Authors response to the reviewers

We thank both reviewers and Tonatiuh Nuñez Ramirez for their comments, which have helped improve the manuscript. Below are our detailed response to reviewers (in blue) as well as a list of relevant changes in the manuscript (in red). When provided, the pages and line numbers refer to the "diff" version of the manuscript, provided at the end of this file.

In addition to the changes requested by the reviewers, the revised manuscript includes revised results for the LUMIA and CarboScope-Regional systems. The new LUMIA inversion fixes pre-processing errors in the preparation of the observations, and the CarboScope-Regional results were updated to harmonize the settings with those prepared for another study. These new results did not have strong consequences on the conclusions of the paper, but have made necessary a partial rewriting of Sections 4.1 and 4.2.3.

Response to review 1

General comments: the authors presented inverse biogenic fluxes estimates in Europe for over ten years (2006-2015) using the regional inversion technique as opposed to global inversions. These flux estimates were done under the protocol of using the same fossil fuel emission and a common database of in-situ CO2 observations, while the transport models, inversion approaches, the choices of observation and prior biogenic fluxes differed. The authors concluded that at the continental scale, the European ecosystems are a relatively small sink (-0.21±0.2 PgC/year), consistent with the results demonstrated by the global inversions in the previous studies. This conclusion is quite out of my expectation and quite different than my experience. I have a few comments below that may potentially change the conclusions and maybe improve the results.

The manuscript is generally well written. The analysis is well presented. The authors provided a reasonable interpretation of their results and are definitely aware of the caveats of the study.

However, after reviewing the submission guidelines of ACP and GMD. I found this MS is a better fit for a GMD publication. I am listing what I found below for the authors' reference.

> ACP research articles:
>
> "...Research articles must include substantial advances and general implications for the scientific understanding of atmospheric chemistry and physics. Manuscripts that report substantial new measurement results, but where the implications for atmospheric chemistry and physics are less developed, may be considered for publication as measurement reports (see below)...."
>
> GMD aims and scope:
>
> "...model experiment descriptions, including experimental details and project protocols;…"

ACP is one of the reference journals for publishing inverse modelling studies and for inversion inter-comparisons (e.g. www.atmos-chem-phys.net/13/9039/2013/, www.atmos-chem-phys.net/18/3047/2018/, www.atmos-chem-phys.net/16/1289/2016/, www.atmos-chem-phys.net/15/12765/2015). Our paper is in line with these studies and helps assessing the robustness of their findings (beyond the sensitivity tests or the diagnostics that individual inverse modellers usually run themselves). We derive conclusions regarding the European NEE that are of general interest for the scientific community (such as the one raised above by the reviewer himself regarding the European mean sink), and even though our study domain is Europe, some of the conclusions are relevant for inversions focusing on other parts of the world (the same models are used in different regions). We therefore disagree with the reviewer's comment and think that ACP is a perfectly well suited target journal for our study.

This work has collected the inversion results from different groups without a well-designed protocol. It is very challenging to fully understand the cause of the spread of the ensemble and the underlying uncertainties, which limits the scientific advances of this work. The results overall match the previous studies which were drawn from the global inversions. There are a few technical aspects that could be improved, and I will list them as follows. I have quite a lot of faith that the author should deliver better inverse estimates if those technical improvements are implemented.

The protocol was designed according to the objectives of the intercomparison. Here, it was not specifically to 'deliver better inverse estimates' (we rather consider it a task for each of the participating modeller to constantly improve their system and their estimates), but to provide a snapshot of the state of the art, and, more importantly, to explore the full range of uncertainties resulting out of an ensemble of sensible assumptions for the inversion set-up. ¶

Another objective for designing the protocol was to keep the extra work load for participating in the intercomparison low to allow for a large number of participants (only two groups were funded for this work).

Eventually the EUROCOM project has and will lead to improvements in the inversion systems, but that is not specifically the aim of this paper.

This and other comments have lead to clarifications of the manuscript, in particular in the introduction (lines 86-91 and 97-103) and the presentation of the protocol (lines 165-196). The two new paragraphs in the discussion also reply to that comment.

Given that it is an important paper for the EUROCOM project and can potentially provide some insight for future experiments, it is worth a publication. However, it's not clear to me how the authors handle boundary conditions and the related uncertainty in the inversions. The authors should have better clarifications on this aspect before re-submission.

The boundary conditions are specific to each system, and briefly described in Section 3.3.1 (which provides references for a more complete description for each system). We have reorganized the text of this section (and renamed it) to make it clearer.

It is not possible to use the same boundary condition in all systems since they rely on different types of transport models or configurations (for instance, CTE does not have boundaries since it is a global model, FLEXINVERT and LUMIA both rely on FLEXPART for regional transport, but FLEXINVERT uses global FLEXPART simulations while LUMIA uses regional ones, so essentially, the BC used in one inversion would not be relevant for the other one).¶

The uncertainty associated to the boundary condition (for the systems that have one) is generally considered part of the observational uncertainty (except in CTE (global) and NAME (which optimizes the BC as well)) and described in Section 3.3.3. Not all systems explicitly differentiate the uncertainty associated to the boundary condition from other uncertainties in the observation space (model error, observation error, etc.), as it is not easy to quantify them individually, but references are provided to papers describing the individual systems and justifying in more details these choices of uncertainties.

This comment has lead to a minor reorganization of Section 3.3.1 (the informations regarding the BC used by the six systems has been grouped together in one paragraph).

For an ACP publication, however, I would recommend the following improvements to start with. The authors built the results on top of a set of ensemble inversion results. The only protocol, for now, is to the same fossil fuel emission and a common database of in-situ CO2 observations. To understand what caused the large spread of the ensemble, fixing at least one of the model components is required.

Actually, to really understand what caused the large spread of the ensemble, an ensemble of ensemble would be needed, fixing, in each ensemble, one of the numerous parameters (prior fluxes, parameters of the prior uncertainty, control resolution, etc.) of the inversions which are potential sources of uncertainty. ¶

That's said, the authors should at least have one set of results using the same transport model, same prior fluxes, same boundary conditions, or the same observations. without the common setup, it limits the depth of the scientific understanding of this work.

By experience, we know that many parameters are uncertain and yield a significant portion of the resulting uncertainty. Weighting each contribution is a very expensive (computationally) and long exercise requiring lots of analysis. This was clearly out of the scope and capabilities of this first analysis of the inter-comparison.

Furthermore, in practice, many of the key parameters driving the uncertainty (the transport model, the formulation of the control vector, etc.) cannot be imposed to all inversion systems. The observation selection, the treatment of the boundary conditions (see above) and the model errors should depend on the transport model etc.

Our choice of a very loose protocol makes it more difficult to systematically analyse the influence of specific components of the inversion systems, but the results provide a realistic range for the European NEE, which was one of the main objectives of the intercomparison.

This and other comments have lead to clarifications of the manuscript, in particular in the introduction (lines 86-91 and 97-103) and the presentation of the protocol (lines 165-196). The two new paragraphs in the discussion also reply to that comment.

The regional inverse results in the MS do not appear to have better constraints than the global inversions. Technically, regional inversions can be driven by mesoscale transport. All of the experiments in the MS were driven with reanalysis or forecast data at ~101 to 102 km, and most of the meteorological forcing data do not have TKE. I suspect that these limitations are the main reasons that lead to unexpected equivalent results to the global inversions. I strongly recommend the working groups to use the mesoscale model output as the meteorological forcing for future experiments.

There were some mistakes in Table 2, which have been corrected in the revised manuscript and may have lead to some misunderstanding: several of the experiments were in fact driven by higher resolution meteorology (IFS operational forecast and UK Met-Office model).

That said, the equivalence of results with global inversions is not that unexpected: global inversions are supposed to derive accurate large-scale constraints. The advantage of regional inversions is their capacity to assimilate data from more sites, and derive constraints at higher resolution where and when the observation coverage is sufficient. As mentioned in the discussion, the regional inversions used here and the global inversions to which they are compared are also not at the same stage of maturity.

This comment has led to changes in Table 2 (correction of errors and harmonization of the denomination of the meteo data)

As I mentioned before, the boundary conditions need to be stated in more detail.

See our answer regarding this above.

The authors mentioned that the locations of some observations are very challenging for transport models in section 4.1. It may be a good idea to remove those challenging sites before inversions to avoid large transport errors. A detailed model-data mismatch would be more appreciated and help to improve the inverse flux estimates.

In future inversions, it will indeed be necessary for some systems to pay more careful attention to these challenging sites (especially HEI and GIF). This can be done by either excluding them or by applying a stricter data selection. Several systems account for this problem already by inflating the observation uncertainty, so that these sites do not have large impact on the results, but others, the impact of these sites can be strong, locally. This issue prevents us from analyzing the fluxes in details in the vicinity of these sites, but at the scale of the domain, or even of the larger regions used in the paper, other source of uncertainty dominate, and we are confident that this has only a minor impact on our conclusions at this stage.

It is difficult to propose much more detailed model-data mismatch than what is already in the paper and SI: there are too many inversions, too many sites and too many years to show everything, and the problematic sites or time of the day/year are not the same for all systems, so it is not even feasible to highlight specifically remarkable time series.

In part in response to this comment, a table showing statistics of the fit to observed concentrations has been moved from SI to the main manuscript (Table 3 in the revised manuscript)

Specific comments:

1. Line 25, spell “NBP” out.

We have fixed this.

2. Line 187, remove of “full”

We have fixed this.

3. Figure 2, do the authors know why VPRM looks so different than others?

Yes, this is in fact already explained in the manuscript (end of Section 3.2.1). VPRM is a diagnostic model (it assimilates eddy-covariance flux observations using a very simplified biosphere model, that can lead to a near zero respiration in winter in large parts of the domain). The CarboScope Regional (CSR) inversion accounts for this by adjusting an annual bias correction in addition to the 3-hourly NEE (this is one example where a stricter protocol forcing some systems to use a prior for which they are not designed might further degrade the results, see above).

4. Table 2, the author can one row of the choice of observation for each group.

We have added a row in Table 2. Note that this is a simplified description, and does not include the site selection (for space limitations, and because it is easy to get it from Figure 6). A full overview of the data selection is provided in Figure SI2.

5. Line 485, I would be not surprised by the smaller error reduction of the annual results than that of the monthly results due to the annual net NEE is close to zero.

Indeed, but we are not saying that it is surprising, we are just stating it as a result here, and we believe it is worthwhile doing so.

6. Line 704, change km2 to km 2

We have fixed this.

Response to review 2

This paper presents the first results of the EuroCom project, with an inter-comparison of net terrestrial ecosystem exchange estimated by 6 regional inversion systems following a flexible protocol that allows maximum participation and sensitivities to different transport, priors and number of in situ observations assimilated.

The manuscript is well written and the explanations are clear overall. The paper would benefit by explaining in more detail what is the purpose of comparing such wide range of diverse systems which produce such large spread in the optimized fluxes. What do we learn from such exercise? It seems that one of the messages is that all these differences in the configuration of the inversion systems have a large influence on the resulting optimized fluxes, i.e. assumptions in prior uncertainty estimation and data assimilation methodology, transport model, temporal/spatial resolution, boundary conditions, number of observations used, ocean fluxes, etc.

As explained in our reply to reviewer #1, the main aim is, at this stage, to assess the range of uncertainties that comes out from regional inversions (beyond the typical sensitivity tests that individual modellers typically run). This requires an ensemble maximizing the variability of the inversion setups. On the contrary, identifying the specific contribution of individual settings/design choices in the inversions requires a very controlled protocol (essentially only varying that parameter we want to estimate the influence of). The two aims are therefore difficult to reconcile in a single study, and we focused on the first one.

This and other comments have lead to clarifications of the manuscript, in particular in the introduction (lines 86-91 and 97-103) and the presentation of the protocol (lines 165-196).

The two new paragraphs in the discussion (Sec. 5 and 5.3) were also written specifically to address this comment.

The results point that regional inversions using the currently available in situ data in Europe are not able to properly constrain the NEE at regional scale, and the spread between different optimized fluxes is as large as the mean or median flux.

This comparison with the mean or median flux is not totally relevant: the net flux (NEE) is relatively small, but the gross fluxes (GPP and respiration) are very large, and the uncertainty that arises from our ensemble should rather be compared to the uncertainty on these terms. The inversions do manage to reduce the range of estimates compared to the priors (and more generally compared to bottom-up models) regarding key features such as the shape of the seasonal cycle. The lack of convergence regarding the annual budget and the IAV at the continental scale are also interesting findings in themselves, as it was not initially expected and will certainly motivate further developments and exchanges between the inverse modellers, that would not have happened if that exercise hadn't been performed.

The other aspect that could be improved is the presentation of the uncertainty from each individual inversion system. I have not seen the posterior uncertainty in any of the plots. It would be useful to add this information in the bar plots where the estimate from each system is compared.

It is actually difficult to obtain such a metric for all systems (see the lack of posterior uncertainties e.g. in Peylin et al., 2013, BGD). Not all of them can technically or practically compute it, and those who can usually do not compute it in ways that are easily comparable. Low or reduced rank inversion approaches can access it through analytical compations but variational inversions must be coupled to complex minimization schemes or Monte Carlo experiments to produce an approximation of the theoretical posterior uncertainty (Kadygrov et al., 2015, ACP). The uncertainty reduction that is sometimes computed in inversions is a diagnostic that is useful in some contexts (e.g. for network design studies with OSSEs using a single inversion framework), but it is highly theoretical (it strongly relies in all the statistical assumptions made by the inversion system) and it could be misleading (Henne et al., 2016, ACP).

<u>Specific comments/questions:</u>

1. The ICOS network is currently not a high density in-situ surface observation network (with 19 stations run by 12 countries as described in line 64).

The "high density" qualifier is subjective, but ICOS is clearly one of the densest continental-scale networks available to date.

We have replaced "high-density" by the slightly more neutral "dense".

2. The paper only addresses large regional budgets at subcontinental scale, not country scale budgets. Why not look at budgets for a relatively large country where there are enough observations, e.g. France or Germany to demonstrate the capability at country scale?

Although in some parts of Europe, the network might be dense enough for allowing some national-scale NEE estimation, this is rather the exception than the rule. Furthermore in those countries where it might be feasible, we do not think that our setups are the most appropriate: the resolution of our inversions (transport model, control vector, prior and other fluxes) is still coarse in contrast to the size of European countries. It might in fact be possible to obtain better or at least more consistent results at that country scale, using the same observational data but an experimental setup dedicated to that scale (smaller domain, higher-resolution fluxes, dedicated prior, etc.). The country scale is politically/societally sensitive therefore we prefer to remain on the cautious side (it should also be . The scale presented in the paper is the one at which we think our inversions are the most relevant.

3. The use of mesoscale transport model is not appropriate as mesoscale weather systems have scales of less than 100 km you need higher resolution than 10 to 100 km to resolve them. It would be best to replace mesoscale model with regional model.

As mentioned in our reply to reviewer #1, there were some errors in Table 2, listing the meteorological input data used by the different transport models. We have corrected this.

The resolution of the transport model is limited by the computational cost of high resolution transport simulations (the computational cost increases exponentially with the resolution). Furthermore, the spatial resolution of the model has to be adapted to that of the observation network. With observation sites separated by, at least a few hundreds km, our observation network does not actually provide very fine-scale constraints, so the benefit of using higher resolution transport model would be limited to a few very specific cases (sites with complex orography, or in vicinity of strong point sources).

4. What are the implications of not correcting for errors in the anthropogenic emissions and ocean fluxes? The signal of the anthropogenic emissions vary during the day. So if inversions

use observations at slightly different times of day, the influence of the anthropogenic emission error on the optimized flux will also vary. Could this explain part of the divergence between the optimized fluxes from the different inversion systems?

The impact of the ocean fluxes on the observations is very small, on the order of the observation uncertainty. It is unlikely that it can explain a significant part of the divergences.

The impact of anthropogenic emissions on the observations can definitively be important even though the signal at the stations used here is dominated by the NEE, which is why we all use the same anthropogenic emissions. Errors in these emissions could be affecting the inversions differently, but not just because they use observations at slightly different times, but also because all the transport models do not represent the sensitivity of the observation to the emissions with the exact same accuracy, and because not all inversions construct their observation error in the same way.

<u>Minor comments:</u>

-Line 26: Define NBP

Done

-Line56: . . .that does not smooth . . .

Done

-Line 119: Replace “the find” by “finding”.

Done

-Line 124: Please define “model error”, “representation error” and “aggregation error”.

The model error is the error due to lack of accuracy of the transport model (i.e. the error it would make even if it was driven by the true fluxes). The representation error is due to the representation of discrete or fine-resolution processes (fluxes, observations) on a coarser model grid. The aggregation error accounts for errors due to the control of fluxes at a coarser resolution than that of the transport model.

We have slightly simplified the sentence to remove the aggregation error (which isn’t formally accounted for by most of the inversions)

-Line 186: Remove the extra “full”.

Done

-Lines 193-203: Temporal resolution of ORCHIDEE prior is missing.

The temporal resolution of the ORCHIDEE product is 3-hours. We have clarified this.

-Lines 193-209: Spatial resolution of the prior from ORCHIDEE and PJ-GUESS is missing.

The spatial resolution is 0.5°, we have clarified this.

-Line 228: PgC/months? Shouldn’t the units be PgC/month?

Yes, we have corrected this.

-Line 233: Please include resolution of EDGARv4.3 inventory

The original resolution is 0.1°, but the product was regridded to 0.5° for the simulations. This has now been specified in the manuscript.

-Lines 241-243: Biomass burning emissions can be large over the summer in the Mediterranean region over the summer (e.g. 2007 and 2015). Could this also explain part of the large divergence between optimized fluxes in that region?

It may play a role, but given that the two systems that included a biomass burning product are at the two opposite extreme of the interval for summer emissions in Southern Europe, it is unlikely that this flux explains a large part of the differences. The Southern Europe region includes parts of the domain that are not well constrained by the observations (Greece and the Balkans), and on the other hand, there are many sites in Spain, which are not easy to represent. Furthermore, the divergences between the prior are the strongest in that part of our domain.

-Lines 253-257: The Takahashi et al. (2009) climatology will underestimate the ocean sink for the period 2006-2015, so this will explain part of the discrepancy between NAME-HB and other inversion systems. Other relatively out-of-date ocean data sets might lead to also an underestimation of the ocean sink. Does this mean that in those systems that do not correct the ocean fluxes, the error in the ocean sink will be attributed to NEE?¶

Yes, this is an inevitable feature of inversions: errors in non-optimized components of the inversion map into the optimized solution. It is however unlikely that the choice of ocean fluxes explain any significant fraction of the differences between the inversions: the impact of ocean fluxes (regardless of which ocean flux dataset is used) at the observation sites is very small and nearly negligible compared to that of the anthropogenic and biogenic fluxes (see figure below).

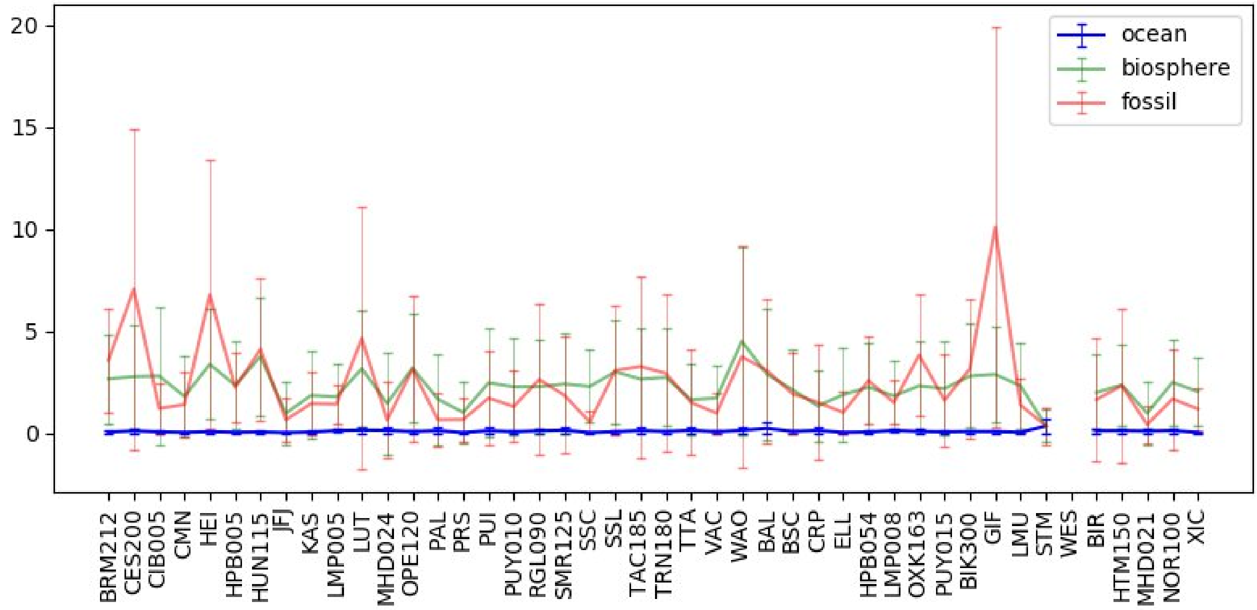

Figure 1: Mean absolute impact of the ocean (CarboScopev1.5), biosphere (LPJ-GUESS prior) and fossil flux estimates at the observation sites, in the LUMIA simulations. The error-bars mark the standard deviation over the 10 year simulation period.

-Line 304: What is the Rödenbeck approach?

In four of the systems (LUMIA, CHIMERE, CSR and NAME), the background concentration corresponds to the transport to the observation sites of a boundary condition at the edge of the domain. In the Rödenbeck approach, used in CSR and LUMIA, it is the global transport model from which that boundary condition is extracted which is used to transport it to the observation sites (while in CHIMERE and NAME, this is done directly by the regional transport model). The approach is described and justified in details in (Rödenbeck et al., 2009), we do not think that it is relevant to re-explain it in our manuscript.

-Line405: Remove extra bracket after Radon.

Done

-Line 409-411: What about the representation error associated with resolution of transport model? Shouldn't it be part of the observation uncertainty? Representation errors tend to be very large at sites close to anthropogenic emissions.

Yes, it is part of the "uncertainty of the forward transport model". The paragraph has been clarified.

-Line 416: How much does the observation uncertainty vary from site to site?

That part of the manuscript was in fact incorrect. The average uncertainty is on the order of 2 ppm and a minimum uncertainty of 1 ppm has been enforced for each obs. The average uncertainty ranges from 1.02 ppm (Mace Head) to 4 ppm (Puijo) for the year 2015. But these mean values hide larger variations: the uncertainty only reaches a maximum of 2.55 ppm at Lampedusa while it goes up to 32.7 ppm at Cabauw.

The manuscript has been corrected.

-Line 427: Is FLEXINVERT+ the only inversion system in which the uncertainties in the fossil fuel emission estimates contribute to the observation error? It seems to be this is an important uncertainty to consider given that most in situ sites in Europe are affected by anthropogenic emissions.

It is the only system to account for it explicitly, but the way other systems account for it when constructing their observation uncertainty vector. The uncertainties in FLEXINVERT are actually in the lower range of the ensemble.

-Line 479: "if" should be "it".

Done

-Line 486: "diagnostics" should be "diagnostic".

Done

-Figure 6: The differences between the prior and posterior appear to be very small at most sites. Does this mean that the prescription of the prior uncertainty is too small and transport uncertainty too large?

Not really. Larger prior uncertainties (or lower observation error) would lead to a better posterior fit to the observations, but not necessarily to more realistic flux adjustments (the inversion systems can "over-fit" the data, i.e. adjust fluxes to compensate the various model or observation errors). The balance of prior and observation uncertainties are set by the modellers in accordance to their experience with their inversion systems.

-Lines 600-603: The three sentences starting with “The figure. . .” would fit better in the caption of Figure 9.

Indeed, we have fixed this.

-Line 608: The Central European NEE is only robust in the sign of the budget, but not the magnitude (as shown in lower righ most panel in Figure 9).

The “most robust” qualifier is given in comparison to the other regions.

-Figure 9: It is not possible to read the figure caption.

We have edited the figure to improve the readability, and repeated part of the information in the caption.

-Line 687: “an constraint” should be “a constraint”.

We have corrected this

-Line 700: “the our four regions” should be “the four regions”.

We have corrected this

-Lines 698-700: If there is a deterioration in the optimized fluxes with respect to the prior fluxes in data sparse regions, doesn’t this mean that the assumptions in the inversion are not correct? One would expect that the optimized fluxes are always better or the same than the prior fluxes (where there are no observations).

This expectation would be verified if the inversion were all constrained with perfectly adequate uncertainty statistics (observations and prior uncertainties), which cannot be the case. In data sparse areas, the inversion strongly relies on the prior error covariance matrices to extrapolate the information from the few available observations. If these covariances poorly match the actual spatial distribution of NEE, the extrapolation can be highly erroneous. In regions where there are a lot of data, the NEE is more directly constrained by the data, and therefore its estimation is more robust.

-Line 730: The use of high resolution is relative. It’s probably best if you specify therange of spatial and temporal scales resolved by the regional inversion systems. A resolution of 0.5 degrees is not considered by most as high spatial resolution. Temporal resolution is not high either if observations are filtered in time to short afternoon and nighttime windows.

The high spatial and temporal resolution is indeed relative to the atmospheric inversion field (global inversions run at best at a 2° resolution).

We have modified the lines according to the reviewer’s suggestion

-Line 731: Given the spread of the optimized fluxes is so large, can the data be used as a validation data set?

The spread of bottom-up models is actually even larger. While none of the inversions should be used individually as a validation dataset, the ensemble statistics (mean/median and spread/standard-deviation provide a good representation of the likely interval of NEE that can be derived from atmospheric CO_2 observations.

We have nonetheless removed the sentence as it indeed can be confusing.

-Lines 735-736: Please include the associated uncertainty of the posterior estimate.

We have added it.

Response to short comment 1

I believe this study is very relevant for assessing the consistency of regional scale NEE estimates from regional inverse modeling systems.

I have some comments regarding the display of information:

In figure 1, it is difficult to obtain information on the temporal density and continuity of measurements from the size of the dots and there is no key. An additional figure similar to figure 2 in Kountouris et al. (2018) or figure 2 in Rödenbeck et al. (2003) would be more useful.

The main purpose of the figure is to show the location of the observation sites and the definition of the regions used further in the paper. The size of the dots is really only qualitative, as the amount of constraint each site provides depends on the specific data selection and uncertainty attribution in each system.

Figure 2 and 3 would be more useful if it also included the sub-continental regions (at least as supplementary information).

We now provide regional versions of Figure 2 and 3 in supplementary information

Figure 6 could be extend to also include other metrics, specifically correlations and standard deviation. I believe it would be more useful not to divide the plots by model but by metric in order to facilitate the model comparison. Order of the sites on the x-axis could be by latitude, longitude or altitude to observe if there are gradients.

We have tried various presentations of the model-data mismatches and settled on that one as it suited best our needs (presenting a synthetic overview of the model-data mismatches).

The revised paper now include one additional table (Table 3, which was previously only provided as SI), which partly addresses that comment.

It was my impression that since we are optimizing towards real data we are missing an assessment on how realistic the fluxes really are. Here I believe comparison of grid scale fluxes to Eddy Covariance measurements would be useful as well as comparison of spatial patterns (e.g. the spatial correlation and gradients) with satellite vegetation fluorescence products.

In theory, we agree that the comparison with EC data would be useful, but it is not as simple as just comparing the EC and inverse-modelling derived fluxes: they are representative of very different spatial scales (a few km2 for EC data, a few thousands of km2 for inversions), therefore either the inversions need to be subsampled or the EC data needs to be upscaled, neither of which is straightforward. Furthermore, EC fluxes come with their own uncertainties and biases. Doing the comparison properly could practically be a study on its own. Likewise, the comparison with fluorescence data and inversion fluxes should only be done in a dedicated study (fluorescence is a proxy of GPP, not of NEE, the direct comparison between the two is not possible).

Furthermore, the use of dense measurement networks has the aim to distinguish small scale flux patterns. However, most of the analyses were at continental scale. I believe more analyses and discussion of the fluxes at the sub-continental scales is needed both for seasonal and interannual variability. At the interannual time scale, it would be useful to know if the variability shown by the inverse models reflects heat waves, droughts, cold spells, etc. If they are able to detect land use change or errors in the anthropogenic emissions.

Here again, we agree that some of these analyses could be interesting, but we do not think that they should be part of this paper. Our inversions span a period of 10 years and the whole European continent. The list of potential climatic anomalies that could be studied is endless, and the paper is already long. This type of analysis is better done in papers dedicated to studying these specific climatic events, and crossing more information sources than just inversions. Our study can then help assessing the relevance of inversions in that context.

Finally, the study only recommends increasing the density of the observation network, particularly in Southern and Eastern Europe. However, no analyses were made on the effects of the modeler's choices, e.g. measurement and prior errors, data selection, use of ocean fluxes. This choices could provide further recommendation for the development of these regional inverse modeling systems.

We have partly replied to this in our answers to the anonymous reviewers: the goal of this study is to identify and quantify discrepancies between the inversion results (and therefore estimate the robustness of regional inversions). Identifying the contribution of different modeler's choices to that overall uncertainty is definitely one of the aims of the EUROCOM project, but it is outside the scope of this paper (it is difficult to do both: estimating the overall uncertainty calls for an ensemble as diverse as possible, while estimating the contribution of one specific choice requires a more controlled protocol).

This and other comments have lead to clarifications of the manuscript, in particular in the introduction (lines 86-91 and 97-103) and the presentation of the protocol (lines 165-196). The two new paragraphs in the discussion also reply to that comment.

References

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., Van Der Laan-Luijkx, I. T., & Zhang, X. (2013). Global atmospheric carbon budget: Results from an ensemble of atmospheric CO2 inversions. Biogeosciences, 10(10), 6699–6720. https://doi.org/10.5194/bg-10-6699-2013

Henne, S., Brunner, D., Oney, B., Leuenberger, M., Eugster, W., Bamberger, I., Meinhardt, F., Steinbacher, M., & Emmenegger, L. (2016). Validation of the Swiss methane emission inventory by atmospheric observations and inverse modelling. Atmospheric Chemistry and Physics, 16(6), 3683–3710. https://doi.org/10.5194/acp-16-3683-2016

Kadygrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C., & Ciais, P. (2015). On the potential of the ICOS atmospheric CO 2 measurement network for estimating the biogenic CO 2 budget of Europe. Atmospheric Chemistry and Physics, 15(22), 12765–12787. https://doi.org/10.5194/acp-15-12765-2015

The regional EUROpean atmospheric transport inversion COMparison, EUROCOM: first results on European wide terrestrial carbon fluxes for the period 2006-2015

Guillaume Monteil[1], Grégoire Broquet[2], Marko Scholze[1], Matthew Lang[2], Ute Karstens[3], Christoph Gerbig[4], Frank-Thomas Koch[8,4], Naomi E. Smith[6], Rona L. Thompson[7], Ingrid T. Luijkx[6], Emily White[5], Antoon Meesters[9], Philippe Ciais[2], Anita L. Ganesan[5], Alistair Manning[5], Michael Mischurow[1], Wouter Peters[10], Philippe Peylin[2], Jerôme Tarniewicz[2], Matt Rigby[5], Christian Rödenbeck[4], Alex Vermeulen[3], and Evie M. Walton[5]

[1]Dep. of Physical Geography and Ecosystem Science, Lund University, Sweden
[2]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[3]ICOS Carbon Portal, Lund University, Sweden
[4]Max Planck Institute for Biogeochemistry, Jena, Germany
[5]University of Bristol, Bristol, United Kingdom
[6]Meteorology and Air Quality, Wageningen University and Research, Wageningen, the Netherlands
[7]NILU – Norsk Institutt for Luftforskning, Kjeller, Norway
[8]Deutscher Wetterdienst, Germany
[9]Vrije Universiteit Amsterdam, The Netherlands
[10]Centre for Isotope Research, University of Groningen, Groningen, the Netherlands

Correspondence: G. Monteil (guillaume.monteil@nateko.lu.se)

Abstract. Atmospheric inversions have been used for the past two decades to derive large scale constraints on the sources and sinks of CO_2 into the atmosphere. The development of ~~high density~~ dense in-situ surface observation networks, such as ICOS in Europe, enables in theory inversions at a resolution close to the country scale in Europe. This has led to the development of many regional inversion systems capable of assimilating these high-resolution data, in Europe and elsewhere. The EUROCOM project (EUROpean atmospheric transport inversion COMparison) is a collaboration between seven European research institutes, which aims at producing a collective assessment of the net carbon flux between the terrestrial ecosystems and the atmosphere in Europe for the period 2006-2015. It aims in particular at investigating the capacity of the inversions to deliver consistent flux estimates from the country scale up to the continental scale.

The project participants were provided with a common database of in-situ observed CO_2 concentrations (including the observation sites that are now part of the ICOS network), and were tasked with providing their best estimate of the net terrestrial carbon flux for that period, and for a large domain covering the entire European Union. The inversion systems differ by the transport model, the inversion approach and the choice of observation and prior constraints, enabling us to widely explore the space of uncertainties.

This paper describes the intercomparison protocol and the participating systems, and it presents the first results from a reference set of inversions, at the continental scale and in four large regions. At the continental scale, the regional inversions

support the assumption that European ecosystems are a relatively small sink (-0.21±0.2 PgC/year). We find that the convergence of the regional inversions at this scale is not better than that obtained in state-of-the-art global inversions. However, more robust results are obtained for sub-regions within Europe, and in these areas with dense observational coverage, the objective of delivering robust country scale flux estimates appears achievable in the near future.

1 Introduction

The carbon budget of Europe has been explored in several large scale synthesis studies, such as the CarboEurope-Integrated Project (Schulze et al., 2009) and the REgional Carbon Cycle Assessment and Processes project (RECCAP; Luyssaert et al., 2012), to name a few. Although these have helped refining the knowledge of the European carbon cycle, large uncertainties remain regarding the quantification of the flux between terrestrial ecosystems and the atmosphere, usually quantified as the Net Ecosystem Exchange (NEE), i.e. the sum of emissions (TER, i.e. autotrophic and heterotrophic respiration) and uptake (GPP, i.e. photosynthesis) of carbon by ecosystems to and from the atmosphere, or alternatively NBP (Net Biome Production), which includes the impact of ecosystem disturbances (fires, land use change, etc.). For instance, Luyssaert et al. (2012) report average estimates of European land carbon sink in a -200 to -360 TgC/year range for the years 2001-2005, depending on the estimation method, and each of these estimates are provided with large uncertainties (with 1-sigma relative uncertainties of 50 to 100%). Confronting the ensemble of results from different syntheses, Reuter et al. (2017) report annual land-atmosphere flux ranging from -400±420 TgC/year up to -1030±470 TgC/year in the 2000s. Beyond the annual long-term budget, the year to year annual flux variations are also poorly known (Bastos et al., 2016). In practice, the lack of a robust and precise quantification of the natural CO_2 fluxes in Europe limits our ability to understand the links between the NEE flux and external forcings such as e.g. meteorological variability (including the impact of extreme events like droughts and cold spells) and trends (Ciais et al., 2005; Maignan et al., 2008) or land use change (Naudts et al., 2016), and to forecast the evolution of the land sink in Europe, in the context of global climate change.

Despite the large uncertainties, there is a growing demand from the policy makers and the society in general for more accurate and relevant numbers, such as estimates of the national budgets of CO_2 fluxes, these demands being reinforced by the Paris Agreement. For instance, the European Commission (under the VERIFY and CHE H2020 projects) is supporting the development of observation based monitoring systems for estimating CO_2 fluxes at national to sub-national scales, with a clear interest in both land ecosystems fluxes and the anthropogenic emissions.

Atmospheric transport inversions rely on transport models and statistical methodologies to derive the most likely estimates of CO_2 fluxes given large datasets of observed atmospheric CO_2 concentrations and a prior information provided in general by ecosystem models. Global inversion systems, using coarse resolution global transport models (typically $>2°$), have so far been the dominant tool for producing top-down estimates of NEE fluxes. The coordination of the inverse modelling community through intercomparison exercises with ≈10 global inverse modelling systems, such as that conducted in the frame of the TRANSCOM and RECCAP projects (Law et al., 1996; Gurney et al., 2002; Patra et al., 2008; Peylin et al., 2013) have been valuable for understanding the strengths and weaknesses of global inversions and to characterise the real uncertainty of the different estimates. However, despite this long term effort, global inversions remain limited by the coarse resolution of the transport models they rely on, as these do not allow a proper representation of observation sites in regions with complex orography or nearby large anthropogenic CO_2 emissions and do not reproduce the high-resolution spatial variability of the CO_2 concentrations that is captured by dense networks.

Regional scale inversions started to emerge about a decade ago. They rely on mesoscale transport models (at 1° down to 10 km resolution), capable of better representing the spatial and temporal variability of concentrations observed by dense networks of CO_2 observations, such as that of the Integrated Carbon Observation System (ICOS) in Europe. In particular the models should be able to account for CO_2 fluxes at a scale that does not smoothes smooth too much the hot spots of fossil fuel CO_2 emissions in cities and industrial areas. They demonstrated some potential to solve for continental to subcontinental budgets at the monthly scale (e.g. Peters et al. (2007); Rödenbeck et al. (2009); Schuh et al. (2010); Gourdji et al. (2012); Broquet et al. (2013); Meesters et al. (2012)). However, the sparse efforts for routinely producing until recently, there has only been limited effort to routinely perform regional inversion estimates (beyond the scope of specific studies), as well as (up until recently) in Europe (partly owing to the difficult access to long-term time series of quality-controlled CO_2 datafrom many sites in Europe, were limiting their development. For those reasons, most synthesis studies up to now kept relying on results). Most published synthesis studies have therefore relied on European NEE estimates from global scale inversionsfor European NEE, based on networks of global global networks of mostly background sites.

The ICOS atmospheric network (icos-atc.lsce.ipsl.fr) is now operational and its number of stations should regularly increase from the current 19 labelled stations towards at least 34 stations37 sites, run by currently twelve and hopefully in the future more European member states. Precursor networks such as those set-up in the framework of the CarboEurope and GHG-Europe projects (Ramonet et al., 2010) and the ICOS preparatory phase provide a robust basis for regional inversions during the pre-ICOS decade. The ICOS Carbon Portal (www.icos-cp.eu) has been set-up to support the exchange of observational data and elaborated products related to the carbon cycle, such as CO_2 fossil fuel flux maps. In addition to in-situ data, the development of satellite observations of CO_2 following the launch of GOSAT (Kuze et al., 2009) in 2009 and OCO-2 in 2014 (Crisp et al., 2004) should further densify the observation coverage, in particular with the foreseen European constellation of CO_2 high resolution imagers of the Copernicus Anthropogenic CO_2 Monitoring mission (CO2M; Pinty et al., 2017), starting from 2025. The use of mesoscale transport model will then become necessary to fully exploit the potential of these large datasets with observations at high spatial and temporal resolution.

In this context, the EUROCOM (EUROpean atmospheric transport inversion COMparison) project aims to coordinate at coordinating a European effort to improve the knowledge on the NEE based on an ensemble of long-term European scale inversions (i.e. covering geographical Europe). The participating groups were tasked with performing an ensemble of mesoscale CO_2 inversions of the European NEE for the period 2006-2015, following a protocol described further in this document. A large dataset of surface CO_2 observations, combining measurements from several European networks and individual research stations was compiled and provided to the participants. A total project involved the participation of seven research groupsparticipated in the project, producing , which have produced an ensemble of more than (to date) 10 12 inversions (including sensitivity experiments), estimating the European NEE for the period 2006-2015 following a protocol described further in this document.

The EUROCOM project is therefore one of the first regional inversion intercomparisons, and the first one at such a scale dedicated to the European NEE.

~~This paper presents the protocol of the intercomparison and a first set of analyses of the results. The inversions were provided by six different groups, with six inversion systems~~The first task of the project, to which this paper is dedicated, is to assess the capacity of regional inversions to robustly estimate European NEE. We focus on key diagnostics which are typically looked at in synthesis studies, such as the annual and monthly budgets of NEE and the inter-annual variability, for the whole Europe and for large regions. We use the results from an ensemble of six inversions (one for each participating system), covering a large spectrum of inversion characteristics (prior constraints, inversion technique, transport models, etc.): PYVAR-CHIMERE (Broquet et al., 2011; Fortems-Cheiney et al., 2019, developed at LSCE, France); LUMIA (Lund University Modular Inversion Algorithm) (Monteil and Scholze, 2019), developed at Lund University (Sweden) as part of the EUROCOM project; CarboScope-Regional (Kountouris et al., 2018a, b, developed at MPI-Jena, Germany); FLEXINVERT ~~+~~(Thompson and Stohl, 2014, from NILU, Norway)); NAME-HB (White et al., 2019, from the University of Bristol, United Kingdom) and CarbonTracker Europe (Peters et al., 2010; van der Laan-Luijkx et al., 2017), from the University of Wageningen, the Netherlands.

The ~~analysis focuses on assessing whether these regional inversions help to better characterise the annual to monthly budgets of NEE for the whole Europe. It also provides first insights on the robustness of the sub-continental flux estimates. The advantages and current limitations of regional inversions, compared to global ones, are also discussed~~EUROCOM project extends beyond the scope of this paper. Forthcoming studies ~~will provide a more in-depth analysis of the whole inversion ensemble , with the aim to better understand the strengths and weaknesses of the regional inversion, characterise their sources of uncertainties and attempt at~~ using a larger ensemble of inversions (including additional sensitivity experiments) will focus on quantifying and reducing specific aspect of the uncertainty, supporting the improvement of both the regional inversion techniques and the design of the European observation network.

The manuscript is organised in five sections. Section 2 briefly summarises the theoretical background behind atmospheric transport inversions. Section 3 details the inversion protocol, the participating inverse modelling systems, and the input products (fluxes and observations) shared within EUROCOM for conducting the inversions. Results are presented in Section 4 and then discussed in Section 5. Finally, Section 6 summarises the paper and provides some remarks on the future of the EUROCOM collaboration, and on regional inverse modelling in general.

2 Inverse modelling methodology / terminology

The theoretical framework of the atmospheric inverse transport modelling has been extensively detailed in past publications (e.g. Enting, 2002; Rayner et al., 2018). Here we only give a brief overview of the basic principles, to facilitate the comprehension of the paper for readers unfamiliar with the approach and to remind of some of the components discussed in detail in Section 3.

Bayesian atmospheric inversions rely on the fact that observed spatio-temporal gradients of CO_2 in the atmosphere reflect the distribution of carbon exchanges between the atmosphere and other carbon reservoirs. The link between the net CO_2 exchange at the surface and the CO_2 concentrations in the atmosphere is established by a forward atmospheric transport model. A first set of modelled CO_2 concentrations ($\mathbf{y^m} = H(\mathbf{x})$) is computed at the time and location of real observations ($\mathbf{y^o}$), based on a prior assumption of what the CO_2 fluxes are ($\mathbf{x_b}$). The mismatch between the modelled and observed concentrations ($\delta\mathbf{y} = H(\mathbf{x}) - \mathbf{y^o}$) is used to derive a correction $\delta\mathbf{x}$ to the prior flux estimate $\mathbf{x_b}$. The posterior flux estimate ($\mathbf{x} = \mathbf{x_b} + \delta\mathbf{x}$) then represent the best statistical compromise between fitting the observations and limiting the departures to the prior, accounting for the statistical distribution of uncertainties in both observations and prior fluxes.

The vector $\mathbf{x}$ is called the control vector. It contains all the parameters that the inversion can adjust. In our case it contains at least the terrestrial ecosystem component of the CO_2 fluxes. It can also contain other adjusted parameters such as bias or boundary concentration terms. The operator H, which establishes the deterministic relationship between a given control vector $\mathbf{x}$ and the corresponding modelled concentrations $\mathbf{y^m}$ is called the observation operator. It encompasses the transport model, but also the impact on the modelled concentrations of any input of the transport model that is not further adjusted in the inversions (prescribed anthropogenic emissions, boundary conditions, etc.).

Following the Bayesian approach and using classical Gaussian errors hypothesis the problem reduces to the find finding the posterior control vector $\mathbf{x_a}$ that minimises the cost function $J(\mathbf{x})$, defined as:

$$J(\mathbf{x}) = \underbrace{\frac{\mathbf{1}}{\mathbf{2}}\delta\mathbf{x}^{\mathbf{T}}\mathbf{B}^{-1}\delta\mathbf{x}}_{J_b} + \underbrace{\frac{\mathbf{1}}{\mathbf{2}}\delta\mathbf{y}^{\mathbf{T}}\mathbf{R}^{-1}\delta\mathbf{y}}_{J_{obs}} \qquad (1)$$

The prior error covariance matrix $\mathbf{B}$ contains a representation of the uncertainties on the prior control vector $\mathbf{x_b}$ and the error covariance matrix $\mathbf{R}$ contains an estimation of the uncertainties in the model data mismatches $\delta\mathbf{y}$ (observational uncertainty , and uncertainties from the observational error, which combines the measurement uncertainties (uncertainty on the observations $\mathbf{y}$) and an the uncertainties associated to the observation operator such as model error, representation error and aggregation error H: the representation error (due to the comparison of point concentration measurements with gridded model concentration) and the model error (uncertainty in non-optimized model parameters, such as the boundary and initial conditions, and the non-optimized fluxes, and uncertainty in the model physics). Departing from the prior control vector $\mathbf{x_b}$ increases J_b, and improving the fit to the observations reduces J_{obs}. $\mathbf{B}$ and $\mathbf{R}$ modulate the relative weight of each departure to the prior and to the observations in J.

The exact specifications of $\mathbf{B}$ and $\mathbf{R}$ affect to a certain extent the outcome of an inversion. For practical reasons, the error covariance matrix for the observations, $\mathbf{R}$ is usually defined as a diagonal matrix with the measurement and model uncertainty (σ) for each observation site specified on the diagonal. Potential error correlation between observations are typically dealt with by limiting the density of observations or inflating their individual uncertainties. The diagonal elements of the prior error covariance matrix $\mathbf{B}$ contains the uncertainties on the prior control parameters (typically here the NEE at the grid scale). The off-diagonal elements, corresponding to the covariances between uncertainties in different control parameters, are difficult to specify because the uncertainties in the NEE estimates have hardly been characterised and quantified (Kountouris et al., 2015). They are however a critical component of the inversion as they determine how independently from each other the different components of the control vector can be adjusted. The inversions in this study follow different implementations of this general methodology, listed in Section 3.3.2.

The optimal control vector $\mathbf{x_a}$ can be solved for using different solution methods. Here we only briefly recall the methods employed by the systems in this study (variational and sequential ensemble approaches, and Markov Chain Monte Carlo), more information on these methods is given in Rayner et al. (2018) and references therein.

The variational method minimises $J(\mathbf{x})$ based on iterative gradient descent methods. Efficient implementations of this method rely either on the availability of the adjoint of the transport model or pre-computed transport Jacobian matrices representing the sensitivity of the observation vector to the control vector. The Monte Carlo approach directly samples the cost function, and in the case of the Markov chain Monte Carlo (MCMC) approach, the samples form a Markov chain, i.e. each sample is not obtained independently, but rather a perturbation of the last previously accepted sample. This allows non-Gaussian PDFs to be used in the inversion, and allows the specification of uncertainties to be explored in so-called "hierarchical" Bayesian frameworks (Ganesan et al., 2014; Lunt et al., 2016). Finally, the Ensemble Kalman Filtering (EnKF) directly derives $\mathbf{x^a}$ following its analytical formulation based on the reduction of the dimensions of the problem through the split of the inversion into sequential windows, and based on on the computation of the matrices involved in the EnKF formulation through an ensemble Monte Carlo approach.

3 Protocol and participating models

~~The main product requested from the participating groups was~~ Given the overall objective of the study (to assess the robustness of inversion-derived European flux estimates), we deliberately opted for a relatively loose protocol: the participants were requested to use a common set of anthropogenic CO_2 emissions (fossil fuel combustion, cement production and large-scale fires) and to use only atmospheric observations from a common dataset, prepared specifically for the EUROCOM project (Section 3.1). They were requested to provide a monthly gridded estimate of the net land-atmosphere CO_2 exchange (Net Ecosystem Exchange, NEE) over the period 2006 to 2015, covering at least the area 15°W-35°E by 33°N-73°N, at a 0.5° by 0.5° spatial resolution (independently of the actual resolution of the inversions).

~~The participants were to a certain extent free to choose their "best" inversion set-up except for a few restrictions and guidelines set out in the EUROCOM inversion intercomparison protocol. The only mandatory requirement for all inversion system was to use a common dataset of anthropogenic CO_2 emissions (fossil fuel combustion, cement production and large-scale fires) as detailed below (Section 3.2.2) and to use only atmospheric observations from a common dataset, prepared specifically for the EUROCOM project (Section 3.1). The precise data selection within that~~ A set of fluxes (prior NEE, anthropogenic and ocean fluxes) was made available to the modellers through a data repository hosted at the ICOS Carbon Portal, along with the common observation database, but except for the imposed anthropogenic emissions, the participants were essentially free to choose the characteristics of their inversions. In particular they could perform further selections on the observations database (selection of observation sites ~~,~~ and selection of observations to use at each site) and ~~the definition of observation uncertainties were also left to the modellers.~~

choose their prior NEE and ocean flux estimates. The treatment of boundary conditions ~~, of meteorological input data, the use of an ocean flux~~ and the precise specification of uncertainties (~~on the prior and on the observations) were~~ prior and observation uncertainties) were also left to the modellers. ~~A set of fluxes (prior NEE, anthropogenic and ocean fluxes) was made available to the modellers through a data repository hosted at the ICOS Carbon Portal, along with the common observation database.~~

~~Note that we use the term NEE (sum of photosynthesis and ecosystem respiration) for the posterior fluxes throughout the paper, because this is what the priori flux estimates from the terrestrial ecosystem models represent. However, strictly speaking, the inversions optimise the flux that is not explained by the prescribed anthropogenic (and ocean) fluxes. This includes the effect of ecosystem disturbances (land use, land management, biotic effects) but also any errors in the prescribed fluxes~~The inversions in our ensemble are constructed with the aim to maximize the diversity of the systems, and hence to obtain a better estimate of the overall uncertainty of inversion results. A stricter protocol, fixing e.g. the prior fluxes and the prior uncertainties would facilitate the interpretation of results but would artificially decrease their spread. Furthermore, some parameters should not be prescribed. For instance, the observations are selected based on the capacity of the underlying transport model to reproduce them, and the type of boundary condition is also dependent on the transport models, which differ across the systems. It is however clear that further steps will be needed, with a stricter protocol, to fully understand the specific causes of the discrepancies that were obtained.

3.1 Common atmospheric observation database

A comprehensive data set of atmospheric CO_2 concentration observations in Europe was compiled as input for the inversion systems, on the basis of the GLOBALVIEWplus v3.2 Observation Package (ObsPack), a product compiled and coordinated at NOAA's Earth System Research Lab together with the ICOS Carbon Portal (Cooperative Global Atmospheric Data Integration Project, 2017). The data set was further extended by including measurements that had been collected in several national and EU-funded projects, like CarboEurope-IP, GHG-Europe, and during the preparatory phase of the Integrated Carbon Observation System (ICOS) Research Infrastructure. Finally, for two stations, the data were obtained from the World Data Center for Greenhouse Gases (https://gaw.kishou.go.jp/).

Compared to the original GLOBALVIEWplus product, we added time series from nine measurement stations and partly complemented time series at two stations. The data sets were harmonised with respect to format and sampling interval, and provided in the ObsPack format (Masarie et al., 2014). The original datasets and data providers of the time series are reported in Table 1, and the locations of the observation sites are also shown in Figure 1.

The majority of sites (35 out of 39) sample concentrations continuously (i.e. hourly or more frequent); 18 sites are tall towers (intake height > 50m), some with observations available at different levels, in which case only the upper level was used (as more difficult for the transport models to represent concentration gradients close to the ground).

The modellers were free to refine the observation selection according to the the ability of their inversion systems to simulate specific stations, and in particular to use their preferred approach to select data within a day (i.e. use of all the observations within a time frame or use of an average of the observations, etc.). The precise observation selection approaches are discussed further in Section 3.3.3, and a full ~~full~~ comparison of the observation assimilated by each system is provided in Figures SI1 and SI2.

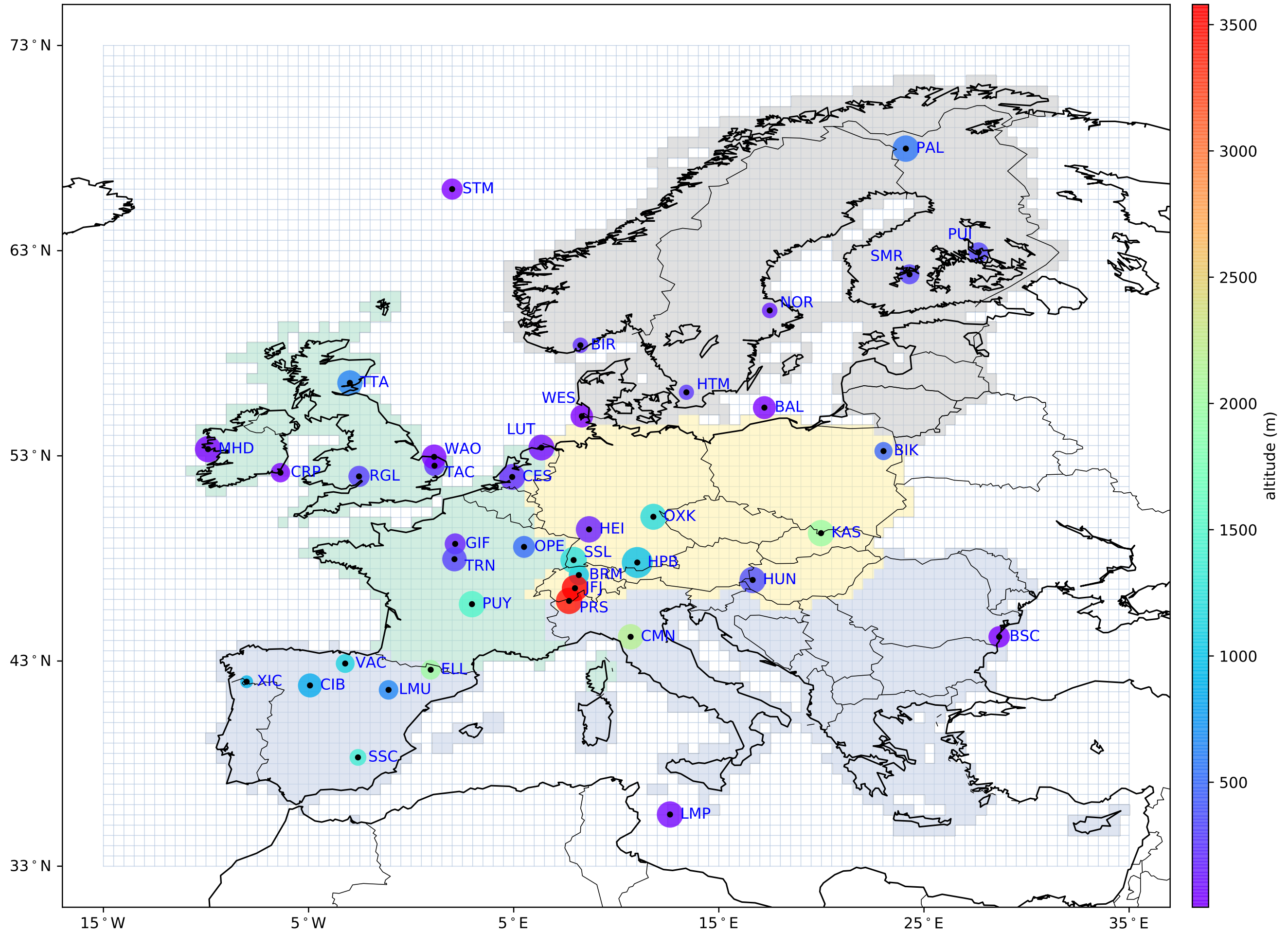

Figure 1. EUROCOM domain (pale blue grid with the 0.5° resolution) and location of the observation sites. The size of the dots is proportional to the number of months with at least one observation available (in the common observation database, not all observations are used in the inversions), and the colour map shows the altitude of the sites (height above ground + sampling height). The four regions used in the analysis are also represented: Western Europe (green), Southern Europe (blue), Central Europe (yellow) and Northern Europe (grey).

Code	Station Name	Lat (°N)	Lon (°E)	Alt (m.a.s.l.)	Intake (m.a.g.l.)	C/F	Period	Dataset	Data Provider
BAL	Baltic Sea	55.350	17.220	3	25	F	2006-2011	GV+ v3.2	NOAA
BIK	Bialystok	53.232	23.027	183	300	C	2006-2007	preICOS	MPI-BGC
BIR	Birkenes	58.389	8.252	219	2	C	2015	GV+ v3.2	NILU
BRM	Beromunster	47.190	8.175	797	212	C	2012-2015	GV+ v3.2	Uni.Bern
BSC	Black Sea Coast	44.178	28.665	0	5	F	2006-2011	GV+ v3.2	NOAA
CES	Cabauw	51.971	4.927	-1	200	C	2006-2015	GV+ v3.2	ECN
CIB	Centro de Investigacion de la Baja Atmosfere	41.810	-4.930	845	5	F	2009-2015	GV+ v3.2	NOAA
CMN	Monte Cimone	44.180	10.700	2165	12	C	2006-2015	WDCGG	IAFMCC
CRP	Carnsore Point	52.180	-6.370	9	14	C	2010-2013	preICOS	EPA
ELL	Estany Llong	42.575	0.955	2002	3	F	2008-2015	GV+ v3.2	ICTA-ICP
GIF	Gif sur Yvette	48.710	2.148	160	7	C	2006-2009	preICOS	LSCE
HEI	Heidelberg	49.417	8.674	116	30	C	2006-2015	GV+ v3.2	UHEI
HPB	Hohenpeissenberg	47.801	11.024	985	5	F	2006-2015	GV+ v3.2	NOAA
HPB	Hohenpeissenberg	47.801	11.010	934	131	C	2015	GV+ v3.2	DWD-HPB
HTM	Hyltemossa	56.098	13.419	115	150	C	2015	GV+ v3.2	Uni.Lund-CEC
HUN	Hegyhátsál	46.950	16.650	248	115	C	2006-2015	GV+ v3.2	HMS
JFJ	Jungfraujoch	46.550	7.987	3570	10	C	2006-2015	GV+ v3.2	KUP
JFJ	Jungfraujoch	46.550	7.987	3570	10	C	2010-2015	GV+ v3.2	EMPA
KAS	Kasprowy	49.232	19.982	1989	5	C	2006-2015	GV+ v3.2	AGH
LMP	Lampedusa	35.510	12.610	45	5	F	2006-2015	GV+ v3.2	NOAA
LMP	Lampedusa	35.520	12.620	45	8	C	2006-2012	preICOS	ENEA
LMU	La Muela	41.594	-1.100	571	79	C	2006-2009	preICOS	ICTA-ICP
LUT	Lutjewad	53.404	6.353	1	60	C	2006-2015	GV+ v3.2	Uni.Groningen
MHD	Mace Head	53.326	-9.904	5	15	C	2006-2015	GV+ v3.2	LSCE
NOR	Norunda	60.086	17.479	46	101	C	2015	GV+ v3.2	Uni.Lund-CEC
OPE	Observatoire Pérenne de l'Environnement	48.562	5.504	390	120	C	2011-2015	preICOS	LSCE
OXK	Ochsenkopf	50.030	11.808	1022	163	F	2006-2015	GV+ v3.2	NOAA
OXK	Ochsenkopf	50.030	11.808	1022	163	C	2006-2007	preICOS	MPI-BGC
PAL	Pallas	67.973	24.116	565	5	C	2006-2015	GV+ v3.2	FMI
PRS	Plateau Rosa	45.930	7.700	3480	10	C	2006-2015	GV+ v3.2	RSE
PUI	Pujio	62.910	27.655	232	84	C	2011-2014	preICOS	FMI
PUY	Puy de Dôme	45.772	2.966	1465	15	C	2006-2015	GV+ v3.2	LSCE
RGL	Ridge Hill	51.998	-2.540	204	90	C	2012-2015	GV+ v3.2	Uni.Bristol
SMR	Smear/Hyytiala	61.847	24.295	181	125	C	2012-2015	GV+ v3.2	UHELS
SSC	Sierra de Segura	38.303	-2.590	1349	20	C	2014-2015	GV+ v3.2	ICTA-ICP
SSL	Schauinsland	47.920	7.920	1205	12	C	2006-2015	GV+ v3.2	UBA
STM	Station M	66.000	2.000	0	7	F	2006-2015	GV+ v3.2	NOAA
TAC	Tacolneston	52.518	1.139	56	185	C	2013-2015	GV+ v3.2	Uni.Bristol
TRN	Trainou	47.965	2.112	131	180	C	2006-2015	preICOS	LSCE
TTA	Angus tall tower	56.555	-2.986	400	222	C	2013-2015	GV+ v3.2	Uni.Bristol
VAC	Valderejo	42.879	-3.214	1102	20	C	2013-2015	GV+ v3.2	ICTA-ICP
WAO	Weybourne	52.950	1.122	20	10	C	2007-2015	GV+ v3.2	UEA
WES	Westerland	54.930	8.320	12	9	C	2006-2015	WDCGG	UBA

Table 1. Observation sites used in the inversions. Datasets with in-situ continuous (C) as well as flask (F) measurements were taken from GLOBALVIEWplus ObsPack, WDCGG, and the EU-funded projects CarboEurope-IP, GHG-Europe and ICOS preparatory phase (all indicated as preICOS).

3.2 Prior and prescribed CO_2 fluxes

All groups split the total surface–atmosphere CO_2 flux in three or four categories: biosphere (NEE, optimised), ocean (sea-atmosphere CO_2 exchanges, prescribed or optimised), anthropogenic (prescribed) and biomass burning (prescribed, used by LUMIA and FLEXINVERT+).

Note that we use the term NEE (sum of photosynthesis and ecosystem respiration) for the posterior fluxes over land throughout the paper, because this is what the prior flux estimates from the terrestrial ecosystem models represent. However, strictly speaking, the inversions optimize the flux that is not explained by the prescribed anthropogenic (and sometimes ocean) fluxes. This includes the effect of ecosystem disturbances (land use, land management, biotic effects) but also projection of errors in the prescribed fluxes.

3.2.1 Terrestrial-ecosystem fluxes

Atmospheric inversions usually rely on NEE simulations from terrestrial ecosystem models to provide the prior value of the NEE component of the control vector (as defined above in Section 2). Within EUROCOM four different simulations of gross (GPP and ecosystem respiration) and net (NEE) terrestrial biosphere fluxes were included: three from process-based models (ORCHIDEE, LPJ-GUESS and SiBCASA), and one from a diagnostic model (VPRM). Two of the four models (ORCHIDEE and LPJ-GUESS) are providing input for the Global Carbon Project annual global CO_2 assessment (Le Quéré et al., 2018).

- **ORCHIDEE (used by PYVAR-CHIMERE, FLEXINVERT+ and NAME-HB) :** ORCHIDEE (Krinner et al., 2005)) computes carbon, water and energy fluxes between the land surface and the atmosphere and within the soil-plant continuum. The model computes the Gross Primary Productivity with the assimilation of carbon based on Farquhar et al. (1980) for C3 plants. Land cover changes (including deforestation, regrowth and cropland dynamic) were prescribed using annual land cover maps derived from the harmonised land use data set (Hurtt et al., 2011) combined with the ESA-CCI land cover products. The ORCHIDEE simulation used here has been produced at a global, 0.5° resolution with 3-hourly output.

- **LPJ-GUESS (used by LUMIA):** LPJ-GUESS (Smith et al., 2014) combines process-based descriptions of terrestrial ecosystem structure (vegetation composition, biomass and height) and function (energy absorption, carbon and nitrogen cycling). Vegetation is dynamically simulated as a series of replicate patches, in which individuals of each simulated plant functional type (or species) compete for the available resources of light and water, as prescribed by the climate data. LPJ-GUESS includes an interactive nitrogen cycle. The simulation used here is forced using the WFDEI meteorological data set (Weedon et al., 2014) and produces 3-hourly output of gross and net carbon fluxes at a 0.5° horizontal resolution.

- **SiBCASA (used by CTE):** SiBCASA (Schaefer et al., 2008) combines the parameterisation of the Simple Biosphere model (SiB) with the biogeochemistry of the Carnegie-Ames Stanford Approach (CASA) calculating the exchange of water, carbon and energy between 25 soil layers, plants, and the atmosphere. The rate of photosynthesis is found using the Ball-Berry-Woodrow model of stomatal conductance (Ball et al., 1987), and C3 and C4 vegetation types are treated separately in the kinetic enzyme model of Farquhar et al. (1980). The simulation used here is forced using meteorological inputs from ERA-Interim, and run ~~it on~~ with a 10 minute time step and a spatial resolution of 1x1 degrees. The actual temporal resolution used in the inversion is 3 hours.

– **VPRM (used by CarboScope-Regional)**: VPRM (Mahadevan et al., 2008) calculates photosynthetic uptake based on a light-use efficiency approach and temperature dependent ecosystem respiration. It uses ECMWF operational meteorological data for radiation and temperature, the SYNMAP land cover classification (Jung et al., 2006), as well as MODIS derived EVI (enhanced vegetation index) and LSWI (land surface water index). Model parameters were optimised for Europe using eddy covariance measurements made during 2007 from 47 sites (Kountouris et al., 2015). The VPRM simulation used here has been produced at a 0.25 degree spatial and hourly ~~temporally~~ temporal resolution.

The mean seasonal cycle and the inter-annual variability of these NEE simulations are shown in Figure 2. Among the notable features is the annual mean NEE of VPRM, which is much lower (≈-1.1 PgC/year) than that of the three other models (ranging from -0.1 to -0.4 PgC/year). VPRM is known to produce a too large uptake (Oney et al., 2017), which can be explained by the optimisation of this diagnostic model against flux measurements from one year. The year to year variations of the annual budget are significant (≈0.1 PgC/year) but not always in phase between the four models. For the mean seasonal cycle, the peak to peak amplitude differs significantly between the models with the smallest amplitude obtained with LPJ-GUESS (around 0.4 PgC/~~months~~month) and the largest with ORCHIDEE (around 0.8 PgC/month). Another visible feature is the phasing of the seasonal cycle in LPJ-GUESS with an earlier CO_2 peak uptake than the other three models (May versus June) and a peak release in August. This phase difference has already been described by Peng et al. (2015).

3.2.2 Anthropogenic emissions

The anthropogenic emissions from combustion of fossil fuels and biofuels, and from cement production are based on a pre-release of the EDGARv4.3 inventory for the base year 2010 (Janssens-Maenhout et al., 2019) and were provided as a 0.5°, hourly resolution product. This specific dataset includes additional information on the fuel mix per emission sector (Janssens-Maenhout, pers. comm.) and thus allows for a temporal scaling of the gridded annual emissions for individual years (2006–2015) according to year-to-year changes of fuel consumption data at national level (bp2, 2016), following the approach of Steinbach et al. (2011). A further temporal disaggregation into hourly emissions is based on specific temporal factors (seasonal, weekly, and daily cycles) for different emission sectors (Denier van der Gon et al., 2011). The seasonality and inter-annual variability of this anthropogenic emissions prior are also reported in Figure 2 (in black).

Agricultural waste burning is already included in the version of the EDGAR v4.3 anthropogenic emission inventory that we are using. Also, large scale biomass burning emissions are negligible in Europe (of the order of 0.01 PgC/year), therefore we decided that no extra biomass burning emission data set should be used in the inversions. Nevertheless, two models (LUMIA and FLEXINVERT) included a prescribed biomass burning source, based on the Global Fire Emission Database v4 (Giglio et al., 2013).

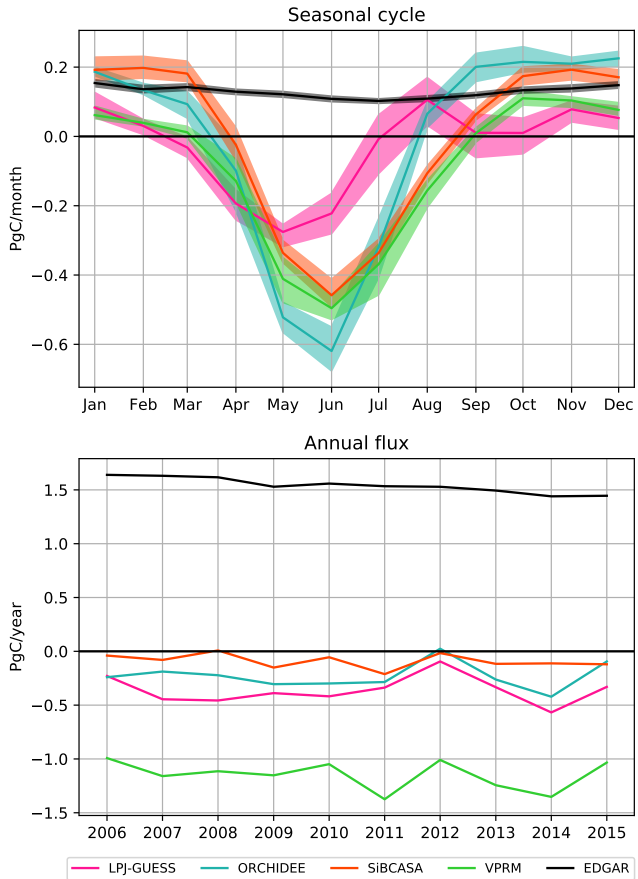

Figure 2. Seasonal cycle (top) and inter-annual variability (bottom) of the prior NEE (colored lines/shades) and prescribed anthropogenic flux (black) used in the inversions, for geographical Europe (see definition in Section 3). The solid lines in the upper plot represents the mean seasonal cycle over the 10 years of the study, while the shaded envelopes show the min/max values over the same period.

3.2.3 Ocean fluxes

The role of the ocean flux in causing spatial CO_2 gradients between stations at the European scale is very minor in regard to the magnitude of other fluxes (below -0.1 Pgc/year). Therefore modelling groups were free to choose which ocean fluxes to use.

Two groups (LUMIA and FLEXINVERT+) used ocean fluxes from the CarboScope surface-ocean pCO_2 interpolation (oc_v1.6 and oc_v1.4 respectively) (Rödenbeck et al., 2013). The CarboScope interpolation provides temporally and spatially resolved estimates of the global sea-air CO_2 flux. Fluxes are estimated by fitting a simple data-driven diagnostic model of ocean mixed-layer biogeochemistry to surface-ocean CO_2 partial pressure data from the SOCAT database. NAME-HB used a climatological prior from Takahashi et al. (2009), which is based on a climatology of surface ocean pCO_2 constructed using measurements taken between 1970 and 2008. The CarboScope-Regional inversion used an ocean flux estimate taken from the Mikaloff Fletcher et al. (2007) global oceanic air-sea CO_2 inversion and CarbonTracker Europe optimized prior fluxes from the ocean inversion of Jacobson et al. (2007). Finally, PYVAR-CHIMERE used a null ocean prior, but allowed the inversion to adjust it.

3.3 Inversion systems

The six inversion systems encompass a wide range of mesoscale regional transport models (with both Lagrangian and Eulerian models) and of approaches for the inversion (variational, ensemble and MCMC methods). The systems also differ by the definition of the boundary conditions, the selection of the observations to be assimilated, the definition of the control vector and the parameterisation of uncertainty covariance matrices. Table 2 presents an overview of the participating systems characteristics.

3.3.1 Transport models and boundary conditions

Four out of the six inversions rely on transport from Lagrangian transport models (LUMIA, FLEXINVERT+, Carboscope-Regional and NAME-HB), while the two others

The six systems cover a diversity of models and model setups. Two systems (PYVAR-CHIMERE and CTE) rely on Eulerian models. This distinction between Eulerian and Lagrangian models is important as it has practical consequences on how the boundary conditions (initial CO_2 concentrations and impact of CO_2 fluxes outside the regional domain) can be imposed, but also on how the sensitivity to surface fluxes is defined.

In Eulerian models, the transport models. The atmosphere is represented by a 3D grid (latitude, longitude and height). The CO_2 concentration is defined at each grid point and is altered at each time step by the CO_2 sources and sinks (i.e. the inversion control vector) in the surface layer, and by the air mass exchanges between the grid cells (at all layers). Boundary conditions are provided in the form of an initial CO_2 field and , when needed (in regional models), as a set of prescribed The other systems (LUMIA, FLEXINVERT, CarboScope-Regional and NAME-HB) all rely on Lagrangian transport models. In these systems, a Lagrangian transport models is used to compute, for each observation, a response function (footprint), i.e. a Jacobian matrix containing the sensitivity of the observed concentration to surface fluxes. The change in CO_2 concentrations at the edges of the domain. Two inversion systems rely on Eulerian models: resulting from the surface fluxes are simply the dot product of each footprint by the corresponding (slice of) the flux vector.

Beyond the Lagrangian/Eulerian distinction, the models differ by the underlying meteorological data used, and by the domain extent:

- The CHIMERE model (used in the PYVAR-CHIMERE relies on the CHIMERE model . CHIMERE system) is a regional Eulerian Chemistry transport model (Menut et al., 2013), forced with ECMWF operational forecasts. The simulations are performed at a horizontal resolution of 0.5° and with 29 vertical levels up to 300 hPa, for the exact EUROCOM domain (as described at the beginning of Section 3). Background concentrations are obtained from the transport (by CHIMERE) of a fixed boundary condition interpolated from the CAMS global inversions of Chevallier et al. (2010).

– The CTE inversions rely inversion relies on the global Eulerian transport model TM5 (Huijnen et al., 2010), driven by air mass transport from the ECMWF ERA-Interim reanalysis. TM5 is here ran run at a global resolution of $3^\circ \times 2^\circ$, with a nested $1^\circ \times 1^\circ$ zoom over Europe (21°W-39°E, 12-66°N), and 25 vertical sigma-pressure levels.

In the four other systems, Lagrangian transport models are used to compute, for each observation, a response function (footprint), i.e. a Jacobian matrix containing the sensitivity of the observed concentration to surface fluxes. The change in CO_2 concentrations resulting from the surface fluxes are simply the dot product of each footprint by the corresponding (slice of) the flux vector.

Lagrangian models typically simulate the dispersion backwards in time from each observation point of a large number of air trajectories (the approaches to do so differ between the models). The aggregated residence time of the air in individual surface grid boxes is taken as a proxy for the sensitivity of the observation point to surface processes in each of these grid boxes. The footprints are necessarily limited in time (each covers a period of at most a few weeks before each observation), and in most instances also in space (unless a global Lagrangian model is used). A "background" term representing the contribution of fluxes outside the space/time domain of the footprint needs to be added to represent the total modelled CO_2 concentration.

The four inversions relying on such pre-computed footprints differ by the actual Lagrangian models used, but also by the approach used to compute the footprints (the definition of the surface layer) and by the type of background information used:

– The CarboScope-Regional system (Kountouris et al., 2018a) relies on footprints from the STILT model (Lin et al., 2003). STILT footprints are computed for the exact EUROCOM domain, at a horizontal resolution of 0.25°, and at a hourly temporal resolution, and they cover a period of 10 days prior to each observation. STILT is driven by short-term forecasts of the ECMWF-IFS model at 0.25° resolution. The surface layer (up to which surface fluxes are mixed instantaneously) is defined as half the height of the planetary boundary layer, at any given time. The background concentrations are computed directly at each observation site by a global, coarse resolution CarboScope CO_2 inversion (Rödenbeck et al., 2003), following the 2-step approach described in Rödenbeck et al. (2009).

– In LUMIA, footprints covering the EUROCOM domain at a 0.5°, 3-hourly resolution were generated with the FLEXPART 10.0 model (Pisso et al., 2019), driven by ECMWF ERA-Interim meteorology. The footprints cover a period of seven days prior to each observation and the surface layer is defined as the atmosphere below 100 m a.g.l.. The background concentrations are also computed following the Rödenbeck et al. (2009) approach, but this time a global TM5-4DVAR inversion is used for computing the background concentrations (Monteil and Scholze, 2019).

– The FLEXINVERT + inversion (Thompson and Stohl, 2014) also relies on footprints from the FLEXPART model, but driven by ECMWF operational forecasts. In contrast to CarboScope-Regional and LUMIA, the footprints are computed globally, on a 0.5° hourly grid, and cover a period of five days before each observation. Since the footprints are global, the background (from the perspective of the transport model) results only from the transport to the observation sites of

~~the initial CO_2 distribution (i.e. the CO_2 distribution at the start of the period covered by each footprint). This initial concentration is calculated as a weighted average of a global CO_2 distribution sampled where and when the FLEXPART trajectories are terminated, and this global CO_2 distribution is based on a bivariate interpolation of observed CO_2 mixing ratios from NOAA sites globally, with monthly resolved fields. Note that for this system, the domain of the transport model is larger than that of the inversion itself.~~

– The NAME-HB system (White et al., 2019) uses footprints from the NAME Lagrangian particle dispersion model. NAME is driven by 3-hourly meteorology from the UK Met Office's Unified Model (Cullen, 1993), at a spatial resolution which changes in time and is 0.233° latitude by 0.352° longitude before mid 2014. The footprints are defined on a large regional domain, ranging from 97.9°W; 10.729°N to 39.38°E; 79.057°N, with a spatial resolution of 0.233° × 0.352° (it covers the eastern half of North America, Europe and the Northern half of Africa). The footprints are computed for a period of 30 days before each observation, at a 2-hourly temporal resolution in the first 24 hours, and the remaining 29 days are integrated. The surface layer is defined as the layer below a height of 40 m. ~~The background~~

Except for CTE which relies on a global model (although it runs at a very coarse resolution outside Europe), all the other systems need boundary conditions (also called background concentrations), which represent the contribution of fluxes outside the space/time domain of the simulation:

– In PYVAR-CHIMERE and in NAME-HB, the background concentrations (BC) correspond to the transport of a boundary condition defined at the edge of the domain, by the transport model used in the inversion. In PYVAR-CHIMERE, the boundary condition is provided by a CAMS global inversion (Chevallier et al., 2010) and in NAME-HB it is derived from a global CO_2 simulation with the MOZART transport model (Palmer et al., 2018) ~~. The MOZART CO_2 field is~~ (sampled at the time when and location where the NAME trajectories leave the NAME domain).

– CarboScope Regional and LUMIA both implement the 2-step approach described in Rödenbeck et al. (2009). In short, the background concentrations correspond to the transport to the observation points of a boundary condition, taken from a global, coarse resolution inversion, by the global transport model used in that global inversion. CarboScope-Regional relies on TM3 for its global inversion, and LUMIA relies on the TM5-4DVAR model.

– In FLEXINVERT, the footprints are global, therefore the background (from the perspective of the transport model) results only from the transport to the observation sites of the initial CO_2 distribution (i.e. the CO_2 distribution at the start of the period covered by each footprint). This initial concentration is calculated as a weighted average of a global CO_2 distribution sampled where and when the FLEXPART trajectories are terminated, and this global CO_2 distribution is based on a bivariate interpolation of observed CO_2 mixing ratios from NOAA sites globally, with monthly resolved fields. Note that for this system, the domain of the transport model is larger than that of the inversion itself.

The boundary conditions are specific to each system: from a technical point of view, the differences in domain extent and in the types of couplings with the boundary/background would make it hard to impose a common BC, but also, the boundary condition is an uncertain term: allowing a diversity of implementations is a way to maximize the exploration of the uncertainties in our intercomparison.

3.3.2 Inversion approaches

Four out of the six systems (PYVAR-CHIMERE, LUMIA, CarboScope-Regional and FLEXINVERT+) implement a variational inversion approach, in which the minimum of the cost function $J(\mathbf{x})$ (Eq. 1) is searched for iteratively. The CTE inversion (Peters et al., 2007; van der Laan-Luijkx et al., 2017) employs an ensemble Kalman smoother with 150 members and a 5-week fixed-lag assimilation window. The NAME-HB inversion uses the MCMC method (Rigby et al., 2011; Ganesan et al., 2014; Lunt et al., 2016; White et al., 2019). In short, this method samples the parameter space and proposals for parameter values are accepted or rejected according to some rules based on the likelihood of the proposal.

Regardless of the inversion technique used, all the groups were asked to provide optimised NEE fluxes at a monthly, 0.5° resolution on the EUROCOM domain. However, the precise control vector optimised in some of the inversions differ from this requested product:

- In PYVAR-CHIMERE, the NEE is optimised at a 6-hourly resolution on each grid cell (on the standard EUROCOM grid), starting from a prior NEE estimate from the ORCHIDEE model (See Section 3.2.1). In addition, the inversion also adjusts the ocean flux estimate, starting from a null prior. The prior uncertainty for each control vector element is proportional to the respiration in the corresponding grid cell (according to the same ORCHIDEE simulation) and further ~~scale~~ scaled to obtain an average uncertainty at the 0.5° and 1 day scale of 2.27 μmol.CO_2/m^2/s (after ~~Kountouris et al. (2018a)~~ Kountouris et al. (2018b)).

- The LUMIA inversion controls the NEE fluxes monthly, on the standard EUROCOM grid, starting from prior NEE from the LPJ-GUESS model. The prior uncertainty is set to 50% of the prior control vector (i.e. the prior NEE), with a minimum uncertainty ~~of~~ set to 1% of the grid point with the largest uncertainty, to avoid zero-uncertainty when NEE is close to zero. The decadal inversion was decomposed in ten 14-month inversions, from which the first and last month were not used.

- In the CarboScope-Regional system, the NEE fluxes are optimised 3-hourly at a 0.5° resolution in the EUROCOM domain, based on a prior NEE estimate from the VPRM model. ~~In addition, the control vector contains a bias term, which scales uniformly the map of annual total respiration.~~ The uncertainty on the prior NEE is set to a uniform value of 2.27 μmol.CO_2/m^2/s~~and the uncertainty on the bias term is adjusted so that the total uncertainty integrated over the domain is 0.3 PgC/year~~, using a spatial error structure with a hyperbolic correlation shape. The setup is identical to the "~~BVR~~nBVH/" case in Kountouris et al. (2018a). The decadal inversion period was divided in three periods (2006-2007, 2008-2011, and 2012-2015).

- FLEXINVERT+ controls the NEE per country × Plant Functional Type (116 control variables per time step across Europe, with PFTs based on those in the CLM model). The fluxes are optimised for 6-hourly periods (0-6,6-12,12-

18,18-0 local time), averaged over five days. The prior NEE flux is based on the ORCHIDEE simulation described in Section 3.2.1, and the uncertainties are set proportional to this prior NEE. The transport model in FLEXINVERT+ is global, therefore the flux estimates used in the inversions are defined over the entire globe. However, the inversion only adjusts NEE within the EUROCOM domain.

- In NAME-HB, the domain of NAME has been split into eight boxes: four "background" boxes outside the EUROCOM domain, and four "foreground" boxes within the EUROCOM domain. The latter were further divided based on a PFT map used in the JULES vegetation model (Still et al., 2009), which includes six PFTs. The inversion optimizes separately the gross primary production (GPP, i.e. the uptake of carbon by plants) and the heterotrophic respiration (TER, with NEE=GPP+TER). The flux components are optimized at a variable temporal resolution, with a maximum resolution of one day (see White et al. (2019) for further details). The oceanic flux is prescribed (based on the Takahashi et al. (2009) climatological pCO_2 estimate), but the background concentrations are part of the control vector and are therefore adjusted during the inversion. Therefore, there are 56 elements in the control vector, 4 elements to optimise the background concentrations, 4×2 elements to optimise the "background" regions for each of GPP and TER, 4×5 elements for the PFT-regions for GPP (as one of the 6 PFTs is not applicable to GPP) and finally 4×6 elements for the PFT-regions for TER. The uncertainties are set to 100% of the prior for GPP and TER, and to 3% of the initial value for the background terms.

- in CTE, the NEE and ocean fluxes are optimised globally on a weekly time resolution in a 5 week lagged window. The global domain is split in 11 TRANSCOM regions, which are further decomposed in ecoregions corresponding to 19 ecosystem types. The fluxes are optimised on 1x1 degrees resolution for the Northern Hemisphere land regions, and by ecoregion and ocean region for the rest of the world. The prior NEE is taken from the SiBCASA simulation described in Section 3.2.1 and the prior oceanic flux is based on Jacobson et al. (2007).

In the three systems that optimise NEE at the pixel scale (LUMIA, PYVAR-CHIMERE and CarboScope-Regional), the spatial resolution of the control vector is in practice further limited by the use of distance based spatial and temporal covariances in the flux covariance matrices ($\mathbf{B}$ in Equation 1), which in effect smoothes the results by preventing the inversion from adjusting neighbouring pixels totally independently. The values of 100 km (CarboScope-Regional) and 200 km (PYVAR-CHIMERE and LUMIA) used for the spatial covariance lengths correspond well to the diagnostics of comparisons between the ecosystem simulations and flux eddy covariance measurements (Kountouris et al., 2018a). These systems and FLEXINVERT+ also assume temporal error covariances of one month at each grid cell.

The NAME-HB and FLEXINVERT+ inversions only control a limited number of PFTs in each region, which means that pixels in the same region and corresponding to the same PFT have a correlation coefficient of 1. Finally, CTE follows an intermediate approach. The flux uncertainties of Northern Hemisphere land pixels within a same ecoregion are correlated ~~following~~ with a variable spatial covariance length to reflect the observation network density (200 km in Europe), and the uncertainties of grid boxes corresponding to different ecoregions are assumed uncorrelated. For the rest of the world, the uncertainties are coupled within each TRANSCOM region decreasing exponentially with distance. The chosen prior standard deviation is 80% on land parameters, and 40% on ocean parameters (van der Laan-Luijkx et al., 2017).

3.3.3 Observation vectors and errors

All the inversions use observations from the stations listed in Table 1. Each participant was, however, free to refine their selection of observations (both in terms of number of sites assimilated and of data selection at each site) to adapt it to the skills of their own inversion system. In practice, five of the six inversions used data from nearly all the observation sites. NAME-HB used only a restricted list of 15 sites (see Figure SI1)

Most of the systems assimilate instantaneous or 1-hour averages of the measurements, taken, when there are several vertical levels of measurements, at the top level of the stations, as it is the least sensitive to very local surface fluxes. NAME-HB assimilates 2-hourly observations (average of the observed concentrations in each 2-hourly interval). Due to the traditional limitations of transport models in terms of representation of the orography and simulation of the vertical mixing (Broquet et al., 2011), most of the systems use observations at low altitude sites during the afternoon only, and observations at high altitude sites during night time only (vertical gradients of CO_2 near to the surface are notoriously difficult to simulate accurately, so observations when the vertical gradients are expected to be the lower are preferred)

- PYVAR-CHIMERE assimilates 1-hour averages of the continuous or flask measurements over specific time windows that depend on the altitude of the stations above the sea level. The selection window is 12:00-18:00 UTC time for stations below 1000 m a.s.l. and 0:00-6:00 local time for stations above 1000 m a.s.l. (following the analysis and choices by Broquet et al. (2011)). The observation errors are set-up as a function of stations, of the height of the station level above the ground and of and season, following the estimates by Broquet et al. (2011, 2013), based on comparison of simulations and measurements of Radon~~)~~. Their standard deviation for the 1-hour averages ranges from 3 to 17 ppm.
- In LUMIA, observations from sites with continuous observations are selected based on the "`dataset_time_window_utc`" flag in the metadata of the observation files. That corresponds, for most sites, to a 11:00 to 15:00 UTC time range, and to a 23:00 to 03:00 UTC time range for mountain sites. At sites with only flask observations, all samples were used. The observation uncertainties are ~~defined~~ set as the quadratic sum of the measurement uncertainties, of the uncertainty ~~of~~ associated to the foreground transport model (i.e. FLEXPART) and ~~of the uncertainty on~~ to the background concentrations. The measurement uncertainties are taken from the data files when available, and a minimum uncertainty of 0.3 ppm is enforced. Foreground transport model uncertainties are computed by performing two similar forward model runs, with TM5 and LUMIA (i.e. FLEXPART + background concentrations from TM5), configured such that the only difference is the model used to compute the transport within the EUROCOM domain (since the two models run at very different resolutions, this provides a reasonable proxy for the representation error). The uncertainties on the background concentrations are set as the standard deviation of the vertical profile of background CO_2 concentrations around each observation (see Monteil and Scholze (2019) for details about the approach). ~~The~~ Finally, a minimum value of 1 ppm was enforced for the combined uncertainty. On average, the combined uncertainty is on the or-

der of 2 ppm (with site-averages ranging from 1.02 (MHD) to 4 ppm , on average(PUI)), but for individual observations it can be as high as 30 ppm.

- CarboScope-Regional assimilates observations, between 11:00 and 16:00 UTC for tall-towers, ground-based or coastal stations, and from 23:00 to 04:00 UTC for mountain stations (the time intervals refer to the beginning of the observation hour). A base representation error of 1.5 ppm was assumed for tall towers, coastal and mountain. For ground based continental sites it was raised to 2.0 ppm, and to 4 ppm for Heidelberg, which is in a urban environment. For sites that provide hourly observations, an error inflation was applied (e.g. for tall towers: 1.5 ppm $\times\sqrt{6 \text{ obs/day} \times 7 \text{ day/week}} =$ 9.7 ppm).

- In FLEXINVERT+, observations were assimilated hourly between 12:00 and 16:00 local time for sites below 1000 m.a.s.l. and between 00:00 and 04:00 for sites higher than 1000 m.a.s.l. The observation uncertainties are calculated as the quadratic sum of the measurement errors (with a minimum of 0.5 ppm), the uncertainty on the initial mixing ratio, assumed to be 1 ppm and the contribution of uncertainties in the fossil fuel emission estimates and in the NEE fluxes from outside the domain, both transported by FLEXPART to the observation sites. The total observation-space uncertainties typically range between 1 and 3 ppm.

- In NAME-HB, observations are filtered based on a combination of two metrics. One is the ratio of the NAME footprint magnitude in the 25 grid boxes closest to the measurement site. If this ratio is high it indicates that a large proportion of the air arriving at a measurement site is from very local sources and may not be resolved by the model. The second metric is the lapse rate modelled by NAME, which is the change of temperature with height and is a measure of atmospheric stability. A high lapse rate suggests very stable atmospheric conditions and may also indicate that there is a lot of local influence on the measurement. With these criteria, some data outside the usual daytime time constraints can be included and daytime data that is not collected during favourable conditions can be removed. In practice however, most of the data included is during the daytime. The measurement uncertainties are taken from the data providers and averaged over the month for each measurement site to give a fixed monthly value. The observation uncertainty is adjusted during the inversion but initially it is the sum of the average measurement uncertainty and a model uncertainty of 3 ppm.

- In CarbonTracker-Europe, flags from data providers are used to screen for representative observations (usually equivalent to the afternoon hours for typical sites and night time hours for mountain sites). A model-data mismatch based on the station category (tower, flask, etc.) is assigned to each site, accounting for both measurement errors and modelling errors at that site. If the difference between the forecast and observation is greater than three times that assigned model-data mismatch, the observations is not used in the inversion.

The range of uncertainties varies a lot across the systems, and can range from one up to tens of ppm. It reflects the different types of coupling between global and regional transport models, and the different range of diagnostics available for each group to quantify their uncertainties. The precise impact of these differences in prescribed observation uncertainties will be analysed in a follow up study. sectionResults

4 Results

4.1 Fit to the observations

Before presenting the posterior NEEfrom the six inversion systems , we first briefly analyse the reduction of the misfits to the CO_2 observation observations assimilated by each inversion. The aims are to first check that all inversions actually reduce the observation misfits improve the model fits to the observations (which is a basic diagnostics diagnostic of atmospheric inversions) but also to determine whether some sites are particularly problematic for some or all of the inversions.

For each inversion, a comparison of Comparisons between the prior and posterior bias and root mean square (RMS) differences (denoted RMS errors i.e. RMSE) between the time series of measured and simulated data computed at each assimilated site, and between the full sets of measured and simulated data, is shown are shown for each inversion in Figure 3. For each inversion, the observation sites are , for each site (sorted according to the corresponding prior bias. prior bias, for each inversion) and for the whole ensemble of assimilated observations.The expectation in this analysis is that all the systems should show a reduction of the misfits (both in terms of mean bias and RMSE), which indeed is what happens. Ideally the posterior misfits should also be close to unbiased (i.e. with respect to the prescribed observational uncertainties).

All inversion satisfy that expectation, and lead to a posterior bias lower than 0.5 ppm (the prior biases are also close to zero for LUMIA, CTE and FLEXINVERTthe inversions do indeed satisfy these expectations. The mean posterior biases range from -0.32 ppm (PYVAR-CHIMERE) to +), and all lead to a net reduction of the 0.04 ppm (NAME-HB). The largest bias reductions are obtained for the inversions that had the largest prior biases (CarboScope-Regional, with a mean bias reduced from -0.91 ppm to -0.18 ppm, and NAME-HB, with a mean bias reduced from -0.87 ppm to +0.04 ppm). The spread of the residuals , with RMSE ranging from 2.9 ppm (CTE) to 5.4 ppm(CarboScope-Regional). is also reduced in all the inversions, with strongest improvements obtained by NAME-HB and (from 4.85 ppm to 2.97 ppm), CarboScope-Regional both start from a relatively larger average negative prior bias (respectively -0.94 and -0.71 ppm), whereas the other systems all start from prior biases ranging between -0.26 ppm(PYVAR-CHIMERE) and 0.27 ppm (CTE). In the case of from 6.11 ppm to 4.70 ppm) and FLEXINVERT (from 5.85 ppm to 4.54 ppm). The best overall posterior fit is however obtained by CTE, with a posterior RMSE of 2.90 ppm and a posterior bias of +0.01 ppm.

The larger prior bias in CarboScope-Regional , this is easily explained by the substantially larger prior CO_2 sink in of the VPRM prior (Section 3.2.1), while since . On the contrary, NAME-HB uses the same ORCHIDEE prior as other inversions (PYVAR-CHIMERE, FLEXINVERT+) ,) so its prior bias must have a different origin (background, transport or oceanic flux). Note that the also that NAME-HB inversion only covers a reduced 5-years period (2011-2015), which limits its comparability with the other inversions. The comparatively low RMSE obtained in the CTE inversion (including in the prior step), despite it using a lower resolution transport model shows that the resolution of the transport (and of the underlying meteorological data) is not the main limitation to fitting the observations.

At the site scale, the decrease of the misfits is rather moderate, up to 30% but , mostly below 20%, without a clear distinction between low altitude and high-altitude sites or between the models. Each inversion occasionally leads to local degradation of the fit (increase of the bias or RMSE). Such degradation can occur when the inversions do not have enough independent degrees of freedom to reconcile contradictory constraints from several sites. This can be because the spatial resolution of the control vector is too low compared to the density of sites, in which case it does not necessarily impact negatively the accuracy of the solution. But it can also be an indication that a site is not well represented by the transport model and could as a consequence introduce a local bias in the posterior flux. Some sites tend to be systematically misrepresented systematically be poorly fitted by the inversions (including in the posterior step), in particular those in the vicinity of large urban areas (with large anthropogenic emissions), such as HEI and GIF. Note that this is accounted for in several of the inversions, but not all, by inflating the model representation errors (which allows the model to degrade the fit to the observations, at a low "cost"). Besides these two sites, if it does not appear that the distribution of the fits is systematic. Especially, there is no major difference between the representation of fit to mountain-top (with night-time observations assimilated) and plain sites.

Error statistics computed on 1-month and 1-year hourly to annual averages of the observations (Table SI1) show larger RMS error reductions in most models (up to 42% , and 3) show considerably larger RMSE reduction for monthly and annual averages (respectively 33% respectively on the whole observation ensemble , in FLEXINVERT+). However, the RMS error reduction for annual averages are generally smaller than that and 34% ensemble median RMSE reduction) than for hourly and daily averages (20% and 23%, respectively). The only exception is NAME-HB, for which the error reduction is on the same order at all time scales (from 41% for hourly averages to 49% for monthly averages). This likely reflects the fact that, except for NAME-HB, all inversions use prior error covariance matrices implementing a temporal correlation length of one month between the flux adjustments, regardless of the actual temporal resolution of the inversions (See 2).

The monthly statistics are the most consistent accross the ensemble (the larger error reduction obtained in CarboScope-Regional and NAME-HB are easily explained by their larger prior bias), which suggests that results at the monthly scale are likely more robustthan at annual to multi-annual scalesthis is the temporal scale for which the comparison is the most robust.

This comparison of the residuals is an important technical diagnostics, but does not indicate how realistic the posterior fluxes are, and should not be interpreted as a ranking of the inversions. Especially, a good posterior representation of the observations is only a sign that the inversion had enough independent degrees of freedom to match the observed concentrations, but does not mean that the observations are sufficient to robustly constrain the control vector, or that the underlying transport model is accurate.

4.2 Posterior European-scale NEE

The monthly prior and posterior NEE from the six inversions, integrated over the whole European domain (as defined in Section 3) and over the 10-year period of the intercomparison are displayed in Figure 4. The dominant feature in Figure 4 is the systematic differences between the seasonal cycles, i.e. each inversion shows a similar pattern in the seasonal cycle (large/small amplitude or timing of peak values) for each year of the simulation period.

Overall, the posterior fluxes remain within or close to the range of values defined by the different priors. In the Sections sections 4.2.1 and 4.2.2, we compare the prior and posterior mean fluxes and their variability, at the annual and monthly time scales. In Section 4.2.3 we have a first look at the sub-continental scale. Results at the grid scale are provided for completeness in SI, but are not further discussed in this paper.

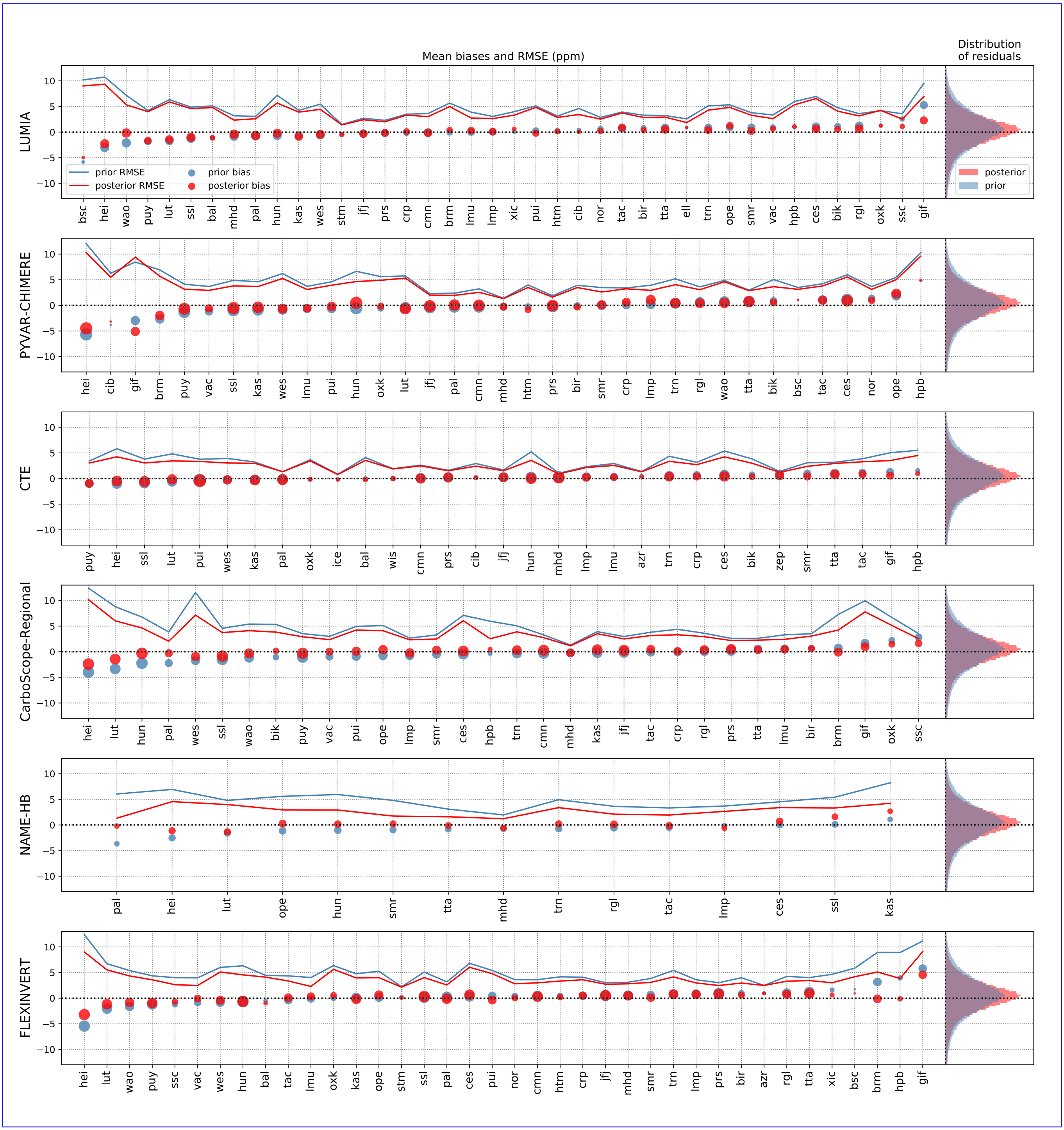

Figure 3. Prior (blue) and posterior (red) mean bias (dots) and RMSE (solid lines) at each observation site, for each inversion. The size of the dots is proportional to the number of assimilated observations. The histograms at the right of each subplot shows the prior and posterior distribution of fit residuals for the entire ensembles of assimilated observations.

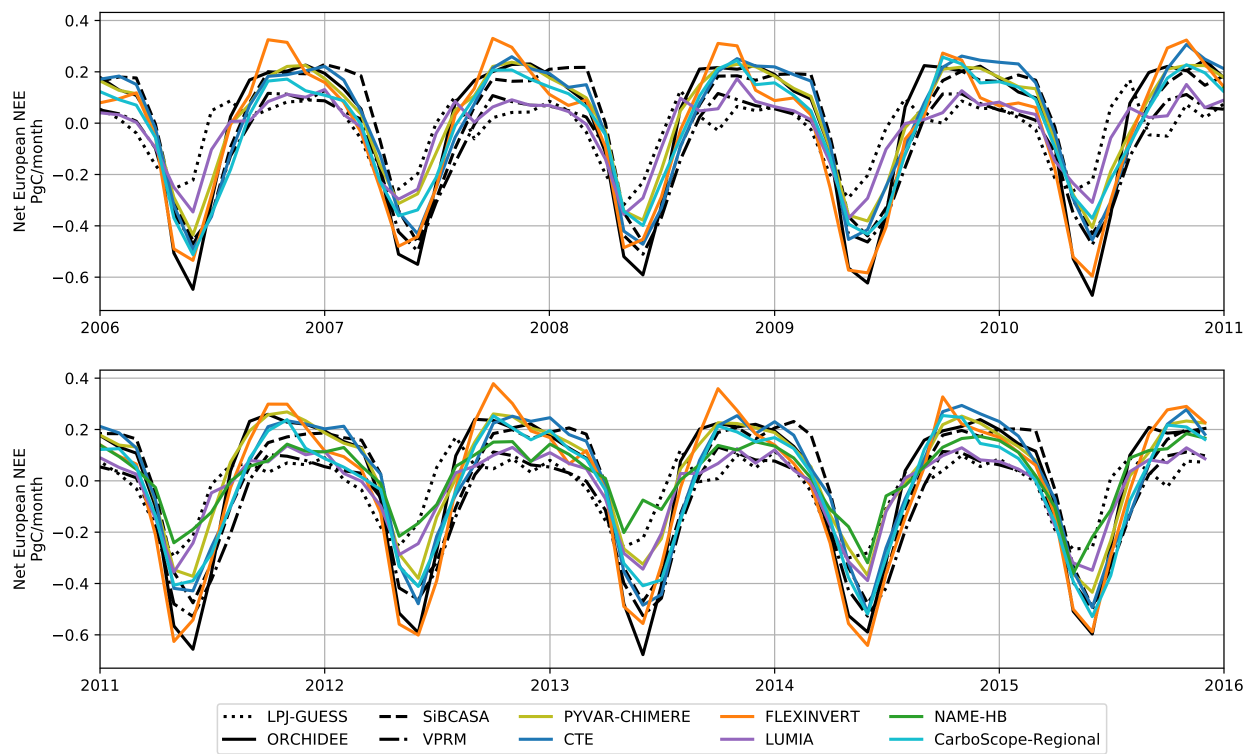

Figure 4. Monthly posterior fluxes, aggregated on the entire domain.

4.2.1 Long term mean and variability of the annual NEE budget

Prior and posterior estimates of the annual budgets of the NEE over the European domain, as well as their mean and standard deviation over the inversion period are reported in Figure 5.

We find an ensemble mean posterior estimate of the 10-year average NEE of ~~-0.19~~ -0.09 PgC/year (-0.16 PgC/year when excluding NAME-HB, which only covers the last five years), with values ranging from a net source of 0.28 PgC/year (PYVAR-CHIMERE) to a net sink of ~~-0.61~~ -0.41 PgC/year (FLEXINVERT). LUMIA and CarboScope-Regional ~~). Besides~~ estimate net sinks (-0.36 PgC/year and -0.32 PgC/year, respectively) and the CTE inversion yields an almost neutral NEE budget (+0.02 PgC/year). Finally, besides PYVAR-CHIMERE, only the NAME-HB system finds the European ecosystems on average to be a net source of CO_2 to the atmosphere over our simulation period. ~~The CTE inversion yields an almost neutral NEE budget (+0.02 PgC/year), and LUMIA and FLEXINVERT+ both derive a sink of ≈-0.4 PgC/year.~~ Overall, the range of estimates from our inversions (~~0.9~~ 0.7 PgC/year) is ~~slightly~~ narrower than that of the priors (1.06 PgC/year between VPRM and SiBCASA).

~~The last column of the Table in Figure 5 shows the standard deviation of each annual NEE estimate, which we use as metrics for their~~ A large fraction of the average spread is due to systematic offsets between the optimized fluxes: the standard variation of the annual flux obtained in each inversion (taken as a metric for the inter-annual variability (IAV)~~. It is lower than 0.17~~

PgC/year for almost all the estimates; FLEXINVERT+ is an exception with a annual NEE) is generally much smaller than the spread of the ensemble. The standard deviation of 0.31 the annual NEE ranges from 0.11 PgC/year . The differences in IAV are therefore, with the possible exception of FLEXINVERT+, only a small contributor to the posterior range of annual budgets ((LUMIA) to 0.33 PgC/year (FLEXINVERT), while the ensemble spread of the annual NEE is on the standard deviation of the ensemble is , on average, 0.34 PgC each year for the posterior NEE). This further highlights that differences between annual budgets of the inversions are primarily driven by the differences in long-term average, and that the latter is not robustly constrained in our set of regional inversions. order of 0.8 PgC/year.

In the case of the CarboScope-Regional inversion, an obvious source for an offset from the other inversions is the prior flux from VPRM, which is much more negative than the other three priors. However, the differences between the three inversions using the NEE field from ORCHIDEE as a prior flux (PYVAR-CHIMERE, FLEXINVERT + and NAME-HB) show that the biases between prior estimates can, at best, only partially explain the offsets in posterior estimates.

The annual anomalies of NEE are compared in Figure 6, and the colors of the cells in Figure 5 also scaled to these anomalies (with the long term mean of each estimate taken as a reference). The ensemble spread of the posterior anomalies is generally much larger than that of the prior, although one system (FLEXINVERT+) is contributing the most to such spread (i.e. in 2009 and 2014 for instance). The spread also strongly varies from year to year, from a minimum spread of 0.16 0.21 PgC/year in 2012 2007 to a maximum of 0.67 PgC/year in 2009. In order to provide a metrics less sensitive to potential model outliers, the medians of the prior (blue) and posterior (red) anomalies are shown in the Figure 6.

The difference between the median prior and posterior anomalies is large at the start of the inversion period, with a median prior flux anomaly of +0.09 PgC/year in 2006 corrected to a median negative anomaly of -0.14 PgC/year by the inversions. Posterior anomalies in 2006 range between -0.2 and -0.1 PgC/year, except for one inversion which reaches 0.3 PgC/year. Therefore, this correction of the NEE by (most of) the inversions in 2006 seems relatively robust. On the contrary, one clear positive anomaly (≈There are a few consistent features, such as a clear positive anomaly in 2012, already present in the priors (median of +0.2 0.19 PgC/year) is present in 2012 in the prior fluxes and is and further confirmed by all the inversions . While it is limited to 2012 in the priors , it already starts in 2011 in some of the inversions (PYVAR-CHIMERE and LUMIA) and extends to 2013 in most of the inversions (particularly the three inversions using ORCHIDEE fluxes as prior: NAME-HB, PYVAR-CHIMERE and FLEXINVERTthe inversions (median value of +). The positive 2012 anomaly is also followed by an almost equivalent negative anomaly in 2014 (-0.2 0.16 PgC/year in the year). In contrast to the priors , -0.13 however, the inversions point to a continuation of this anomaly in 2013 (+0.10 PgC/year in the posteriors) with this time however, a large spread between the posteriors (from -0.55 year), and do not confirm the negative anomaly present in most priors in 2014. The inversions also point to negative anomalies in 2006, 2009 and 2015 (respectively -0.16 PgC/yearfor FLEXINVERT to +0.11 , -0.19 PgC/year for CTE). Another relatively large (0.14 and -0.1 PgC/year) divergence between the prior and posterior anomaliesis found in 2009, but because of the very large spread in the inversion ensemble that year, we do not consider it very robust. For most of the other years during this 10-year period, the median for both the prior and posterior estimates of the NEE generally indicate small anomaliesthat are clearly outside the range of prior anomalies, although the spread of the ensemble is rather large for each of these three years (> 0.5 PgC/year). Overall however, the posterior median anomalies remain within or close to the range of prior anomalies for most of the 10-years period, but individual inversions can diverge a lot from the rest of the ensemble of posterior estimates, like FLEXINVERT ~~+~~ before 2010 and in 2014.

In summary, ~~a few robust features in terms of flux anomalies seem to be captured by the inversion ensemble (the 2006 and 2012-2014 anomalies~~the inversions clearly reduce the interval of estimates regarding the mean annual NEE of the European domain (compared to the ensemble of priors), but ~~the analysis of their drivers is~~do not robustly capture the inter-annual variability of that annual flux. Analyzing the differences in annual IAV is, however, complicated by the ~~aggregation over the~~ large spatial and temporal scales over which the fluxes are averaged: the observation network is not homogeneous, and the inversions may constrain some ~~regions~~ parts of the domain or times of the year better than others. In the following sections (4.2.2 to 4.2.3) we analyse the inversion results at finer temporal and spatial scales.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Mean	Std
PYVAR-CHIMERE	0.12	0.29	0.34	0.02	0.24	0.53	0.51	0.47	0.27	0.05	0.28	0.18
CTE	-0.16	-0.01	-0.12	0.09	0.28	-0.01	0.25	-0.12	0.14	-0.10	0.02	0.16
FLEXINVERT	-0.13	-0.23	-0.19	-1.01	-0.47	-0.45	-0.34	-0.08	-0.96	-0.23	-0.41	0.33
LUMIA	-0.40	-0.36	-0.31	-0.54	-0.26	-0.24	-0.20	-0.36	-0.47	-0.45	-0.36	0.11
NAME-HB						0.09	0.35	0.40	0.07	0.17	0.22	0.15
CarboScope-Regional	-0.69	-0.27	-0.04	-0.42	0.00	-0.36	-0.16	-0.29	-0.34	-0.65	-0.32	0.23
LPJ-GUESS	-0.23	-0.45	-0.46	-0.39	-0.42	-0.34	-0.09	-0.33	-0.57	-0.33	-0.36	0.13
ORCHIDEE	-0.24	-0.19	-0.22	-0.31	-0.30	-0.29	0.02	-0.26	-0.42	-0.09	-0.23	0.12
SiBCASA	-0.04	-0.08	0.01	-0.15	-0.06	-0.21	-0.02	-0.12	-0.11	-0.12	-0.09	0.07
VPRM	-0.99	-1.16	-1.11	-1.15	-1.05	-1.37	-1.01	-1.24	-1.35	-1.03	-1.15	0.14
Median Posteriors	-0.16	-0.23	-0.12	-0.42	0.00	-0.13	0.05	-0.10	-0.13	-0.16	-0.15	0.17
Median Priors	-0.23	-0.32	-0.34	-0.35	-0.36	-0.31	-0.06	-0.30	-0.49	-0.23	-0.29	0.13

Figure 5. Annual NEE budget (positive for a source to the atmosphere and negative for a sink) for the six inversions (upper section of the array), for the four priors (middle section) and median prior and posterior fluxes (lower section). The two last columns show respectively the mean NEE and the standard deviation of the NEE estimate, for each simulation. The colors of the cells indicate the strength and direction of the annual NEE anomaly (respective to the mean of each row, see Figure 6 for the actual values).

~~Prior (blue) and posterior (red) mean bias (dots) and RMSE (solid lines) at each observation site, for each inversion. The size of the dots is proportional to the number of assimilated observations. The last two points on the right, and their associated error bars represent respectively the mean bias and RMSD taken over the ensemble of assimilated observations.~~

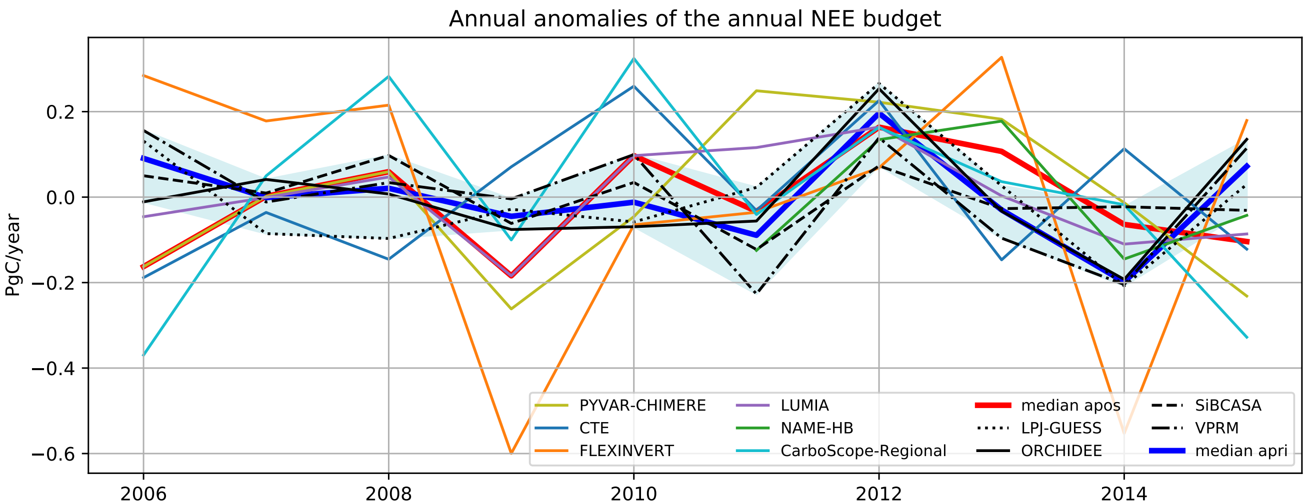

Figure 6. NEE anomalies of the six inversion posteriors and of the four priors. The median of the prior anomalies is shown as a thick blue solid line, that of the posteriors is shown as a thick red solid line. The blue shaded area shows the envelope of prior anomalies.

4.2.2 Seasonal variability of NEE

The mean monthly posterior NEE estimates for the six inversions together with the prior fluxes are shown in Figure 7. At first glance, the spread of the posterior fluxes matches approximately that of the prior estimations, with very similar mean spring (May-June) uptakes, ranging from -0.24 (NAME-HB) to -0.55 PgC/month (FLEXINVERT+) in the posteriors and from -0.28 to -0.62 PgC/month in the priors. Winter posterior emissions are slightly higher (from +0.13 (LUMIA) to +0.32 PgC/month in FLEXINVERT+) than the priors (+0.11 to +0.23 PgC/month). As a result, the median seasonal cycles are also very similar, with a similar phasing and a seasonal cycle amplitude of $\approx$ 0.55 PgC.

This similarity between the prior and posterior ensembles hides more significant differences at the level of individual ensemble members. The phasing of the seasonal cycle is very consistent among the inversions, with terrestrial ecosystem becoming a CO_2 sink (flux sign switch around April and August and with a peak uptake in June). On the contrary, the bottom-up simulations used as priors have four relatively distinct seasonal patterns (see also Figure 2).

For instance, LPJ-GUESS simulates an early peak CO_2 uptake in May, which is not confirmed by the inversions (only NAME-HB yields to a similar peak). LPJ-GUESS simulates a NEE alternating between being a neutral flux and a positive but small ($\approx$0.1 PgC/month) net CO_2 source between July and March. This is most of the time outside or at the edge of the range of flux estimates derived from the inversions. The strong peak carbon uptake in June in ORCHIDEE (-0.62 PgC/month in June) clearly exceeds the lower boundary of the posterior ensemble (-0.55 PgC/month ~~, in the FLEXINVERT+ inversion, which is itself an outlier among the ensemble~~in FLEXINVERT). The positive NEE found by ORCHIDEE at the end of the summer (0.13 PgC/month in August and September) is also contradicted by the inversions ($\approx$0.04 PgC/month ensemble median in these two months). The phasing of the seasonal cycles in VPRM and SiBCASA are in good agreement with that of the inversions. The winter NEE estimate in VPRM (≈0.07 PgC/month between October and March) is lower than suggested by the inversions (ensemble median of 0.16 PgC/month), and on the contrary, the inversions point to a lower NEE than found by SiBCASA in the first three months of the year (0.19 PgC/month between January and March, compared to a corresponding ensemble median of 0.09 PgC/month).

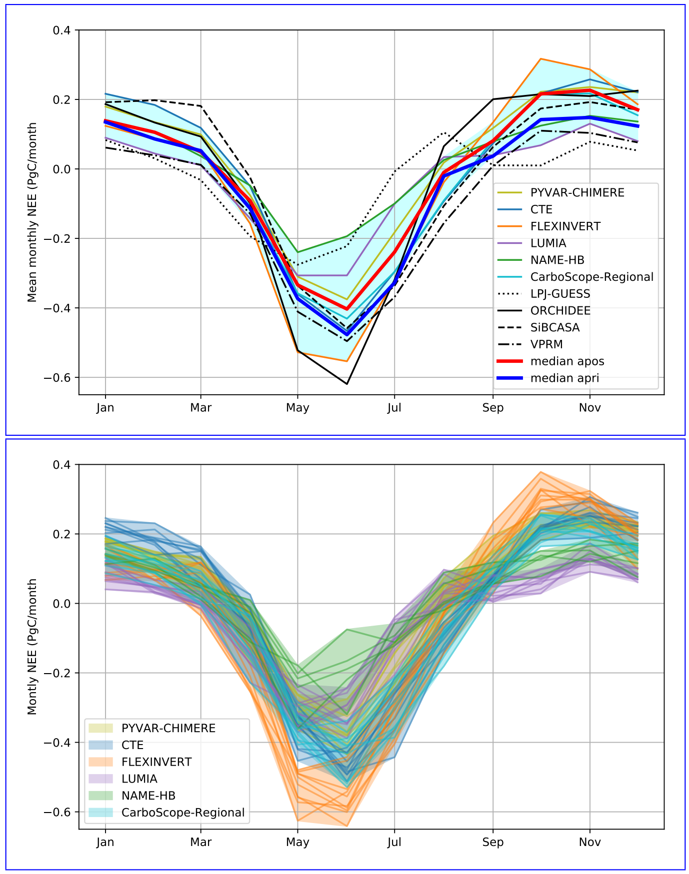

Figure 7. left: average seasonal cycles of the prior and posterior estimates. The prior and posterior ensemble median are represented as thick solid lines, and the spread of the posterior ensemble is shown as a shaded area. Right: variability of the seasonal cycle during the ten years of the inversion (the shaded areas represent the range of monthly NEE and the solid lines correspond to the individual years)

The variability of the seasonal ~~cycle for each inversion is~~ cycles obtained with each system are illustrated in the right hand plot of Figure 7.~~The solid lines show the posterior NEE for each year and inversion, and the shaded areas represent the variability of the seasonal cycle inferred by each inversion system.~~ The systematic differences between the inversion systems dominate the picture, and far exceed the monthly IAV within each inversion during the peak growth period (May-June) and during the fall (October-November). The peak to peak amplitude of the mean seasonal cycle inferred by the different inversions varies between 0.4 PgC/month for NAME-HB and 0.9 PgC/month for FLEXINVERT~~;~~.

Figure 8 focuses on the monthly anomalies relative to the average seasonal cycle that drive these main differences. The figure provides more insights to explain the IAV of the annual budgets, discussed in Section 4.2.1. For instance, the negative annual flux anomaly ~~in four inversions in 2006~~ (i.e. enhanced sink) found by four inversions in 2006 is explained by a stronger than usual carbon uptake in the summer of 2006. The negative NEE anomaly remains throughout most of the fall and winter of 2006-2007 ~~(up to -0.025 PgC/month during the period May to December 2006)~~ and becomes even more pronounced in March 2007, after which it switches sign. The 2012 anomaly is on the contrary spread over the entire year in almost all the inversions. It is however already well described by the priors, the inversions here provide a confirmation.

In some instances it may be possible to relate these NEE ~~monthly~~ anomalies to climate anomalies. For example, the summer 2006 in Europe was marked by a heat wave lasting for most of the month of July, and was followed by a particularly mild winter, which could explain the relatively stronger carbon sink from June 2006 to May 2007 (Rebetez et al., 2009). However, the size of our domain is assumedly much larger than the spatial extent of most potential climate anomalies, which complicates this type of analysis. We therefore briefly delve in the spatial distribution of the flux adjustments in the following section.

4.2.3 Spatial variability

Analysing the spatial variations of the fluxes may reveal robust local signals in areas where the transport models are more reliable and where the observation network is denser. It can also help to better interpret the results in terms of underlying processes in a large region such as Europe where the ecosystems and climate are highly heterogeneous. However, getting robust signals at regional scales is challenging due to the limited spatial resolution of the transport models and to the relative simplicity and large scales of the error correlations used for characterising the prior flux uncertainties. A detailed analysis of the regional signals will be published in a follow-up article. Here, we only provide a brief overview of the spatial distribution of the NEE adjustments to provide a first assessment of the potential of regional inversions to analyse subcontinental scale NEE variations and to support the previous analysis of the anomalies at the European scale.

We aggregate the fluxes in four large regions: Northern Europe (Scandinavia, Finland and the Baltic states), Southern Europe (the Iberian Peninsula, Italy, Greece, Romania and the Balkan states), Western Europe (France, Benelux, UK and Ireland) and Central Europe (the remaining countries, up to the Eastern border of Poland). The regions are pictured in Figure 1. These four regions correspond roughly to four climate zones (Nordic, Mediterranean, oceanic and continental) and exclude parts at the edge of our domain which are not sampled by the observation network (North Africa, Turkey, far east of Europe).

Average regional monthly budgets for both prior and posterior estimates are shown for each region in the upper row of Figure 9~~. The figure also shows the median of~~ , along with the prior and posterior ~~ensembles (respectively as thick blue and red lines). Finally, the spread of the posterior ensemble is highlighted (blue shaded area). The second row of plots show average prior and posterior regional annual budgets~~ensemble median and spread. Some of the systematic differences between the posterior seasonal cycles already noted at the European domain scale are present in all or most of the regions. This is in particular the case for the lower amplitude of the NAME-HB seasonal cycle and the autumn positive NEE peak in the FLEXINVERT + inversion. But others, such as the positive bias of the PYVAR-CHIMERE posterior (i.e. 0.28 PgC/year, see Figure 5), can be more clearly attributed to one specific region, like Southern Europe.

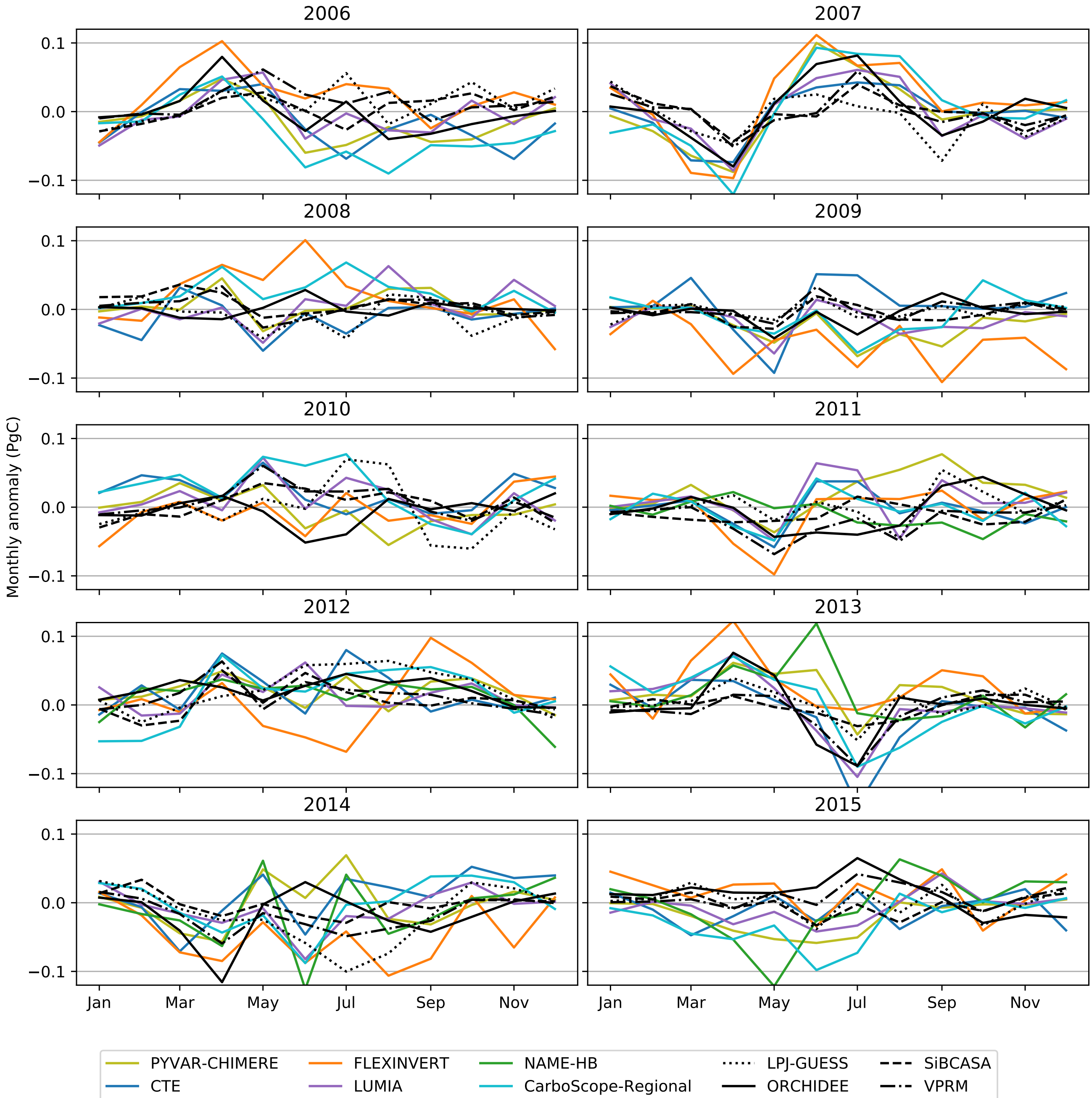

Figure 8. Monthly prior and posterior anomalies of the seasonal cycle (each simulation compared to its own average seasonal cycle (left hand plot of Figure 7)

Central Europe

NEE is most robustly estimated in the Central Europe region, which is not surprising because it is the region most densely sampled by the observation network. The median prior and posterior fluxes are nearly identical, but the spread of the posterior ensemble is generally narrower than that of the prior fluxes. In particular, the LPJ-GUESS NEE estimate is clearly outside the range of posteriors in the summer (it points to a peak uptake of -0.04 PgC/month in June, half of the -0.08 PgC/month ensemble median).

In terms of net annual budget, the inversions fall in two categories: CarboScope-Regional and LUMIA point to a sink of -0.12 PgC/year, all the other inversion systems yield a close to zero annual budget. The similarity in the annual budget from these four inversions is, however, most likely by coincidence because the seasonal distribution of the fluxes is rather different (FLEXINVERT ~~+~~ points to a summer uptake 30% larger than that found in the NAME-HB inversion, compensated by larger winter emissions).

Western Europe

Western Europe is also well sampled by the observation network, but because of the dominating westerly winds in our domain it is more sensitive to boundary conditions than the Central Europe region. The spread of the prior fluxes (0.02 to 0.04 PgC/month) is narrower than in Central Europe and is not further reduced by the inversions. In summer, the NAME-HB inversion suggests a reduced carbon uptake (-0.02 PgC/month in June, compared to a prior ensemble mean of -0.06 PgC/month), but as mentioned earlier, this is a systematic feature of that inversion, not specific to Western Europe. In Fall (October to December), two inversions point to a much stronger positive flux than the priors and the other inversion systems (up to +0.75 PgC/month in November, double the value of the posterior ensemble mean of +0.35 PgC/month). As a result, there is little convergence between the annual budgets, which range between a net sink of -0.12 PgC/year (CarboScope-Regional) to a source of 0.06 PgC/year (CTE).

Southern Europe

The strongest ~~correction~~ corrections to the prior fluxes are obtained in Southern Europe. The median value of the posterior estimates points to a ≈30% reduction of the summer CO_2 uptake compared to the median of the prior fluxes. The spread of the posterior ensemble is larger than in the other regions (~~0.03 to 0.1~~ -0.08 to 0.14 PgC/month) but the region is also where the spread of the prior interval is the largest (up to 0.13 PgC/month in July).

The shape of the LPJ-GUESS seasonal cycle is different from that of the other models, with two periods of negative NEE (February-June and October), and a peak carbon flux to the atmosphere in August. For most of the year, it remains outside the range of posterior scenarios, and is therefore not compatible with the atmospheric observations.

The seasonal cycles of the three other prior fluxes are in phase with that of the inversion ensemble, but the amplitude of the summer uptake in ORCHIDEE and SiBCASA is larger than that inferred by the inversions, and the peak of carbon emissions simulated by ORCHIDEE in August and September is also corrected by the inversions (respectively 0.04 and 0.07 PgC/month, compared to maximum ensemble posterior values of 0.02 and 0.04 PgC/month).

Northern Europe

In Northern Europe, for most of the year the range of posterior estimates is larger than that of the prior fluxes. All the simulations (including both prior and posterior) are well in phase, with a summer peak uptake in June/July and a stable winter flux between October and March. The inversions contradict the near-zero flux of the VPRM prior from August to December. The size of the summer uptake varies by a factor three, between the -0.04 PgC/month as estimated by NAME in June and a corresponding value of -0.13 Pgc/month estimated by the FLEXINVERT ~~+~~ inversion. The prior and posterior median are however nearly identical. Three inversions (CarboScope-Regional, FLEXINVERT ~~+~~ and LUMIA) yield a clear annual net carbon sink (~~-0.09~~ -0.06 to -0.14 PgC/year) for this region, however, the agreement on the size of the annual budget by CarboScope-Regional and FLEXINVERT ~~+~~ is again by coincidence, as they distribute the fluxes very differently throughout the year.

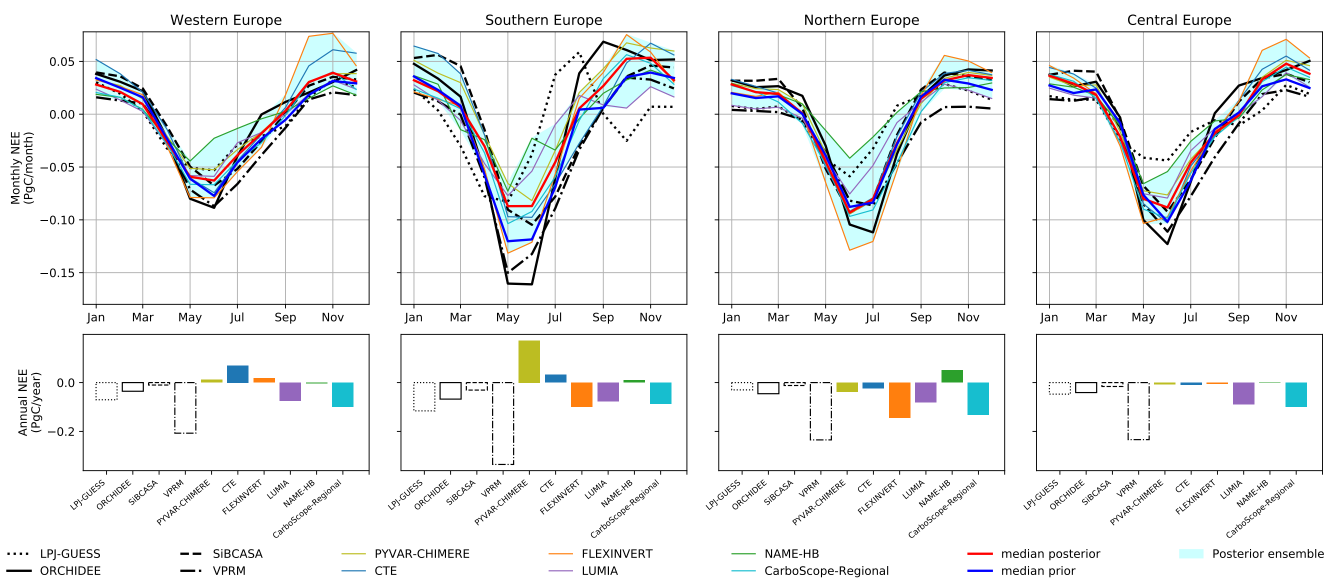

Figure 9. Upper row: Mean prior (black lines) and posterior (coloured lines) seasonal cycle of the terrestrial carbon flux in the four regions highlighted in Figure 1; prior and posterior ensemble median (thick blue and red lines) and posterior spread (blue shaded area). Lower row: mean annual net terrestrial carbon flux for these same regions

5 Discussion

We have presented an overview of the results from the first set of inversions from the EUROCOM project. The ensemble includes one inversion for each of the participating systems and is designed so as to maximize the exploration of uncertainties, by covering a large diversity of inversion settings (prior, boundary condition, type of optimization technique and of optimized vector, etc.). This approach yields a more realistic approximation of the uncertainties, and enables identifying the dominant features of that uncertainty, better than what could be obtained from e.g. a set of sensitivity experiments using a single inversion system, or than the theoretical uncertainty estimates that are sometimes computed along with the inversions. The trade-off of that approach is that it doesn't easily permit to determine the causes for the discrepancies between the inversions.

In the subsections below, we further discuss several aspects of our results, and put them in perspective with the previously published literature. First we discuss the annual budget of NEE in Europe, which has been a debated topic. Then in Section 5.2 we discuss the aspects of our results that are at this stage the most robust, and for which regional inversions seem the most relevant. Finally we briefly discuss the future perspectives of regional inversions, and of the EUROCOM intercomparison exercise.

5.1 How well can regional-scale inversions constrain the annual budget of European NEE?

The annual budget of NEE is a key metric to characterise the amount of carbon absorbed by the European ecosystems, since it balances the releases in winter and at night (by ecosystem respiration) with the uptakes during daytime, mostly in spring and summer (by photosynthesis). Annual to multi-annual budgets are an important measure to quantify the impact of environmental conditions such as ecosystem management, disturbances and climate extremes on the terrestrial carbon cycle.

The annual budget has notably been synthesised in Reuter et al. (2017): on the one hand, global inversions that assimilate only surface observations showed the geographical Europe as a moderate to rather small carbon sink (≈-0.4 PgC/year) on multi-annual scales; and that, on the other hand, inversions constrained by satellite retrievals of total column atmospheric CO_2 (XCO_2) consistently infer that it is a much larger sink, on the order of -1 PgC/year. More recent studies suggest a smaller uptake: Scholze et al. (2019) find a mean sink of -0.3 $\pm$ 0.08 PgC/year by assimilating three datasets (namely in-situ atmospheric CO_2 and remotely sensed soil moisture and vegetation optical depth) into their Carbon Cycle Data Assimilation System. Similarly, Crowell et al. (2019) estimate a mean sink of -0.25 $\pm$ 0.46 PgC/year from the ensemble of an intercomparison of atmospheric inversions based on XCO_2 observations from OCO-2. These estimates correspond to a geographical European domain which extends eastwards to the Ural, and which is much larger than the domain studied here. The areas of highest uptake in these satellite inversions are located in the eastern part of Europe, i.e. east of our EUROCOM domain.

From our ensemble of inversions we find a median sink of ~~-0.21~~ -0.15 PgC/year, relatively constant from year to year and with no significant trend over the ten years of the period studied. Our study therefore tends to support the hypothesis that ecosystems in the European domain studied here are a weak carbon sink. Because of the differences in the domain extent, our inversions cannot close the controversy. But they indicate that, if there is a strong land sink over Europe (on the order of 1 PgC/year), then most of it has to be located in Eastern Europe, beyond the extent of our dense observation network.

Figure 10 provides results for our European domain (long term mean and IAV) from a set of state of the art global inversions that assimilate only surface observations and which cover the time period studied here. They correspond to the set of global inversions used for the Global Carbon Project annual analyses (Le Quéré et al., 2018). The range of mean annual NEE obtained from these global inversions is about half that obtained from our regional inversions (0.8 PgC/year), which suggest that, at this scale, our regional inversions do not constrain the annual NEE better than global inversions. The spread between these 4 state-of-the-art global inversions selected for the GCP synthesis actually corresponds to the outcome of a long process of improvement and selection of inversion configurations, as reflected by the very large spread of 1.8 PgC/year obtained from the inter-comparisons by Peylin et al. (2013). Therefore, one can expect the process of inter-comparing regional scale inversions started here with the EUROCOM project to yield a much-refined estimate of the annual to multi-annual budgets in the coming years. ~~We note here again, that our inversion protocol was intentionally very loose, to allow for more systems to participate and hence to maximise the exploration of the space of uncertainties. It is therefore expected that the range of estimates would be large, and we consider it as a rather conservative representation of the true uncertainties.~~

The slightly narrower spread of the global inversions nonetheless questions specific aspects of the regional inversions, which may prevent them from providing more precise estimates of the continental-scale fluxes. Part of the constraint on the European NEE in global inversions comes from the observed large scale atmospheric gradients between stations located in the Atlantic Ocean and Asia. These ~~constraint~~ constraints are only incompletely transferred to regional inversion via their boundary conditions and the shorter scale gradients captured by the continental observations may not be sufficient to characterise the continental carbon balance. Unless the surface network is extended to cover sufficiently the Eastern and Southern parts of the domain, it might be useful to impose ~~an~~ a constraint on the large scale gradients to the regional inversions. However this also ~~mean~~ means that the relevant scale for regional inversions is possibly much smaller. The next section focuses thus on the spatial and temporal scales where our ensemble of inversions leads to robust and consistent results.

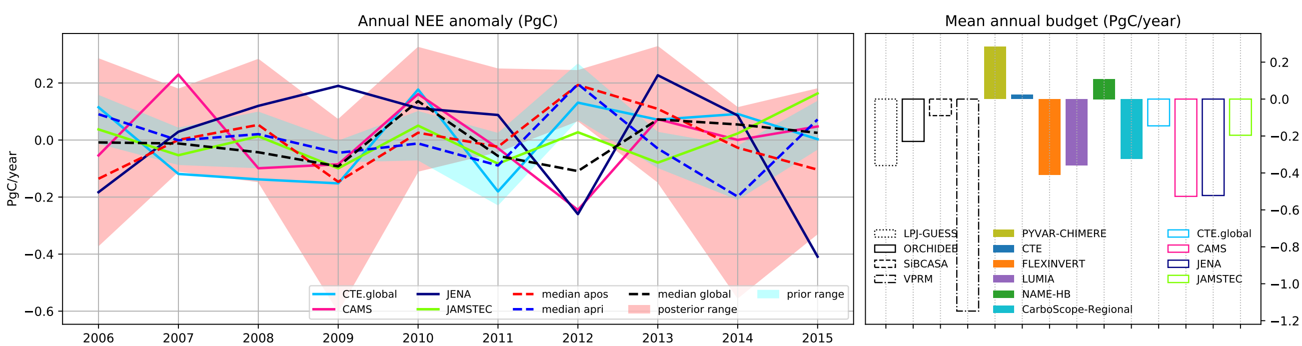

Figure 10. Comparison of the EUROCOM inversion ensemble with global inversions from the Global Carbon Project (Le Quéré et al., 2018). The right hand plot shows the annual NEE anomalies, with our prior and posterior ensembles shown as shaded areas, for the clarity of the figure. The right hand plot shows the mean annual budgets. The results from the global inversions were extracted for the exact EUROCOM domain on the global carbon atlas (www.globalcarbonatlas.org).

5.2 New insights on the European land carbon flux

While the net annual carbon flux is an important metric, focusing on it can give an overly pessimistic view of the results, especially integrated over the whole European domain for which the observational coverage is not homogeneous. In fact, a large share of the discrepancies between the inversions can be attributed to regions located at the eastern and southern borders of our domain that are not part of the four regions discussed in Section 4.2.3 (Russia, Ukraine, Turkey and North Africa, but also large swathes of the "Southern Europe" region).

Indeed, the ensemble of inversions leads to a narrower range of estimates than the ensemble of priors for regions with a dense network such as Central Europe (0.21 PgC/year of difference between the priors, vs. ~~0.13~~ 0.10 PgC/year between the posterior estimates). In contrast, the range of the ensemble of the inversions is almost three times the range of the prior estimates (0.33 PgC/year between the optimised annual NEE, vs. 0.12 PgC/year between the prior estimates) in the parts of the domain that are outside the our four regions (see above), despite these regions being only rarely downwind of the observation network.This means that although the posterior annual estimates at the continental (whole domain) scale may not be more robust than in e.g. the GCP global inversions (as discussed in 5.1), the regional inversions in our intercomparison are capable of resolving annual fluxes at the scale of large countries (e.g. 0.8 to 1.6 million km2 km^2 as for our four regions), provided that the observational network is dense enough.

The size of the spread between the posterior monthly flux estimates varies by a factor five throughout the year (at the continental scale). Monthly fluxes are usually well resolved in the first nine months of the year except for June, but this larger spread in June is due to one single inversion (NAME-HB), which among other differences uses a much reduced observation network (See Section 4.1) and covers only the last five years of our simulation period. The larger range among the estimates at the end of the year is more problematic and points to a problem of the inversions to robustly resolve the winter NEE, i.e. the the terrestrial respiration flux. Some speculative explanations could be larger systematic transport errors (winter concentrations are more difficult to represent, in particular at the highest latitudes, where the boundary layer remains extremely shallow and the vertical atmospheric stratification is high when the days are very short), and/or larger relative differences in the prescribed prior flux uncertainties between the inversions (uncertainties are overall smaller in winter because of the lack of photosynthesis).

In three regions (Western, Central and Northern Europe), the prior and posterior ensemble median of the seasonal cycle are almost identical, meaning that the inversions mostly provide a confirmation of the prior knowledge (see Figure 9). However, the differences between our ensemble median (i.e. best-informed guess) and each individual prior are sometimes large. For instance, the inversions consistently yield a summer uptake twice larger in Central Europe compared to the one computed by LPJ-GUESS. The results present therefore useful information for bottom-up modellers as they can be used to identify local or regional shortcomings in their models. This is also true when looking at the inversion results for Southern Europe. Although the posterior estimates are not as consistent with each other as in the other regions, we nevertheless can clearly identify some shortcomings in the priors because they are far out of the ensemble spread (e.g. summer summer uptake by ORCHIDEE and SiBCASA, the double peak from LPJ-GUESS).

In summary, the relative lack of convergence of the annual fluxes at the scale of the continent hides more robust features at monthly and smaller regional scales, especially in Central and Western Europe, where the observation network is the densest. The divergences between the inversions regarding the winter fluxes will need to be investigated through a targeted effort. Nonetheless, the aim of optimising fluxes at country scale appears achievable for the large countries in areas with a dense network. In the later years, the density of observation sites in Northern Europe has increased a lot, so it is expected that the spread between the posterior estimates in this region can be reduced significantly in the future.

5.3 Future prospects of regional inversions and of the EUROCOM intercomparison

Our inversion protocol was intentionally very loose, to allow for the participation of a wide range of inversion systems and hence maximise the exploration of the space of uncertainties. It is therefore expected that the range of estimates would be large, and we consider it as a rather conservative representation of the actual uncertainties. The analysis of the results offers a few clues:

- The resolution of the transport model does not appear to be a major limitation (the best fit to the observations is actually obtained by the inversion using the lowest resolution transport model, see Table 3).
- Systematic patterns such as the differences in seasonal cycle or the average net annual NEE also seem relatively unaffected by the changes in the sites selections that occur thoughout the 10-years period, as the availability of data evolves (Figure 7, right).
- The choice of a prior necessarily has an impact, but can only explain a small fraction of the differences between the inversions (the differences between the three inversions using the ORCHIDEE prior are on the same order as the differences with the inversions using different priors).

The explanations for the spread of the ensemble are likely multi-factorial and specific to each feature of the spread. A traditional approach to assess the quality of inversion results would be to verify whether the posterior fluxes also enable improvements in the fit to independent (i.e. non-assimilated) observed CO_2 concentrations. This approach has however its own limitations. The goodness of fit to observations at one location cannot be generalized to the entire inversion domain/period. Using the few available (not already assimilated) concentration data (e.g. from aircraft measurement campaigns) could in fact deliver the incorrect message that one system is systematically over or under-performing the others. For the same reason, a validation by comparison with in-situ flux measurements is risky, as the differences in scales between the fluxes optimized by the inversions and those measured by e.g. eddy-covariance sites are very large.

The inter-comparison exercise presented here does not enable a direct validation of the optimized fluxes, but it provides an estimation of how large the uncertainties are, and it helps identifying its dominant features in a more systematic manner than what could be done with just one inversion system. To conclusively determine the drivers of that uncertainty and discriminate their relative contributions, targeted efforts will be needed, primarily in the form of sensitivity experiments using a stricter protocol and focusing on one driver of uncertainty at a time. The knowledge gained from the first phase of the intercomparison will help design these new experiments, which should rapidly lead to an improved convergence of the ensemble.

6 Conclusions and future of the EUROCOM project

The EUROCOM project delivered a set of European NEE estimates at ~~high temporal and spatial~~ a 0.5°, monthly resolution at the disposal to the scientific community (Monteil et al., 2020). The data can be used as comparison and validation dataset for both bottom-up and inverse modellers. The input datasets (observations and prior fluxes) remain available for inverse modelling groups willing to submit additional inversions, and we expect the size and robustness of the ensemble to grow over time. An extension of the inversions until 2019 is currently ongoing.

Our best posterior estimate (ensemble median) of the long-term mean annual terrestrial European NEE of ~~-0.21~~ -0.15±0.08 PgC/year over the years 2006-2015 is comparable to the median value of -0.3±0.11 PgC/year from our prior estimates as well as recent estimates from other studies (e.g. Scholze et al., 2019; Crowell et al., 2019), albeit for a slightly different domain). Since our domain here does not cover the European part of Russia, the area that is postulated to contribute most to the large European carbon sink (see e.g. Reuter et al., 2017)), we cannot resolve this controversy here with our intercomparison.

We deliberately kept the requirements in the intercomparison protocol (i.e. use of prescribed common data sets or inversion set ups) to a minimum (namely, prescribed fossil fuel emissions and common domain) to encourage the participation of voluntary contributions from regional atmospheric inverse modelling groups~~in this EUROCOM intercomparison project~~. Such an intercomparison approach, where a large number of parameters influencing results of the inversions vary from one system to another, presents the advantage that the resulting distribution of results provides a good approximation of the distribution of uncertainties on the net European terrestrial carbon flux. Indeed, the analysis shows that no inversion is clearly more or less valid than the others and depending on the focus metrics, each can be an outlier. Such a multi-model/multi-inversion system ensemble is the best approach for providing robust estimates of the European carbon budget.

The ~~robustly modelled features~~ most robustly modelled feature in our ensemble ~~are mainly the IAV and~~ is the mean seasonality of the annual CO_2 sink in regions with a dense observational network, i.e. mainly central and western Europe illustrating the usefulness of a coordinated infrastructure such as ICOS in delivering high-quality observations. The coverage of the observational network in some regions of Europe is still limited, which is clearly reflected in a larger spread in the annual and monthly budgets in these regions within our ensemble. Observations from satellites, such as OCO-2 or the upcoming CO2M, may help in increasing the coverage but they have their own limitations (prone to clouds and aerosols, limited coverage during the winter season if based on passive optical instruments).

The mean annual terrestrial NEE itself is not strongly constrained by the observations and we find a spread of ~~0.8~~ 0.7 PgC/year within our ensemble. As mentioned above, this is partly because of the high freedom in the choice of settings. This freedom in the choice of settings makes it rather challenging to fully understand the causes of the spread in the ensemble results and the underlying uncertainties. We will investigate these differences in more detail and evaluate some of the specific parameters involved in the inversion set-ups in a forthcoming paper. Eventually, this will lead to a much better quality of the regional inversion estimates that could not have been possible without such an intercomparison exercise.

Currently, the main benefit of regional inversions over global ones does not appear to be at the scale of the continent, but rather at finer spatial scales, in regions well covered by the observation networks. The observation network seems sufficiently dense to envision robust country-scale estimates of the carbon balance (at least for the largest countries) in Western and Central Europe. Recent expansions of the networks both in Northern and Southern Europe should also enable a significant reduction of the spread between the inversions in the near future.

Author contributions. The intercomparison was collectively designed in the frame of the EUROCOM project coordinated by GB and MS with support from UK, GM, ML and PP. GM wrote the paper together with GB and MS. UK coordinated the construction of the common atmospheric observation database together with JT who compiled the pre-ICOS observations), processed the anthropogenic emission dataset, coordinated the exchange of observations, prior datasets and inversion results through the ICOS Carbon Portal, and supported the analysis of the results. GM designed and performed the LUMIA inversions. ML performed the PYVAR-CHIMERE inversions. CR provided the ocean flux used in some of the inversions, and together with CG designed the CarboScope-Regional system. FTK performed the CarboScope-Regional inversions. RLT developed the FLEXINVERT system and performed the FLEXINVERT inversions. EDW, AJM, EMW performed NAME model runs and processed NAME output for the high time frequency NAME-HB inversion. EDW, ALG and MR developed the NAME-HB inverse modelling method and code for performing high time frequency inversions and adapted the approach for the EUROCOM domain. ITL and WP designed the CTE inversion system; ITL and NES performed the CTE inversions and provided the SiBCASA prior. MM, PP and CG produced respectively the LPJ-GUESS, ORCHIDEE and VPRM priors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank the Swedish Research Council for funding the EUROCOM project (Dnr 349-2014-6576). MS and GM acknowledge support from the three Swedish strategic research areas ModElling the Regional and Global earth system (MERGE), the e-science collaboration (eSSENCE), and Biodiversity and Ecosystems in a Changing Climate (BECC). GM thanks the NSC at Linköping University (part of the Swedish National Infrastructure for Computing, SNIC) for providing computer resources for the LUMIA simulations. GB and JT acknowledge the support by L. Rivier (for the transmission of pre-ICOS data), F. Chevallier and I. Pison (for the continuous development and maintenance of the PYVAR-CHIMERE system). FTK, CR, and CG thank the Deutsches Klimarechenzentrum (DKRZ) for use of the high-performance computing facilities. NES and ~~ITvdLL~~ ITL benefited from computing resources from Netherlands Organization for Scientific Research (NWO; grant no. SH-312, 17616). NES was funded by NWO/OCW for Ruisdael/ICOS-NL. ~~ITvdLL~~ ITL was funded by a NWO Veni grant (016.Veni.171.095). EW was funded by the UK Natural Environment Research Council (NERC) GW4+ Doctoral Training Partnership. ALG is funded through a NERC Independent Research Fellowship NE/L010992/1. MR was supported by NERC grants NE/K002236/1 and NE/S004211/1. RLT was funded by the Research Council of Norway through the research infrastructure project "Integrated Carbon Observation System (ICOS) Norway (grant no. 245927)". We thank Greet Janssens-Meanhout (European Commission, Joint Research Center, Ispra, Italy) for providing the fuel type and category specific version of the EDGAR v4.3 anthropogenic emission data and Hugo Denier van der Gon (TNO, The Netherlands) for making available the temporal emission profiles. We thank ICOS-ERIC for financial support. We also thank the data providers from the GLOBALVIEW-plus v3.2 product and from the WDCGG for the atmospheric CO_2 observations used in the inversions. Finally, we thanl the ICOS Atmospheric Thematic Center for providing the CO_2 measurements from CarboEurope-IP, GHG-Europe, and ICOS Preparatory Phase.

References

BP: Statistical Review of World Energy 2016, Tech. rep., http://www.bp.com/content/dam/bp/excel/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-workbook.xls, 2016.

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions BT - Progress in Photosynthesis Research: Volume 4 Proceedings of the VIIth International Congress on Photosynthesis Pr, pp. 221–224, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-017-0519-6_48, https://doi.org/10.1007/978-94-017-0519-6{_}48, 1987.

Bastos, A., Janssens, I. A., Gouveia, C. M., Trigo, R. M., Ciais, P., Chevallier, F., Peñuelas, J., Rödenbeck, C., Piao, S., Friedlingstein, P., and Running, S. W.: European land CO2 sink influenced by NAO and East-Atlantic Pattern coupling, Nature Communications, 7, 10 315, https://doi.org/10.1038/ncomms10315, https://doi.org/10.1038/ncomms10315, 2016.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, Journal of Geophysical Research: Atmospheres, 116, n/a–n/a, https://doi.org/10.1029/2011JD016202, http://doi.wiley.com/10.1029/2011JD016202, 2011.

Broquet, G., Chevallier, F., Bréon, F.-M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmospheric Chemistry and Physics, 13, 9039–9056, https://doi.org/10.5194/acp-13-9039-2013, https://www.atmos-chem-phys.net/13/9039/2013/, 2013.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, a. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. a., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. a., Vermeulen, a. T., Wofsy, S., and Worthy, D.: CO$_2$ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, Journal of Geophysical Research, 115, D21 307, https://doi.org/10.1029/2010JD013887, http://doi.wiley.com/10.1029/2010JD013887, 2010.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, 437, 529–533, https://doi.org/10.1038/nature03972, https://doi.org/10.1038/nature03972, 2005.

Cooperative Global Atmospheric Data Integration Project: Multi-laboratory compilation of atmospheric carbon dioxide data for the period 1957-2016; obspack_co2_1_GLOBALVIEWplus_v3.2_2017-11-02, https://doi.org/10.15138/G3704H, http://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php?id=obspack{_}co2{_}1{_}GLOBALVIEWplus{_}v3.2{_}2017-11-02, 2017.

Crisp, D., Atlas, R., Breon, F.-M., Brown, L., Burrows, J., Ciais, P., Connor, B., Doney, S., Fung, I., Jacob, D., Miller, C., O'Brien, D., Pawson, S., Randerson, J., Rayner, P., Salawitch, R., Sander, S., Sen, B., Stephens, G., Tans, P., Toon, G., Wennberg, P., Wofsy, S., Yung, Y., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Advances in Space Research, 34, 700–709, https://doi.org/10.1016/J.ASR.2003.08.062, https://www.sciencedirect.com/science/article/pii/S0273117704003539, 2004.

Crowell, S., Baker, D., Schuh, A., Basu, S., Jacobson, A. R., Chevallier, F., Liu, J., Deng, F., Feng, L., McKain, K., Chatterjee, A., Miller, J. B., Stephens, B. B., Eldering, A., Crisp, D., Schimel, D., Nassar, R., O'Dell, C. W., Oda, T., Sweeney, C., Palmer, P. I., and Jones, D. B. A.: The 2015–2016 carbon cycle as seen from OCO-2 and the global in situ network, Atmospheric Chemistry and Physics, 19, 9797–9831, https://doi.org/10.5194/acp-19-9797-2019, https://www.atmos-chem-phys.net/19/9797/2019/, 2019.

Cullen, M. J. P.: The unified forecast/climate model, The Meteorological Magazine, pp. 81–94, 1993.

Denier van der Gon, H. D., Hendriks, C., Kuenen, J., Segers, A., and Visschedijk, A.: Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns, in: EU FP7 MACC deliverable report D_D-EMIS_1.3, https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC{_}TNO{_}del{_}1{_}3{_}v2.pdf, 2011.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, 2002.

Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, Planta, 149, 78–90, https://doi.org/10.1007/BF00386231, https://doi.org/10.1007/BF00386231, 1980.

Fortems-Cheiney, A., Pison, I., Dufour, G., Broquet, G., Berchet, A., Potier, E., Coman, A., Siour, G., and Costantino, L.: Variational regional inverse modeling of reactive species emissions with PYVAR-CHIMERE, Geoscientific Model Development Discussions, pp. 1–22, https://doi.org/10.5194/gmd-2019-186, https://www.geosci-model-dev-discuss.net/gmd-2019-186/, 2019.

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, Atmospheric Chemistry and Physics, 14, 3855–3864, https://doi.org/10.5194/acp-14-3855-2014, https://www.atmos-chem-phys.net/14/3855/2014/, 2014.

Giglio, L., Randerson, J. T., and Van Der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317–328, https://doi.org/10.1002/jgrg.20042, 2013.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO$_{2}$ exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, https://doi.org/10.5194/bg-9-457-2012, https://www.biogeosciences.net/9/457/2012/, 2012.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of annual mean {CO}_2 sources and sinks, Nature, 415, 626–630, 2002.

Huijnen, V., Williams, J., Van Weele, M., Van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., De Laat, J., Boersma, F., Bergamaschi, P., Van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H. W.: The global chemistry transport model TM5: Description and evaluation of the tropospheric chemistry version 3.0, Geoscientific Model Development, 3, 445–473, https://doi.org/10.5194/gmd-3-445-2010, 2010.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change, 109, 117–161, https://doi.org/10.1007/s10584-011-0153-2, 2011.

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochemical Cycles, 21, https://doi.org/10.1029/2005GB002556, https://doi.org/10.1029/2005GB002556, 2007.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., Petrescu, A. M. R., Solazzo, E., and Oreggioni, G. D.: EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012, Earth System Science Data, 11, 959–1002, https://doi.org/10.5194/essd-11-959-2019, https://www.earth-syst-sci-data.net/11/959/2019/, 2019.

Jung, M., Henkel, K., Herold, M., and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sensing of Environment, 101, 534–553, https://doi.org/10.1016/J.RSE.2006.01.020, https://www.sciencedirect.com/science/article/pii/S0034425706000514, 2006.

Kountouris, P., Gerbig, C., Totsche, K.-U., Dolman, A. J., Meesters, A. G. C. A., Broquet, G., Maignan, F., Gioli, B., Montagnani, L., and Helfter, C.: An objective prior error quantification for regional atmospheric inverse applications, Biogeosciences, 12, 7403–7421, https://doi.org/10.5194/bg-12-7403-2015, https://www.biogeosciences.net/12/7403/2015/, 2015.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F., and Heimann, M.: Technical Note: Atmospheric CO2 inversions on the mesoscale using data-driven prior uncertainties: methodology and system evaluation, Atmospheric Chemistry and Physics, 18, 3027–3045, https://doi.org/10.5194/acp-18-3027-2018, https://www.atmos-chem-phys.net/18/3027/2018/, 2018a.

Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F., and Heimann, M.: Atmospheric CO2 inversions on the mesoscale using data-driven prior uncertainties: quantification of the European terrestrial CO2 fluxes, Atmospheric Chemistry and Physics, 18, 3047–3064, https://doi.org/10.5194/acp-18-3047-2018, https://www.atmos-chem-phys.net/18/3047/2018/, 2018b.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, 1–33, https://doi.org/10.1029/2003GB002199, 2005.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied Optics, 48, 6716–6733, https://doi.org/10.1364/AO.48.006716, http://ao.osa.org/abstract.cfm?URI=ao-48-35-6716, 2009.

Law, R. M., Rayner, P. J., Denning, A. S., Erickson, D., Fung, I. Y., Heimann, M., Piper, S. C., Ramonet, M., Taguchi, S., Taylor, J. A., Trudinger, C. M., and Watterson, I. G.: Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO2 inversions, Global Biogeochemical Cycles, 10, 783–796, https://doi.org/10.1029/96GB01892, https://doi.org/10.1029/96GB01892, 1996.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth System Science Data, 10, 2141–2194, https://doi.org/10.5194/essd-10-2141-2018, https://www.earth-syst-sci-data.net/10/2141/2018/, 2018.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD003161, https://doi.org/10.1029/2002JD003161, 2003.

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, Geoscientific Model Development, 9, 3213–3229, https://doi.org/10.5194/gmd-9-3213-2016, https://www.geosci-model-dev.net/9/3213/2016/, 2016.

Luyssaert, S., Abril, G., Andres, R., Bastviken, D., Bellassen, V., Bergamaschi, P., Bousquet, P., Chevallier, F., Ciais, P., Corazza, M., Dechow, R., Erb, K.-H., Etiope, G., Fortems-Cheiney, A., Grassi, G., Hartmann, J., Jung, M., Lathière, J., Lohila, A., Mayorga, E., Moosdorf, N., Njakou, D. S., Otto, J., Papale, D., Peters, W., Peylin, P., Raymond, P., Rödenbeck, C., Saarnio, S., Schulze, E.-D., Szopa, S., Thompson, R., Verkerk, P. J., Vuichard, N., Wang, R., Wattenbach, M., and Zaehle, S.: The European land and inland water CO2, CO, CH4 and N2O balance between 2001 and 2005, Biogeosciences, 9, 3357–3380, https://doi.org/10.5194/bg-9-3357-2012, https://www.biogeosciences.net/9/3357/2012/, 2012.

Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochemical Cycles, 22, https://doi.org/10.1029/2006GB002735, https://doi.org/10.1029/2006GB002735, 2008.

Maignan, F., Bréon, F. M., Vermote, E., Ciais, P., and Viovy, N.: Mild winter and spring 2007 over western Europe led to a widespread early vegetation onset, Geophysical Research Letters, 35, https://doi.org/10.1029/2007GL032472, https://doi.org/10.1029/2007GL032472, 2008.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth System Science Data, 6, 375–384, https://doi.org/10.5194/essd-6-375-2014, https://www.earth-syst-sci-data.net/6/375/2014/, 2014.

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, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017797, https://doi.org/10.1029/2012JD017797, 2012.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geoscientific Model Development, 6, 981–1028, https://doi.org/10.5194/gmd-6-981-2013, https://www.geosci-model-dev.net/6/981/2013/, 2013.

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., and Sarmiento, J. L.: Inverse estimates of the oceanic sources and sinks of natural CO2 and the implied oceanic carbon transport, Global Biogeochemical Cycles, 21, https://doi.org/10.1029/2006GB002751, https://doi.org/10.1029/2006GB002751, 2007.

Monteil, G. and Scholze, M.: Regional CO2 inversions with LUMIA, the Lund University Modular Inversion Algorithm, v1.0, Geoscientific Model Development Discussions, pp. 1–35, https://doi.org/10.5194/gmd-2019-227, https://www.geosci-model-dev-discuss.net/gmd-2019-227/, 2019.

Monteil, G., Broquet, G., Scholze, M., Lang, M., Karstens, U., Gerbig, C., Koch, F.-T., Smith, N., Thompson, R., van der Laan-Luijkx, I., White, E., and Meesters, A.: EUROCOM ensemble of inversion results for 2006-2015, https://doi.org/10.18160/G068-1T09, https://meta.icos-cp.eu/collections/wlymZyMC5vTHtvk0_yFnDnlf, 2020.

Naudts, K., Chen, Y., McGrath, M. J., Ryder, J., Valade, A., Otto, J., and Luyssaert, S.: Europe's forest management did not mitigate climate warming, Science, 351, 597 LP – 600, https://doi.org/10.1126/science.aad7270, http://science.sciencemag.org/content/351/6273/597.abstract, 2016.

Oney, B., Gruber, N., Henne, S., Leuenberger, M., and Brunner, D.: A CO-based method to determine the regional biospheric signal in atmospheric CO2, Tellus B: Chemical and Physical Meteorology, 69, 1353 388, https://doi.org/10.1080/16000889.2017.1353388, https://doi.org/10.1080/16000889.2017.1353388, 2017.

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., Nemitz, 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, Atmospheric Chemistry and Physics, 18, 11 753–11 777, https://doi.org/10.5194/acp-18-11753-2018, https://www.atmos-chem-phys.net/18/11753/2018/, 2018.

Patra, P. K., Law, R. M., Peters, W., Rödenbeck, C., Takigawa, M., Aulagnier, C., Baker, I., Bergmann, D. J., Bousquet, P., Brandt, J., Bruhwiler, L., Cameron-Smith, P. J., Christensen, J. H., Delage, F., Denning, A. S., Fan, S., Geels, C., Houweling, S., Imasu, R., Karstens, U., Kawa, S. R., Kleist, J., Krol, M. C., Lin, S.-J., Lokupitiya, R., Maki, T., Maksyutov, S., Niwa, Y., Onishi, R., Parazoo, N., Pieterse, G., Rivier, L., Satoh, M., Serrar, S., Taguchi, S., Vautard, R., Vermeulen, A. T., and Zhu, Z.: TransCom model simulations of hourly atmospheric CO2: Analysis of synoptic-scale variations for the period 2002–2003, Global Biogeochemical Cycles, 22, https://doi.org/10.1029/2007GB003081, https://doi.org/10.1029/2007GB003081, 2008.

Peng, S., Ciais, P., Chevallier, F., Peylin, P., Cadule, P., Sitch, S., Piao, S., Ahlström, A., Huntingford, C., Levy, P., Li, X., Liu, Y., Lomas, M., Poulter, B., Viovy, N., Wang, T., Wang, X., Zaehle, S., Zeng, N., Zhao, F., and Zhao, H.: Benchmarking the seasonal cycle of CO2 fluxes simulated by terrestrial ecosystem models, Global Biogeochemical Cycles, 29, 46–64, https://doi.org/10.1002/2014GB004931, https://doi.org/10.1002/2014GB004931, 2015.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker., Proceedings of the National Academy of Sciences of the United States of America, 104, 18 925–18 930, https://doi.org/10.1073/pnas.0708986104, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2141884{&}tool=pmcentrez{&}rendertype=abstract, 2007.

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, Global Change Biology, 16, 1317–1337, https://doi.org/10.1111/j.1365-2486.2009.02078.x, http://doi.wiley.com/10.1111/j.1365-2486.2009.02078.x, 2010.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., Van Der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: Results from an ensemble of atmospheric CO2 inversions, Biogeosciences, 10, 6699–6720, https://doi.org/10.5194/bg-10-6699-2013, 2013.

Pinty, B., Janssens-Maenhout, G., Dowell, M., Zunker, H., Brunhes, T., Ciais, P., Dee, D., Denier van der Gon, H., Dolman, H., Drinkwater, M., Engelen, R., Heimann, M., Holmlund, K., Husband, R., Kentarchos, A., Meijer, Y., Palmer, P., and Scholze, M.: An operational anthropogenic CO2 emissions monitoring & verification support capacity - Baseline requirements, model components and functional architecture, European Commission Joint Research Center, EUR 28738 EN, https://doi.org/10.2760/08644, http://edgar.jrc.ec.europa.eu/news{_}docs/Report{_}Copernicus{_}CO2{_}Monitoring{_}TaskForce{_}2017.pdf, 2017.

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.3, Geoscientific Model Development Discussions, pp. 1–67, https://doi.org/10.5194/gmd-2018-333, https://www.geosci-model-dev-discuss.net/gmd-2018-333/, 2019.

Ramonet, M., Ciais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmid, M., Simmonds, P., Xueref-Rémy, I., and Necki, J.: A recent build-up of atmospheric CO2 over Europe. Part 1: observed signals and possible explanations, Tellus B: Chemical and Physical Meteorology, 62, 1–13, https://doi.org/10.1111/j.1600-0889.2009.00442.x, https://www.tandfonline.com/doi/full/10.1111/j.1600-0889.2009.00442.x, 2010.

Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of Data Assimilation applied to biogeochemistry, Atmospheric Chemistry and Physics Discussions, pp. 1–32, https://doi.org/10.5194/acp-2018-1081, https://www.atmos-chem-phys-discuss.net/acp-2018-1081/, 2018.

Rebetez, M., Dupont, O., and Giroud, M.: An analysis of the July 2006 heatwave extent in Europe compared to the record year of 2003, Theoretical and Applied Climatology, 95, 1–7, https://doi.org/10.1007/s00704-007-0370-9, https://doi.org/10.1007/s00704-007-0370-9, 2009.

Reuter, M., Buchwitz, M., Hilker, M., Heymann, J., Bovensmann, H., Burr Ows, J. P., Houweling, S., Liu, Y. Y., Nassar, R., Chevallier, F., Ciais, P., Marshall, J., and Reichstein, M.: How much CO2 is taken up by the European terrestrial biosphere?, Bulletin of the American Meteorological Society, 98, 665–671, https://doi.org/10.1175/BAMS-D-15-00310.1, 2017.

Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, Atmospheric Chemistry and Physics, 11, 9887–9898, https://doi.org/10.5194/acp-11-9887-2011, https://www.atmos-chem-phys.net/11/9887/2011/, 2011.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO$_2$ flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmospheric Chemistry and Physics, 3, 1919–1964, 2003.

Rödenbeck, C., Gerbig, C., Trusilova, K., and Heimann, M.: A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, Atmospheric Chemistry and Physics Discussions, 9, 1727–1756, https://doi.org/10.5194/acpd-9-1727-2009, 2009.

Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., and Heimann, M.: Global surface-ocean pCO2 and sea-Air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Ocean Science, 9, 193–216, https://doi.org/10.5194/os-9-193-2013, 2013.

Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A.: Combined simple biosphere/carnegie-ames-stanford approach terrestrial carbon cycle model, Journal of Geophysical Research: Biogeosciences, 113, 1–13, https://doi.org/10.1029/2007JG000603, 2008.

Scholze, M., Kaminski, T., Knorr, W., Vossbeck, M., Wu, M., Ferrazzoli, P., Kerry, Y., Mialon, A., Richaume, P., Rodriguez-Fernandez, N., Vittucci, C., Wigneron, J.-P., Mecklenburg, S., and Drusch, M.: Mean European carbon sink over 2010-15 estimated by simultaneous assimilation of atmospheric CO2, soil moisture and vegetation optical depth, 2019.

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, 1625–1644, https://doi.org/10.5194/bg-7-1625-2010, https://www.biogeosciences.net/7/1625/2010/, 2010.

Schulze, E. D., Luyssaert, S., Ciais, P., Freibauer, A., Janssens, I. A., Soussana, J. F., Smith, P., Grace, J., Levin, I., Thiruchittampalam, B., Heimann, M., Dolman, A. J., Valentini, R., Bousquet, P., Peylin, P., Peters, W., Rödenbeck, C., Etiope, G., Vuichard, N., Wattenbach, M., Nabuurs, G. J., Poussi, Z., Nieschulze, J., and Gash, J. H.: Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance, Nature Geoscience, 2, 842–850, https://doi.org/10.1038/ngeo686, 2009.

Smith, B., Wärlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014, 2014.

Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H.: The CO2 release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, Atmospheric Chemistry and Physics, 11, 6855–6870, https://doi.org/10.5194/acp-11-6855-2011, https://www.atmos-chem-phys.net/11/6855/2011/, 2011.

Still, C., Berry, J., Collatz, G., and Defries, R.: ISLSCP II C4 Vegetation Percentage, https://doi.org/10.3334/ORNLDAAC/932, http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds{_}id=932, 2009.

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., Schuster, U., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Johannessen, T., Olsen, A., Bellerby, R., Wong, C., Delille, B., Bates, N., and de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 554–577, https://doi.org/10.1016/J.DSR2.2008.12.009, https://www.sciencedirect.com/science/article/pii/S0967064508004311, 2009.

Thompson, R. L. and Stohl, a.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geoscientific Model Development Discussions, 7, 3751–3801, https://doi.org/10.5194/gmdd-7-3751-2014, http://www.geosci-model-dev-discuss.net/7/3751/2014/, 2014.

van der Laan-Luijkx, I. T., van der Velde, I. R., van der Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Zhang, H. F., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., and Peters, W.: The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001–2015, Geoscientific Model Development, 10, 2785–2800, https://doi.org/10.5194/gmd-10-2785-2017, https://www.geosci-model-dev.net/10/2785/2017/, 2017.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505–7514, https://doi.org/10.1002/2014WR015638.Received, 2014.

White, E. D., Rigby, M., Lunt, M. F., Smallman, T. L., Comyn-Platt, E., Manning, A. J., Ganesan, A. L., O'Doherty, S., Stavert, A. R., Stanley, K., Williams, M., Levy, P., Ramonet, M., Forster, G. L., Manning, A. C., and Palmer, P. I.: Quantifying the UK's carbon dioxide flux: an atmospheric inverse modelling approach using a regional measurement network, Atmospheric Chemistry and Physics, 19, 4345–4365, https://doi.org/10.5194/acp-19-4345-2019, https://www.atmos-chem-phys.net/19/4345/2019/, 2019.

Inversion systems (references) Inversion system	PYVAR-CHIMERE (Broquet et al., 2011; Fortems-Cheiney et al., 2019) PYVAR-CHIMERE (LSCE)	LUMIA (Monteil and Scholze, 2019) LUMIA (Lund University)	FLEXINVERT+ (Thompson and Stohl, 2014) FLEXINVERT (NILU)	CarboScope-Regional (Kountouris et al., 2018a) CarboScope-Regional (MPI-BGC-Jena)	CarbonTracker Europe (Peters et al., 2010; van der Laan-Luijkx et al., 2017) CarbonTracker Europe (WUR)	NAME-HB (White et al., 2019) NAME-HB (Uni. Bristol)
Institute Reference	LSCE Broquet et al. (2011); Fortems-Cheiney et al. (2019)	Lund University Monteil and Scholze (2019)	NILU Thompson and Stohl (2014)	MPI-BGC-Jena Kountouris et al. (2018a)	Wageningen Univ. Peters et al. (2010); van der Laan-Luijkx et al. (2017)	Univ. Bristol White et al. (2019)
Method	Variational	Variational	Variational	Variational	EnKF	MCMC
Transport model	CHIMERE (Eulerian)	FLEXPART (Lagrangian)	FLEXPART (Lagrangian)	STILT (Lagrangian)	TM5 (Eulerian)	NAME (Lagrangian)
Meteorological forcing	ECMWF ERA operational forecasts	ECMWF ERA-Interim reanalysis	ECMWF operational forecasts	Short term forecasts of ECMWF-IFS at 0.25° resolution ECMWF operational forecasts	ECMWF ERA-Interim reanalysis	UK Met Office's Unified Model (Cullen, 1993)
Background / Boundary condition	Prescribed At at domain edge from a CAMS LMDZ inversion	Prescribed At at obs. location from a TM5-4DVAR inversion	interpolation of NOAA data + transport of prescribed fluxes outside the EUROCOM domain	Prescribed At at obs. location from a global CarboScope inversion	None (global inversion)	Optimised at domain edge, from a MOZART simulation prior
transport and inversion domain	31.5°N to 74°N; 15.5°W to 35°E	33°N to 73°N; 15°W to 35°E	Global transport, inversion on a 30°-75°N, -15°-35°E domain	33°M to 73°N, 15°W to 35°E	Global, zoom over Europe (21°W-39°E, 12-66°N)	10.729°N to 79.057°N; 97.9°W to 39.38°E
Inversion spatial resolution	$0.5° \times 0.5°$	$0.5° \times 0.5°$	PFTs × countries	$0.5° \times 0.5°$	$1° \times 1°$ over Europe, $3° \times 2°$ globally	Large regions × PFTs
Inversion temporal resolution	6 hours	1 month	12 hours	3 hours	weekly	variable (max 1 day)
Prior estimate of NEE	ORCHIDEE	LPJ-GUESS	ORCHIDEE	VPRM	SiBCASA	ORCHIDEE
Correlation (spatial, temporal) scales of the prior uncertainty	200 km, 1 month	200 km, 1 month	No spatial correlation between PFT/-country regions, 1 month	100 km, 1 month	200 km with no correlation between different PFTs, 5 weeks	No correlations (large regions)
Ocean fluxes	part of the control vector (6 hour and 0.5° resolution), null prior	Prescribed (Rödenbeck et al., 2013)	Prescribed (Rödenbeck et al., 2013)	Prescribed (Mikaloff Fletcher et al., 2007)	Optimised (Jacobson et al., 2007)	Prescribed (Takahashi et al., 2009)
Observation selection[1]	12:00 to 18:00 below 1000 m.a.s.l. and 0:00 to 6:00 above	11:00 to 15:00 LT low altitude, 23:00 to 3:00 LT high altitude	12:00 to 16:00 LT below 1000 m.a.s.l. and 00:00 to 04:00 LT above	11:00 to 16:00 UTC low altitude and 23:00 to 4:00 UTC at mountain stations	11:00 to 15:00 LT low altitude, 23:00 to 3:00 LT at high altitude	based on transport model performance

Table 2. Overview of the inverse modelling systems and configuration of the inversions. [1] The classification of low/high altitude varies on a case-by-case basis in some systems, see Figure SI2 for a more complete information.

	Annual			Monthly			Daily			Hourly		
	prior	post.	reduct.	prior	post.	reduct.	prior	post.	reduct.	prior	post.	reduct.
LUMIA	1.57	1.07	31 %	2.79	2.04	27 %	4.78	4.16	13 %	4.76	4.11	14 %
PYVAR-CHIMERE	1.80	1.51	16 %	2.71	2.00	26 %	4.50	3.69	18 %	4.85	4.10	16 %
CTE	0.88	0.57	36 %	1.57	1.13	28 %	3.28	2.52	23 %	3.30	2.70	18 %
CarboScope-Regional	1.56	0.73	53 %	2.54	1.33	48 %	4.75	3.32	30 %	5.39	4.18	23 %
NAME-HB	1.32	0.76	42 %	2.97	1.50	49 %	4.67	2.65	43 %	4.77	2.81	41 %
FLEXINVERT	1.63	1.10	33 %	2.88	1.80	37 %	5.04	3.87	23 %	5.32	4.19	21 %
Median	1.56	0.92	34 %	2.75	1.65	33 %	4.71	3.51	23 %	4.81	4.10	20 %

Table 3. Prior and posterior RMS difference (ppm) and percentage RMS difference reduction between modelled and observed concentrations, averaged hourly to annually. The average is first performed site by site, and then the site values are averaged together, with weights proportional to the lenght of each time series.