support System Potential Evapotranspiration Dataset for Great Britain (1961-2015) [CHESS-PE]. doi: 10.5285/8baf805d-39ce-4dac-b224-c926ada353b7

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., ... Morel, A. (2011, June). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899–9904. doi: 10.1073/pnas.1019576108

Sahoo, B., & Mayya, Y. (2010, August). Two dimensional diffusion theory of trace gas emission into soil chambers for flux measurements. *Agricultural and Forest Meteorology*, *150*(9), 1211–1224. doi: 10.1016/j.agrformet.2010.05.009

Santoro, O., M.; Cartus. (2021). ESA Biomass Climate Change Initiative (Biomass$_c$ci): Global datasets of forest above-ground biomass for the years

2010, 2017 and 2018, v2 [Dataset]. *Centre for Environmental Data Analysis*.

Smallman, T., Exbrayat, J.-F., Mencuccini, M., Bloom, A., & Williams, M. (2017). Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests [Journal Article]. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 528–545.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., ... Smith, J. (2008, February). Greenhouse gas mitigation in agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1492), 789–813. doi: 10.1098/rstb.2007.2184

Smith, P., & Smith, P. (2004, June). Monitoring and verification of soil carbon changes under Article 3.4 of the Kyoto Protocol. *Soil Use and Management*, *20*(2), 264–270. doi: 10.1079/SUM2004239

Stehfest, E., & Bouwman, L. (2006, July). N2O and NO emission from agricultural fields and soils under natural vegetation: Summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems*, *74*(3), 207–228. doi: 10.1007/s10705-006-9000-7

Van Oijen, M., Cameron, D., Levy, P. E., & Preston, R. (2017, August). Correcting errors from spatial upscaling of nonlinear greenhouse gas flux models. *Environmental Modelling & Software*, *94*, 157–165. doi: 10.1016/j.envsoft.2017.03.023

Xu, L., Furtaw, M. D., Madsen, R. A., Garcia, R. L., Anderson, D. J., & McDermitt, D. K. (2006). On maintaining pressure equilibrium between a soil CO2 flux chamber and the ambient air. *Journal of Geophysical Research*, *111*(D8). doi: 10.1029/2005JD006435