We are very grateful for the helpful comments of the referees, which we have used to improve the manuscript as described below.

Please note:
In the original text the CARDAMOM-DALEC modelling framework was referred to as CARDAMOM, however to avoid confusion about how this framework works it is now referred to as DALEC in the manuscript. Therefore, the model name has been changed to DALEC in the author comments and DALEC is used in the responses below.

Anonymous Referee #1

Reviewer comment:
A main point of critique is a missing validation or at least evaluation of the inversion results and posterior fluxes. This is of course not an easy task but at least some basic evaluation tests should have been performed. This could be done by comparing modelled CO2 vertical profiles using the posterior fluxes against aircraft measurements or, if not available, ground-based observations withheld from the inversion. Also, the resolution of the posterior fluxes might already be high enough to compare them directly with eddy covariance based observations. Such an evaluation is missing completely. It is therefore not clear how 'trustworthy' the posterior fluxes are and, also, which one of the two inversions based on different priors performs better than the other. This is a crucial point currently missing in the manuscript and should be added before publication.

Author response:
The posterior fluxes have been used to model mole fractions at Weybourne. Table 5 (Table 1 in response supplement PDF) includes the fit to the data of modelled mole fractions using prior and posterior fluxes for all 4 (DALEC, JULES, gross flux, net flux) inversions for the WAO site. Fig. S14 (Fig. 1 in response) shows the residual mole fractions for all 4 inversions. Fit to the data is improved using the posterior vs prior fluxes.

Change to manuscript:
To test our posterior results against data that has not been included in the inversion, the posterior fluxes have been used to simulate mole fractions at Weybourne Atmospheric Observatory (see Fig. 1 for location in relation to the other sites and Table 1 for site information). The statistics of fit to the data are given in italics in Table 5 and show an improvement in R2 of 0.18 with the DALEC inversion and 0.13 with the JULES inversion, an improvement in RMSE of 1.09 ppm with the DALEC inversion and 0.75 ppm with the JULES inversion and an improvement in the mean bias of 0.64 ppm in the DALEC inversion and 0.56 in the JULES inversion. These results show that the a posteriori fluxes improve the fit to the data at a measurement station not included in the inversion. The results are very similar between the two inversions at this site, but suggest that the DALEC inversion may perform slightly better, at least in this region of the UK. Figure S14 shows the residual mole fractions at Weybourne for each of the inversions carried out in this work.

Reviewer comment:
L 43: Are flux measurements really localized down to centimetres? Probably not.

Author response:
Agreed, changed to metres.

Change to manuscript:
However, they are relatively localised estimates (metres to hectares), which are challenging to scale up to national levels.

Reviewer comment:
L 46: What do you mean by ‘are driven by observational data to varying degree of detail’?

Author response:
Here we were trying to convey that different observation-based products may be assimilated into biosphere models at different temporal and spatial resolutions. Wording is changed to clarify.

Change to manuscript:
Such models describe processes to varying degrees of complexity, with poorly described errors and are driven by observational data at differing temporal and spatial resolutions; hence predictions of biogenic greenhouse gas (GHG) fluxes have poorly quantified biases and can vary significantly between models (Todd-Brown et al., 2013; Atkin et al., 2015).

Reviewer comment:
L 55/56: Indeed, inversions are a valuable tool, but they are also not free from errors. It would be good to mention here sources for uncertainties in inversions and put these into perspective.

Author response:
Added the text below

Change to manuscript:
However, errors in atmospheric transport, unknown uncertainties related to the prior fluxes and issues surrounding the under-determined nature of the problem are all limitations of this approach.

Reviewer comment:
L 62: This is of course not true: Using these measurements in an inversion framework is not an independent way of providing estimates GHG emissions because the inverse modelling system requires prior emission fields as an input. Hence it is not independent of bottom-up inventories. This sentence needs to be rephrased in the manuscript.

Author response:
Rephrased as

Change to manuscript:
To support this legislation, a continuous and automated measurement network has been established (Stanley et al., 2018; Stavert et al., 2018) with the goal of providing estimates of GHG emissions using methods that are complementary to those used to compile the UK’s bottom-up emissions inventory, reported annually to the UNFCCC.

Reviewer comment:
L 74: Why is it rarely the case that model uncertainties are well characterized? This is also related to the comment on L 55/56.

Author response:
Explanation has been added

Change to manuscript:

In practice, this is rarely the case because, for example, uncertainties related to the atmospheric transport model are poorly understood and uncertainties related to biospheric flux estimates from models are largely unknown.

Reviewer comment:
L 80/81: But using Gaussian PDFs is only a choice made by the user, there is no mathematical need for Gaussian PDFs, one can use any PDF to describe prior knowledge. So why is that a problem here?

Author response:
In an analytical inversion there is a mathematical need for Gaussian PDFs and the majority of previous studies will use Gaussian PDFs. Very few existing set-ups can accommodate non- Gaussian PDFs.

Change to manuscript:
Furthermore, for reasons of mathematical and computational convenience, Gaussian probability density functions (PDFs) are commonly used to describe prior knowledge (e.g. Miller et al., 2014).

Reviewer comment:
L 83/84: Why is the size of the diurnal cycle a problem and how does it matter if you solve for monthly fluxes?

Author response:
Being able to simulate the diurnal cycle of the observations as closely as possible is very important if you want to make the most of high frequency data. Figure S2 (Fig. 2 in response) has been adapted in the revised manuscript and shows what the simulated mole fractions look like if monthly fluxes are used. They are much smaller in magnitude and 12 hours out of phase with the data. Table S1 (Table 2 in response supplement PDF) shows an inversion result using monthly fluxes and it is not at all realistic. Therefore, not taking the diurnal cycle into account has a large impact on the inversion.

Change to manuscript:
The choice of 24-hour disaggregation balanced considerations of computational efficiency and simulation accuracy. For certain months and sites, we carried out a set of tests to determine how sensitive our simulated mole fractions and inversion results were, when footprints were disaggregated for the first 12 to 72 hours prior to each measurement (Figure S2; Table S1). Assuming that the 72-hour simulations were the most accurate, we found little degradation in performance by using only 48 or 24 hours disaggregation, when compared to the other uncertainties in the system (e.g. differences between fluxes derived using the 24, 48 and 72 hour simulations were smaller than the 90% confidence interval). However, when only 12 hours was used (or fully integrated footprints), the modelled diurnal cycle was out of phase with the observations.

Reviewer comment:
L 156/157: Shouldn't this 'surface-exchange' height be dependent of actual meteorological conditions and vary for instance with boundary layer height or the strength of vertical mixing?

Author response:

Our team has tested various different schemes (e.g. different “surface heights” using with similar inversion frameworks in Manning et al., 2011 and Arnold et al., 2018), and results seem to be relatively insensitive to this choice.

Reviewer comment:
Sec 2.2.2: Have you done some sensitivity tests on how to handle the boundary conditions? It would be interesting to see how the results change if you don’t include the boundary conditions in the control vector. There is of course a trade off between getting the boundary conditions right and using as much of the observational information as possible to constrain the surface fluxes. In principle, the boundary conditions are nuisance variables, which obviously influence your results but are in themselves not very interesting.

Author response:
We have tested this by carrying out the DALEC inversion with boundary conditions that have been perturbed by +/- 1ppm a priori. The results are shown in Table 3 in the response supplement PDF and a summary of the test has been added to the text.

Change to manuscript:
A sensitivity test where 1ppm is added or taken away from the mole fractions at the domain edges indicates that in June a $\pm$1ppm change translates to a 1-3% change in the inversion result and in December a $\pm$1ppm change translates to a 7-11% change in the inversion result. These changes are substantially smaller than the posterior uncertainty.

Reviewer comment:
L 235 and Sections 2.3.2 and 2.3.3: The wording is maybe a bit misleading here. First you say, that ocean and anthropogenic fluxes are subtracted, and thus treated as perfectly known. Then you explain that these are prior fluxes. Usually a prior flux is a flux that gets updated through the inversion yielding a posterior flux. But you do not do that here. I suggest to reword Sections 2.3.2 and 2.3.3 and not use the word ‘prior’ for ocean and anthropogenic fluxes. Also, I wonder how well the ocean fluxes are really know in that area, especially if you take the Takahashi fluxes (representing open ocean fluxes) as an estimate of the UK coastal ocean fluxes!

Author response:
Changed wording around anthropogenic and ocean fluxes. The ocean component is small so the ocean fluxes do not have a significant impact on the results.

Change to manuscript:
Since the oceanic flux component is small, the comparatively low temporal and spatial resolution of these flux estimates does not significantly impact the inversion results.

Reviewer comment:
L 293-295: Does that mean that only MODIS LAI is assimilated? That also means that you are assimilating model output (since MODIS LAI is not a measured or even observed quantity).

Author response:
It means that MODIS LAI, a temporally and spatially explicit estimate of biomass and a spatially explicit estimate of soil carbon are assimilated. Yes, these are model outputs and the wording has been changed to clarify this.

Change to manuscript:

Observation-derived information used in the current analysis are satellite-based remotely sensed time series of Leaf Area Index (LAI) (MODIS; MOD15A2 LAI-8 day version 5, http://lpdacc.usgs.gov/), a prior estimate of above ground biomass (Thurner et al., 2014) and a prior estimate of soil organic matter (Hiederer and Köchy, 2012).

Reviewer comment:
L 303: It seems that both biosphere models use MODIS LAI data in some way. How independent are then the estimates of JULES and DALEC?

Author response:
There are two key differences between the models in terms of LAI. Firstly, DALEC uses temporally explicit estimates whereas JULES uses a climatology, which means that DALEC is able to capture the interannual variability. Secondly, DALEC assimilates MODIS LAI within a calibration process to simulate its own LAI. This means that DALEC can shift away from MODIS LAI estimates which, through combination with other data, the framework finds to be unlikely. Aside from these points, the models also have very different structures (physiological representation) and parameterisation (ecosystem traits), for example very different descriptions of the carbon assimilation and respiration mechanisms, so they would give different estimates of GPP for the same LAI anyway. We have reworded the description of DALEC and the paragraph about the differences between the models to clarify these points.

Change to manuscript:
[DALEC description]
DALEC is a simplified terrestrial C-cycle model (Smallman et al., 2017). DALEC uses location specific ensembles of process parameters and initial conditions retrieved using the CARDAMOM model-data fusion approach (Bloom et al., 2016). CARDAMOM uses a Bayesian approach within a Metropolis-Hastings MCMC algorithm to compare model states and flux estimates against observational information to determine the likelihood of potential parameter sets guiding the parameterisation processes at pixel scale. DALEC simulates the ecosystem carbon balance, including uptake of CO_2 via photosynthesis, CO_2 loss via respiration, mortality and decomposition processes, and carbon flows between ecosystem pools (non-structural carbohydrates, foliage, fine roots, wood, fine litter, coarse woody debris and soil organic matter). GPP, or photosynthesis, is estimated using the aggregated canopy model (ACM; Williams et al., 1997) while autotrophic respiration is estimated as a fixed fraction of GPP. Canopy phenology is determined by a Growing Season Index (GSI) model as a function of temperature, day length and vapour pressur deficit (proxy for water stress). Mortality and decomposition processes follow first order kinetic equations (i.e. a daily fractional loss of the C stock in question). The decomposition parameters are modified based on an exponential temperature sensitivity parameter. The current version of DALEC used here does not include a representation of the water cycle; rather, water stress is parameterised through a sensitivity to high vapour pressure deficit as part of the GSI phenology model. Comprehensive descriptions of CARDAMOM can be found in Bloom et al. (2016) and DALEC in Smallman et al. (2017).

DALEC estimates carbon fluxes at a weekly time step and 25 km $\times$ 25 km spatial resolution. The weekly time step information was downscaled to 2-hourly intervals, assuming that each day repeated throughout each week. Downscaling of GPP fluxes was assumed to be distributed through the daylight period based on intensity of incoming shortwave radiation. Respiration fluxes were downscaled across the full diurnal cycle assuming exponential temperature sensitivity (code for downscaling is available from the authors on request).

Observation-derived information used in the current analysis are satellite-based remotely sensed time series of Leaf Area Index (LAI) (MODIS; MOD15A2 LAI-8 day version 5,

http://lpdacc.usgs.gov/), a prior estimate of above ground biomass (Thurner et al., 2014) and a prior estimate of soil organic matter (Hiederer and Köchy, 2012). Meteorological drivers were taken from ERA-Interim reanalysis. Ecosystem disturbance due to forest clearances were imposed using Global Forest Watch information (Hansen et al., 2013). CARDAMOM-DALEC differs from typical land surface models in using these data to generate a probabalistic model parameterisations and initial conditions estimates for each pixel, with no a priori assumptions about plant functional types, nor steady states.

[Description of model differences]
In order to understand some of these seasonal differences it is useful to compare the processes taking place in each model. Sections 2.3.1 and 2.3.2 provide detailed descriptions of each model and we give an overview of the main differences here. The DALEC explicitly simulates the soil and litter stocks, growth and turnover processes. LAI is estimated by DALEC at a weekly time step; DALEC was calibrated using MODIS LAI estimates at the correct time and location of the analysis, explained in Sect. 2.3.1. In the JULES system, soil and litter carbon stocks are fixed values for each grid cell, calibrated from 1990-2000, and a fixed climatology of MODIS LAI and canopy height is used. Therefore, DALEC has interannual variability in LAI and soil carbon stocks and can adjust the parameters to find the most likely estimates in combination with other data, whereas these parameters remain constant in JULES. This is potentially advantageous for DALEC, although the use of a climatology in JULES means that noise in the MODIS LAI estimates will be averaged out. Since, LAI and soil and litter carbon stocks are fixed in JULES, variability in TER and GPP fluxes are governed by meteorology, primarily temperature but also significant signals from photosynthetically active radiation and precipitation via the soil moisture. Meteorology drives the JULES model at a 2-hourly timestep as opposed to a weekly time-step in DALEC. Therefore, in the 2-hourly DALEC product used here, the diurnal range is not explicitly simulated and is the result of a downscaling process from a weekly resolution. This downscaling is done based on light and temperature curves as explained in Sect. 2.3.1. In DALEC, the autotrophic respiration is parameterised as a fixed fraction of the GPP for a given site but varies between sites, roughly ranging from 0.3 to 0.7. In JULES, the autotrophic respiration is the sum of plant maintenance and growth respiration terms, which are calculated separately as process based functions of the GPP, the maximum rate of carboxylation and leaf nitrogen content (Clark et al., 2011). Typically, the autotrophic respiration in JULES is roughly 0.1-0.25 of the GPP. Therefore, there are some large differences between the model structures and parameterisations, particularly in how the respiration fluxes are simulated. This could be leading to too small a diurnal range in DALEC TER and too large a diurnal range in JULES TER.

Reviewer comment:
L 364: How did you determine the length of the burn-in period and does the number of iterations include the burn-in period?

Author response:
The number of iterations does not include the burn-in period. This has been clarified in the text. The length of the burn-in was determined by visual inspection of the chains and a conservatively large estimate was used to ensure that it was sufficient.

Change to manuscript:
The algorithm had a burn-in period of 5×10^4 iterations and was then run for an additional 2×10^5 iterations to appropriately explore the posterior distribution.

Reviewer comment:

Sec 2.4.3 and Eq (7): What is x and y here? How many basis functions do you have in total and how does the Jacobian look like? Maybe you can add an equation for the Jacobian: H= del .../ del ...

Author response:
X and y have been described already in Eq. 3. There are 19 spatial basis functions, this has been clarified in the text. Parts of Section 2.4.3 about the Jacobian matrix have been reworded to clarify how the Jacobian is set up.

Change to manuscript:
A scaling factor is solved in the inversion, scaling GPP and TER within the 4 outer regions and within maps of five or six PFTs in the sub-domain: broadleaf tree, needleleaf tree, C3 grasses, C4 grasses, shrubland and, in the case of TER, bare soil. Therefore, there are 19 spatial basis functions.

H has dimensions m (number of data points) by n (number of parameters).

To create this linear model, we multiplied the footprints by the prior GPP and TER fluxes separately, then multiplied these by the fractional map of basis functions (described in Sect. 2.4.2) and summed over the domain. The boundary conditions were broken down by four further basis functions for each edge of the domain as explained in Sect. 2.2.2. The parameters vector, x, consisted of a set of scaling factors that multiplied the fluxes or boundary conditions. Multiplying the sensitivity matrix by the prior estimate of x, a vector of ones, yields the prior modelled mole fraction time-series at a site.

Reviewer comment:
L 393: The word ‘tested’ is not correct here, ‘applied’ would fit better. In any case it would be good to add a few sentences on testing you set up in e.g. an identical twin experiment.

Author response:
We propose not to change the wording. We feel that the presented synthetic tests and agreement between two inversions shows sufficient evidence that our system performs well. Furthermore, many more synthetic tests have been carried out, which are too numerous to show here, but will be citable by the time of publication.

Reviewer comment:
Section 3.1: This goes back to my main comment on evaluating the inversions. Can you say which result is more realistic?

Author response:
From the validation study, DALEC performs marginally better than JULES but not significantly, therefore it is difficult to say which is best. Some suggestion of this has been added to Section 3.4:

Change to manuscript:
[About the fits to data included in the inversion]
Overall, the fits are relatively similar between the DALEC and JULES inversions implying that the two inversions perform similarly well.

[About the fits to the validation data set]
Again, the results are very similar between the two inversions but suggest that the DALEC inversion may perform slightly better, at least in this region of the UK.

Reviewer comment:
L 414: How can soil and litter carbon stocks be fixed in JULES? I wonder a model with fixed carbon stocks can provide decent estimates of the actual respiration fluxes.

Author response:
The main driver of temporal variability in soil respiration in JULES is the soil temperature given the Q10 and the second largest driver is soil moisture. It is true that an element of temporal variability is lost using a fixed soil carbon. However, this would be the second or third most important driver of temporal variability. Also, there is a lack of such observations of this behaviour (e.g. changes in soil carbon stocks over time driving changes in soil respiration), hence using a fixed map, which gives a reasonable spatial distribution, was the preferred option here.

Reviewer comment:
L 480: Do you mean 'underestimating' the net summer flux compared to the true flux? And if so, how do you know the true flux?

Author response:
We meant compared to the posterior flux.

Change to manuscript:
Generally, the models underestimate the net summer flux compared to the posterior flux (to the greatest extent in 2014), although the summer estimate from the JULES inversion in 2013 is not statistically different from the prior.

Reviewer comment:
L 516: Do you mean here the posterior fluxes from the inversions or the prior fluxes from the two different models? Maybe stick to a common notation/terminology for the fluxes, e.g. prior fluxes and posterior fluxes throughout the manuscript and not refer to them just by model name.

Author response:
We meant posterior fluxes from the inversions. The sentence has been rewritten.

Change to manuscript:
However, the net sink in the JULES inversion is larger than the DALEC inversion in Scotland, south Wales, Northern Ireland and south-west England.

Reviewer comment:
Sec 3.4: This section presents some posterior diagnostics of the inversions and presents a first step towards an evaluation of the inversions. What do the different fits to the data mean for the inversions and posterior fluxes?

Author response:
Added more detail to Section 3.4

Change to manuscript:
Agreement between the data and the posterior simulated mole fractions at the measurement sites used to constrain the inversion is greatly improved compared to prior simulated mole fractions, with R2 values increasing by a minimum of 0.24 and up to 0.5 (to give values ranging between 0.53 and 0.71) and root mean square error (RMSE) decreasing by at least

1.35 ppm and up to 2.6 ppm (to give values ranging between 1.26 ppm and 2.71 ppm). Table 5 shows all statistics for the prior and posterior mole fractions compared to the observations of atmospheric CO2 concentrations. Overall, the fits are relatively similar between the DALEC and JULES inversions, implying that the two inversions perform similarly well by these metrics. In terms of R2, the best fit to the data is observed at Heathfield in the DALEC inversion and Angus in the JULES inversion. In terms of RMSE, the best fit to the data is observed at Angus in the DALEC inversion and Mace Head in the JULES inversion. The smallest posterior mean bias is observed at Angus in the DALEC inversion and Ridge Hill in the JULES inversion. Therefore, there are some small spatial differences in how well each of the inversions is able to fit the data but no clear indication of which areas of posterior flux might be subject to the largest improvement in either inversion. Figures S12 and S13 show the residual mole fractions in 2014 and indicate that residuals are somewhat larger during the summer than the winter

Reviewer comment:
L 588-590: Agricultural activities should somehow (implicitly) be accounted for by the biosphere models through the use of MODIS LAI, which should capture events such as harvest in the LAI.

Author response:
There are a couple of reasons why MODIS LAI is not able to capture events such as harvest.
1) MODIS LAI uses constant reflectance fractions so even though crops will have a different reflectance fraction to forest, for example, the fact that surface reflectance of crops change dramatically with senescence and post-harvest litter on the soil surface is not accounted for. This will partially mean that there will be errors introduced into MODIS's LAI estimates for crops.
2) Even with the correct LAI, crops are different from other ecosystems in terms of their C flux due to the removal of biomass which would otherwise decompose. This is in addition to harvest itself introducing litter into the system, which decomposes at a different time than it would if it was "naturally" occurring. In neither DALEC or JULES was a model structure which allows for C redistribution within the ecosystem / or harvest removal processes represented which means that the models cannot reproduce the associated fluxes.
This is therefore another area where performing an inversion can help to resolve these features that are hard to identify with MODIS LAI.

Reviewer comment:
L 622-626: Again, this hypothesis means that you trust your inversion results but it is not clear on which basis you trust the inversions. This hypothesis should be supported by a more substantial evaluation of the inversions.

Author response:
Addressed in Anonymous Referee #1's first comment.

Anonymous Referee #3

Reviewer comment:
On p.6, the authors assess the impact of the timespan for which footprints are disaggregated, and they conclude that the effect of going from 24h to 72h of time-disaggregation is negligible. However, I do not find the provided evidence for this statement very convincing. Indeed, the mole fractions calculated at Ridge Hill coincide nicely for 24h,

48h and 72h, but is this result representative for the entire UK? And how do the results compare to a simulation without time-disaggregation? The expected range of net annual biospheric fluxes has shifted significantly between the simulations with 24h of aggregation and with 72h of disaggregation, compared to the prior. It would be helpful if the results for no and for 48h of time- disaggregated footprints are included in table S1 as well. In addition, the uncertainty of the obtained net flux seems to be unaffected by the timespan of the disaggregated footprint. Can you comment on this?

Author response:
Figure S2 (Fig. 2 in response) has now been replotted, showing the results for no time-disaggregation using monthly fluxes, 6 hours back and 12 hours back, as well as the original 24, 48 and 72 hours. Fig. S2 (Fig. 2 in response) now also shows the modelled mole fractions at Tacolneston, showing the same result, indicating that it is representative of multiple stations in the UK. Table S1 (Table 2 in response supplement PDF) now includes results with no time disaggregation, along with disaggregated footprints going 12, 24, 48 and 72 hours back. Although there are differences between 24, 48 and 72 hours back, these are not statistically significant (within the 90%ile range). However, for 12 hours or inversions using the integrated footprints, the inversions are statistically different from the time-disaggregated results. Compared to the seasonal cycle or the differences between the JULES and DALEC inversions in Fig. 5, the differences between 24, 48 and 72 are small. The uncertainty reduction will primarily depend on the magnitude of the sensitivity (in addition to the number of measurements and the measurement/model uncertainty). Since we examine the net flux here, it is consistent that time-integrated footprint and the time-disaggregated footprints lead to similar uncertainty reduction in the net flux.

As an aside, the inversion result for the 72 hours back test has changed slightly. In the previous version, we carried out the 72 hours back inversion using data that had the anthropogenic and ocean forward modelled using 72 hours back footprints. However, we have now changed this for consistency, so all of these tests use data with fixed components removed using 24 hours back footprints.

Change to manuscript:
The choice of 24-hour disaggregation balanced considerations of computational efficiency and simulation accuracy. For certain months and sites, we carried out a set of tests to determine how sensitive our simulated mole fractions and inversion results were, when footprints were disaggregated for the first 12 to 72 hours prior to each measurement (Figure S2; Table S1). Assuming that the 72-hour simulations were the most accurate, we found little degradation in performance by using only 48 or 24 hours disaggregation, when compared to the other uncertainties in the system (e.g. differences between fluxes derived using the 24, 48 and 72 hour simulations were smaller than the 90% confidence interval). However, when only 12 hours was used (or fully integrated footprints), the modelled diurnal cycle was out of phase with the observations.

Reviewer comment:
On p.12 the basis functions for the inversion are shortly discussed. This part could be made more clear by including a figure showing the clustering of the scaling factors.

Author response:
We have added a figure (Fig. S6 – Fig. 3 in response) showing the spatial basis functions used in the inversion (based on PFTs). We hope this clarifies this discussion.

Change to manuscript:

Within the West-Central Europe area (the hatched region in Fig. S1), a map of the fraction of different plant functional types (PFTs) in each grid cell has been used to further break down the region (Fig. S6).

Reviewer comment:
The inversion framework allows to scale the prior estimates for respiration and GPP separately. In a synthetic test it is shown that this approach indeed allows to compensate for biases in either respiration or GPP, which is obviously not possible with the NEE inversion. To strengthen the point of the superior behavior of the TER+GPP inversion over the NEE inversion, it would be nice to instead include a synthetic test with a truth that less obviously favors one approach over the other, e.g. by using a combination of the JULES and DALEC fluxes as truth.

Author response:
To address the reviewer's comment, the synthetic test in Fig. S5 (Fig. 4 in response) has been changed to use synthetic data created using the DALEC biosphere fluxes, and the NAME model, while JULES is used for the prior fluxes.

Change to manuscript:
To demonstrate this, we have carried out a synthetic test (Fig. S5) in which we have investigated the ability of our inversion system to solve for a "true" flux, created using the DALEC prior fluxes and NAME simulations, in an inversion that used the JULES fluxes as the prior. Figure S5(a) shows that monthly posterior fluxes for the inversion where GPP and TER are separated agree with the "true" flux within estimated uncertainties in 16 out of 24 months. In contrast, whilst the posterior fluxes for the inversion where NEE is scaled has changed significantly from the prior, it is not in agreement with the "true" flux except in July 2013 and August and September 2014. The posterior diurnal cycles of GPP, TER and NEE, which are shown as an average for June 2014 in Fig. S5(b) and Fig. S5(c), highlight the differences in diurnal cycle between the two models. The inversion that can adjust the two sources separately leads to higher night-time fluxes, which are closer to the "true" flux than the prior. On the other hand, the inversion where NEE is scaled can only stretch or shrink the diurnal cycle in one direction, increasing both the daytime sink and night-time source, or decreasing them, together. In this case, they have decreased, which does bring the net June 2014 flux in Fig. S5(a) closer to the "true" June 2014 flux but cannot go far enough to reconcile the monthly fluxes.

Reviewer comment:
A different anthropogenic flux map (EDGAR) is used outside the UK. Are these fluxes in agreement with the NAEI fluxes for the UK?

Author response:
The fluxes are not in agreement (NAEI is 482 Tg/yr over the UK in 2013, 439 Tg/y in 2014, whilst EDGAR is 540 Tg/yr over UK in both years as we use the 2010 values throughout the inversion). However, our tests show that this does not have a significant impact on the derived fluxes from the UK.

Change to manuscript:
Within the UK, the NAEI and EDGAR fluxes differ by around 15% (540 Tg/yr for EDGAR, 460 Tg/yr for NAEI). We do not find that our derived UK fluxes are significantly affected by perturbations of this magnitude applied to anthropogenic emissions outside the UK.

Reviewer comment:

The data statistics in table S2 cover the year 2014 only. Are the statistics for 2013 similar, and why are they not included? It would be nice to see these statistics for the entire range of presented inversion results, i.e. both 2013 and 2014. Moreover, for the inversion with JULES prior the bias increases for 2 out of 6 measurement sites. Can you comment on this? In addition, I find it peculiar that the prior RMSE and prior bias differ between the GPP-TER and the NEE inversion. How did you evaluate these statistics? I would also suggest to include the statistics for the NEE-inversion in table S2 (maybe put them in a smaller font and/or grey) and to move the table to the main paper.

Author response:
This table has been changed to give statistics calculated across all months of the inversion. The table has also been merged with the table of statistics from the NEE inversion and moved to the main paper (Table 5 – Table 1 in response supplement PDF). The reason why the JULES prior bias increased for some sites and for the prior statistics to be different between the two inversions, is because we initially calculated the statistics for the prior using the posterior baseline (we did this because the prior mole fractions were strongly influenced by biases in the baseline mole fractions, however, we now realise that this was confusing, as it made it look as though the inversion could be making the fit to the data worse, which of course is not possible). We have now changed this to use the prior baseline to avoid confusion. Therefore, the plots of residual mole fractions Figs. S12-13 (Figs 5-6 in response) have also been updated to use the prior baseline with prior modelled mole fractions.

Change to manuscript:
Agreement between the data and the posterior simulated mole fractions at the measurement sites used to constrain the inversion is greatly improved compared to prior simulated mole fractions, with R2 values increasing by a minimum of 0.24 and up to 0.5 (to give values ranging between 0.53 and 0.71) and root mean square error (RMSE) decreasing by at least 1.35 ppm and up to 2.6 ppm (to give values ranging between 1.26 ppm and 2.71 ppm). Table 5 shows all statistics for the prior and posterior mole fractions compared to the observations of atmospheric CO2 concentrations.

Reviewer comment:
The optimized state vector and hyper-parameters are not discussed, while it would be interesting to have at least a look at the number of time periods resolved in the inversion, how scaling factors for GPP and TER vary, and to what extent the boundary conditions required scaling during the inversion. Please include such a discussion.

Author response:
A figure has been added to the supplement (Fig. S11 – Fig. 7 in response) showing the mean and maximum number of time periods that each month is broken down into for each spatial basis function for each month. The lines below have been added to the main text. In terms of the scaling factors, we do not think it is so useful to look at the scaling factors and hyperparameters in isolation and the figures showing the posterior fluxes for GPP and TER (Figs. S7-S8) already show this information in a more meaningful way.

Change to manuscript:
Fig. S11 shows that the inversion typically scaled the fluxes within 4 or 5 temporal regions per month, although for some parameters in some months scaling factors were found up to roughly a daily resolution.

Reviewer comment:

Due to the use of satellite remote-sensing data in DALEC, one would expect a benefit of using this prior over the JULES model, at least with respect to spatial distribution of the fluxes. This point does not seem to be addressed.

Author response:
In fact, both models use remote sensing LAI data from MODIS and the main difference in their use of this data is not in the spatial distribution but that DALEC uses a time-varying product with interannual variablilty and JULES uses a climatology. This does potentially give DALEC an advantage, however DALEC could also suffer due to the greater noise in the MODIS estimates which will be averaged out in the climatology. It is difficult to say what impact this has on the inversion compared to the other differences between the models. Importantly, the inversion helps to bring the two estimates together, despite the possible advantages of using DALEC. Some clarification has been added to the introduction and to Section 3.1

Change to manuscript:
[Introduction]
Gross primary productivity (GPP) and terrestrial ecosystem respiration (TER) estimates from the Joint UK Land Environment Simulator (JULES) and Data Assimilation Linked Ecosystem Carbon (DALEC) are used as prior flux constraints. JULES is a state-of-the-art physically based, process driven model that estimates the energy, water and carbon fluxes at the land-atmosphere boundary and uses a variety of observation-derived products describing physical parameters as inputs (Best et al., 2011; Clark et al., 2011). DALEC, on the other hand, is a simplified terrestrial C-cycle model which is calibrated independently at each location retrieving both process parameters and initial conditions using the CARbon DAta MOdel fraMework (CARDAMOM) model-data fusion system. CARDAMOM ingests satellite based remotely sensed estimates of the state of terrestrial ecosystems (Bloom and Williams, 2015; Bloom et al., 2016; Smallman et al., 2017).

[Section 3.1]
Therefore, DALEC has interannual variability in LAI and soil carbon stocks and can adjust the parameters to find the most likely estimates in combination with other data, whereas these parameters remain constant in JULES. This is potentially advantageous for DALEC, although the use of a climatology in JULES means that noise in the MODIS LAI estimates will be averaged out.

Reviewer comment:
Figure 6 can be improved by adding two additional subfigures that show the difference between DALEC and JULES posterior fluxes. Also the addition of maps showing the related uncertainty in posterior NEE would be helpful.

Author response:
Figure 6 (Fig. 8 in response) has been adapted to show the difference between DALEC and JULES. We don't think that the maps of posterior uncertainty add much beyond the uncertainties shown on the monthly fluxes in Fig. 5.

Reviewer comment:
P.2, the abbreviation GHG is used before it is introduced.

Author response:
Done

Reviewer comment:
P.5, please include an indication of the uncertainty related to the CO2 measurements

Author response:
Done – added to Data selection and model uncertainty section (2.2.1)

Change to manuscript:
Monthly average measurement uncertainty is around 0.9 ppm.

Reviewer comment:
P.10, please mention the spatiotemporal resolution of the prior ocean flux estimates.

Author response:
This was already mentioned in Table 3 (now Table 2).

Reviewer comment:
P.11, to reduce initial confusion, it might be useful to move the paragraph starting at line 350 (about the variable time instead of space dimension) to before equation 4.

Author response:
We think moving this paragraph to before the equation would move the equation too far away from the description of its key elements however, in the paragraph you mention, we have clarified that k relates to equation 4.

Change to manuscript:
Therefore, in this case k in Eq. 4 is more specifically the unknown number of time periods resolved in the inversion, which is important because CO2 fluxes vary strongly in time and have high uncertainty in their temporal variation.

Reviewer comment:
Please be consistent in the colors you use: e.g. the colors for DALEC and JULES are reversed between figures 2 and 5.

Author response:
Done

Reviewer comment:
The section numbering in chapter 4 should be updated, now 4.2 comes after 4.3

Author response:
Done

Quantifying the UK's Carbon Dioxide Flux: An atmospheric inverse modelling approach using a regional measurement network

Emily D. White[1], Matthew Rigby[1], Mark F. Lunt[1,2] Anita L. Ganesan[3], Alistair J. Manning[4], Simon O'Doherty[1], Ann R. Stavert[1,5], Kieran Stanley[1], T. Luke Smallman[2,6], Mathew Williams[2,6], Edward Comyn-Platt[7], Peter Levy[8], Michel Ramonet[9], Grant L. Forster[10,11], Andrew C. Manning[10], Paul I. Palmer[2]

[1]School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK
[2]School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3FF, UK
[3]School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK
[4]Hadley Centre, Met. Office, Exeter, EX1 3PB, UK
[5]Climate Science Centre, CSIRO Oceans and Atmosphere, Aspendale VIC 3195, Australia
[6]National Centre for Earth Observation, Edinburgh, EH9 3FF, UK
[7]Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK
[8]Centre for Ecology and Hydrology (Edinburgh Research Station), Penicuik, E26 0QB, UK
[9]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, Gif-sur-Yvette, 91198, France
[10]Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK
[11]National Centre for Atmospheric Science, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Correspondence to: Emily White (emily.white@bristol.ac.uk), Matthew Rigby (matt.rigby@bristol.ac.uk)

Abstract. We present a method to derive atmospheric-observation-based estimates of carbon dioxide (CO_2) fluxes at the national scale, demonstrated using data from a network of surface tall tower sites across the UK and Ireland over the period 2013-2014. The inversion is carried out using simulations from a Lagrangian chemical transport model and an innovative hierarchical Bayesian Markov chain Monte Carlo (MCMC) framework, which addresses some of the traditional problems faced by inverse modelling studies, such as subjectivity in the specification of model and prior uncertainties. Biospheric fluxes related to gross primary productivity and terrestrial ecosystem respiration are solved separately in the inversion and then combined a posteriori to determine net ecosystem exchange of CO_2. Two different models, DALEC and JULES, provide prior estimates for these fluxes. We carry out separate inversions to assess the impact of these different priors on the posterior flux estimates and evaluate the differences between the prior and posterior estimates in terms of missing model components. The Numerical Atmospheric dispersion Modelling Environment (NAME) is used to relate fluxes to the measurements taken across the regional network. Posterior CO_2 estimates from the two inversions agree within estimated uncertainties, despite large differences in the prior fluxes from the different models. With our method, averaging results from 2013 and 2014, we find a total annual net biospheric flux for the UK of -8 ± 79 Tg CO_2 yr^{-1} (DALEC prior) and -64 ± 85 Tg CO_2 yr^{-1} (JULES prior), where negative values represent an uptake of CO_2. These biospheric CO_2 estimates show that annual UK biospheric sources and sinks are roughly in balance. These annual mean estimates consistently indicate a greater net release of CO_2 than the prior estimates, which show much more pronounced uptake in summer months.

Deleted: Mathew Williams[2],
Formatted: Superscript
Deleted: Platt[6]
Deleted: Levy[7]
Deleted: Ramonet[8]
Deleted: .
Deleted: 3JN
Deleted: [6]Centre
Deleted: Ecology and Hydrology, Wallingford, OX10 8BB
Deleted: [8]Laboratoire
Deleted: primary productivity.
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: -ve
Deleted: are
Deleted: higher
Deleted: the
Deleted:

1 Introduction

There are significant uncertainties in the magnitude and spatiotemporal distribution of global carbon dioxide (CO_2) fluxes to and from the atmosphere, particularly those due to terrestrial ecosystems (Le Quéré et al., 2018). Reliable methods for quantifying carbon budgets at policy relevant scales (i.e. national or sub-national) will be important to accurately and transparently evaluate each country's progress towards achieving their Nationally Determined Contributions (NDCs) made following the Paris Agreement (UNFCCC, 2015).

Regional terrestrial carbon fluxes can be estimated using a range of observational, computational and inventory-based methods. These include "bottom-up" approaches such as the up-scaling of direct flux measurements made using eddy covariance or chamber systems (Baldocchi and Wilson, 2001) and models of atmosphere-biosphere CO_2 exchange. Flux measurements are important for understanding the small-scale processes responsible for carbon fluxes. However, they are relatively localised estimates (metres to hectares), which are challenging to scale up to national levels. Biosphere models and land surface models (LSMs) can be used to estimate carbon fluxes using coupled representations of biogeophysical and biogeochemical processes, driven by observations of meteorology and ecosystem parameters (Potter, 1999; Clark et al., 2011; Bloom et al., 2016). Such models describe processes to varying degrees of complexity, with poorly described errors and are driven by observational data at differing temporal and spatial resolutions; hence predictions of biogenic greenhouse gas (GHG) fluxes have poorly quantified biases and can vary significantly between models (Todd-Brown et al., 2013; Atkin et al., 2015).

Atmospheric inverse modelling is a "top-down" approach that provides an alternative to the bottom-up approaches. Inversions have been used to indirectly estimate country-scale (e.g. Matross et al., 2006; Schuh et al., 2010; Meesters et al., 2012) and continental (e.g. Gerbig et al., 2003; Peters et al., 2010; Rivier et al., 2010) biospheric CO_2 budgets using atmospheric mole fraction observations, where the contribution of anthropogenic fluxes to the observations has been removed. In this approach, a model of atmospheric transport relates spatiotemporally resolved surface fluxes of biospheric CO_2 to atmospheric measurements of CO_2 mole fractions. Biospheric fluxes derived from bottom-up approaches are often used as prior estimates in the inversion. Since atmospheric observations are sensitive to fluxes spanning tens to hundreds of kilometres (Gerbig et al., 2009), inverse methods are a valuable tool for examining national fluxes and evaluating estimates of surface exchange of CO_2 at larger spatial scales. However, errors in atmospheric transport, unknown uncertainties related to the prior fluxes and issues surrounding the under-determined nature of the problem are all limitations of this approach.

The United Kingdom (UK) government has set legally binding targets to curb greenhouse gas (GHG) emissions in an attempt to prevent dangerous levels of climate change. The Climate Change Act 2008 (The UK government, 2008) commits the UK

Deleted: biogenic systems

Deleted: if we are

Deleted: centimetres

Deleted: kilometres

Deleted: to varying degrees of detail

Deleted:

Deleted: described above. It has been used to indirectly estimate country-scale (e.g. Matross et al., 2006; Schuh et al., 2010; Meesters et al., 2012)

Deleted: (e.g. Gerbig et al., 2003; Peters et al., 2010; Rivier et al., 2010)…

Deleted: is used to relate

to 80% cuts in GHG emissions, from 1990 levels, by 2050. To support this legislation, a continuous and automated measurement network has been established (Stanley et al., 2018; Stavert et al., 2018) with the goal of providing estimates of GHG emissions using methods that are complementary to those used to compile the UK's bottom-up emissions inventory reported annually to the United Nations Framework Convention on Climate Change (UNFCCC). Previous studies have used data from the UK Deriving Emissions related to Climate Change (UK-DECC) network to infer emissions of methane, nitrous oxide and HFC-134a from the UK (Manning et al., 2011; Ganesan et al., 2015; Say et al., 2016). These studies found varying levels of agreement with bottom-up inventory methods, where estimates of GHG emissions are made using reported statistics from various sectors (e.g. road transport, power generation, etc.). Here we use the DECC network and two additional sites from the Greenhouse gAs Uk and Global Emissions (GAUGE) programme (Palmer et al., 2018) to estimate biospheric fluxes of CO_2. Whilst anthropogenic emissions, which are the remit of the UK inventory, are not estimated in this study, these biosphere estimates represent the first step towards a framework for estimating the complete UK CO_2 budget.

Atmospheric inverse modelling of GHGs using Bayesian methods presents some known challenges. Robust uncertainty quantification in Bayesian frameworks can be difficult as they require that uncertainties in the prior flux estimate, and uncertainties in the atmospheric transport model's ability to simulate the data, are well characterised. In practice, this is rarely the case because, for example, uncertainties related to the atmospheric transport model are poorly understood and uncertainties related to biospheric flux estimates from models are largely unknown. Various studies have investigated the use of data-driven uncertainty estimation (Michalak, 2004; Berchet et al., 2013; Ganesan et al., 2014; Kountouris et al., 2018b). Inversions are also known to suffer from "aggregation errors". One type of aggregation error arises from the way in which areas of the flux domain are grouped together to decrease the number of unknowns, because usually there are not sufficient data to solve for fluxes in each model grid cell (Kaminski et al., 2001). Furthermore, for reasons of mathematical and computational convenience, Gaussian probability density functions (PDFs) are commonly used to describe prior knowledge (e.g. Miller et al., 2014). However, Gaussian assumptions can lead to unphysical solutions in the case of atmospheric GHG emissions or uptake processes, as they permit both positive and negative solutions.

CO_2 presents further complications over other GHGs in that atmosphere-biosphere CO_2 exchange has a diurnal flux cycle that is significantly larger than the net flux, and has strong, spatially varying surface sources and sinks. Gerbig et al. (2003) was one of the first to develop an analysis framework for regional scale CO_2 flux inversions. The study sets out the need to explicitly simulate the diurnal cycle of biospheric fluxes and highlights the importance of high spatial and temporal resolution data when addressing the unique problems of representation and aggregation errors caused by the highly varying nature of CO_2 fluxes in both space and time. Inverse modelling studies of CO_2 flux typically assume that anthropogenic fluxes are "fixed" in the inversion (e.g. Meesters et al., 2012; Kountouris et al., 2018a). This is based on the assumption that uncertainties in anthropogenic fluxes are low compared to those of the biospheric fluxes. However, it has been suggested that this may not necessarily be the case (Peylin et al., 2011).

Deleted: independent of
Deleted: anthropogenic
Deleted: that must be
Deleted: UK Parliament and submitted to the
Deleted: it represents
Deleted: various
Formatted: English (UK)
Formatted: English (UK)
Deleted: (e.g. Michalak, 2004; Berchet et al., 2013; Ganesan et al., 2014; Kountouris et al., 2018b)
Deleted: can be
Deleted: (Miller et al., 2014)
Deleted: (Gerbig et al., 2003)
Deleted: (e.g. Meesters et al., 2012; Kountouris et al., 2018a)

Here we outline a framework for evaluating the net biospheric CO_2 exchange (net ecosystem exchange, NEE) from a small to medium sized country (the UK covers an area of around 250,000 km^2) using the high-resolution regional, Lagrangian transport model, the Numerical Atmospheric dispersion Modelling Environment (NAME, Jones et al., 2006). To address many of the problems outlined above, we use an adapted form of a hierarchical Bayesian, trans-dimensional Markov chain Monte Carlo (MCMC) inversion (Rigby et al., 2011; Ganesan et al., 2014; Lunt et al., 2016). In the hierarchical Bayesian framework presented in Ganesan et al. (2014), "hyperparameters" that define the prior flux and model-data "mismatch" uncertainty PDFs are included in the inversion, which is solved using a Metropolis-Hastings MCMC algorithm (e.g. Rigby et al., 2011). This hierarchical approach has been shown to lead to more robust posterior uncertainty quantification in Bayesian frameworks where prior uncertainties are not well characterised (Ganesan et al., 2014). Lunt et al. (2016) built on this method, developing a "trans-dimensional" framework that accounted for the uncertainty in the definition of basis functions (the way in which flux grid cells are aggregated), and allowed this to propagate through to the posterior estimate.

Gross primary productivity (GPP) and terrestrial ecosystem respiration (TER) estimates from the Joint UK Land Environment Simulator (JULES) and Data Assimilation Linked Ecosystem Carbon (DALEC) models are used as prior flux constraints. JULES is a state-of-the-art physically based, process driven model that estimates the energy, water and carbon fluxes at the land-atmosphere boundary and uses a variety of observation-derived products describing physical parameters as inputs (Best et al., 2011; Clark et al., 2011). DALEC, on the other hand, is a simplified terrestrial C-cycle model which is calibrated independently at each location retrieving both process parameters and initial conditions using the CARbon DAta MOdel fraMework (CARDAMOM) model-data fusion system. CARDAMOM ingests satellite based remotely sensed estimates of the state of terrestrial ecosystems (Bloom and Williams, 2015; Bloom et al., 2016; Smallman et al., 2017).

Below, we first describe our approach for modelling biospheric CO_2 fluxes, including several novel aspects compared to previous work in this area. We then investigate the impact of using two different models that simulate biospheric fluxes (JULES and DALEC) within our proposed inverse framework and discuss the discrepancies between the prior and posterior flux estimates.

2 Method

The main components of a regional atmospheric inverse modelling framework are the atmospheric CO_2 mole fraction data itself, a model of atmospheric transport including a set of boundary conditions at the edge of the regional domain and some initial information or "first guess" of regional CO_2 fluxes. These components are combined in an inversion set-up with a mechanism for dealing with uncertainties in the inputs. To make the problem computationally manageable, the regional domain

Deleted: Jones et al., 2006)

Deleted: Lunt et al. (2016)

Deleted: CARbon DAta MOdel fraMework (CARDAMOM)

Deleted: CARDAMOM

Deleted: framework ingesting

Deleted: to retrieve process parameters for the DALEC carbon balance model

Deleted: CARDAMOM

is often decomposed into a number of basis functions, describing a spatial grouping of grid cells within which fluxes are scaled up or down. The selection of these basis functions constitutes a further key element of the atmospheric inverse problem.

2.1 Site location and measurements

This study focuses on the years 2013 and 2014. During this period, atmospheric CO_2 mole fractions were continuously measured at six sites across the UK and Republic of Ireland (see Table 1 for site information and Fig. 1 for the location of the sites). Four of these sites originally formed the UK-DECC network and are described in Stanley et al. (2018), whilst two were developed under the GAUGE programme and are described in Stavert et al. (2018). The site at Mace Head, Republic of Ireland, is a coastal, 10 m above ground level (a.g.l.), station situated primarily to measure concentrations of background air arriving at the site from the Atlantic Ocean. The Laboratoire des Sciences du Climat et de l'Environnement (LSCE) is responsible for making CO_2 measurements at this site from a 23 m.a.g.l. inlet (see Vardag et al., 2014 for a full site description). All of the UK sites are tall-tower stations (with inlets ranging from 42 to 248 m.a.g.l), designed to measure elevated greenhouse gas mole fractions as air is transported over the surface in the UK and Europe.

Continuous CO_2 measurements are made at all stations using cavity ring-down spectrometers (CRDS: Picarro G2301 or G2401). CRDS data are corrected for daily linear instrumental drift using standard gases and for instrumental non-linearity using calibration gases, spanning a range of above and below ambient mole fractions, on a monthly basis (Stanley et al., 2018). Calibration and standard gases are of natural composition and calibrated at the GasLab Max Planck Institute for Biogeochemistry, Jena, or the World Calibration Centre for CO_2 at Empa, linking them to the World Meteorological Organisation (WMO) X2007 scale (Stanley et al., 2018; Stavert et al., 2018). At sites with multiple inlets, measurements are taken for the same length of time at each inlet, each hour. This means that measurements at each height at Bilsdale and Tacolneston (with 3 inlets) are taken continuously for roughly 20 minutes every hour, and at Heathfield and Ridge Hill (with 2 inlets) measurements are taken continuously for roughly 30 minutes at each inlet every hour. For the purposes of the inverse modelling carried out in this study, the continuous CRDS data are used from the highest inlets and averaged to a 2-hour time resolution. Further information about the instruments, measurement protocol and uncertainty estimates can be found in Stanley et al. (2018) and Stavert et al., (2018).

2.2 Atmospheric transport model

In this work we use a Lagrangian particle dispersion model (LPDM), NAME, which tracks thousands of particles back in time from observation locations. The model determines the locations where air masses interacted with the surface, and therefore where surface CO_2 sources and sinks could contribute to a CO_2 concentration measurement. The model provides a gridded sensitivity of each mole fraction observation to the potential flux from each grid cell and this is often referred to as the "footprint" of a particular observation (for further details, see e.g. Manning et al., 2011).

Deleted: Four of these sites originally formed the UK-DECC network and are described in (Stanley et al., 2018)

Deleted: (Stavert et al., 2018)

Deleted: inlet (see Vardag et al., 2014

Field Code Changed

Deleted: Manning et al., 2011)

At each two-hourly measurement time step, the model releases 20,000 particles, which are tracked back in time for 30 days, so that by the end of this period the majority of particles will have left the model domain (Fig. S1). Since most CO_2 flux to the atmosphere occurs at the surface, we record the instances where the particles are in the lowest 40m of the atmosphere and assume that this represents the sensitivity of observed mole fractions to surface fluxes in the inversion domain. The domain used to calculate atmospheric transport covers most of Europe, the east coast of North and Central America and North Africa (-97.9˚ – 39.38˚ longitude and 10.729˚ – 79.057˚ latitude). The spatial resolution of the meteorological analysis dataset used to drive the model, from the Met Office Unified Model (Cullen, 1993), was 0.233˚ by 0.352˚ (roughly 25 by 25 km over the UK).

In many previous inverse modelling studies using LPDMs (e.g. Manning et al., 2011; Thompson and Stohl, 2014; Steinkamp et al., 2017) the footprint is assumed to be equal to the integrated air history over the duration of the simulation (e.g. 30 days, as in Fig. 1). Based on the assumption that fluxes have not changed substantially during the 30-day period, the integrated footprint can be multiplied by the prior flux and summed over all the grid cells in the domain to create a time series of modelled mole fractions at each measurement site. However, many CO_2 inverse modelling studies using other LPDMs have disaggregated footprints back in time, capturing changes in surface sensitivity on timescales shorter than the duration of the simulation, thereby attempting to account for diurnal variation in CO_2 fluxes (Denning et al., 1996; Gerbig et al., 2003; Gourdji et al., 2010). Thus far, a disaggregation such as this has not been used in NAME simulations, so we describe our method here.

In our simulations, we determined the footprint for 2-hourly average periods back in time for the first 24 hours before the observation, and then replaced the first 24 hours of integrated sensitivities with these time-disaggregated footprints. Mole fractions were simulated by multiplying these footprints by biospheric flux estimates for the corresponding time, so that the variability in the source or sink of CO_2 was represented in the modelled observations. This is demonstrated in Eq. 1, which yields the modelled mole fraction, y_t, for one 2-hourly measurement time step, t, at one measurement site.

$$y_t = \sum_{i=0}^{12}\sum_{j=0}^{n} fp_{t-i,j} \times q_{t-i,j} + \sum_{j=0}^{n} fp_{remainder_j} \times q_{month_j} \qquad (1)$$

Here i denotes the number of 2-hour periods back in time before the particle release at time t; j represents the grid cell where n is the maximum number of grid cells; $fp_{t-i,j}$ is one grid cell of the two-dimensional time-disaggregated footprint for that time; $q_{t-i,j}$ is one grid cell of the two-dimensional, two-hourly flux field corresponding to the time the particles were interacting with the surface; $fp_{remainder}$ is the remaining 29 day footprint and q_{month} is the monthly average flux. The choice of 24-hour disaggregation balanced considerations of computational efficiency and simulation accuracy. For certain months and sites we carried out a set of tests to determine how sensitive our simulated mole fractions and inversion results were when footprints were disaggregated for the first 12 to 72 hours prior to each measurement (Figure S2; Table S1). Assuming that the

Deleted: (e.g. Manning et al., 2011; Steinkamp et al., 2017; Thompson and Stohl, 2014)

Deleted: fp_{t-i}

Deleted: q_{t-i} is

Deleted: We find that the diurnal cycle of CO_2 flux strongly impacts the mole fraction observations in the first 24 hours of transport before an observation is made. However, the use of time-integrated footprints, multiplied by average fluxes for the remainder of the period incurs only minor errors in our simulation. An investigation of the impact of going further back in time (e.g. 72 hours) revealed some small differences in daily maximum and minimum CO_2 concentration on some days in a forward model run (Fig. S2). Nevertheless, posterior UK net biosphere fluxes were in agreement, within the estimated posterior uncertainties, between an inversion using footprints disaggregated for 24 hours and one using footprints disaggregated for 72 hours (Table S1).

72-hour simulations were the most accurate, we found little degradation in performance by using only 48 or 24 hours disaggregation, when compared to the other uncertainties in the system (e.g. differences between fluxes derived using the 24, 48 and 72 hour simulations were smaller than the 90% confidence interval). However, when only 12 hours was used (or fully integrated footprints), the modelled diurnal cycle was out of phase with the observations.

2.2.1 Data selection and model uncertainty

LPDMs are known to perform poorly under certain meteorological conditions. In particular, it is often assumed that model-data mismatch should be smallest during periods when the boundary layer is relatively well mixed. A common approach is to only include daytime data in the inversion (e.g. Meesters et al., 2012; Steinkamp et al., 2017; Kountouris et al., 2018a) or separate morning and afternoon averages (e.g. Matross et al., 2006). To make use of as much high frequency measurement information as possible, we use a filter based on two metrics to remove times of high atmospheric stability and/or stagnant conditions. The first metric is based on calculating the ratio of the NAME footprint magnitude in the 25 grid boxes in the immediate vicinity of the measurement station to the total for all of the grid boxes in the domain. A high ratio indicates times when a significant fraction of air influencing the observation point originates from very local sources, which may not be resolved by the model (Lunt et al., 2016). The second metric is based on the modelled lapse rate at each site, which is a measure of atmospheric stability. A high lapse rate suggests very stable conditions, which would be conducive for significant local influence. Thresholds for each of these criteria were chosen to preserve as much data as possible, whilst retaining only points that the model was (somewhat subjectively) found to resolve well. In practice, the filter retained many more daytime than night time points (see Fig. S3 for an analysis of the data removed in 2014) and inversion results were mostly similar to when only daytime data were used, however differences were seen in some months when stagnant conditions occurred for several daytime periods (Fig. S4).

Model uncertainty (or model-data mismatch) has a measurement uncertainty component and a component that takes into account the ability of the model to represent real atmospheric conditions. The measurement uncertainty was assumed to be equal to the standard deviation of the measurements over the 2-hour period to give an estimate of measurement repeatability and a measure of the sub-model-timescale variability in the observations. The 2-hourly measurement uncertainty was then averaged over the month to ensure that measurements of high concentrations were not de-weighted, as they are more likely to have greater variability and therefore a larger standard deviation. Monthly average measurement uncertainty is around 0.9 ppm. The measurement uncertainty is combined with a range of prior values for model uncertainty (as this is a poorly constrained quantity) and together the model-measurement uncertainty is one of the hyper-parameters solved in the inversion (further explained in Sect. 2.4.1).

Deleted: (e.g. Meesters et al., 2012; Steinkamp et al., 2017; Kountouris et al., 2018a)

Deleted: (SD)

Deleted: SD.

2.2.2 Boundary conditions

The footprints from the LPDM only take into consideration the influence on the observations of sources intercepted within the model domain. Therefore, an estimate of the mole fraction at the boundary must be made and incorporated into the simulated mole fractions. To estimate spatial and temporal gradients in these boundary conditions we use the global Eulerian Model for OZone And Related chemical Tracers (MOZART, Emmons et al., 2010). The model was run using GEOS-5 meteorology (Rienecker et al., 2011) and global biospheric fluxes from the NASA-CASA biosphere model (Potter, 1999), global ocean fluxes from Takahashi et al. (2009) and global anthropogenic fluxes from the Emission Database for Global Atmospheric Research (EDGAR, EC-JRC/PBL, 2011). When particles leave the NAME model domain, we record the time and location of the exit point. We then use MOZART to find the concentration of CO_2 at these locations to serve as prior boundary conditions. The global MOZART initial mole fraction field for January 2014 was scaled before commencing the 2014 MOZART run to match the surface South Pole value to the mean NOAA January 2014 flask value (Dlugokencky et al., 2018). This scaling factor was also applied to any pre-January 2014 MOZART output to prevent any discontinuities in the boundary mole fraction fields. The mole fraction at each domain edge (N, E, S, W) is then scaled up or down during the inversion to account for uncertainties in the MOZART boundary conditions (Lunt et al., 2016). A sensitivity test where 1ppm is added or taken away from the mole fractions at the domain edges indicates that in June a ±1ppm change translates to a 1-3% change in the inversion result and in December a ±1ppm change translates to a 7-11% change in the inversion result. These changes are substantially smaller than the posterior uncertainty.

Deleted: Emmons et al., 2010)

2.3 Prior information

In this work, we used model analyses to provide prior information about biospheric fluxes. Two models (DALEC and JULES) were used to assess how much influence the choice of biospheric prior has on the outcome of the inversion. The NAME model was used to simulate the contribution of anthropogenic and oceanic fluxes to the data, and this contribution was removed from the observations prior to the inversion. The fluxes used for this calculation are described below. The spatial and temporal resolution of the prior information and fixed fluxes are summarised in Table 2 and emissions from each source over the UK are shown in Figure 2.

Deleted: CARDAMOM

Deleted: is

Deleted: 3

In a synthetic data study in which biospheric CO_2 was inferred, Tolk et al., (2011) found that separately solving for positive fluxes (autotrophic and heterotrophic respiration combined, TER) and negative fluxes (GPP) in atmospheric inversions provided a better fit to the atmospheric mole fraction data than inversions that scaled NEE only. Equation 2 describes the relationship between these three variables:

$$NEE = TER - GPP \qquad (2)$$

This separation has been applied in various studies demonstrating model set-ups with synthetic data, for example: geostatistical approaches (Göckede et al., 2010), ensemble Kalman filter methods (Zupanski et al., 2007; Lokupitiya et al., 2008) and Bayesian methods (Schuh et al., 2009). However, this separation is not routinely used in CO_2 inversions, as there are only a limited number of real data studies where it has been implemented (e.g. Gerbig et al., 2003; Matross et al., 2006; Schuh et al., 2010; Meesters et al., 2012).

In this inversion, we separately solved for TER and GPP, and then combined them a posteriori to determine NEE. Similarly to the studies cited above, we find closer agreement with the data than if NEE were scaled directly. Furthermore, we note that, if only one factor is used to scale both TER and GPP, it is impossible for the inversion to respond to a prior that has, for example, too strong a sink but a source of the correct magnitude. To demonstrate this, we have carried out a synthetic test (Fig. S5) in which we have investigated the ability of our inversion system to solve for a "true" flux, created using the DALEC prior fluxes and NAME simulations, in an inversion that used the JULES fluxes as the prior. Figure S5(a) shows that monthly posterior fluxes for the inversion where GPP and TER are separated agree with the "true" flux within estimated uncertainties in 16 out of 24 months. In contrast, whilst the posterior fluxes for the inversion where NEE is scaled has changed significantly from the prior, it is not in agreement with the "true" flux except in July 2013 and August and September 2014. The posterior diurnal cycles of GPP, TER and NEE, which are shown as an average for June 2014 in Fig. S5(b) and Fig. S5(c), highlight the differences in diurnal cycle between the two models. The inversion that can adjust the two sources separately leads to higher night-time fluxes, which are closer to the "true" flux than the prior. On the other hand, the inversion where NEE is scaled can only stretch or shrink the diurnal cycle in one direction, increasing both the daytime sink and night-time source, or decreasing them, together. In this case, they have decreased, which does bring the net June 2014 flux in Fig. S5(a) closer to the "true" June 2014 flux but cannot go far enough to reconcile these monthly fluxes.

Given the results of our synthetic test, separating GPP and TER in the inversion appears to be an important improvement on scaling NEE directly and it is what we have implemented here. However, in addition to the main inversions presented in this paper, where GPP and TER are separated, we have carried out two further inversions for JULES and DALEC where only NEE is scaled. The results of these additional inversions are discussed in Sect. 4.1.

2.3.1 DALEC biospheric fluxes

DALEC is a simplified terrestrial C-cycle model (Smallman et al., 2017) that uses location specific ensembles of process parameters and initial conditions retrieved using the CARDAMOM model-data fusion approach (Bloom et al., 2016). CARDAMOM uses a Bayesian approach within a Metropolis-Hastings MCMC algorithm to compare model states and flux estimates against observational information to determine the likelihood of potential parameter sets guiding the parameterisation processes at pixel scale. DALEC simulates the ecosystem carbon balance, including uptake of CO_2 via photosynthesis, CO_2

Deleted: (e.g. Gerbig et al., 2003; Matross et al., 2006; Schuh et al., 2010; Meesters et al., 2012)
Deleted: resolve
Deleted: CARDAMOM
Deleted: but with a GPP that has been halved
Deleted: full CARDAMOM
Deleted: a
Deleted: the seasonal cycle of
Deleted: is able to change phase and manages to approach
Deleted: .
Deleted: seasonal cycle of
Deleted: can only move towards
Deleted: by shallowing the amplitude without changing the phase, because reducing the sink means the source must also be reduced. This can also be seen in the
Deleted:).
Deleted: separates
Deleted: is able to correctly estimate the diurnal cycle of the "true" GPP flux…
Deleted: allows it to find a better estimate for
Deleted: of NEE than the case where NEE is scaled. Whilst
Deleted: is an extreme case, a comparison of the CARDAMOM and JULES estimates demonstrates large relative differences in TER and GPP…
Deleted: would be "hard wired" if only NEE were scaled (Figure
Deleted: b)).
Deleted: simple
Deleted: CARDAMOM
Deleted: **CARDAMOM**
Deleted: The CARbon DAta MOdel fraMework (CARDAMOM) uses a model-data fusion approach to retrieve location specific ensembles of parameters for the DALEC model (Bloom et al., 2016). CARDAMOM uses a Bayesian approach within a Metropolis-Hastings MCMC algorithm to compare model states and flux estimates against observational information to determine the likelihood of potential parameter sets guiding the parameterisation processes. DALEC simulates the ecosystem carbon balance including uptake of CO_2 via photosynthesis, CO_2 loss via respiration, mortality and decomposition processes, including carbon flows between ecosystem pools (non-structural carbohydrates, foliage, fine roots, wood, fine litter, coarse woody debris and soil organic matter). GPP or photosynthesis is estimated using the aggregated canopy model (ACM; Williams et al., 1997) while autotrophic respiration is estimated as a fixed fraction of GPP. Mortality and decomposition processes follow first order kinetic equations (i.e. a daily fractional loss of the C stock in question). The mortality and decomposition parameters are modified based on an exponential temperature sensitivity parameter. Note that the current version of DALEC does not include a representation of the water cycle, rather water stress is ... [1]

loss via respiration, mortality and decomposition processes, and carbon flows between ecosystem pools (non-structural carbohydrates, foliage, fine roots, wood, fine litter, coarse woody debris and soil organic matter). GPP, or photosynthesis, is estimated using the aggregated canopy model (ACM; Williams et al., 1997) while autotrophic respiration is estimated as a fixed fraction of GPP. Canopy phenology is determined by a Growing Season Index (GSI) model as a function of temperature, day length and vapour pressur deficit (proxy for water stress). Mortality and decomposition processes follow first order kinetic equations (i.e. a daily fractional loss of the C stock in question). The decomposition parameters are modified based on an exponential temperature sensitivity parameter. The current version of DALEC used here does not include a representation of the water cycle; rather, water stress is parameterised through a sensitivity to high vapour pressure deficit as part of the GSI phenology model. Comprehensive descriptions of CARDAMOM can be found in Bloom et al. (2016) and DALEC in Smallman et al. (2017).

DALEC estimates carbon fluxes at a weekly time step and 25 km × 25 km spatial resolution. The weekly time step information was downscaled to 2-hourly intervals, assuming that each day repeated throughout each week. Downscaling of GPP fluxes was assumed to be distributed through the daylight period based on intensity of incoming shortwave radiation. Respiration fluxes were downscaled across the full diurnal cycle assuming exponential temperature sensitivity (code for downscaling is available from the authors on request).

Observation-derived information used in the current analysis are satellite-based remotely sensed time series of Leaf Area Index (LAI) (MODIS; MOD15A2 LAI-8 day version 5, http://lpdacc.usgs.gov/), a prior estimate of above ground biomass (Thurner et al., 2014) and a prior estimate of soil organic matter (Hiederer and Köchy, 2012). Meteorological drivers were taken from ERA-Interim reanalysis. Ecosystem disturbance due to forest clearances were imposed using Global Forest Watch information (Hansen et al., 2013). CARDAMOM-DALEC differs from typical land surface models in using these data to generate probabalistic model parameterisations and initial conditions estimates for each pixel, with no a priori assumptions about plant functional types, nor steady states.

Deleted: Observational

Deleted: /)

2.3.2 JULES biospheric fluxes

JULES is a process driven Land Surface Model (LSM) that estimates the energy, water and carbon fluxes at the land-atmosphere boundary (Best et al., 2011; Clark et al., 2011). We used JULES version 4.6 driven with the WATCH Forcing Data methodology applied to Era-Interim reanalysis data (WFDEI) meteorology (Weedon et al., 2014) which were interpolated to a 0.25° × 0.25° grid (Schellekens et al., 2017). We prescribed the land cover for 9 surface types and the vegetation phenology for 5 plant functional types (PFTs) using MODIS monthly LAI climatology and fixed MODIS land cover and canopy height data (Berry, et al., 2009). The soil thermal and hydrology physics are described using the JULES implementation of the Brooks and Corey formulation (Marthews et al., 2015) with the soil properties sourced from the Harmonised World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009). Soil carbon was calculated as the equilibrium balance between litter-fall and soil

Deleted: The Joint UK Land Environment Simulator (JULES)

respiration for the period 1990-2000 using the formulation of (Mariscal, 2015). The full JULES configuration and science options are available for download from the Met Office science repository (https://code.metoffice.gov.uk/trac/roses-u/browser/a/x/0/9/1/trunk?rev=75249).

2.3.3 Anthropogenic fluxes

Estimates of fluxes due to anthropogenic activity within the UK were obtained from the National Atmospheric Emissions Inventory (NAEI, http://naei.beis.gov.uk). The NAEI provides a yearly estimate of emissions, which we have disaggregated into a 2-hourly product, based on temporal patterns in activity data, varying on diurnal, weekly and seasonal scales. The inventory emissions were disaggregated according to the UNECE/CORINAIR Selected Nomenclature for sources of Air Pollution (SNAP) sectors (UNECE/EMEP, 2001). Figure 2(d) shows the seasonal and diurnal cycle for this inventory, summed over the UK, for 2014. Outside the UK, anthropogenic emissions come from EDGAR v4.2 FT2010 inventory data for 2010 (EC-JRC/PBL, 2011). This is a fixed 2D map that is used throughout the inversion period. Within the UK, the NAEI and EDGAR fluxes differ by around 15% (540 Tg yr^{-1} for EDGAR, 460 Tg yr^{-1} for NAEI). We do not find that our derived UK fluxes are significantly affected by perturbations of this magnitude applied to anthropogenic emissions outside the UK.

2.3.4 Ocean fluxes

Ocean flux estimates are from Takahashi et al. (2009). They are based on a climatology of surface ocean pCO_2 constructed using measurements taken between 1970 and 2008. The monthly UK coastal ocean flux (defined as the UK's exclusive economic zone) from this product is plotted in Fig. 2(e). Since the oceanic flux component is small, the comparatively low temporal and spatial resolution of these flux estimates does not significantly impact the inversion results.

2.4 Inverse method

2.4.1 Hierarchical Bayesian trans-dimensional inversion

Like many atmospheric inverse methods, our framework is based on traditional Bayesian statistics, given by Eq. 3:

$$\rho(\boldsymbol{x}|\boldsymbol{y}) = \frac{\rho(\boldsymbol{y}|\boldsymbol{x})\rho(\boldsymbol{x})}{\rho(\boldsymbol{y})} \qquad (3)$$

where $\boldsymbol{y}$ is a vector containing the observations and $\boldsymbol{x}$ is a vector of the parameters to be estimated (such as the flux and boundary condition scaling). The traditional Bayesian approach requires that decisions about the form of the prior PDF, $\rho(\boldsymbol{x})$, and likelihood function, $\rho(\boldsymbol{y}|\boldsymbol{x})$, are made a priori. These pre-defined decisions have the potential to strongly influence the form of the posterior PDF in an inversion (Ganesan et al., 2014). Instead, we introduce a second "level" to the traditional Bayes equation, to account for the fact that initial parameter uncertainty estimates are themselves uncertain. This is known as a

Deleted: 2

Deleted: prior

Deleted: the Emission Database for Global Atmospheric Research (

Deleted:)

Deleted: 3

Deleted: Prior ocean

Deleted: simply

Deleted: grid cells closest to the UK, up to a maximum distance of 500km…

Deleted: Bayesian

"hierarchical" Bayes framework where additional parameters, known as hyper-parameters, are used to describe the uncertainties in the prior and the model.

Alongside the additional hyper-parameters $\boldsymbol{\theta}$, we also introduce an additional term, k, that describes the size of the inversion grid, following the trans-dimensional inversion approach described in Lunt et al. (2016). In this approach, the number of basis functions to be solved is not fixed a priori and hence $\boldsymbol{x}$ has an unknown length. The number of unknowns is itself a parameter to be solved for in the inversion, with the uncertainty in this term propagating through to the posterior parameter estimates, more fully accounting for the uncertainties that are only tacitly implied within a traditional Bayesian approach. The full trans-dimensional hierarchical Bayesian equation that is solved in our inversion thus becomes:

$$\rho(\boldsymbol{x},\boldsymbol{\theta},k|\boldsymbol{y}) \propto \rho(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta},k)\rho(\boldsymbol{x}|\boldsymbol{\theta},k)\rho(k)\rho(\boldsymbol{\theta}) \qquad (4)$$

where θ is a set of hyper-parameters describing the uncertainty on $\boldsymbol{x}$ (σ_x), the model-measurement error (σ_y), and the correlation timescale in the model-measurement covariance matrix (τ). These hyper-parameters are summarised in Table 3 along with the prior PDFs used to describe them in this inversion set-up.

In this study, we have adapted the trans-dimensional method to keep a fixed set of regional basis functions (described in Sect. 2.4.3) but allow the inversions to have a variable *time* rather than *space* dimension. We perform our inversion calculations over one month at a time, but with the trans-dimensional case in *time* we find multiple scaling factors for each fixed region over the course of the inversion, down to a minimum daily resolution. Therefore, in this case k in Eq. 4 is more specifically the unknown number of time periods resolved in the inversion, which is important because CO_2 fluxes vary strongly in time and have high uncertainty in their temporal variation.

In general, there is no analytical solution to our hierarchical Bayesian equation, so we approximate the posterior solution using a reversible jump Metropolis-Hastings MCMC algorithm (Metropolis et al., 1953; Green, 1995; Tarantola, 2005; Lunt et al., 2016). The algorithm explores the possible values for each parameter by making a new proposal for a parameter value at each step of a "chain" of possible values. Proposals are accepted or rejected based on a comparison between the "current" and "proposed" state's fit to the data (likelihood ratio), deviation from the prior PDF (prior ratio), and a term governing the probability of generating the proposed state versus the reverse proposal (proposal ratio). More favourable parameter values or model states are always accepted; however, less favourable parameter values or model states can be randomly accepted in order to fully explore the full posterior PDF. The algorithm had a burn-in period of 5×10^4 iterations and was then run for an additional 2×10^5 iterations to appropriately explore the posterior distribution. At the end of the algorithm a chain of all accepted parameter values is stored (if a proposal is rejected the chain will spend longer at the previously accepted value). A histogram of this chain describes a posterior PDF for each parameter so that statistics such as the mean, median and standard

Deleted: Lunt et al. (2016)

Deleted: parameters estimate

Deleted: which

Deleted: 2

Formatted: Font color: Black

Formatted: Font color: Black

Deleted: for the highly variable temporal nature of CO_2 fluxes

Formatted: Font color: Black

Deleted: the

deviation can be calculated. The trace of each chain was examined qualitatively to ensure that the algorithm had been run for a sufficient number of iterations to converge on a result.

2.4.2 Basis functions

Our domain is split into five spatial regions separating West-Central Europe from North-East, South-East, South-West and North-West regions, shown in Fig. S1. Within the West-Central Europe area (the hatched region in Fig. S1), a map of the fraction of different plant functional types (PFTs) in each grid cell has been used to further break down the region (Fig. S6). This is the same PFT map used in the JULES biospheric simulation (see Sect. 2.3.2). A scaling factor is solved in the inversion, scaling GPP and TER within the 4 outer regions and within maps of five or six PFTs in the sub-domain: broadleaf tree, needleleaf tree, C3 grasses, C4 grasses, shrubland and, in the case of TER, bare soil. Therefore there are 19 spatial basis functions in total.

Deleted: hashed

Deleted: .

2.4.3 Definition of Jacobian matrix

Footprints from NAME, prior fluxes, boundary conditions and basis functions are all combined into a matrix of partial derivatives, alternatively described as a "Jacobian" or "sensitivity" matrix, that describes the change in mole fraction with respect to a change in each of the input parameters. This is the "model" in the inversion set-up, denoted **H** in the description of the linear forward model (Eq. 7), where $\boldsymbol{\varepsilon}$ is the mismatch between modelled "observations" and what has actually been measured in the atmosphere. **H** has dimensions m (number of data points) by n (number of parameters).

Deleted: derivitives

Deleted: basis functions

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\varepsilon} \qquad (7)$$

To create this linear model, we multiplied the footprints by the prior GPP and TER fluxes separately, then multiplied these by the fractional map of basis functions (described in Sect. 2.4.2) and summed over the domain. The boundary conditions were broken down by four further basis functions for each edge of the domain as explained in Sect. 2.2.2. The parameters vector, $\boldsymbol{x}$, consisted of a set of scaling factors that multiplied the fluxes or boundary conditions. Multiplying the sensitivity matrix by the prior estimate of $\boldsymbol{x}$, a vector of ones, yields the prior modelled mole fraction time-series at a site. Therefore, during our inversion, we are updating this vector of ones as a scaling factor, to scale up or down emissions for each PFT and biospheric component to better agree with the data. Whilst in theory we have posterior information about the gross GPP and TER biospheric components separately, we combine this into a net ecosystem exchange (NEE) flux estimate, as we believe this to be more robust (Tolk et al., 2011). Therefore, throughout this paper we discuss posterior NEE estimates, however the results of the separate sources can be found in the supplement in Fig. S7-S9.

Deleted: area covered by each basis function.

Deleted: S5-

3 Results

We have applied our CO_2 inversion set-up to UK biospheric CO_2 flux estimation using output from two different models of biospheric flux as a prior constraint in two inversions. We first describe differences between the output from the two prior models, then present the UK flux estimates found with this method, along with the spatial distribution of posterior fluxes.

3.1 Differences between DALEC and JULES

The CO_2 fluxes from DALEC and JULES differ both temporally and spatially. Figure 2 (a-c) shows UK fluxes of GPP, TER and NEE from the two models. Most notable differences are seen in TER where JULES has a large diurnal range whereas DALEC has a small diurnal range. Averaged to monthly resolution, the fluxes are relatively similar although DALEC has a higher TER flux from July to October. Diurnal ranges for GPP are more similar in magnitude, however JULES exhibits a stronger sink in spring with maximum uptake in June. DALEC has maximum uptake in July and exhibits a stronger sink in autumn. Combining these two fluxes, we can see that the profile of NEE for both models is quite different. The daily maximum source from JULES remains relatively constant throughout the year, whereas the daily maximum source in DALEC follows a similar seasonal cycle to the daily maximum sink (albeit with a smaller magnitude). Monthly net fluxes are similar between both models for much of the year although JULES has stronger uptake between March and June.

In order to understand some of these seasonal differences it is useful to compare the processes taking place in each model. Sections 2.3.1 and 2.3.2 provide detailed descriptions of each model and we give an overview of the main differences here. DALEC explicitly simulates the soil and litter stocks, growth and turnover processes. LAI is estimated by DALEC at a weekly time step; DALEC was calibrated using MODIS LAI estimates at the correct time and location of the analysis, explained in Sect. 2.3.1. In the JULES system, soil and litter carbon stocks are fixed values for each grid cell, calibrated from 1990-2000, and a fixed climatology of MODIS LAI and canopy height is used. Therefore, DALEC has interannual variability in LAI and soil carbon stocks and can adjust the parameters to find the most likely estimates in combination with other data, whereas these parameters remain constant in JULES. This is potentially advantageous for DALEC, although the use of a climatology in JULES means that noise in the MODIS LAI estimates will be averaged out. Since, LAI and soil and litter carbon stocks are fixed in JULES, variability in TER and GPP fluxes are governed by meteorology, primarily temperature but also significant signals from photosynthetically active radiation and precipitation via the soil moisture. Meteorology drives the JULES model at a 2-hourly timestep as opposed to a weekly time-step in DALEC. Therefore, in the 2-hourly DALEC product used here, the diurnal range is not explicitly simulated and is the result of a downscaling process from a weekly resolution. This downscaling is done based on light and temperature curves as explained in Sect. 2.3.1. In DALEC, the autotrophic respiration is parameterised as a fixed fraction of the GPP for a given site but varies between sites, roughly ranging from 0.3 to 0.7. In JULES, the autotrophic respiration is the sum of plant maintenance and growth respiration terms, which are calculated separately as process based functions of the GPP, the maximum rate of carboxylation and leaf nitrogen content (Clark et al.,

Deleted: tested

Deleted: and find estimates for UK biospheric CO_2 flux from these

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: we can

Deleted: The CARDAMOM system

Deleted: retrieved from the

Deleted: model (which

Deleted:) and updated at each weekly time step

Deleted: it uses

Deleted: from JULES

Deleted: (PAR)

Deleted: This gives CARDAMOM interannual variability in LAI and soil carbon stocks, whereas these parameters remain constant in JULES year to year. However, meteorology

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: downscaled

Deleted: In both models, the autotrophic respiration is a fixed fraction of the GPP, roughly ranging from 0.1 to 0.5. Therefore, there are some large differences between the model processes

2011). Typically, the autotrophic respiration in JULES is roughly 0.1-0.25 of the GPP. Therefore, there are some large differences between the model structures and parameterisations, particularly in how the respiration fluxes are simulated. This could be leading to too small a diurnal range in DALEC TER and too large a diurnal range in JULES TER.

Figures 3 and 4 show spatial maps of GPP, TER and NEE from both models averaged over winter (December, January, February) and summer (June, July, August) months. The pattern of TER is similar for both models, however JULES always has a stronger source over Northern Ireland and DALEC has a stronger source in east England. In winter there are only small spatial variations in DALEC GPP fluxes, whereas JULES has its largest uptake in south-west England and Wales. In summer, the models are roughly in agreement in the size of the sink in Wales and the majority of England, however JULES has a stronger sink in Scotland and Northern Ireland and DALEC has a stronger sink in central and south-east England. The differences between the models in GPP and TER lead to fairly different winter NEE flux maps. DALEC is a net source everywhere in winter, with areas of strongest net source in southern Scotland, east and central England. JULES is a small net winter sink in Northern Ireland, Wales, and south and central England. Summer NEE fluxes are similar between the models, although JULES has a stronger net sink in Scotland and Northern Ireland.

3.2 Posterior net UK biospheric CO_2 flux 2013-2014

We have derived estimates for annual NEE from the UK using CO_2 flux output from the two different models of biospheric flux as prior information (Fig. 5 – orange and blue bars for DALEC and JULES respectively): $-13 \pm^{90}_{87}$ Tg CO_2 yr^{-1} (DALEC prior) and $-76 \pm^{91}_{90}$ Tg CO_2 yr^{-1} (JULES prior) in 2013 and $-2 \pm^{70}_{68}$ Tg CO_2 yr^{-1} (DALEC) and $-51 \pm^{80}_{78}$ Tg CO_2 yr^{-1} (JULES) in 2014. These annual net flux estimates from both models agree within the estimated uncertainties and mean values are higher than their respective priors in both cases. The uncertainties straddle the zero net flux line implying that the UK is roughly in balance between sources and sinks of biospheric CO_2. However, according to the inversion using JULES, a net biospheric source is less likely than in the inversion using DALEC. When added to the anthropogenic and ocean fluxes that remained fixed during the inversion we produce the following estimates for annual total net CO_2 release from the UK (Fig. 5 – yellow and green bars for DALEC and JULES respectively): $448 \pm^{90}_{87}$ Tg CO_2 yr^{-1} (DALEC prior) and $386 \pm^{91}_{90}$ Tg CO_2 yr^{-1} (JULES prior) in 2013 and $418 \pm^{70}_{68}$ Tg CO_2 yr^{-1} (DALEC prior) and $369 \pm^{80}_{78}$ Tg CO_2 yr^{-1} (JULES prior) in 2014. While we are assuming that anthropogenic and ocean fluxes are perfectly known, the uncertainties on these fluxes are comparatively small (Peylin et al., 2011). When the anthropogenic source was varied by $\pm 10\%$, a conservatively large estimate of these uncertainties, we find posterior biospheric flux estimates using the DALEC prior that still suggest a balanced biosphere, and posterior flux estimates using the JULES prior that suggest a small net sink at the lowest end of the possibilities explored here (see Fig. S10). All mean annual posterior estimates, regardless of the anthropogenic source used, suggest the prior net biospheric flux is underestimated, i.e. posterior biospheric uptake of CO_2 is smaller than predicted by the models. However, this is less statistically significant with the 2013 inversion using the DALEC prior.

Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: blue and
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted:
Deleted: CARDAMOM
Deleted: green and
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: S11
Deleted: CARDAMOM

The monthly posterior UK estimates using both models (Fig. 5) mostly agree well with each other within the uncertainties, however they are both notably different from the prior estimates, especially in 2014. The posterior total UK flux estimate, achieved by adding the posterior NEE fluxes to anthropogenic and coastal ocean fluxes, shows that, according to the DALEC inversion, the UK may not be a net sink of CO_2 at any time of year in 2013 and 2014. However, the JULES inversion suggests the UK is a net sink of CO_2 in June of both years.

Posterior seasonal cycle amplitudes are generally smaller than the prior amplitudes, except in the DALEC inversion in 2014. Table 4 gives the posterior maximum and minimum values of NEE, leading to seasonal cycle amplitudes of 469 Tg CO_2 yr^{-1} and 578 Tg CO_2 yr^{-1} for 2013 and 633 Tg CO_2 yr^{-1} and 737 Tg CO_2 yr^{-1} for 2014, for the DALEC and JULES inversions respectively. These values are 90% and 76% of the prior amplitudes in 2013 and 123% and 85% of the prior amplitudes in 2014.

The largest differences between the prior and posterior are seen in spring and summer for both models. Posterior UK NEE estimates from the DALEC inversion are in agreement with the prior for 11 months: during the first half of 2013, in the majority of winter months (December, January, February) and in June 2014. When the DALEC inversion posterior UK NEE estimates are not in agreement with the prior, they are usually larger, with a maximum difference in 2013 of $235 \pm^{92}_{91}$ Tg CO_2 yr^{-1} in August and a maximum difference in 2014 of $551 \pm^{80}_{84}$ Tg CO_2 yr^{-1} in July, although in spring (March, April, May) 2014 they tend to be smaller than the prior, with a maximum difference of $-194 \pm^{64}_{60}$ Tg CO_2 yr^{-1} in April. Posterior UK NEE from the JULES inversion agrees with the prior for nine months during the two-year period, the majority of which are between November and February. Otherwise, the posterior estimate from the JULES inversion is larger than the prior with a maximum difference in 2013 of $318 \pm^{70}_{71}$ Tg CO_2 yr^{-1} in April and a maximum difference in 2014 of $407 \pm^{76}_{72}$ Tg CO_2 yr^{-1} in July.

Looking at the spring and summer differences more closely, we find that the JULES model has a systematically lower net spring flux than the posterior, and the DALEC model is either in agreement with or higher than the posterior estimate of the net spring flux. Generally, the models underestimate the net summer flux compared to the posterior flux (to the greatest extent in 2014), although the summer estimate from the JULES inversion in 2013 is not statistically different from the prior. The average spring difference between the posterior and the prior for the DALEC inversion is $-2 \pm^{89}_{88}$ Tg CO_2 yr^{-1} in 2013 and $-133 \pm^{67}_{63}$ Tg CO_2 yr^{-1} in 2014, whereas for the JULES inversion it is 219 ± 87 Tg CO_2 yr^{-1} in 2013 and $164 \pm^{67}_{65}$ Tg CO_2 yr^{-1} in 2014. The average summer difference for the DALEC inversion is $135 \pm^{111}_{108}$ Tg CO_2 yr^{-1} in 2013 and $263 \pm^{82}_{83}$ Tg CO_2 yr^{-1} in 2014, whereas for the JULES inversion it is $94 \pm^{104}_{107}$ Tg CO_2 yr^{-1} in 2013 and 312 ± 85 Tg CO_2 yr^{-1} in 2014. The prior sink in June as estimated by the JULES model is nearly twice that of DALEC and posterior estimates tend to agree with the DALEC prior in this month.

Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: 5
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: are underestimating
Deleted:)
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM

Figure S9(c) shows the daily minimum and maximum in the posterior net biospheric estimates for 2014. It is worth bearing in mind at this point that while the temporal resolution of the inversion is flexible, it can go down to a minimum resolution of one day (as explained in Sect. 2.4.1). Therefore, the diurnal profile of TER and GPP for each model is imposed, however it can be scaled up or down from day to day. Fig. S11 shows that the inversion typically scaled the fluxes within 4 or 5 temporal regions per month, although for some parameters in some months scaling factors were found up to roughly a daily resolution. For both inversions, the posterior NEE flux has a similar profile. Compared to Fig. 2(c) the inversion tends to a seasonal cycle in daily maximum uptake that resembles that of the JULES model prior, with a turning point in maximum uptake occurring abruptly between June and July, a steep gradient in spring and a shallow gradient in autumn. On the other hand, the seasonal cycle in daily maximum source resembles that of the DALEC model prior, which has a stronger seasonal variation compared to that of the JULES model prior, albeit with a larger amplitude. This would suggest that the underestimation in net spring flux seen in the JULES prior is generally due to the model underestimating the spring source, rather than overestimating the spring sink. It also suggests that the overestimation in net summer flux in the DALEC prior is possibly a combination of the model overestimating the summer sink and underestimating the summer source. The overestimation in the net summer flux in JULES is more likely to be due to an underestimation of the summer source. However, as diurnal fluxes vary on a scale nearly an order of magnitude larger than that of the monthly fluxes, it is clear that any relatively small changes in the maximum source or sink will have a relatively large effect on the daily net flux. Therefore, the monthly net flux is the more robust result here and we are not able to confidently draw conclusions from the sub-monthly results.

Deleted: S8

Deleted: CARDAMOM

Deleted: CARDAMOM

3.3 Posterior spatial distribution of biospheric fluxes

Figure 6 shows mean posterior net biospheric fluxes (NEE) for winter 2013 and summer 2014 from both the DALEC and JULES inversions. In winter 2013, posterior NEE fluxes from the DALEC inversion are fairly heterogeneous and are largest over south-west Scotland and east and central England. This posterior spatial distribution is roughly similar to the prior. From the inversion using JULES prior fluxes, the posterior net biospheric flux is much smoother than it is for the inversion using DALEC. It is largest in north-west England and almost zero in east England. The whole of south/central England, Wales, and Northern Ireland have increased posterior winter fluxes compared to the prior, turning these areas from a net sink in the prior to a net source in the posterior.

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

In summer 2014, NEE fluxes from the two inversions display many similarities, with areas of net source in east, central (extending further south in the JULES inversion) and north-west England and areas of net sink elsewhere. However, the net sink in the JULES inversion is larger than the DALEC inversion in Scotland, south Wales, Northern Ireland and south-west England. This differs from the prior flux maps, which have only very small areas of small net uptake in central England in DALEC and in east England in JULES. Both the DALEC and JULES posterior fluxes generally display reduced uptake compared to the prior, except in north Wales.

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

3.4 Model-data comparison

Agreement between the data and the posterior simulated mole fractions at the measurement sites used to constrain the inversion is greatly improved compared to prior simulated mole fractions, with R^2 values increasing by a minimum of 0.24 and up to 0.5 (to give values ranging between 0.53 and 0.71) and root mean square error (RMSE) decreasing by at least 1.35 ppm and up to 2.6 ppm (to give values ranging between 1.26 ppm and 2.71 ppm). Table 5 shows all statistics for the prior and posterior mole fractions compared to the observations of atmospheric CO_2 concentrations. Overall, the fits are relatively similar between the DALEC and JULES inversions, implying that the two inversions perform similarly well by these metrics. In terms of R^2, the best fit to the data is observed at Heathfield in the DALEC inversion and Angus in the JULES inversion. In terms of RMSE, the best fit to the data is observed at Angus in the DALEC inversion and Mace Head in the JULES inversion. The smallest posterior mean bias is observed at Angus in the DALEC inversion and Ridge Hill in the JULES inversion. Therefore, there are some small spatial differences in how well each of the inversions is able to fit the data but no clear indication of which areas of posterior flux might be subject to the largest improvement in either inversion. Figures S12 and S13 show the residual mole fractions in 2014 and indicate that residuals are somewhat larger during the summer than the winter.

To test our posterior results against data that has not been included in the inversion, the posterior fluxes have been used to simulate mole fractions at Weybourne Atmospheric Observatory (see Fig. 1 for location in relation to the other sites and Table 1 for site information). The statistics of fit to the data are given in italics in Table 5 and show an improvement in R^2 of 0.18 with the DALEC inversion and 0.13 with the JULES inversion, an improvement in RMSE of 1.09 ppm with the DALEC inversion and 0.75 ppm with the JULES inversion and an improvement in the mean bias of 0.64 ppm in the DALEC inversion and 0.56 in the JULES inversion. These results show that the a posteriori fluxes improve the fit to the data at a measurement station not included in the inversion. The results are very similar between the two inversions at this site, but suggest that the DALEC inversion may perform slightly better, at least in this region of the UK. Figure S14 shows the residual mole fractions at Weybourne for each of the inversions carried out in this work.

4 Discussion

4.1 Inversion performance

Solving for both TER and GPP separately allows the JULES-prior and DALEC-prior inversions to converge to a similar posterior solution. Using two very different prior NEE flux estimates, we produce two similar posterior NEE flux estimates that have a similar seasonal amplitude, and agree on the majority of monthly and all annual fluxes within the estimated uncertainties. This indicates that our posterior estimates are driven by the data rather than determined by the prior. However, when we carry out the same inversion but scale NEE (Fig. S15) we find the two posterior flux estimates do not converge on a common result. The posterior seasonal cycles remain relatively unchanged compared to the prior and annual net biospheric

Deleted: Simulated
Deleted: are
Deleted: after the inversion
Deleted: 29
Deleted: 51
Deleted: 54
Deleted: 74
Deleted: 0.29
Deleted: 51
Deleted: 34
Deleted: 56
Deleted: S2
Deleted: concentration.
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: CARDAMOM
Deleted: S9
Deleted: S10
Deleted: . The Figures show
Formatted: Font color: Auto
Deleted: CARDAMOM
Deleted: results
Deleted: S12

flux estimates tend to be similar to, or larger than, the prior. These annual net biospheric flux estimates are therefore 3 – 39 times smaller than the inversion that separates GPP and TER, meaning the posterior estimates from the two types of inversions do not overlap, even within estimated uncertainties. Evaluating the statistics of how well the NEE inversions fit the data (Table 5), we find they do not perform as well as the separate GPP and TER inversion, both at the sites included in the inversion and at the validation site, WAO. However, this is to be expected to some degree, because separating the two sources gives the inversion more degrees of freedom to fit the data.

As recommended by Tolk et al. (2011), we are only hoping to achieve an improved estimate for the net fluxes here rather than the gross GPP and TER fluxes themselves. The posterior gross fluxes are included in the supplement (Fig. S7-S9) but due to the correlation between the spatial and temporal distribution of GPP and TER they have not been presented in the main text. This can be seen in summer and winter flux maps (Fig. S7 and S8) and in the posterior annual flux estimates in Fig. S9(d), in particular where JULES TER and GPP show similarly large differences from the prior. This could also be a result of the imposed diurnal cycle, as it would appear the posterior TER flux in the JULES inversion is tending to a higher daily minimum, matching that of the DALEC prior, and may ultimately be trying to move towards a smaller diurnal variation in TER. However, because the whole diurnal cycle must be scaled, the daily maximum TER must also increase and may mean the GPP must increase, causing increased uptake, to compensate for the increased source from TER. Allowing flexibility on sub-daily timescales may lead to similar estimates of GPP and TER between the two inversions with different priors. However, questions remain over whether there is enough temporal information for this to be the case.

The fact that common monthly and annual posterior net biospheric flux estimates are reached when the prior biospheric fluxes are spatially and temporally different would suggest that the choice of prior is not necessarily a major factor in guiding the inversion result for our network, when GPP and TER are scaled separately. In this respect, it is also particularly encouraging that the seasonal cycles in the posterior diurnal range are similar for both inversions (Fig. S9(c)).

4.2 Differences between prior and posterior NEE estimates

The posterior seasonal cycle in both inversions differs significantly from the prior. This implies that the biospheric models used to obtain prior GPP and TER fluxes are either over- or under-estimating the strength of some processes, or they are omitting some processes altogether. The largest differences between the posterior solution and the prior model output are seen in spring and summer. In Sect. 3.2 we have shown that spring differences arise from an overestimation of the net spring uptake of CO_2 in the JULES model and a correct/underestimation of the net spring uptake in DALEC. However, in summer (particularly in 2014), the posterior net UK fluxes are higher than both priors in July and August.

One process that occurs during the months July and August is crop harvest. Harvest is not directly resolved in either of the models of the biosphere used in this work, thereby providing a possible explanation for the differences between the posterior

Deleted: S3

Deleted: S6-S8

Deleted: S5

Deleted: S6

Deleted: S8

Deleted: CARDAMOM

Deleted: S8

Deleted: 3

Deleted: flux in the JULES model and a correct/overestimation of the net spring flux

Deleted: CARDAMOM

Deleted: This

Deleted: taken into account

and prior in these months. Harvest typically occurs between July and September and arable agricultural land covered 26% of the UK in 2013 and 2014 (DEFRA, 2014, 2015), so there is potential for unaccounted activity in this area to cause large changes to net CO_2 fluxes. The areas of net source in summer (shown in Fig. 6) do also coincide with areas of large-scale agriculture (e.g. east and central England). Crop harvest potentially changes the biosphere in the following ways: firstly, crops mature en masse, leading to an abrupt loss of productivity. Secondly, during harvest there is an abrupt removal of biomass and input of harvest residues on the field. This increases litter input that is readily available for decomposition, increasing heterotrophic respiration. Thirdly, when the field is ploughed the soil is disturbed, which can increase heterotrophic respiration. Finally, when the crop is no longer covering the soil surface this layer can become drier and the energy balance is altered. In Smallman et al. (2014), the reduction in atmospheric CO_2 concentration due to crop uptake is reported for 2006 to 2008 and an abrupt increase in atmospheric CO_2 can be seen between June (peak source) and August, where CO_2 uptake from crops is halted as a result of harvest. Harvest may explain the abrupt shift from net sink to net zero / net source observed between July and August in DALEC in 2013 and June and July in both models in 2014. The earlier time in 2014 does coincide with a year of early harvest (DEFRA, 2015) although this may well be fortuitous. Later in the summer, there may be some plant regrowth in ploughed fields leading to increased GPP. This would be consistent with the shallower gradient observed in net biospheric fluxes between September and October 2013 in the DALEC posterior estimate, between August and September 2014 in the JULES posterior estimate and the decrease in net flux observed between July and September 2014 in the DALEC posterior estimate.

If agricultural activity is the source of the July, August, September difference between prior and posterior UK NEE estimates, then it could amount to emissions of 4 – 10 % of currently reported annual anthropogenic emissions in 2013 and 17 – 19 % in 2014. However, other explanations for this difference could be large uncertainties in the seasonal disaggregation of anthropogenic fluxes, uncertainties in the transport model, or a combination of over and under-estimation of other biospheric processes.

4.3 Implications for UK CO_2 emissions estimates

The results of UK biospheric CO_2 fluxes using our set-up suggest the UK biosphere is roughly in balance, whereas prior estimates from models of the biosphere estimate a net sink. Even when we assume an uncertainty on our anthropogenic fluxes of 10% (a conservative estimate), inversions using both models still give mean posterior estimates that are larger than their respective priors (see Fig. S10). Therefore, when using models of the biosphere to contribute to inventory estimates of CO_2 emissions, care must be taken to attribute sufficient uncertainties to model estimates, otherwise the amount of CO_2 taken up by the biosphere on an annual basis may be overestimated. Methods such as the one described in this paper could provide an important constraint on the UK's biospheric CO_2 fluxes as carbon sequestration processes, such as reforestation, and other land use change activities are increasingly used as policy solutions to contribute to carbon targets.

Deleted: will again

Deleted: This

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: underestimation

Deleted: 2

Deleted: usng

Deleted: S11

Deleted: ,

Deleted: ,

5 Conclusion

We have developed a framework for estimating net biospheric CO_2 fluxes in the UK that takes advantage of recent innovation in atmospheric inverse modelling and a relatively dense regional network of measurement sites. Two inversions are carried out using prior flux estimates from two different models of the biosphere, DALEC and JULES. Fluxes of GPP and TER are scaled separately in the inversions. Despite significant differences in prior biospheric fluxes, we find consistent monthly and annual posterior flux estimates, suggesting that in this study the choice of model to provide biospheric CO_2 flux priors in the inversion is not a major factor in guiding the inversion result with our framework and network. However, given the hypothesised importance of missing process representation from both models, e.g. agriculture, an improved model may result in an improved analysis, reducing uncertainties and biases highlighted in this study.

Similarly to Tolk et al. (2011), we find that the NEE is more robustly derived if GPP and TER are solved separately, and then combined a posteriori. Our results suggest that inversions that scale only NEE could be underestimating net CO_2 fluxes, as we find posterior estimates 3 – 39 times smaller than those obtained using an inversion where GPP and TER are separated.

We find that the UK biosphere is roughly in balance, with annual net fluxes (averaged over the study period) of -8 ± 79 Tg CO_2 yr^{-1} and -64 ± 85 Tg CO_2 yr^{-1} according to the DALEC and JULES inversions respectively. These mean annual fluxes are systematically higher than their respective priors, implying that net biospheric fluxes are underestimated in the models of the biosphere used in this study. The posterior seasonal cycles from both inversions differ significantly from the prior seasonal cycles and generally have a reduced amplitude of 90% and 76% of the prior amplitude in 2013 according to the DALEC and JULES inversions respectively, and 85% of the prior amplitude in 2014 according to the JULES inversion, however the posterior seasonal cycle amplitude from the DALEC inversion in 2014 was increased by 122%. Our results suggest an overestimated net spring flux in the JULES model and an overestimation of the net summer flux in both models of the biosphere. We propose that the difference seen between the prior and posterior flux estimates in summer and early autumn could be a result of the disturbance caused by crop harvest, leading to abrupt reduction in plant CO_2 uptake and increase in respiration sources, as it is not taken into account in either model. However, this hypothesis is just one of a combination of uncertain factors that could lead to the differences seen, so further work would be needed to investigate the importance of crop harvest in UK CO_2 emissions.

The method developed and described here represents a first step towards looking at the UK biospheric CO_2 budget with a hierarchical Bayesian trans-dimensional MCMC inverse modelling framework. Further work is required to robustly constrain biospheric CO_2 fluxes, through comparison with other model set-ups.

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

6 Code availability

Hierarchical Bayesian trans-dimensional MCMC code is available on request from Matthew Rigby (matt.rigby@bristol.ac.uk).

7 Author contributions

Emily White carried out the research. Emily White, Matt Rigby and Alistair Manning designed the research. Mark Lunt and Anita Ganesan developed the model code. Simon O'Doherty, Kieran Stanley, Ann Stavert, Michel Ramonet, Grant Forster and Andrew Manning provided data. T. Luke Smallman, Edward Comyn-Platt, Peter Levy and Mathew Williams provided model output. Emily White, Matt Rigby, Alistair Manning, Mark Lunt, Anita Ganesan, Simon O'Doherty, Kieran Stanley, Ann Stavert, Luke Smallman, Edward Comyn-Platt, Peter Levy and Paul Palmer wrote the text.

Deleted: and

Deleted: Emly

8 Acknowledgements

Emily White acknowledges the support of a NERC GW4+ Doctoral Training Partnership studentship from the Natural Environment Research Council (NE/L002434/1). Observations from BSD and HFD were supported under NERC grant NE/K002236/1. DECC network data are maintained by grant TRN1028/06/2015 from the UK Department of Business, Energy and Industrial Strategy. We are grateful to Gerard Spain, Duncan Brown, Stephen Humphrey and Andy MacDonald, Carole Helfter and Neil Mullinger for their work maintaining the measurements at Mace Head, Tacolneston and Bilsdale measurement sites. Anita Ganesan was funded under a UK Natural Environment Research Council (NERC) Independent Research Fellowship (NE/L010992/1). T. L. Smallman and Mathew Williams were supported by NERC GHG program GREENHOUSE, grant NE/K002619/1 and this study was funded as part of NERC's support of the National Centre for Earth Observation. Paul Palmer gratefully acknowledged funding from NERC under grant reference NE/K002449/1, and gratefully acknowledges his Royal Society Wolfson Research Merit Award.

Deleted: was supported by

Deleted: NERC) Great Western 4+ Doctoral Training Partnership and NERC grant …

Deleted: M014851

Deleted: .

Deleted: was

References

Atkin, O. K., Bloomfield, K. J., Reich, P. B., Tjoelker, M. G., Asner, G. P., Bonal, D., Bönisch, G., Bradford, M. G., Cernusak, L. A., Cosio, E. G., Creek, D., Crous, K. Y., Domingues, T. F., Dukes, J. S., Egerton, J. J. G., Evans, J. R., Farquhar, G. D., Fyllas, N. M., Gauthier, P. P. G., Gloor, E., Gimeno, T. E., Griffin, K. L., Guerrieri, R., Heskel, M. A., Huntingford, C., Ishida, F. Y., Kattge, J., Lambers, H., Liddell, M. J., Lloyd, J., Lusk, C. H., Martin, R. E., Maksimov, A. P., Maximov, T. C., Malhi, Y., Medlyn, B. E., Meir, P., Mercado, L. M., Mirotchnick, N., Ng, D., Niinemets, Ü., O'Sullivan, O. S., Phillips, O. L., Poorter, L., Poot, P., Prentice, I. C., Salinas, N., Rowland, L. M., Ryan, M. G., Sitch, S., Slot, M., Smith, N. G., Turnbull, M. H., VanderWel, M. C., Valladares, F., Veneklaas, E. J., Weerasinghe, L. K., Wirth, C., Wright, I. J., Wythers, K. R., Xiang, J., Xiang, S. and Zaragoza-Castells, J.: Global variability in leaf respiration in relation to climate, plant functional types and leaf traits, New Phytol., 206(2), 614–636, doi:10.1111/nph.13253, 2015.

Baldocchi, D. D. and Wilson, K. B.: Modeling CO_2 and water vapor exchange of a temperate broadleaved forest across hourly to decadal time scales, Ecol. Model., 142(1–2), 155–184, doi:10.1016/S0304-3800(01)00287-3, 2001.

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavrič, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M. and Tarniewicz, J.: Towards better error statistics for atmospheric inversions of methane surface fluxes, Atmos. Chem. Phys., 13(14), 7115–7132, doi:10.5194/acp-13-7115-2013, 2013.

Berry, J. A., Collatz, G. J., Defries, R. S. and Still, C. J.: ISLSCP II C4 Vegetation Percentage, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee., 2009.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4(3), 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Bloom, A. A. and Williams, M.: 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, 2015.

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L. and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, P. Natl. Acad. Sci. USA, 113(5), 1285–1290, doi:10.1073/pnas.1515160113, 2016.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C. and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4(3), 701–722, doi:10.5194/gmd-4-701-2011, 2011.

Cullen, M. J. P.: The unified forecast/climate model, Meteorol. Mag., 122, 81–94, 1993.

DEFRA: Agriculture Statistics in 2013, Department for Environment Food and Rural Affairs, UK government., 2014.

DEFRA: Agriculture Statistics in 2014, Department for Environment Food and Rural Affairs, UK government., 2015.

Denning, A. S., Randall, D. A., Collatz, G. J. and Sellers, P. J.: Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model, Tellus B, 48(4), 543–567, doi:10.1034/j.1600-0889.1996.t01-1-00010.x, 1996.

Dlugokencky, E. J., Lang, P. M., Mund, J. W., Crotwell, A. M., Crotwell, M. J. and Thoning, K. W.: Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968-2017, Version: 2018-07-31. [online] Available from: ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/, 2018.

EC-JRC/PBL: Emissions Database for Global Atmospheric Research, European Comission, Joint Research Centre (JRC)/ Netherlands Environmental Assessment Agency (PBL). [online] Available from: http://edgar.jrc.ec.europa.eu, 2011.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L. and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3(1), 43–67, doi:10.5194/gmd-3-43-2010, 2010.

FAO/IIASA/ISRIC/ISS-CAS/JRC: Harmonized World Soil Database (version 1.1), FAO, Rome, Italy and IIASA, Laxenburg, Austria., 2009.

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., Manning, A. J., Prinn, R. G., Fraser, P. J., Harth, C. M., Kim, K.-R., Krummel, P. B., Li, S., Mühle, J., O'Doherty, S. J., Park, S., Salameh, P. K., Steele, L. P. and Weiss, R. F.: Characterization

of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmos. Chem. Phys., 14(8), 3855–3864, doi:10.5194/acp-14-3855-2014, 2014.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. . E., Sturges, W. T., Moncrieff, J. B. and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmos. Chem. Phys., 15(11), 6393–6406, doi:10.5194/acp-15-6393-2015, 2015.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S. and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res.-Atmos., 108(D24), doi:10.1029/2003JD003770, 2003.

Gerbig, C., Dolman, A. J. and Heimann, M.: On observational and modelling strategies targeted at regional carbon exchange over continents, Biogeosciences, 6(10), 1949–1959, doi:10.5194/bg-6-1949-2009, 2009.

Göckede, M., Michalak, A. M., Vickers, D., Turner, D. P. and Law, B. E.: Atmospheric inverse modeling to constrain regional-scale CO_2 budgets at high spatial and temporal resolution, J. Geophys. Res., 115(D15), doi:10.1029/2009JD012257, 2010.

Gourdji, S. M., Hirsch, A. I., Mueller, K. L., Yadav, V., Andrews, A. E. and Michalak, A. M.: Regional-scale geostatistical inverse modeling of North American CO_2 fluxes: a synthetic data study, Atmos. Chem. Phys., 17, 2010.

Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82(4), 711–732, doi:10.1093/biomet/82.4.711, 1995.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G.: High-Resolution Global Maps of 21st-Century Forest Cover Change, Science, 342(6160), 850–853, doi:10.1126/science.1244693, 2013.

Hiederer, R. and Köchy, M.: Global Soil Organic Carbon Estimates and the Harmonized World Soil Database, Joint Research Centre of the European Commission., 2012.

Jones, A., Thomson, D., Hort, M. and Devenish, B.: The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III, in Air Pollution Modeling and Its Application XVII, edited by C. Borrego and A.-L. Norman, pp. 580–589, Springer US, Boston, MA., 2006.

Kaminski, T., Rayner, P. J., Heimann, M. and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res.-Atmos., 106(D5), 4703–4715, doi:10.1029/2000JD900581, 2001.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F. and Heimann, M.: Atmospheric CO_2 inversions on the mesoscale using data-driven prior uncertainties: quantification of the European terrestrial CO_2 fluxes, Atmospheric Chemistry and Physics, 18(4), 3047–3064, doi:10.5194/acp-18-3047-2018, 2018a.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F. and Heimann, M.: Technical Note: Atmospheric CO_2 inversions on the mesoscale using data-driven prior uncertainties: methodology and system evaluation, Atmos. Chem. Phys., 18(4), 3027–3045, doi:10.5194/acp-18-3027-2018, 2018b.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurtt, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lima, I., Lombardozzi, D., Metzl,

N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., Padin, X. A., Peregon, A., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wiltshire, A. J., Zaehle, S. and Zhu, D.: Global Carbon Budget 2017, Earth. Syst. Sci. Data, 10(1), 405–448, doi:10.5194/essd-10-405-2018, 2018.

Lokupitiya, R. S., Zupanski, D., Denning, A. S., Kawa, S. R., Gurney, K. R. and Zupanski, M.: Estimation of global CO_2 fluxes at regional scale using the maximum likelihood ensemble filter, J. Geophys. Res., 113(D20), doi:10.1029/2007JD009679, 2008.

Lunt, M. F., Rigby, M., Ganesan, A. L. and Manning, A. J.: Estimation of trace gas fluxes with objectively determined basis functions using reversible-jump Markov chain Monte Carlo, Geosci. Model Dev., 9(9), 3213–3229, doi:10.5194/gmd-9-3213-2016, 2016.

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G. and Derwent, R. G.: Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, J. Geophys. Res., 116(D2), doi:10.1029/2010JD014763, 2011.

Mariscal, A.: Dynamique du carbone du sol et rétroactions climatiques, M.S., Univeristé Libre de Bruxelles, Brussels, Belgium., 2015.

Marthews, T. R., Dadson, S. J., Lehner, B., Abele, S. and Gedney, N.: High-resolution global topographic index values for use in large-scale hydrological modelling, Hydrol. Earth Syst. Sc., 19(1), 91–104, doi:10.5194/hess-19-91-2015, 2015.

Matross, D. M., Andrews, A., Pathmathevan, M., Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Lee, J. T., Zhao, C., Bakwin, P. S., William Munger, J. and Hollinger, D. Y.: Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, Tellus B, 58(5), 344–358, doi:10.1111/j.1600-0889.2006.00206.x, 2006.

Meesters, A. G. C. A., Tolk, L. F., Peters, W., Hutjes, R. W. A., Vellinga, O. S., Elbers, J. A., Vermeulen, A. T., van der Laan, S., Neubert, R. E. M., Meijer, H. A. J. and Dolman, A. J.: Inverse carbon dioxide flux estimates for the Netherlands, J. Geophys. Res.-Atmos., 117(D20), doi:10.1029/2012JD017797, 2012.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of State Calculations by Fast Computing Machines, J. Chem. Phys., 21(6), 1087–1092, doi:10.1063/1.1699114, 1953.

Michalak, A. M.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res., 109(D14), doi:10.1029/2003JD004422, 2004.

Miller, S. M., Michalak, A. M. and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geosci. Model Dev., 7(1), 303–315, doi:10.5194/gmd-7-303-2014, 2014.

Palmer, P. I., O'Doherty, S., Allen, G., Bower, K., Bösch, H., Chipperfield, M. P., Connors, S., Dhomse, S., Feng, L., Finch, D. P., Gallagher, M. W., Gloor, E., Gonzi, S., Harris, N. R. P., Helfter, C., Humpage, N., Kerridge, B., Knappett, D., Jones, R. L., Le Breton, M., Lunt, M. F., Manning, A. J., Matthiesen, S., Muller, J. B. A., Mullinger, N., Nemiitz, E., O'Shea, S., Parker, R. J., Percival, C. J., Pitt, J., Riddick, S. N., Rigby, M., Sembhi, H., Siddans, R., Skelton, R. L., Smith, P., Sonderfeld, H., Stanley, K., Stavert, A. R., Wenger, A., White, E., Wilson, C. and Young, D.: A measurement-based verification framework for UK greenhouse gas emissions: an overview of the Greenhouse gAs Uk and Global Emissions (GAUGE) project, Atmos. Chem. Phys. Discuss., 1–52, doi:10.5194/acp-2018-135, 2018.

Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Van Der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M. and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Glob. Change Biol., 16(4), 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x, 2010.

Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P. and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling: model intercomparison, Atmos. Chem. Phys., 11(13), 6607–6622, doi:10.5194/acp-11-6607-2011, 2011.

Potter, C. S.: Terrestrial biomass and the effects of deforestation on the global carbon cycle: Results from a model of primary production using satellite observations, BioScience, 49(10), 769–778, doi:10.2307/1313568, 1999.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M. and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Rigby, M., Manning, A. J. and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, Atmos. Chem. Phys., 11(18), 9887–9898, doi:10.5194/acp-11-9887-2011, 2011.

Rivier, L., Peylin, P., Ciais, P., Gloor, M., Rödenbeck, C., Geels, C., Karstens, U., Bousquet, P., Brandt, J. and Heimann, M.: European CO_2 fluxes from atmospheric inversions using regional and global transport models, Climatic Change, 103(1–2), 93–115, doi:10.1007/s10584-010-9908-4, 2010.

Say, D., Manning, A. J., O'Doherty, S., Rigby, M., Young, D. and Grant, A.: Re-Evaluation of the UK's HFC-134a Emissions Inventory Based on Atmospheric Observations, Environ. Sci. Technol., 50(20), 11129–11136, doi:10.1021/acs.est.6b03630, 2016.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W. and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth. Syst. Sci. Data, 9(2), 389–413, doi:10.5194/essd-9-389-2017, 2017.

Schuh, A. E., Denning, A. S., Uliasz, M. and Corbin, K. D.: Seeing the forest through the trees: Recovering large-scale carbon flux biases in the midst of small-scale variability, J. Geophys. Res., 114(G3), doi:10.1029/2008JG000842, 2009.

Schuh, A. E., Denning, A. S., Corbin, K. D., Baker, I. T., Uliasz, M., Parazoo, N., Andrews, A. E. and Worthy, D. E. J.: A regional high-resolution carbon flux inversion of North America for 2004, Biogeosciences, 7(5), 1625–1644, doi:10.5194/bg-7-1625-2010, 2010.

Smallman, T. L., Williams, M. and Moncrieff, J. B.: Can seasonal and interannual variation in landscape CO_2 fluxes be detected by atmospheric observations of CO_2 concentrations made at a tall tower?, Biogeosciences, 11(3), 735–747, doi:10.5194/bg-11-735-2014, 2014.

Smallman, T. L., Exbrayat, J.-F., Mencuccini, M., Bloom, A. A. and Williams, M.: Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests, J. Geophys. Res.-Biogeo., 122(3), 528–545, doi:10.1002/2016JG003520, 2017.

Stanley, K. M., Grant, A., O'Doherty, S., Young, D., Manning, A. J., Stavert, A. R., Spain, T. G., Salameh, P. K., Harth, C. M., Simmonds, P. G., Sturges, W. T., Oram, D. E. and Derwent, R. G.: Greenhouse gas measurements from a UK network of tall towers: technical description and first results, Atmos. Meas. Tech., 11(3), 1437–1458, doi:10.5194/amt-11-1437-2018, 2018.

Stavert, A. R., O'Doherty, S., Stanley, K., Young, D., Manning, A. J., Lunt, M. F., Rennick, C. and Arnold, T.: UK greenhouse gas measurements at two new tall towers for aiding emissions verification, Atmos. Meas. Tech. Discuss., 1–55, doi:10.5194/amt-2018-140, 2018.

Steinkamp, K., Mikaloff Fletcher, S. E., Brailsford, G., Smale, D., Moore, S., Keller, E. D., Baisden, W. T., Mukai, H. and Stephens, B. B.: Atmospheric CO_2 observations and models suggest strong carbon uptake by forests in New Zealand, Atmos. Chem. Phys., 17(1), 47–76, doi:10.5194/acp-17-47-2017, 2017.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R. and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans, Deep-Sea Res. Pt. II, 56(8–10), 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics., 2005.

The UK government: Climate Change Act 2008, The Stationary Office Limited, London, UK., 2008.

Thompson, R. L. and Stohl, A.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geosci. Model Dev., 7(5), 2223–2242, doi:10.5194/gmd-7-2223-2014, 2014.

Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S. R. and Schmullius, C.: Carbon stock and density of northern boreal and temperate forests: Carbon stock and density of boreal and temperate forests, Global Ecol. Biogeogr., 23(3), 297–310, doi:10.1111/geb.12125, 2014.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G. and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, Biogeosciences, 10(3), 1717–1736, doi:10.5194/bg-10-1717-2013, 2013.

Tolk, L. F., Dolman, A. J., Meesters, A. G. C. A. and Peters, W.: A comparison of different inverse carbon flux estimation approaches for application on a regional domain, Atmos. Chem. Phys., 11(20), 10349–10365, doi:10.5194/acp-11-10349-2011, 2011.

UNECE/EMEP: EMEP/CORINAIR Emission Inventory Guidebook - 3rd edition 2001, European Environment Agency. [online] Available from: https://www.eea.europa.eu/publications/technical_report_2001_3 (Accessed 6 August 2018), 2001.

UNFCCC: Adoption of the Paris Agreement FCCC/CP/2015/L.9/Rev. 1, 2015.

Vardag, S. N., Hammer, S., O'Doherty, S., Spain, T. G., Wastine, B., Jordan, A. and Levin, I.: Comparisons of continuous atmospheric CH_4, CO_2 and N_2O measurements - results from a travelling instrument campaign at Mace Head, Atmos. Chem. Phys., 14(16), 8403–8418, doi:10.5194/acp-14-8403-2014, 2014.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50(9), 7505–7514, doi:10.1002/2014WR015638, 2014.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Shaver, G. R. and Johnson, L. C.: Predicting Gross Primary Productivity in Terrestrial Ecosystems, Ecol. Appl., 7(3), 882–894, doi:10.1890/1051-0761(1997)007[0882:PGPPIT]2.0.CO;2, 1997.

Zupanski, D., Denning, A. S., Uliasz, M., Zupanski, M., Schuh, A. E., Rayner, P. J., Peters, W. and Corbin, K. D.: Carbon flux bias estimation employing Maximum Likelihood Ensemble Filter (MLEF), J. Geophys. Res., 112(D17), doi:10.1029/2006JD008371, 2007.

Formatted: Bibliography

Deleted: **Table 1:** Measurement site information.¶

Table 1: Measurement site information. The location of sites is also shown in Fig. 1. *Weybourne data is used for validation of the results only and is not included in the inversions. LSCE – Laboratoire des Sciences du Climat et de l'Environnement, DECC – Deriving Emissions related to Climate Change, GAUGE – Greenhouse gAs Uk and Global Emissions, UEA – University of East Anglia.

Site	Site code	Location	Inlet Height (m above ground level)	Network
Mace Head	MHD	53.327 °N, 9.904 °W	24	LSCE
Ridge Hill	RGL	51.998 °N, 2.540 °W	90	DECC
Tacolneston	TAC	52.518 °N, 1.139 °E	185	DECC
Heathfield	HFD	50.977 °N, 0.231 °E	100	GAUGE
Bilsdale	BSD	54.359 °N, 1.150 °W	248	GAUGE
Angus	TTA	56.555 °N, 2.986 °W	222	DECC
*Weybourne	WAO	52.950 °N, 1.122 °E	10	UEA

Table 2: Specifications for different prior and fixed fluxes.

	Spatial Resolution	Temporal Resolution
Biogenic fluxes		
JULES	0.25˚ x 0.25˚	2-hourly
DALEC	25 km x 25 km (1˚ x 1˚ outside the UK)	2-hourly
Anthropogenic fluxes		
NAEI (UK)	1 km x 1 km	2-hourly
EDGAR (outside UK)	0.1˚ x 0.1˚	Yearly (using 2010)
Ocean fluxes	4˚ x 5˚	Monthly (climatology)

Formatted Table

Moved (insertion) [1]

Moved (insertion) [2]

Deleted: ¶ ¶ ¶ Table 2

Moved (insertion) [3]

Formatted: Justified, Line spacing: 1.5 lines

Moved (insertion) [4]

Table 3: Probability density functions (PDFs) for parameter and hyper-parameter scaling factors. Mean and standard deviation in fourth and fifth columns relate to lognormal PDFs, lower bound and upper bound relate to uniform PDFs.

Parameter		PDF	Mean / lower bound	Standard deviation / upper bound
Prior uncertainty				
GPP	x_{GPP}	Lognormal	1	1
	$\sigma_{x_{GPP}}$	Uniform	0.1	1.5
TER	x_{TER}	Lognormal	1	1
	$\sigma_{x_{TER}}$	Uniform	0.1	1.5
Boundary conditions	x_{BC}	Lognormal	1	1
	$\sigma_{x_{BC}}$	Uniform	0.01	0.05
Model-measurement representation uncertainty				
Standard deviation	σ_y	Uniform	0.9 ppm	45 ppm
Correlation timescale	τ	Uniform	1 hour	120 hours

Table 4: Posterior UK estimates for the maximum net biospheric source and sink (values also shown in Fig. 5). The month in brackets indicates the month in which the maximum source/sink occurred.

	Year	Max. sink (Tg CO_2 yr^{-1})		Max. source (Tg CO_2 yr^{-1})	
DALEC	2013	$-298 \pm {}^{140}_{136}$	(June)	$171 \pm {}^{94}_{76}$	(January)
	2014	$-360 \pm {}^{87}_{88}$	(June)	$273 \pm {}^{65}_{63}$	(November)
JULES	2013	$-456 \pm {}^{90}_{91}$	(June)	$122 \pm {}^{83}_{78}$	(December)
	2014	$-542 \pm {}^{97}_{100}$	(June)	$195 \pm {}^{65}_{70}$	(October)

Deleted: st. dev.
Formatted Table
Deleted: St. dev.
Deleted: Rtot
Deleted: 3
Deleted: 15
Formatted: Left, Line spacing: single
Deleted: Table 3: Specifications for different priors.¶
Moved up [2]: ¶ Spatial Resolution
Deleted: CARDAMOM ... [2]
Moved up [3]: Anthropogenic fluxes
Deleted: CARDAMOM

Table 5: Prior and posterior fit to data statistics for the inversion period 2013-2014. R^2 and RMSE are calculated monthly and averaged over this period. Values in brackets are the posterior fit statistics for the corresponding net flux inversions. *Weybourne data (from February to December 2013) is used for validation of the results only and is not included in the inversions.

DALEC inversion

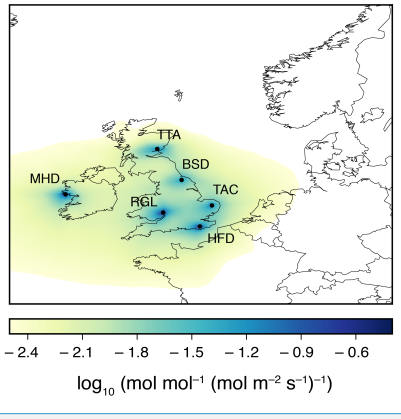
Measurement site	Prior R^2	Posterior R^2	Prior RMSE	Posterior RMSE	Prior mean bias	Posterior mean bias
Mace Head	0.20	0.59	2.88	1.53	-1.19	0.55
		(0.54)		(1.62)		(0.38)
Ridge Hill	0.26	0.67	3.82	2.09	-1.27	-0.10
		(0.61)		(2.30)		(-0.05)
Tacolneston	0.22	0.61	3.92	2.20	-1.63	-0.25
		(0.56)		(2.44)		(-0.28)
Heathfield	0.21	0.71	4.07	1.88	-1.99	0.11
		(0.58)		(2.31)		(0.21)
Bilsdale	0.20	0.60	4.62	2.02	-3.68	-0.52
		(0.55)		(2.23)		(-0.58)
Angus	0.35	0.67	3.09	1.28	-2.35	-0.01
		(0.63)		(1.41)		(0.00)
*Weybourne	0.13	0.31	6.17	5.08	2.89	2.25
		(0.28)		(5.32)		(2.37)

JULES inversion

Measurement site	Prior R^2	Posterior R^2	Prior RMSE	Posterior RMSE	Prior mean bias	Posterior mean bias
Mace Head	0.29	0.66	2.84	1.26	-1.33	0.16
		(0.56)		(1.44)		(-0.01)
Ridge Hill	0.33	0.67	3.86	2.14	-1.14	-0.21
		(0.59)		(2.41)		(-0.05)
Tacolneston	0.24	0.53	4.06	2.71	-1.84	-0.89
		(0.52)		(2.70)		(-0.74)
Heathfield	0.28	0.66	4.07	2.14	-2.43	-0.25
		(0.57)		(2.38)		(-0.23)
Bilsdale	0.33	0.61	4.53	2.10	-3.60	-0.96
		(0.62)		(2.19)		(-0.82)
Angus	0.43	0.67	2.85	1.39	-1.78	0.43

Moved (insertion) [5]
Formatted: Not Superscript/ Subscript
Formatted: Not Superscript/ Subscript
Formatted: Justified

Moved (insertion) [6]
Formatted: Font color: Auto
Formatted: Justified, Line spacing: 1.5 lines
Deleted: ¶



		(0.62)		(1.55)		(0.48)
*Weybourne	0.16	0.29	5.85	5.10	2.63	2.07
		(0.23)		(5.49)		(2.56)

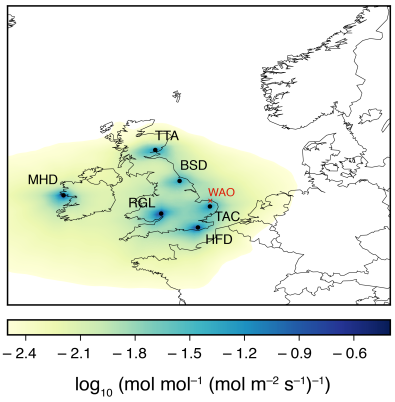


Figure 1: Mean annual NAME footprint for 2014, for each of the six sites. MHD: Mace Head; RGL: Ridge Hill; HFD: Heathfield; TAC: Tacolneston; BSD: Bilsdale; TTA: Angus. WAO shows the location of the Weybourne Atmospheric Observatory, where data has been used to validate the results but has not been included in the inversion (the mean footprint from this station is not plotted).

Formatted: Font: Bold

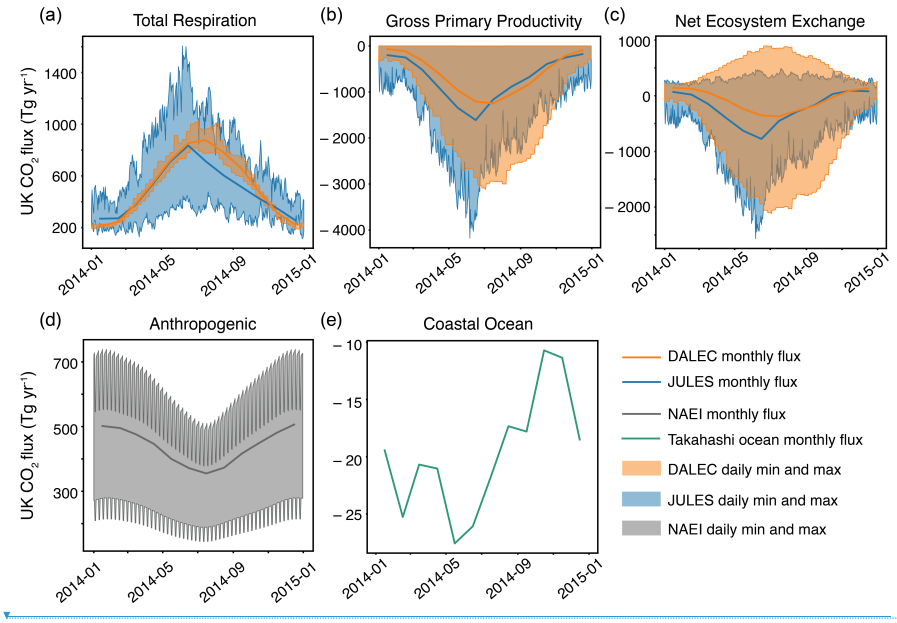


Figure 2: Prior UK fluxes in 2014. (a-c) Comparison of JULES (blue) and DALEC (orange) monthly fluxes and minimum and maximum daily values for TER, GPP and NEE respectively. (d) Monthly anthropogenic fluxes and minimum and maximum daily values from the NAEI inventory within the UK. (e) Monthly coastal ocean net fluxes from the Takahashi et al. (2009) ocean CO_2 flux product.

290

Deleted: ¶

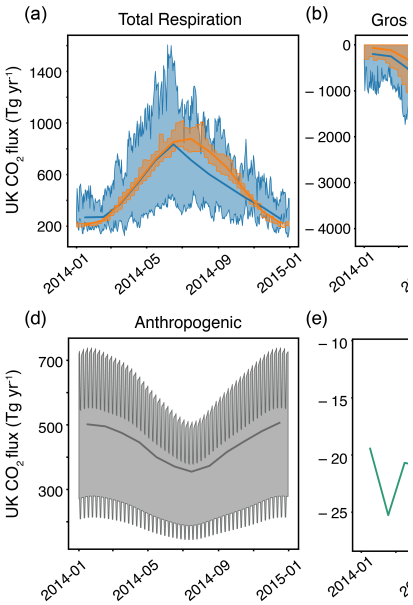


Figure 2: Prior UK fluxes in 2014. (a-c) Comparison of JULES (blue) and CARDAMOM (orange) monthly fluxes and minimum and maximum daily values for TER, GPP and NEE respectively. (d) Monthly anthropogenic fluxes and minimum and maximum daily values from the NAEI inventory within the UK. (e) Monthly coastal ocean net fluxes from the Takahashi et al. (2009) ocean CO_2 flux product.¶

... [3]

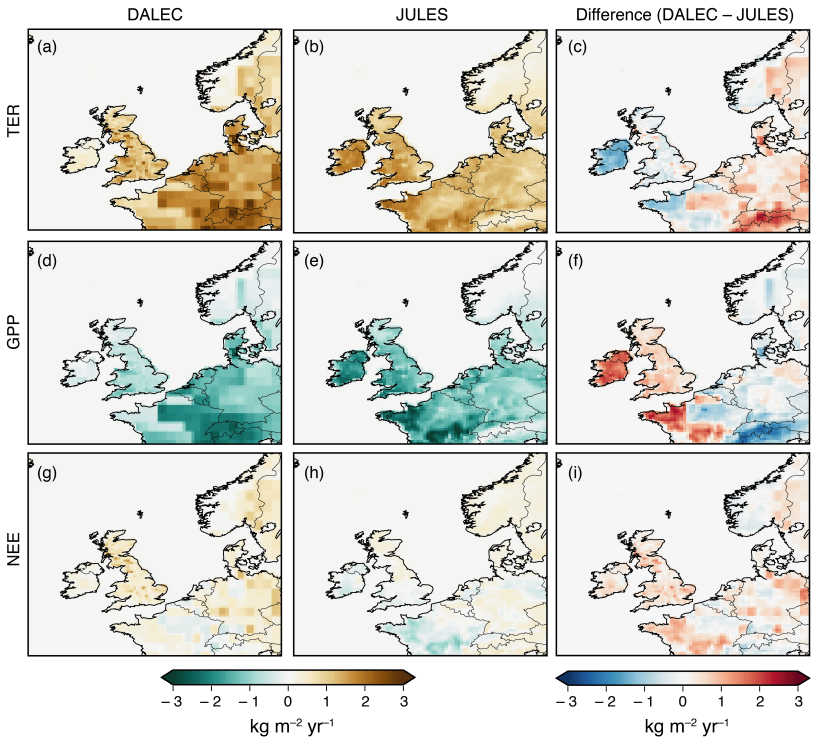


Figure 3: Average prior flux maps for winter 2013 (December 2013, January – February 2014). (a) TER from DALEC; (b) TER from JULES; (c) the difference between DALEC and JULES TER; (d) GPP from DALEC; (e) GPP from JULES; (f) the difference between DALEC and JULES GPP; (g) NEE from DALEC; (h) NEE from JULES; (i) the difference between DALEC and JULES NEE.

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

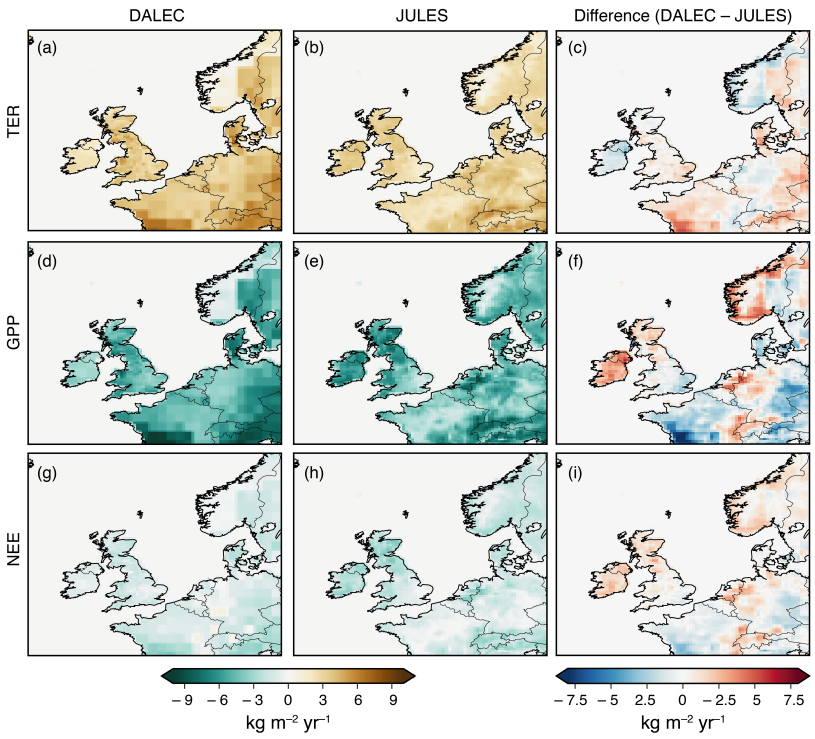
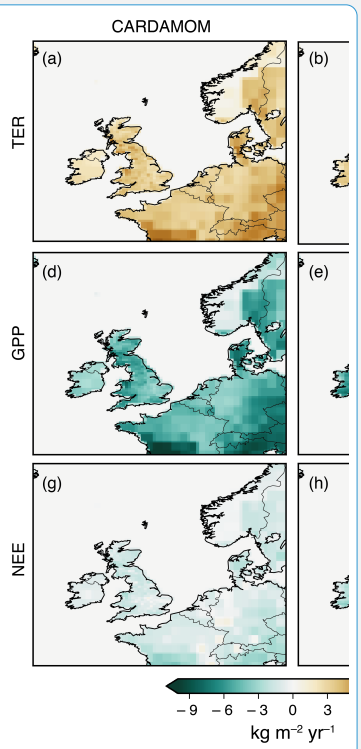


Figure 4: Prior average flux maps for summer 2014 (June – August 2014). (a) TER from DALEC; (b) TER from JULES; (c) the difference between DALEC and JULES TER; (d) GPP from DALEC; (e) GPP from JULES; (f) the difference between DALEC and JULES GPP; (g) NEE from DALEC; (h) NEE from JULES; (i) the difference between DALEC and JULES NEE.



Deleted:

Deleted: CARDAMOM

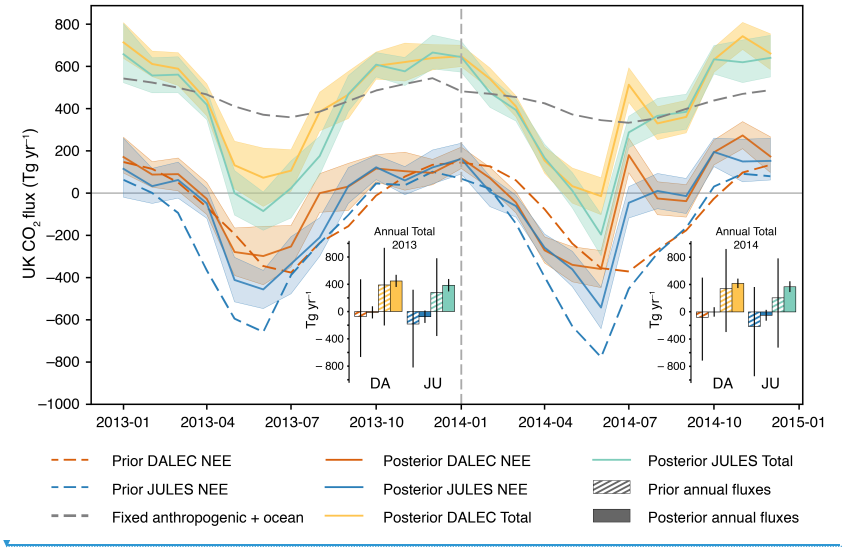
Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM

Deleted: CARDAMOM



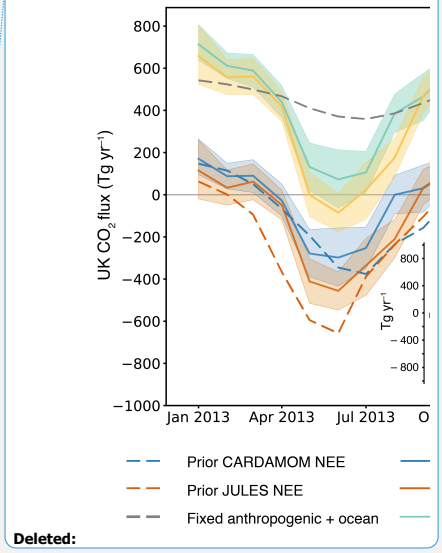


Figure 5: Posterior monthly net UK CO_2 flux (positive is emission to atmosphere). Orange and blue monthly fluxes are posterior net biospheric (NEE) fluxes for DALEC and JULES respectively. Prior biosphere fluxes from DALEC and JULES are shown in dashed orange and blue lines respectively. The fixed anthropogenic and ocean fluxes are denoted by the dark grey dashed line. Yellow and green monthly fluxes are the sum of the posterior NEE fluxes and the fixed anthropogenic and ocean fluxes. Shading represents 5^{th} – 95^{th} percentile. The bar charts represent annual net UK CO_2 flux for 2013 (left) and 2014 (right). Hashed bars denote prior annual fluxes, solid bars denote posterior annual fluxes. The bar colours correspond to the line colours: left hand bars for each model are NEE fluxes, right hand bars for each model are total fluxes (NEE + fixed sources). Uncertainty bars represent 5^{th} – 95^{th} percentile. DA – DALEC. JU – JULES.

Deleted:

Deleted: (+ve

Deleted: .

Deleted: and CARDOMOM

Deleted: CA – CARDAMOM.

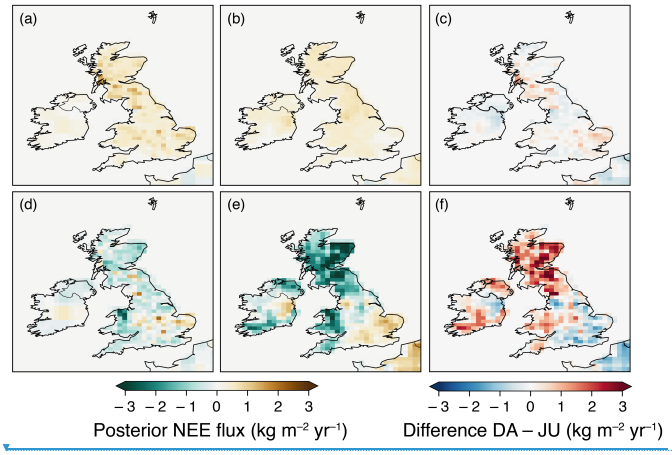
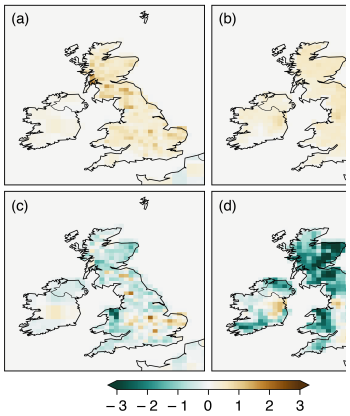


Figure 6: Posterior net biospheric (NEE) flux maps averaged over winter 2013 (December 2013, January – February 2014) and summer 2014 (June – August 2014). (a) Winter NEE flux from DALEC inversion. (b) Winter NEE flux from JULES inversion. (c) Difference between winter NEE flux from DALEC (DA) and JULES (JU) inversions. (d) Summer NEE flux from DALEC inversion. (e) Summer NEE flux from JULES inversion. (f) Difference between summer NEE flux from DALEC (DA) and JULES (JU) inversions.

Deleted:



Deleted: CARDAMOM

Deleted: Summer

Deleted: CARDAMOM inversion.

Deleted: ¶

Supplementary Figures and Tables:

Table S1: Annual UK net biospheric flux for June 2014 as estimated with an inversion using footprints disaggregated for 12, 24, 48 and 72 hours back in time, as well as an inversion using integrated footprints combined with monthly fluxes. DALEC NEE was used as the prior flux in this test.

Prior	-355
Posterior	
Integrated footprints	$79 \pm^{103}_{106}$
12-hour back footprints	$-207 \pm^{85}_{85}$
24-hour back footprints	$-356 \pm^{87}_{88}$
48-hour back footprints	$-382 \pm^{92}_{86}$
72-hour back footprints	$-412 \pm^{110}_{101}$

Deleted: hours back in time
Deleted: compared to
Deleted: disaggregated for 72 hours back in time. CARDAMOM
Formatted Table
Deleted: 355
Formatted: Font: +Body CS (Times New Roman)
Deleted: Posterior (
Formatted Table
Deleted:)
Deleted: - 356 $\pm^{87}_{88}$
Formatted: Font: +Body CS (Times New Roman)
Deleted: Posterior (
Formatted Table
Deleted:)
Deleted: - 395 $\pm^{90}_{91}$
Formatted: Font: +Body CS (Times New Roman)
Moved up [1]: ¶
¶
Table
Deleted: S2: Prior and posterior fit to data statistics for year 2014 (calculated for each month and averaged over the year). Prior modelled mole fractions are added to the posterior baseline before they are compared to the data.¶
CARDAMOM inversion ... [4]
Moved up [5]: Measurement site
Formatted: Not Superscript/ Subscript
Formatted: Not Superscript/ Subscript
Deleted: Mace Head ... [5]
Moved up [4]: ¶
¶
Table
Deleted: S3: Prior and posterior fit to data statistics for year 2014 for the inversion using NEE prior flux as opposed to separating GPP and TER (calculated for each month and averaged over the year). Prior modelled mole fractions are added to the posterior baseline before they are compared to the data.¶
CARDAMOM inversion ... [6]
Moved up [6]: Measurement site
Formatted: Font color: Auto
Deleted: Mace Head ... [7]

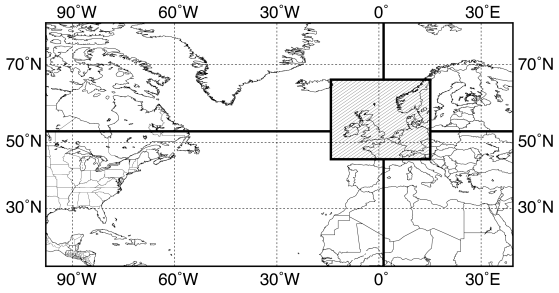
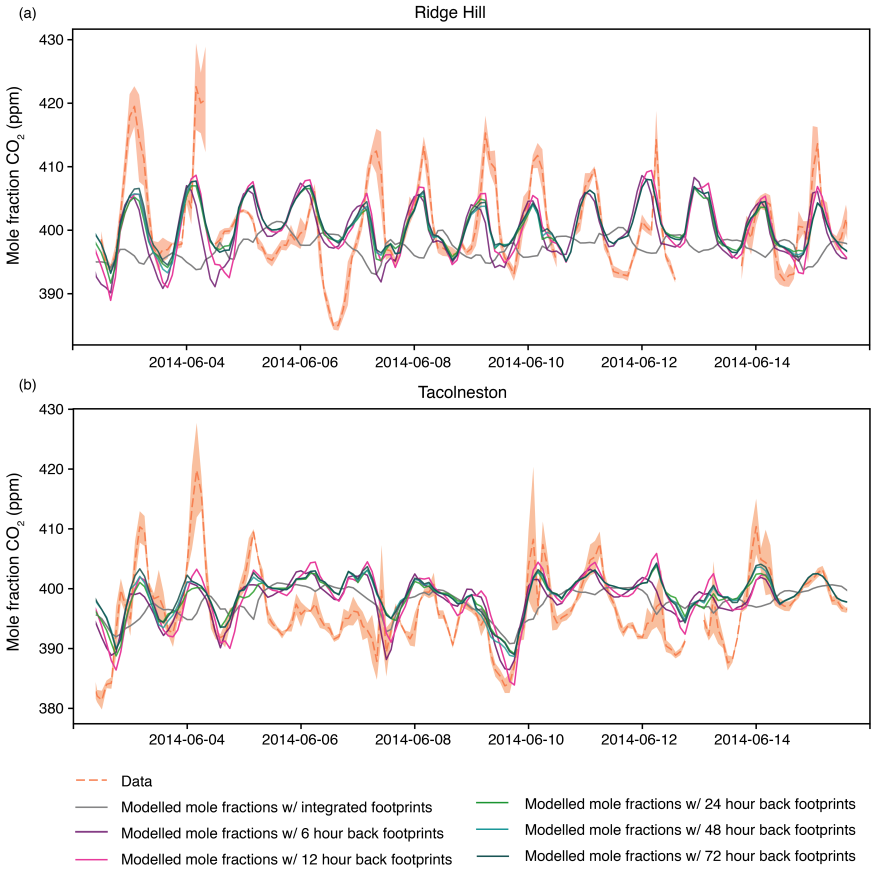
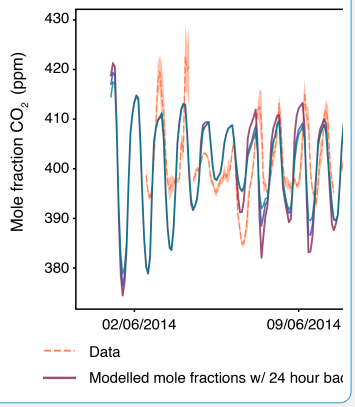


Figure S1: The domain used to calculate NAME footprints. The four edge boxes correspond to four basis functions. The hatched box is the main area of focus for this study and basis functions in this area are based on a fractional map of 6–7 different PFTs (Fig. S6).

Deleted: .





Deleted:

Figure S2: Forward modelled mole fractions at Ridge Hill and Tacolneston for part of June 2014 using DALEC NEE fluxes and NAME footprints that are disaggregated back in time for 6, 12, 24, 48 and 72 hours, as well as using integrated footprints with monthly fluxes. Anthropogenic and ocean fluxes have been forward modelled and removed from the data. Shading on the data represents $\pm$ **1**$\boldsymbol{\sigma}$.

Deleted: for

Deleted: CARDAMOM

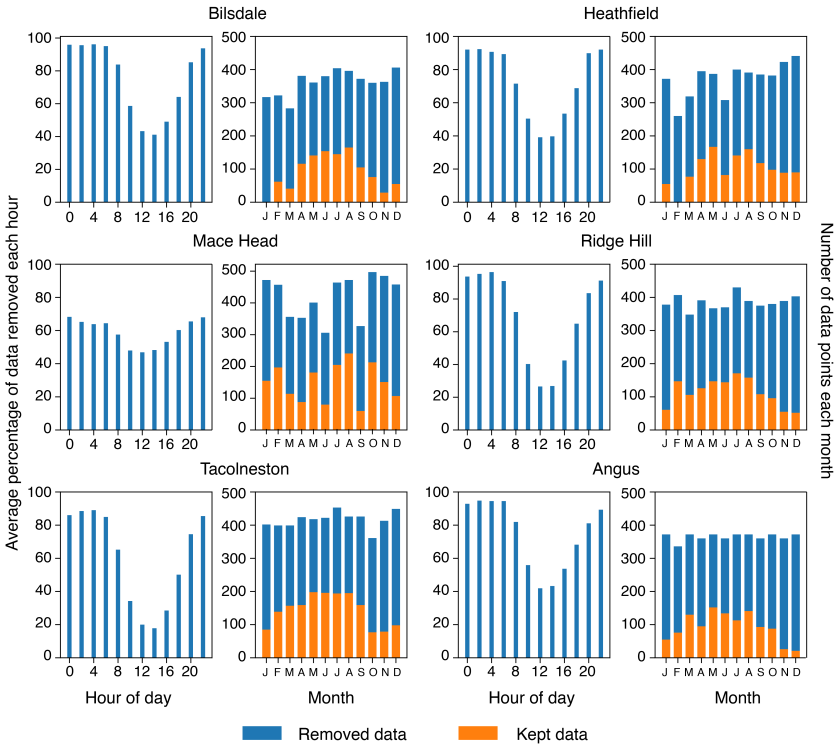


Figure S3: Data filtered out in 2014 using the "local-lapse" filter. Left hand bar charts for each site show the average percentage of data removed for each 2-hour period in the day. Right hand bar charts for each site show the number of data points used in the inversion for each month (orange bars) and the number of data points removed prior to the inversion for each month (blue bars).

Deleted: ¶

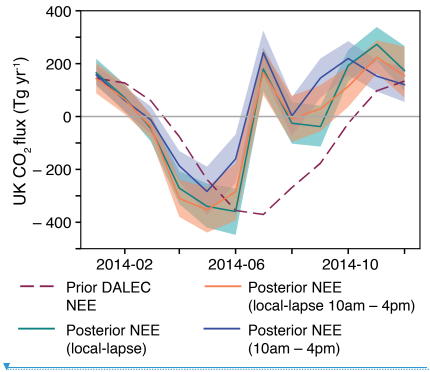
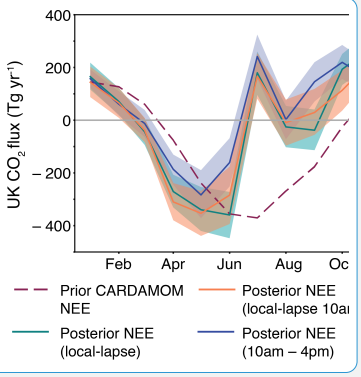
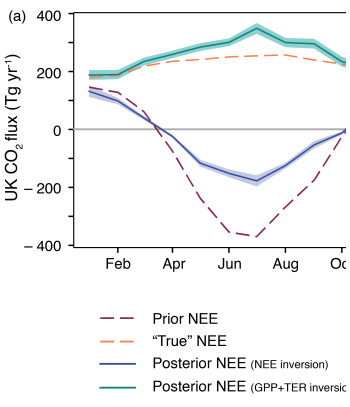


Figure S4: A comparison of the results of three different inversions for 2014 using DALEC prior GPP and TER fluxes and three differently filtered data sets. Local-lapse: the filter used on the final results, a combination of localness and vertical temperature profile metrics. Local-lapse 10am – 4pm: data is filtered with the "local-lapse" filter and then only times between 10am and 4pm are selected. 10am – 4pm: all data between 10am and 4pm is used. Shading represents 5th – 95th percentile.



Deleted:

Deleted: CARDAMOM



Deleted:

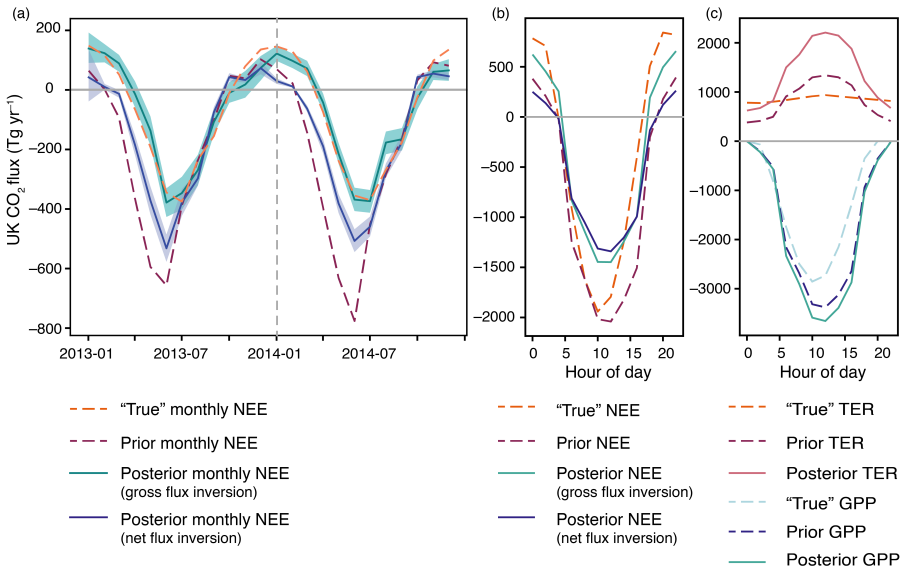


Figure S5: Synthetic test results. Synthetic data was produced using DALEC biospheric fluxes and NAME simulations– the "true" flux. Prior fluxes are provided by JULES. The "NEE inversion" only scales NEE in the inversion. The "GPP+TER inversion" scales GPP and TER separately in the inversion. NEE prior PDF ($\boldsymbol{x_{NEE}}$) has Gaussian uncertainty distribution and its standard deviation hyper-parameter ($\boldsymbol{\sigma_{x_{NEE}}}$) has a uniform distribution with a range reflecting an absolute uncertainty of approximately 40–400 Tg (see Table 3 for the comparable set-up for the separate GPP and TER inversion). (a) shows prior and posterior monthly flux estimates for the UK in 2014 compared to the "true" flux. Shading represents the 5th – 95th percentiles. (b) shows average diurnal cycle in June 2014 for prior and posterior NEE in both inversions, as well as the "true" NEE. (c) shows average diurnal cycle in June 2014 for prior and posterior GPP and TER in the "GPP+TER" inversion, as well as the "true" GPP.

Deleted: CARDAMOM

Deleted: where the GPP was reduced by half

Deleted: 2

Deleted: GPP and TER in the "GPP+TER" inversion, as well as the "true" GPP ("true" TER is equal to prior TER). (c) shows average diurnal cycle in June 2014 for prior and posterior

Deleted: ¶

¶

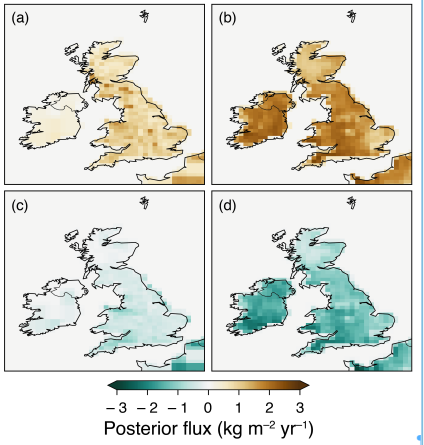


Figure S6

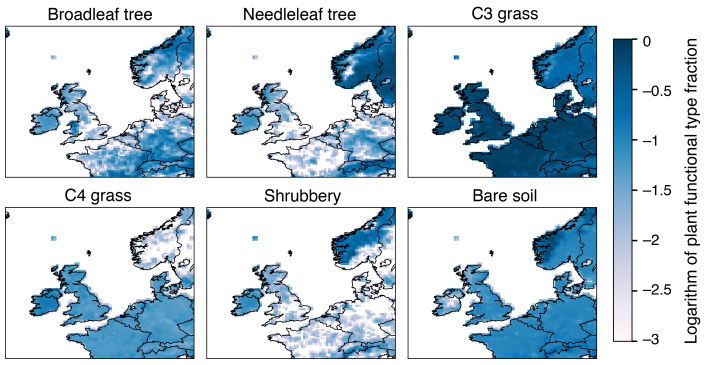


Figure S6: Maps of plant functional type (PFT) fraction for each of the 6 PFTs used as spatial basis functions within the sub-domain. Note the scale is logarithmic.

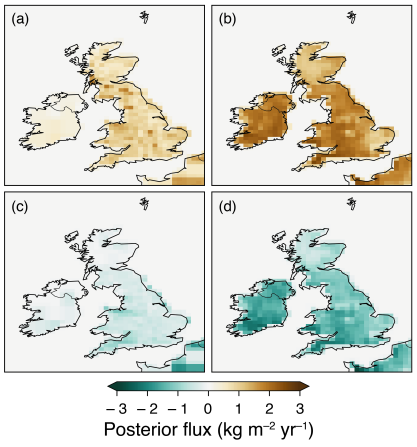


Figure S7: Posterior TER and GPP flux maps averaged over winter 2013 (December 2013, January – February 2014). (a) Winter TER flux from DALEC inversion. (b) Winter TER flux from JULES inversion. (c) Winter GPP flux from DALEC inversion. (d) Winter TER flux from JULES inversion.

Deleted: CARDAMOM

Deleted: CARDAMOM

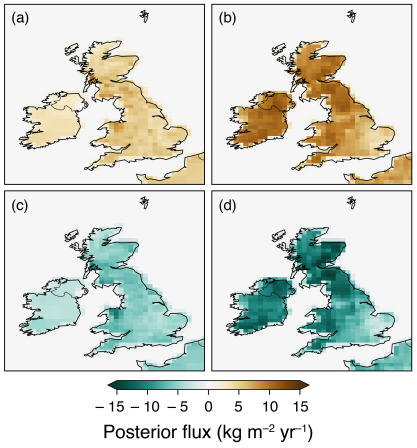


Figure S8: Posterior TER and GPP flux maps averaged over summer 2014 (June – August 2014). (a) Summer TER flux from DALEC inversion. (b) Summer TER flux from JULES inversion. (c) Summer GPP flux from DALEC inversion. (d) Summer GPP flux from JULES inversion.

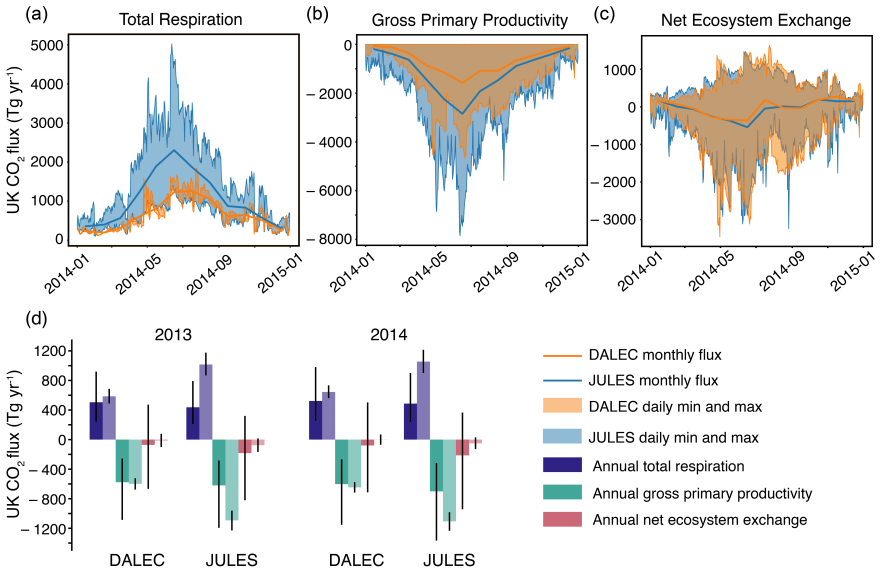
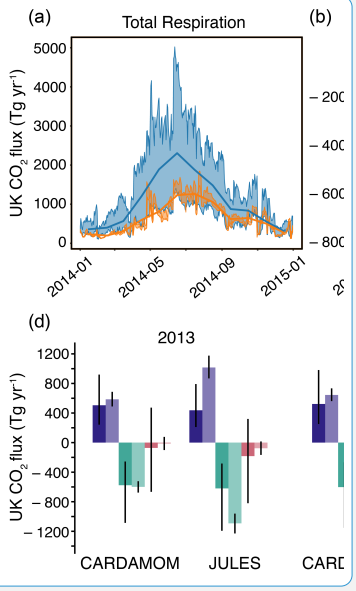


Figure S9: Posterior UK fluxes in 2014. (a-c) Comparison of monthly fluxes and minimum and maximum daily values for TER, GPP and NEE respectively resulting from JULES inversion (blue) and DALEC inversion (orange). (d) Annual CO_2 fluxes for TER, GPP and NEE for 2013 and 2014 from DALEC and JULES inversions. Dark bars denote prior annual fluxes, light bars denote posterior annual fluxes. Uncertainty bars represent 5^{th} – 95^{th} percentile.

Deleted: S7
Deleted: CARDAMOM
Deleted: CARDAMOM
Formatted: Caption



Deleted:

Deleted: S8
Deleted: CARDAMOM
Deleted: CARDAMOM

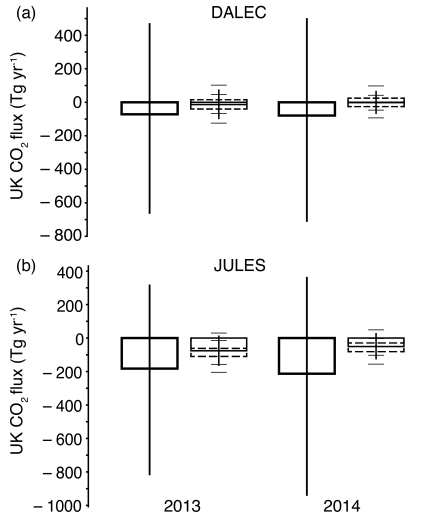
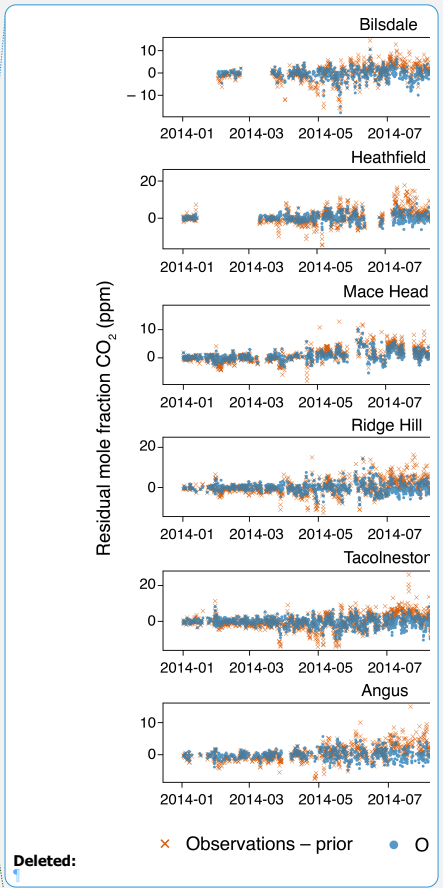


Figure S10: Annual UK NEE flux estimates from DALEC and JULES inversions for 2013 and 2014. Left bars are prior NEE estimates, right bars are posterior NEE estimates. Dashed bars on the posterior estimates represent annual NEE fluxes for inversions that use fixed anthropogenic fluxes multiplied by ± 10%. Uncertainty bars represent 5th – 95th percentile. Solid uncertainty bars on posterior estimates are the uncertainty on the inversions using normal anthropogenic fluxes. Whiskers on the posterior estimates are the uncertainty on the inversions using anthropogenic fluxes multiplied by ± 10%.



Deleted:

Moved (insertion) [7]

Moved (insertion) [8]

Deleted: S9

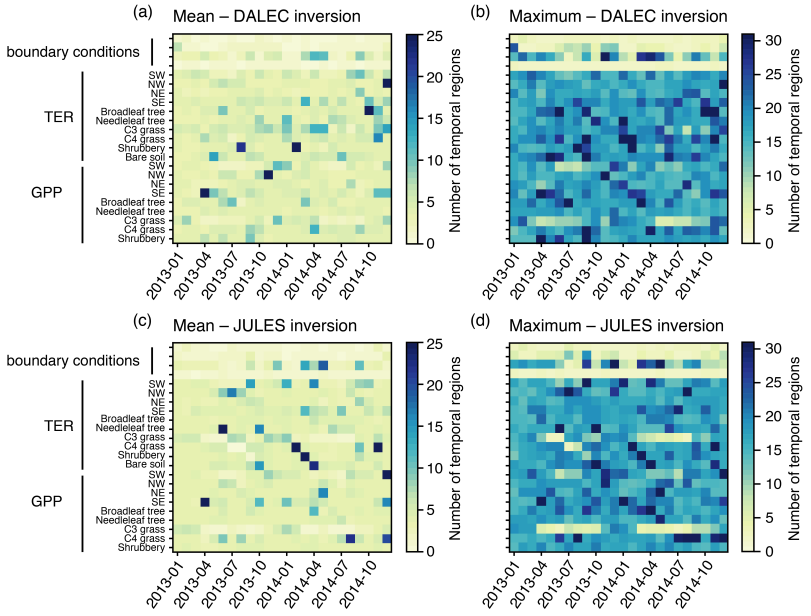


Figure S11: Mean and maximum number of temporal regions each month, taken across the number of algorithm iterations, for each source and spatial region.

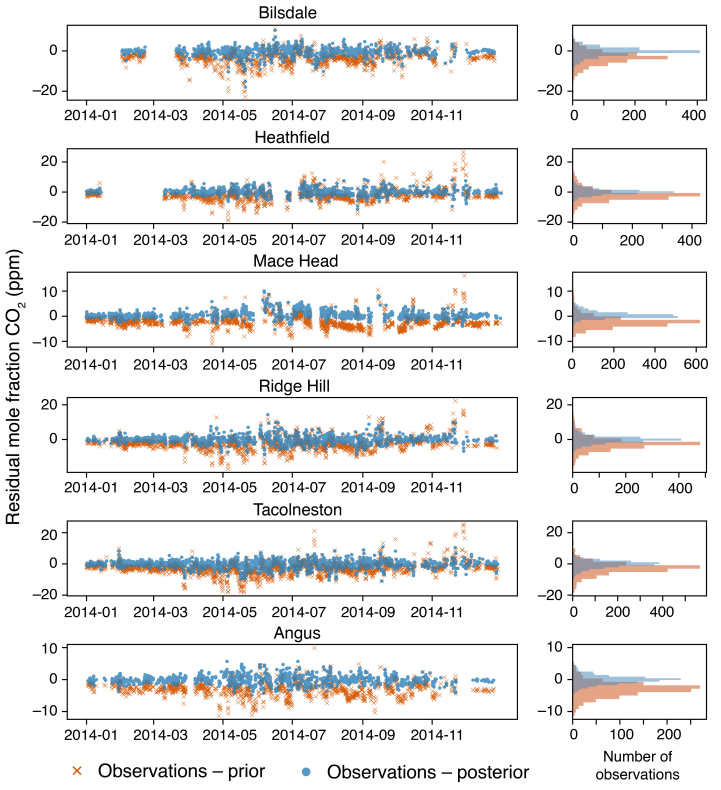


Figure S12: Left: Residual mole fractions for prior and posterior modelled CO_2 concentrations in 2014 using DALEC prior biospheric fluxes. Right: Histogram of prior residuals (orange) and posterior residuals (blue). The mean of the histogram represents the mean bias.

Deleted: CARDAMOM

Deleted: Prior modelled mole fractions are added to the posterior baseline before they are compared to the data.

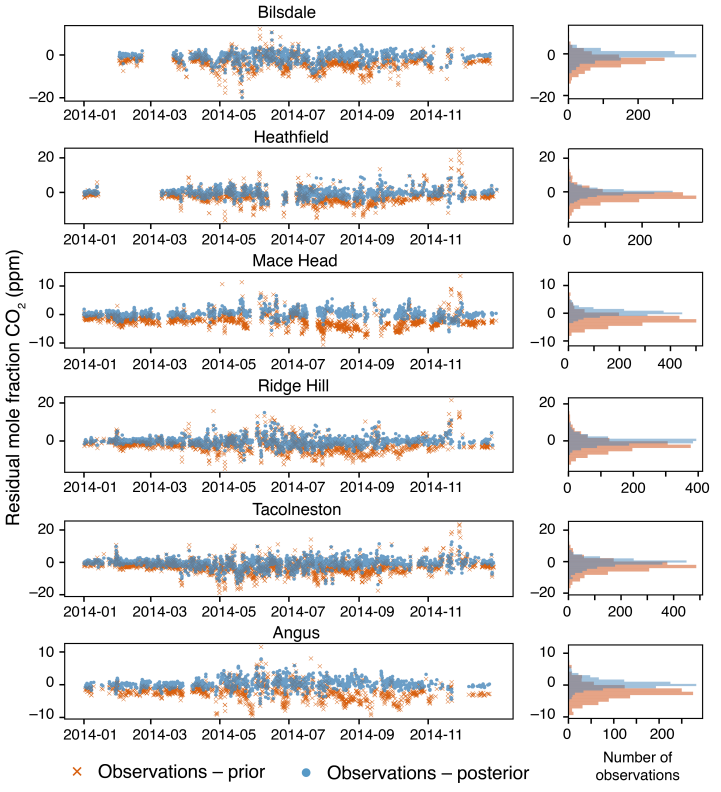
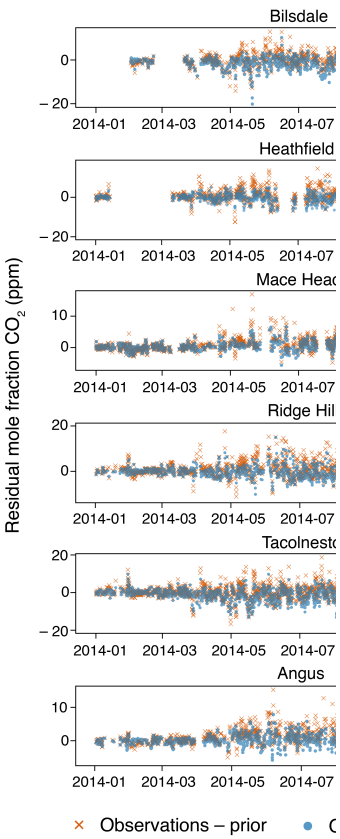


Figure S13: Left: Residual mole fractions for prior and posterior modelled CO_2 concentrations in 2014 using JULES prior biospheric fluxes. Right: Histogram of prior residuals (orange) and posterior residuals (blue). The mean of the histogram represents the mean bias.



Deleted:

Deleted: S10

Deleted: Prior modelled mole fractions are added to the posterior baseline before they are compared to the data.

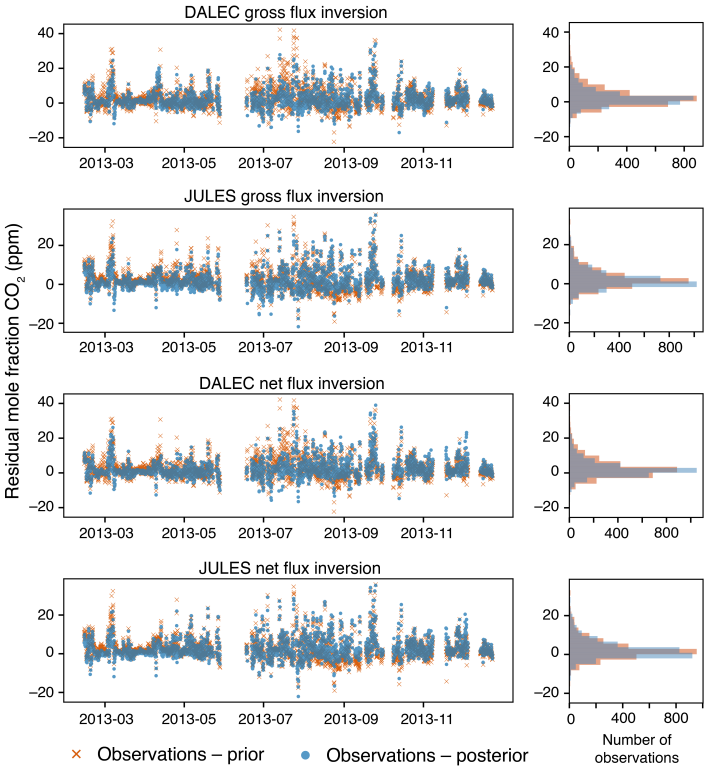


Figure S14: Left: Residual mole fractions for modelled CO_2 concentrations at Weybourne in 2013 using prior DALEC and JULES fluxes, and posterior DALEC and JULES fluxes from both the gross (scaling GPP and TER separately) and net (scaling just NEE) flux inversions. Weybourne data was not included in the inversions. Right: Histogram of residuals. The mean of the histogram represents the mean bias.

Moved up [7]: Left bars are prior NEE estimates, right bars are posterior NEE estimates. Dashed bars on the posterior estimates represent annual NEE fluxes for inversions that use fixed anthropogenic fluxes multiplied by ±

Deleted: ¶

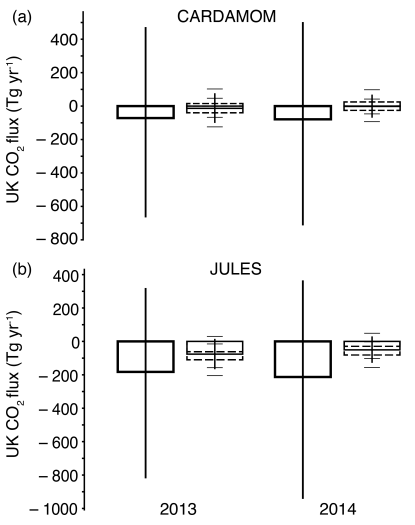


Figure S11: Annual UK NEE flux estimates from CARDAMOM and JULES inversions for 2013 and 2014.

Moved up [8]: Uncertainty bars represent 5th – 95th percentile. Solid uncertainty bars on posterior estimates are the uncertainty on the inversions using normal anthropogenic fluxes. Whiskers on the posterior estimates are the uncertainty on the inversions using anthropogenic fluxes multiplied by ±

Deleted: 10%.

Deleted: 10%.¶ ... [8]

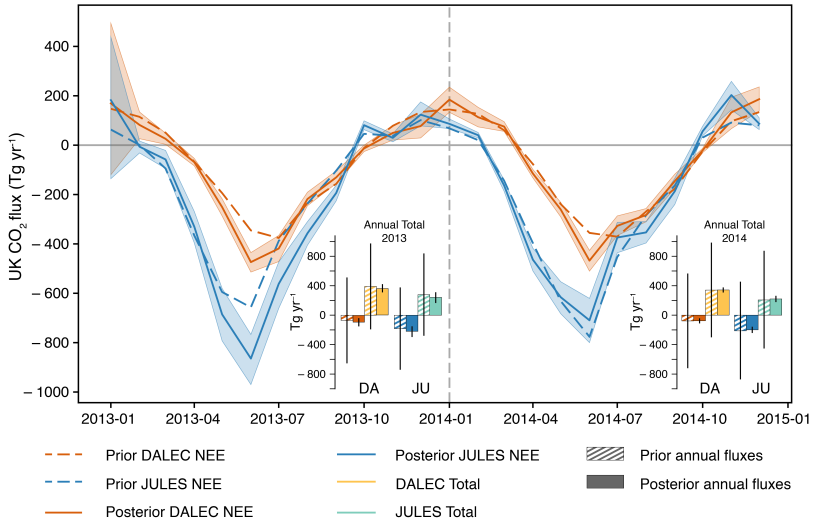


Figure S15: Posterior monthly net UK CO_2 flux (+ve is emission to atmosphere) for the inversion that scales only NEE rather than GPP and TER separately. Orange and blue monthly fluxes are posterior net biospheric (NEE) fluxes for DALEC and JULES respectively. Prior biosphere fluxes from DALEC and JULES are shown in dashed orange and blue lines respectively. Shading represents 5th – 95th percentile. The bar charts represent annual net UK CO_2 flux for 2013 (left) and 2014 (right). Hashed bars denote prior annual fluxes, solid bars denote posterior annual fluxes. The bar colours correspond to the line colours: left hand bars for each model are NEE fluxes, right hand bars for each model are total fluxes (NEE + fixed sources). Uncertainty bars represent 5th – 95th percentile. DA – DALEC. JU – JULES. NEE prior PDF ($\boldsymbol{x_{NEE}}$) has Gaussian uncertainty distribution and its standard deviation hyper-parameter ($\boldsymbol{\sigma_{x_{NEE}}}$) has a uniform distribution with a range reflecting an absolute uncertainty of approximately 40–400 Tg (see Table 3 for the comparable set-up for the separate GPP and TER inversion).

Deleted: S12

Deleted: .

Deleted: and CARDOMOM

Deleted: CA – CARDAMOM.

Deleted: 2

Formatted: Font: 9 pt

Page 9: [1] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 30: [2] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 34: [3] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 39: [4] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 39: [5] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 39: [6] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 39: [7] Deleted	**Emily White**	**2/13/19 11:20:00 PM**
Page 52: [8] Deleted	**Emily White**	**2/13/19 11:20:00 PM**