

Response to all referee comments for ACP manuscript No.: acp-2017-273 [Peter Bergamaschi et al., Inverse modelling of European CH₄ emissions during 2006-2012 using different inverse models and reassessed atmospheric observations]

5 This document describes the detailed point-by-point response to all three referee comments

(1) *comments from Referees [in italics]*

(2) author's response [regular fonts]

(3) [author's changes in manuscript \[blue\]](#).

Furthermore, additional updates of the manuscript are described in section 2

10

1. Response of referee comments

1.1 Anonymous Referee #1

Major comments - "2.1 Wetland hypothesis"

15 *I do not find these arguments convincing. The arguments, as presented, are inconclusive at best. The region where we would expect the largest wetland emissions is Northern Europe, however in this region the inversions consistently point to a reduced seasonal cycle compared to WETCHIMP. The EU-28 seasonal cycle in WETCHIMP is ~10 Tg/yr which is roughly the same as the top-down seasonal cycle in their inversions. But, again, their inversion pointed to a decrease in the seasonal cycle in Northern Europe where the bulk of the wetland emissions should be. So why do we think this is due to wetlands? Because other sources are assumed to be atemporal? The*
20 *authors acknowledge that other sources could have seasonal cycles (e.g., manure emissions are temperature dependent, enteric fermentation could have a seasonal cycle due to variations in the herd size, etc).*

Although the WETCHIMP model ensemble estimates large CH₄ emissions for Northern Europe (1.9 (0.8-3.5) Tg CH₄ yr⁻¹ (mean, minimum, maximum); excluding Norway), this data set estimates significant wetland emissions also for western Europe (1.6 (0.4-3.1) Tg CH₄ yr⁻¹), eastern Europe (0.3 (0.03-0.9) Tg CH₄ yr⁻¹) and southern
25 Europe (0.6 (0.01-1.1) Tg CH₄ yr⁻¹). Excluding Northern Europe, the sum of the WETCHIMP CH₄ emissions for western, eastern, and southern Europe is 2.5 (0.4-5.1) Tg CH₄ yr⁻¹, corresponding to 12.5% (2.2%-25.6%) of the total reported anthropogenic CH₄ emissions for EU-28, which highlights the potential significant contribution of wetland emissions also for western / eastern / southern Europe.

30 While the inversions of TM5-4DVAR, TM5-CTE, TM3-STILT yield indeed a smaller seasonal cycle for Northern Europe compared to the mean of the WETCHIMP models (but similar amplitude for TM5-CTE), they derive significant seasonal cycles also for western / eastern / southern Europe, broadly consistent with the range of seasonal variations of the WETCHIMP ensemble. Our interpretation of this result is that indeed the spatial distribution of wetland emissions of the WETCHIMP ensemble (within Europe) is not fully consistent with the inversion results, but we consider the considerable derived seasonal variation for western / eastern / southern
35 Europe as indication that wetlands could contribute significantly also in these sub-regions.

This interpretation is indeed based on the assumption that anthropogenic CH₄ emissions have only very small seasonal variations. To our knowledge, only very few studies investigating the seasonal variations of the

anthropogenic emissions are available (and have been discussed in the discussion paper). Clearly further studies on this topic will be required.

We will emphasize more clearly in the revised paper the caveats of the hypothesis of significant wetland emissions.

5 We emphasize now more clearly the caveats in the abstract ("However, the contribution of natural sources remains rather uncertain, especially their regional distribution.") and the conclusions ("Furthermore, it needs to be emphasized that wetland inventories have large uncertainties and show large differences in the spatial distribution of CH₄ emissions.")

10 *There is little-to-no discussion of the background used for the region (see next comment), could errors in the background be driving this?*

The global models assimilate also global observations from the NOAA ESRL global cooperative air sampling network. The model simulations outside Europe have been further analyzed for TM5-4DVAR, showing in general very good agreement with observations at global background stations (similar as shown in previous papers, see
15 Bergamaschi et al. [2013], Figure S4a). Therefore, it seems unlikely, that errors in the background are driving the derived seasonal variations of European CH₄ emissions.

We have extended the analysis of the background (see also "2. Further updates of the manuscript"), and evaluated the quality of the background by comparing model simulations with the aircraft observations for events with very low simulated contribution (≤ 3 ppb) from European CH₄ emissions (new Fig. 14S). This
20 analysis suggests that the background calculated by TM5-4DVAR is relatively realistic, while the regional models NAME and STILT have a positive bias in the background. This is now further discussed in section 4.2 ("Evaluation of inverse models"). However, there is no indication for a seasonal bias of the inversions (see updated Figure 6 and discussion in section 4.2).

25 *There is no mention of the methane sink, is the OH correct? If OH were too low then you may have an artificially low seasonal cycle in the global simulations (which would, again, impact the background concentrations).*

The global models apply OH fields that were calibrated against methyl chloroform measurements [Patra et al, 2011; Bergamaschi et al., 2010; Houweling et al., 2014]. Since the global models assimilate global observations, potential deficiencies of the global OH fields are likely to be largely compensated by (artificial) increments of the
30 global fluxes. As mentioned above, e.g. TM5-4DVAR reproduces the measurements at global background stations very well (the performance of other global model at global sites were not further investigated in this study). The impact of different global OH fields on derived European CH₄ emissions has been investigated by Bergamaschi et al. [2010], which showed only a very small impact.

The sensitivity experiment for TM5-4DVAR using different global OH fields [Bergamaschi et al., 2010] has been
35 included in the description of TM5-4DVAR (supplementary material, section 1.1).

It's unclear to this reviewer why the authors did not just perform an inversion with atemporal emissions and compare the posterior seasonality to the prior seasonality. This would show how much of this derived

seasonality comes from the data instead of the prior. It would allow them to say which regions have significant seasonal cycles. The authors could have achieved much of this by looking at the seasonal cycles in their case with homogenous prior emissions.

Also the inversion results from inversion S3 (which was performed without using detailed bottom-up inventories as 'a priori'), show significant seasonal cycles in derived emissions. This confirms that the derived seasonal cycle is driven by the observations, and not by the a priori emissions. This was not mentioned in the discussion paper but will be included in the revised paper.

We have included a new figure (Fig. 5S) with the mean seasonal cycles for all inversions (including S3) and added a short discussion in section 4.1.

10

Major comments - "2.2 Poor description of methods makes it difficult to gain any insight"

The description of the various inversion systems is poor. There is a single paragraph in the main text describing the inversions. There is no mathematical description of the inversions. This is quite surprising since, at it's core, this is an inversion paper.

15 The inverse modelling system are described in the supplementary material (SM), section 1 "Atmospheric models" (summarizing the main elements of each system). Furthermore, all seven inverse models are described comprehensively in separate specific papers (see references in the SM).

For most models used in this study only smaller updates were applied (compared to previously published applications). Therefore, we had chosen to put the model descriptions in the SM (and would prefer to keep this in the SM also in the revised version).

However, we will somewhat extend the general description of the models in the main paper (section 3.2 "Atmospheric models") in the revised version.

We have included the applied boundary conditions (background) in Table 3 and included the information about the optimization of the background in the text. Some further details were added (or updated) in the supplementary material.

At the bare minimum, the author's should state the assumptions for their inversions (e.g., Gaussian errors?).

Most inverse modelling systems applied in this study use Gaussian probability density functions for the uncertainties of the emissions (in case of TM5-4DVAR a 'semi lognormal' pdf is used; see SM section 1.1).

30 We will add the applied pdfs in the model description for those models where this information is missing in the discussion paper.

We have added the information about the applied probability density functions (pdfs) in the detailed model description in the supplementary material.

35 *There is additional text in the supplement (~1 paragraph per model) but it is difficult to synthesize the models. Some of the models are regional but it's not clear where the boundary conditions are coming from.*

It is clearly stated in section 3.2 ("Atmospheric models"; page 5, lines 23-25) where the boundary conditions are coming from:

5 "The regional models use boundary conditions from inversions of the global models (STILT from TM3, COMET from TM5, CHIMERE from LMDZ, or estimate the boundary conditions in the inversions (NAME), using baseline observations at Mace Head as 'a priori' estimates."

Furthermore, the boundary conditions are described also in the SM for all regional models (STILT, NAME, CHIMERE, COMET).

Although already described in the text, we have included now the boundary conditions (background CH₄ mole fractions) also in Table 3.

10

Some of the models are estimating the covariance matrices from the data, some are not.

We assume that the reviewer refers here to the observation covariance matrix. The uncertainties of the observations (diagonal elements of the covariance matrices) include both the measurement error and the model error. Most models use the "working standard repeatability" (see section 2 of main paper) as observation error. However the estimates of the model errors are very different in the different inverse modelling systems (and generally based on simplified assumptions). For most models the assumed uncertainties of the observations is described in SM section 1 - for those models where this information has been missing (CHIMERE, COMET), it will be added.

20 The assumed uncertainties for the observations (including estimates of model errors) are now described for all models in the supplementary material (section 1).

25 *It is extremely difficult for the reader to understand why these inversions are performing differently. For example, it seems that the boundary conditions are coming from global models in the case of some regional models, how independent are these different inversion systems (especially the global/regional ones)? Are we comparing apples to apples?*

The global models providing the boundary conditions for the regional models are generally largely independent from the regional models (apart from the fact that the different models may have some features in common, e.g. use of same or similar meteo data sets).

No change in the manuscript regarding this point.

30

How much of the differences are due to assumptions vs transport vs something else? It's extremely difficult to understand the differences without clearly laying out the key differences between the models.

35 Given the very high complexity of the different inverse modelling systems, it is indeed very difficult to understand where the differences in the derived emissions are coming from. But this is actually not the goal of this study (and would require further specific modelling experiments). The objective of this study is to use the model ensemble to provide more realistic overall uncertainty estimates (from the range of the inverse models) and to evaluate the model performance by validation against independent observations.

The background mole fractions have been identified as one major parameter which can lead to biases in the derived emissions. This is now discussed in more detail in the revised version (section 4.2, see also "2. Further updates of the manuscript").

5 *I would point the authors to the Henne et al. (2016) paper as an example of a paper that does a good job of explicitly highlighting the differences between their inversion systems and allows the readers to actually gain insight from the ensemble of inversions. Table 2 from Henne et al. (2016) is a particularly good example of how one can demonstrate the major differences between inversion frameworks.*

The fundamental difference between the study of Henne et al. (2016) and our study is that Henne et al. use one
10 single inverse modelling system, varying various input parameters / settings of this system as compiled in their Table 2. In contrast, our study uses very different inverse modelling systems, which makes it inherently more difficult to highlight the differences between the systems (which are largely independent systems and which differ in many aspects). Important parameters (model resolution, meteorology, a priori emission inventories, applied station sets are compiled in Tables 1, 2, and 3. We will include also the applied baselines for the regional
15 models in Table 3.

We have included the applied boundary conditions (background) in Table 3.

Also, the phrase "no a priori" is, almost certainly, using incorrect terminology. The posterior probability is
20 *proportional to the product of the likelihood and the prior probability: Posterior probability / Likelihood × Prior probability. Using a homogenous distribution of emissions is still including a prior, it just isn't based on a bottom-up inventory. To actually use "no a priori" would be "Maximum Likelihood Estimation" where one simply finds the parameters that maximize the likelihood term*

In section 3.1 we have described S3 as:

"Inversion S3 was performed without using detailed bottom-up inventories as 'a priori', in order to analyse the
25 constraints of observed atmospheric CH4 on emissions independent of 'a priori' information (using a homogeneous distribution of emissions over land and over the ocean, respectively, as starting point for the inversions in a similar manner as in Bergamaschi et al. [2015])."

The short notion "no a priori" has been only used in Table 2. We will add a footnote in this table to refer the reader to the above description in section 3.1

30 "no a priori" has been changed to "no detailed a priori inventory" and footnote has been added.

Major comments - "2.3 'Novel' Bias method"

This "novel" bias method is, essentially, what an inversion already does. . . They are just plotting the model-data mismatch averaged over different parts of the atmosphere. This is hardly a "novel approach".

35 *(mathematical derivation not repeated here)*

From this, it's quite easy to see how $c_{obs} - c_{mod} = \Delta c_{obs} - \Delta c_{mod}$. So, as I stated above, all the authors have done is plot the model-data mismatch ($c_{obs} - c_{mod}$) averaged over two parts of the atmosphere. It does not strike this reviewer as particularly "novel".

5 We do not agree with the statement of the reviewer that our approach to estimate the bias in the derived emissions is "essentially, what an inversion already does", since we look at independent observations that were not used in the inversion - which is a common method to validate inverse models (see e.g. Michalak et al., [2016]). Commonly, however, such analyses are performed to diagnose qualitatively, if the inverse models have biases.

10 The novel aspect of our method is that we use the baseline in order to extract the signal which comes from the European emissions. Integrating the enhancement of the model simulations compared to the background over the entire boundary layer or the entire column of the lower troposphere (and comparison with the corresponding observed CH₄ enhancement) provides a measure of the total CH₄ emitted by European emission. The ratio of the simulated vs. observed integrated enhancements provides a first order estimate of the relative bias in the model emissions.

15 As explained in section 4.2, the validation against independent aircraft profiles is very important, since the inverse models assimilate only surface observation. Therefore, potential errors in the vertical mixing of the models can introduce significant biases in the derived emission.

Independent from the the comment of the reviewer (with which we do not agree), we have updated the analysis of the model biases (see also "2. Further updates of the manuscript").

20

There are novel approaches that attempt to account for systematic errors in inversions in a rigorous manner. Weak-Constraint 4D-Var (Tremolet, 2006) and Hierarchical Bayesian inference (see Ganesan et al., 2014 and references therein) are two good examples of this.

25 We agree that the "Hierarchical Bayesian inference" is an interesting approach to provide more realistic uncertainty estimates for individual models (i.e. estimates within the individual inverse modelling systems, corresponding to the error bars in our Figure 3). Nevertheless, validation against independent observations will remain indispensable as independent evaluation of the inverse models.

Also the mentioned "Weak-Constraint 4D-Var" is certainly a very interesting technique - but to our knowledge so far only applied in some cases for data assimilations, but not in inverse modelling systems.

30 **No change in the manuscript regarding this point.**

1.2 Anonymous Referee #2

35 *This study presents a multi-model top-down assessment of European methane emissions using the European measurements network. As mentioned, these measurements are performed with the aim to verify bottom-up inventories reported to the UNFCCC. As such this study can be seen as an assessment of where we are in this process, extending the number of years that were reported in a previous assessment. The results highlight the importance of taking into account natural emissions of methane. Combining natural and anthropogenic emissions the reported total for EU-28 ends up in close agreement with the inventories. The study is a useful*

reference, and as such it makes a good contribution to ACP. However, as will be explained below, it also misses some useful opportunities to add value to the previous assessment with the potential to substantially increase the significance of this work. Having gone through the major effort of organizing this model inter-comparison already, the points listed under 'discussion' should receive serious consideration in my opinion.

5 We thank the reviewer for the very positive overall evaluation of our study.

DISCUSSION

In the context of emission verification, testing the EU-28 total is relevant, however, the network probably resolves additional independent pieces of information. The question is how many, and what this means for the capacity of the European network to resolve country scale emissions. This applies not only to average emissions, but also to their trends. One may argue that in the framework of the COP21 climate agreement the ability to evaluate trends is even more important than the average. Looking at the results that are presented, information about trends is clearly visible in the time series, but to my surprise it is not discussed at all. Even if it turns out that these trends are not significant it is useful to quantify and discuss how far we are from this target. It is a bit surprising that the multi-year time dimension, which is the new element of this study compared to the previous one, is left unexplored.

10
15

The anthropogenic CH₄ emissions reported to UNFCCC have indeed decreased between 2006 and 2012 by 11.6%. The models show rather smaller trends (which are in most cases indeed probably not significant). An evaluation of the uncertainties of the trends, however, is very difficult, since this requires information about the error correlations between subsequent years (which is not available). We will include a short discussion of the trends in the revised version.

20

[We have included a short discussion of the CH₄ trends at the end of section 4.1 \(and included the new Figure 7S in the supplementary material\).](#)

A useful attempt is made to assess biases in transport models using vertical profile measurements. However, what is missing is the link between these biases and the inverted emissions. It is mentioned that those models that overestimate PBL average CH₄ should overestimate emissions. In fact, all the ingredients are available to quantify this link and assess the impact of transport biases on emissions. It raises the question why this is not done. Is it an important factor explaining the range of emission that are found or not?

25

Following the suggestion of the reviewer we analyzed the relationship between the estimated relative bias (based on the enhancement compared to the background integrated over the boundary layer) and the model emissions in the area around the regular aircraft profiles sites. The analysis showed significant correlations between model emissions and estimated model bias. We will include this analysis in the revised version.

30

[We have included now an analysis of the correlation between the derived relative bias and the regional model emissions around the aircraft sites \(new Figure 16S in the supplementary material, and discussion at the end of section 4.2\).](#)

35

SPECIFIC COMMENTS

page 4, line 6: Which targets are set by the quality control mentioned here? Are they met?

No specific threshold values have been set. The typical range for the "working standard repeatability" is ~1-4 ppb. Since this "working standard repeatability" is used by the inverse models, measurements with higher "working standard repeatability" are weighted less in the inversion.

No change in the manuscript regarding this point.

5

page 5, line 16: Using constant a priori flux uncertainties also? How do these emissions / uncertainties relate to those of the other scenarios?

10 For inversion S3 very large uncertainties of the homogeneous a priori fluxes were assumed (ranging between 500% and 600% per grid-cell and month; see model description in the supplementary material) in order to give the inversion enough degree of freedom to retrieve regional emission hot spots (which have much higher emissions than the applied homogeneous a priori fluxes). In contrast, the assumed uncertainties per grid cell and months are much smaller for the other scenarios (typically 100%).

We have added a short reference to the supplementary material, where the specific settings for inversion S3 are described in more details for the individual models.

15

page 5, line 24: Do the regional models (apart from NAME) prescribe boundary conditions, or allow further optimization?

20 Apart from NAME, the boundary conditions are further optimized also in CHIMERE, while the other regional models used prescribed boundary conditions. These boundary conditions were derived from optimized concentrations of global inversion systems (STILT: from TM3, COMET: from TM5-4DVAR, CHIMERE: from LMDZ).

The information about the optimization of the boundary conditions been added in section 3.2.

page 8, line 10-15: It would be good to mention some typical numbers here for the bottom up and top down derived seasonal amplitudes (it is not so clear to see from figure 4)

25 We will add the numbers of the derived seasonal amplitudes in the revised version.

Instead of including the numbers, we have added now the new Figure 5S which visualizes the mean seasonal cycles for all scenarios.

page 8, line 30-35: How about the seasonality in the energy sector? (domestic heating etc.)

30 No or only small seasonal variations were found in the limited number of studies investigating natural gas distribution system [Wennberg et al., 2012; McKain et al., 2014]. Wong et al. [2016] argued that "the natural gas distribution pipeline system is pressure-regulated at several points, and leakage should be independent of consumption to first order", but that natural gas storage facilities may have seasonally varying leakage rates, depending on energy demands.

35 Reference to [McKain et al., 2015] has been added.

page 9, line 7: *The difference between the observed vs simulated amplitude of variability (as used in Taylor diagrams for instance) provides a piece of information that is more independent from correlation as the RMS that is used here.*

- 5 Following the suggestion of the reviewer we will analyze also the difference between the observed vs simulated amplitude of variability.

We have extended Figure 8S (previous Figure 6S), including an additional panel with the ratio between modelled and observed standard deviations.

10 **1.3 Anonymous Referee #3**

Summary/General comments:

- The manuscript presents ‘top-down’ optimized methane emissions for Europe for the 2006-2012 time period. A new, harmonized 18 site-monitoring network is used with seven inverse models and four experiments. Optimized emissions are reported (and are overall consistent between top-down and bottom-up), biases are assessed using aircraft data, and the inference of a non-negligible wetland source is intimated. Overall it is interesting and important work to pursue. It is not easy to use this many different model/inverse approaches to one regional question, and this can potentially provide substantially more information and understanding for how to best quantify fluxes with atmospheric observations. This paper is well-placed in ACP.*
- 15

We thank the reviewer for the very positive overall evaluation of our study.

- 20 *However, there are a couple important gaps that need to be addressed before I can recommend publication. Most importantly, the description of different models and inverse methods is somewhat lacking, this should be a central element of this work, and this needs to be improved before I can recommend publication*

- The inverse modelling system are described in the supplementary material, summarizing the main elements of each system. Furthermore, all seven inverse models are described comprehensively in separate specific papers.
- 25 Nevertheless, we will include some further details in the description of the models.

Some further details (e.g., about the a priori probability density functions and assumed observations errors) were added / complemented in the supplementary materials.

30 **Major comments:**

- Models/Inverse methods: There is limited discussion of the different models, and specifically, of the inverse methodology being employed by each model. I understand much of this is referenced to various previous publications, and the supplement does go through each model independently, but it is important for the reader to see more comparative details in this manuscript to be able to understand the differences between models/inversions and possible nuanced causes. A succinct but clear description in its own section of the different inverse approaches used and the subtle “expert-user” choices made to define the inversion would be*
- 35

essential. For example the prior uncertainties and correlations lengths, which are defined differently in the different inversions, could be rather impactful on the results. How were these different priors chosen, and how important is this choice? The authors have conducted multiple experiments – they need to better convey to the reader the differences between the inversions and experiments so we can better assess the meaning of similar/different results. In many ways this could be one of the biggest contributions of this paper.

The specific settings of the individual inverse models are indeed largely "expert-user" choices. For many models the sensitivity of derived emissions on these settings were investigated in more detail (and described in the papers of the individual inverse modelling systems). E.g. for TM5-4DVAR different spatial correlation lengths (between 100 and 300 km) were analyzed [Bergamaschi et al., 2010], showing an overall only very small impact on the derived emissions. In the present study, the philosophy was to prescribe only the basic settings for the inversions, such as a priori emission inventories, observational data sets, and inversion time period.

The main objective of this study is to use the model ensemble to provide more realistic overall uncertainty estimates (from the range of the inverse models), rather than investigating the sensitivity of individual inversion results on specific settings of the individual models. Given the large fundamental differences of the different inverse models (e.g. grid based inversion in TM5-4DVAR compared to optimization of larger pre-defined larger regions and different land-ecosystem types in the TM5-CTE (ensemble Kalman filter), it would not be possible to apply fully consistent settings in the different models.

The different inversions of this study investigate the impact of the different set of stations and the use of 'a priori' information. The different settings for the 4 inversion experiments are summarized in Table 2 and described in section 3.1.

We have included a reference [Bergamaschi et al., 2010] for the sensitivity of derived CH₄ emissions on spatial correlations (and OH fields) for TM5-4DVAR (section 1.1 supplementary material).

Sensitivity of network to domain: Western Europe has the highest density of observation sites, and measurement density (and sensitivity to emissions) falls off rapidly in other regions of Europe. Given this, how appropriate is it to lump the entirety of the domain together? I'd like to see a little more discussion of the sensitivity of the network and therefore dependence of prior/assumptions in some of the domains. Another way to consider this question is how many regions can the network distinguish, and how do these regions compare with geopolitical domains? This impacts my next point.

Indeed the available stations are not evenly distributed across Europe, and the observational coverage is relatively sparse in southern Europe and Scandinavia. The fact that inversion S3 yields similar estimates for the emissions of Northern and Southern Europe (for most models; however lower estimates for NAME) compared to the other inversions (which include the detailed emission inventories as a priori) suggests that nevertheless the limited observations provide also some constraints on the total emissions from these sub-regions. We did not perform specific sensitivity experiments in this study, but we will include some more discussion of the network coverage (and the limited observational constraints in southern Europe) in the revised version.

We have included a short discussion of the network coverage in section 4.1 ("Despite the significantly larger number of European monitoring stations in the present study, however, we emphasize that the available stations do not very well cover the whole EU-28 area. Consequently, the emissions especially from Southern Europe remain poorly constrained.").

Importance of wetlands: I'm not sure if from this analysis alone the authors can conclude substantial wetland source are or are not required to match observations. The largest prior wetland estimate (and seasonality) is in Northern Europe, where there are few observation points and the inverted seasonality is actually smaller than WETCHIMP models. When aggregating all of Europe together, it would appear the added emissions and seasonality from WETCHIMP is helpful in bringing bottom-up and top-down closer together – but given this point of spatial/seasonal errors in the Northern Europe domain I'm not sure this overall improvement is indicative of a better representation or coincidence where the inversion finds large seasonality in other regions of Europe where WETCHIMP models do not expect significant wetland sources. I would think the authors should tone down the statement of wetlands importance in the abstract, and also would like to see further defense of the seasonality signal observed and attribution that it must be wetlands.

Indeed the spatial distribution of wetlands in Europe in the WETCHIMP ensemble is not fully consistent with the results from the inverse models and most inverse models (except TM5-CTE) show a smaller amplitude of the seasonal variations than the mean of the WETCHIMP ensemble. Nevertheless, the WETCHIMP ensemble estimates significant wetland emissions also in western / southern / eastern Europe (2.5 (0.4-5.1) Tg CH₄ yr⁻¹; see also our reply to reviewer #1) and the seasonal cycles derived by 4 models (TM5-4DVAR, TM5-CTE, TM3-STILT, and LMDZ) are broadly consistent with the range of seasonal variations of the WETCHIMP ensemble (although indeed the amplitude of the mean seasonal cycles of WETCHIMP are smaller for western / southern / eastern Europe). We fully agree that the uncertainties of wetland emissions remain very high (as directly evident from the very different spatial distributions of the individual WETCHIMP inventories (see Figure 4S). This has been mentioned in the text, but will be further emphasized in the revised version.

Also inversion S3 (which was performed without using detailed bottom-up inventories as 'a priori'), shows significant seasonal cycles in derived emissions (for EU-28 and all European subregions (but relatively small in southern Europe)), which confirms that the derived seasonal cycles are driven by the observations (and not by the a priori emissions).

We also agree that uncertainties remain in the attribution of the seasonal cycle to wetlands, since some anthropogenic sources may also exhibit some (smaller) seasonal variations (see also our reply to reviewer #1). We will emphasize the caveats of our wetland hypothesis more clearly in the revised version (including the abstract).

We have included a new figure (Figure 5S) with the mean seasonal cycles derived in the different inversions (including S3). Furthermore, we emphasize now more clearly the caveats of our wetland hypothesis in the abstract ("However, the contribution of natural sources remains rather uncertain, especially their regional distribution.") and the conclusions ("Furthermore, it needs to be emphasized that wetland inventories have larger uncertainties and show large differences in the spatial distribution of CH₄ emissions.").

2. Further updates of the manuscript

We have further refined our method to evaluate the model biases (section 4.2):

(1) In order to reduce the impact of potential errors of the background on the calculation of the relative bias (section 4.2) we have increased the threshold of $\Delta_{\text{COBS, BL}}$ and $\Delta_{\text{COBS, COL}}$ from 10 ppb to 20 ppb.

(2) For the evaluation of the enhancements of the measurements vs. background we use for NAME and TM3-STILT now the background evaluated by TM5-4DVAR for the NAME and TM3-STILT domains. For the evaluation

of the simulated CH₄ enhancements of NAME and TM3-STILT, we use now the actual background used in NAME and TM3-STILT (instead of the TM5-4DVAR background used previously). The relative biases derived with these updates for NAME and TM3-STILT are considered much more accurate, and reveal a significant negative regional bias for these two models. The NAME and TM3-STILT backgrounds are evaluated and discussed now in some more detail in the paper (new Figures 14S and 15S).

Furthermore, some updates of the LMDZ inversions were included in the revised version:

- for LMDZ inversion S1 uncertainty estimates have been included.

- LMDZ inversion S3 has been updated (use a priori uncertainty of 600% (instead of 200% used previously)). As a result also mean total CH₄ emissions for the EU-28 have slightly changed (from 26.7 to 26.8 Tg CH₄ yr⁻¹)

In addition, we updated the NOAA AGGI and recent atmospheric CH₄ in the introduction.

Updates of Figures:

Figure 3: Include uncertainty estimates for LMDZ inversion S1. Update LMDZ inversion S3.

Figure 4: Include uncertainty estimates for LMDZ inversion S1.

Figure 5: Include also background CH₄ used in NAME (based on MHD data) and TM3-STILT (based on TM3)

Figure 6: Update evaluation of relative bias using threshold of $\Delta_{\text{COBS, BL}}$ of 20 ppb (instead of 10 ppb). Updated evaluation of model enhancements / relative bias for NAME and TM3-STILT. Linear fits have been removed (as they were not discussed in the paper).

Figure 7: Update evaluation of relative bias using threshold of $\Delta_{\text{COBS, BL}}$ and $\Delta_{\text{COBS, COL}}$ of 20 ppb (instead of 10 ppb). Updated evaluation of model enhancements / relative bias for NAME and TM3-STILT.

Supplementary material

Figure 3S: update S3 inversion LMDZ (use a priori uncertainty of 600% (instead of 200% used previously)).

Figure 5S: new figure with mean seasonal cycles for all scenarios (as suggested by reviewer 1 and 2)

Figure 6S (previous Figure 5S): include uncertainty estimates for LMDZ S4

Figure 7S: new figure with analysis of trends (as suggested by reviewer 2)

Figure 8S (previous Figure 6S): include also analysis of ratio of model standard deviation to observed standard deviation (as suggested by reviewer 2)

Figure 9-12S (previous Figure 7-10S): Update evaluation of enhancements vs background using thresholds of $\Delta_{\text{COBS, BL}} = 20$ ppb and $\Delta_{\text{COBS, COL}} = 20$ ppb (instead of 10 ppb). The integrated enhancements of the measurements vs background (evaluated by TM5-4DVAR) are evaluated now separately for the TM5 zoom domain and the NAME and STILT model domains. Furthermore, for the NAME and TM3-STILT model, the integrated model enhancement is now evaluated using the NAME background (based on MHD baseline data) and TM3 background, respectively.

Figure 13S (previous Figure 11S): Update evaluation of relative bias using threshold of $\Delta_{\text{COBS, COL}}$ of 20 ppb (instead of 10 ppb). Updated evaluation of model enhancements / relative bias for NAME and TM3-STILT. Linear fits have been removed (as they were not discussed in the paper).

5 **New Figure 14S:** Evaluation of CH₄ background for TM5-4DVAR, NAME, and TM3-STILT, comparing model simulations with the aircraft observations for events with very low simulated contribution (≤ 3 ppb) from European CH₄ emissions.

New Figure 15S: Background CH₄ at European monitoring stations for TM5-4DVAR, NAME, and TM3-STILT.

New Figure 16S: Correlation between relative bias and regional model emissions around the aircraft sites (as suggested by reviewer #2).

10

References

- Bergamaschi, P., et al., Atmospheric CH₄ in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, *J Geophys Res-Atmos*, 118(13), 7350-7369, doi: 10.1002/jgrd.50480, 2013.
- 15 Bergamaschi, P., et al., Inverse modeling of European CH₄ emissions 2001-2006, *J. Geophys. Res.*, 115(D22309), doi:10.1029/2010JD014180, 2010.
- Houweling, S., et al., A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, *Atmos. Chem. Phys.*, 14, 3991-4012, doi: 10.5194/acp-14-3991-2014, 2014.
- McKain, K., et al., Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts, *PNAS*, 112, 1941-1946, 2015.
- 20 Michalak, A. M., Randazzo, N. A., and Chevallier, F.: Diagnostic methods for atmospheric inversions of long-lived greenhouse gases, *Atmos. Chem. Phys.*, 17, 7405-7421, <https://doi.org/10.5194/acp-17-7405-2017>, 2017.
- Patra, P. K., et al., TransCom model simulations of CH₄ and related species: linking transport, surface flux and chemical loss with CH₄ variability in the troposphere and lower stratosphere, *Atmos. Chem. Phys.*, 11,
- 25 12813-12837, doi: 10.5194/acp-11-12813-2011, 2011.
- Wennberg, P. O., et al., On the Sources of Methane to the Los Angeles Atmosphere, *Environmental Science & Technology*, 40 46(17), 9282-9289, doi: 10.1021/es301138y, 2012.
- Wong, C. K., T. J. Pongetti, T. Oda, P. Rao, K. R. Gurney, S. Newman, R. M. Duren, C. E. Miller, Y. L. Yung, and S. P. Sander, Monthly trends of methane emissions in Los Angeles from 2011 to 2015 inferred by CLARS-FTS
- 30 observations, *Atmos. Chem. Phys.*, 16(20), 13121-13130, doi: 10.5194/acp-16-13121-2016, 2016.

Inverse modelling of European CH₄ emissions during 2006-2012 using different inverse models and reassessed atmospheric observations

Peter Bergamaschi¹, Ute Karstens^{2,3}, Alistair J. Manning⁴, Marielle Saunois⁵, Aki Tsuruta⁶, Antoine Berchet^{5,7}, Alexander T. Vermeulen^{3,8}, Tim Arnold^{4,9,10}, Greet Janssens-Maenhout¹, Samuel Hammer¹¹,
5 Ingeborg Levin¹¹, Martina Schmidt¹¹, Michel Ramonet⁵, Morgan Lopez⁵, Jost Lavric², Tuula Aalto⁶,
Huilin Chen^{12,13}, Dietrich G. Feist², Christoph Gerbig², László Haszpra^{14,15}, Ove Hermansen¹⁶, Giovanni
Manca¹, John Moncrieff¹⁰, Frank Meinhardt¹⁷, Jaroslaw Necki¹⁸, Michal Galkowski¹⁸, Simon
O'Doherty¹⁹, Nina Paramonova²⁰, Hubertus A. Scheeren¹², Martin Steinbacher⁷, and Ed Dlugokencky²¹

¹ European Commission Joint Research Centre, Ispra (Va), Italy

10 ² Max Planck Institute for Biogeochemistry, Jena, Germany

³ ICOS Carbon Portal, ICOS ERIC, University of Lund, Sweden

⁴ Met Office Exeter, Devon, UK

⁵ Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), CEA-CNRS-UVSQ, Université Paris-Saclay,
F-91191 Gif-sur-Yvette, France

15 ⁶ Finnish Meteorological Institute (FMI), Helsinki, Finland

⁷ Swiss Federal Laboratories for Materials Science and Technology (Empa), Dübendorf, Switzerland

⁸ Energy research Centre of the Netherlands (ECN), Petten, Netherlands

⁹ National Physical Laboratory, Teddington, Middlesex, TW11 0LW, UK

¹⁰ School of GeoSciences, The University of Edinburgh, Edinburgh, EH9 3FF, UK

20 ¹¹ Institut für Umweltphysik, Heidelberg University, Germany

¹² Center for Isotope Research (CIO), University of Groningen, Netherlands

¹³ Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA

¹⁴ Hungarian Meteorological Service, Budapest, Hungary

¹⁵ Research Centre for Astronomy and Earth Sciences, Geodetic and Geophysical Institute, Sopron, Hungary

25 ¹⁶ Norwegian Institute for Air Research (NILU), Norway

¹⁷ Umweltbundesamt, Messstelle Schauinsland, Kirchzarten, Germany

¹⁸ AGH University of Science and Technology, Krakow, Poland

¹⁹ Atmospheric Chemistry Research Group, University of Bristol, Bristol, UK

²⁰ Voeikov Main Geophysical Observatory, St. Petersburg, Russia

30 ²¹ NOAA Earth System Research Laboratory, Global Monitoring Division, Boulder, CO, USA

Correspondence to: Peter Bergamaschi (peter.bergamaschi@ec.europa.eu)

35

40

Abstract. We present inverse modelling ('top-down') estimates of European methane (CH₄) emissions for 2006-2012 based on a new quality-controlled and harmonized in-situ data set from 18 European atmospheric monitoring stations. We applied an ensemble of seven inverse models and performed four inversion experiments, investigating the impact of different sets of stations and the use of 'a priori' information on emissions.

5 The inverse models infer total CH₄ emissions of ~~26.7-8~~ (20.2-29.7) Tg CH₄ yr⁻¹ (mean, 10th and 90th percentiles from all inversions) for the EU-28 for 2006-2012 from the four inversion experiments. For comparison, total anthropogenic CH₄ emissions reported to UNFCCC ('bottom-up', based on statistical data and emissions factors) amount to only 21.3 Tg CH₄ yr⁻¹ (2006) to 18.8 Tg CH₄ yr⁻¹ (2012). A potential explanation for the higher range of 'top-down' estimates compared to 'bottom-up' inventories could be the contribution from natural sources, such as peatlands, wetlands, and wet soils. Based on seven
10 different wetland inventories from the "Wetland and Wetland CH₄ Inter-comparison of Models Project" (WETCHIMP) total wetland emissions of 4.3 (2.3-8.2) CH₄ yr⁻¹ from EU-28 are estimated. The hypothesis of significant natural emissions is supported by the finding that several inverse models yield significant seasonal cycles of derived CH₄ emissions with maxima in summer, while anthropogenic CH₄ emissions are assumed to have much lower seasonal variability. Taking into account the wetland emissions from the WETCHIMP ensemble, the top-down estimates are broadly consistent with the sum of anthropogenic and natural bottom-up inventories. However, the contribution of natural sources remains rather uncertain, especially their regional distribution.

15 Furthermore, we investigate potential biases in the inverse models by comparison with regular aircraft profiles at four European sites and with vertical profiles obtained during the "Infrastructure for Measurement of the European Carbon Cycle (IMECC)" aircraft campaign. We present a novel approach to estimate the biases in the derived emissions, based on the comparison of
20 simulated and measured enhancements of CH₄ compared to the background, integrated over the entire boundary layer and over the lower troposphere. The estimated This analysis identifies average regional biases range between -40% and 20% for several models at the aircraft profile sites in France, Hungary and Poland.

1 Introduction

25 Atmospheric methane (CH₄) is the second most important long-lived anthropogenic greenhouse gas (GHG), after carbon dioxide (CO₂), and contributed ~17% to the direct anthropogenic radiative forcing of all long-lived GHGs in ~~2015~~2016, relative to 1750 (NOAA Annual Greenhouse Gas Index (AGGI) [*Butler and Montzka*, 2017]). The globally averaged CH₄ mole fraction reached a new high of ~~1842.8~~~~1834.0~~ ± 0.5 ppb in ~~2015~~2016 (global average from marine surface sites [*Dlugokencky*, 2017]), more than 2.5 times the pre-industrial level [*WMO*, 2016b]. The increase in atmospheric CH₄ has been monitored by direct
30 atmospheric measurements since the late 1970s [*Blake and Rowland*, 1988; *Cunnold et al.*, 2002; *Dlugokencky et al.*, 1994; *Dlugokencky et al.*, 2011]. Atmospheric growth rates were large in the 1980s, decreased in the 1990s and were close to zero during 1999-2006. Since 2007, atmospheric CH₄ increased again significantly [*Dlugokencky et al.*, 2009; *Nisbet et al.*, 2014; *Rigby et al.*, 2008], at an average growth rate of 5.7 ± 1.1 ppb yr⁻¹ during 2007-2013, and at a further increased rate of ~~11.2~~
~~2.1~~~~10.1~~ ± 2.3 ppb yr⁻¹ during 2014-~~2015~~2016 [*Dlugokencky*, 2017].

35 While the global net balance (global sources minus global sinks) of CH₄ is well defined by the atmospheric measurements of in-situ CH₄ mole fractions at global background stations, the attribution of the observed spatial and temporal variability to specific sources and regions remains very challenging [*Houweling et al.*, 2017; *Kirschke et al.*, 2013; *Saunoy et al.*, 2016]. Global inverse models are widely used to estimate emissions of CH₄ at global/continental scale, using mainly high-accuracy surface measurements at remote stations (e.g. [*Bergamaschi et al.*, 2013; *Bousquet et al.*, 2006; *Mikaloff Fletcher et al.*, 2004a; b; *Saunoy et al.*, 2016]). In addition, satellite retrievals of GHGs have also been used in a number of studies. In particular,
40 near-IR retrievals from SCIAMACHY and GOSAT providing column average mole fractions (XCH₄) have been demonstrated

to provide additional information on the emissions at regional scales [Alexe et al., 2015; Bergamaschi et al., 2009; Wecht et al., 2014]. However, current satellite retrievals may still have biases and their use in atmospheric models is at present limited by the shortcomings of models in realistically simulating the stratosphere, especially at higher latitudes [Alexe et al., 2015; Locatelli et al., 2015]. Furthermore, integration over the entire column implies that the signal from the CH₄ variability in the planetary boundary layer (which is directly related to the regional emissions) is reduced in the retrieved XCH₄.

In contrast, in-situ measurements at regional surface monitoring stations can directly monitor the atmospheric mole fractions within the boundary layer, providing strong constraints on regional emissions. Such regional monitoring stations have been set up in the last years especially in the United States [Andrews et al., 2014] and Europe (e.g., [Levin et al., 1999; Lopez et al., 2015; Popa et al., 2010; Schmidt et al., 2014; Vermeulen et al., 2011]). The measurements from these stations were used in a number of inverse modelling studies to estimate emissions at regional and national scales [Bergamaschi et al., 2010; Bergamaschi et al., 2015; Ganesan et al., 2015; Henne et al., 2016; Kort et al., 2008; Manning et al., 2011; Miller et al., 2013]. A specific objective of these studies is the verification of 'bottom-up' emission inventories reported under the United Nations Framework Convention on Climate Change (UNFCCC), which are based on statistical activity data and measured or estimated emission factors [IPCC, 2006]. For many CH₄ source sectors (e.g., fossil fuels, waste, agriculture), emission factors exhibit large spatial, temporal and site-to-site variability (e.g., Brandt et al. [2014]), which inherently limits the capability of bottom-up approaches to provide accurate total emissions. Particular challenges are the representation of 'high-emitters' or 'super-emitters' in bottom-up inventories [Zavala-Araiza et al., 2015], but also of minor source categories (e.g., abandoned coal mines or landfill sites), which, if not properly accounted for, may result in incorrect inventories. Independent verification using atmospheric measurements and inverse modelling is therefore considered essential to ensure the environmental integrity of reported emissions [Levin et al., 2011; National Academy of Science, 2010; Nisbet and Weiss, 2010; Weiss and Prinn, 2011] and has been suggested to be used for the envisaged 'transparency framework' under the Paris agreement [WMO, 2016a].

Inverse modelling ('top-down') is a mass-balance approach, providing information from the integrated emissions from all sources. However, the quality of the derived emissions critically depends on the quality and density of measurements, and the quality of the atmospheric models used. In particular, when aiming at verification of bottom-up inventories, thorough validation of inverse models and realistic uncertainty estimates of the top-down emissions are essential.

Bergamaschi et al. [2015] showed that the range of the derived total CH₄ emissions from north-western and eastern Europe using four different inverse modelling systems, was considerably larger than the uncertainty estimates of the individual models. While the latter typically use Bayes' theory to calculate the reduction of assumed 'a priori' emission uncertainties by assimilating measurements (propagating estimated observation and model errors to the estimated emissions), an ensemble of inverse models may provide more realistic overall uncertainty estimates, since estimates of model errors are often based on strongly simplified assumptions and do not represent the total uncertainty. Furthermore, validation of the inverse models against independent observations not used in the inversion is important to assess the quality of the inversions.

Here, we present a new analysis, estimating European CH₄ emissions over the time period 2006-2012 using seven different inverse models. We apply a new, quality-controlled and harmonized data set of in-situ measurements from 18 European atmospheric monitoring stations generated within the European FP7 project InGOS ("Integrated non-CO₂ Greenhouse gas Observing System"). The InGOS data set is complemented by measurements from additional European and global discrete air sampling sites. Compared to the previous paper by Bergamaschi et al. [2015], which analysed 2006-2007, this study extends the target period (2006-2012), takes advantage of the larger and more stringently quality-controlled observational data set, and includes additional inverse models. Furthermore, we present a more comprehensive validation of model results using an extended set of aircraft observations, aiming at a more quantitative assessment of the overall errors. Finally we examine in more detail the potential contribution of natural emissions (such as peatlands, wetlands, or wet soils) using seven different

wetland inventories from the "Wetland and Wetland CH₄ Inter-comparison of Models Project" (WETCHIMP) [Melton *et al.*, 2013; Wania *et al.*, 2013].

2 Atmospheric measurements

The European monitoring stations used in this study are compiled in Table 1 and their locations are shown in Figure 1. The core data set is from 18 stations with in-situ CH₄ measurements. These measurements have been rigorously quality-controlled within the InGOS project. The quality control includes regular measurements of so-called target gases that monitor instrument performance and long-term stability [Hammer *et al.*, 2013; Lopez *et al.*, 2015; Schmidt *et al.*, 2014; WMO, 1993]. The instrument precision has been evaluated as 24 h moving 1 σ standard deviation of bracketing working standards (denoted as "working standard repeatability"). A suite of other quality measures and error contributions, uncertainty in non-linearity corrections, potentially causing systematic biases between stations, have been investigated [Vermeulen, 2016], however, they have not been used in the inversions. The in-situ measurements are reported as hourly average dry air mole fractions (in units of nmol mol⁻¹, abbreviated as ppb), including the standard deviation of all individual measurements within one hour.

At most stations, the measurements have been performed using gas chromatography (GC) systems equipped with flame ionization detectors (FID). At the station Pallas (PAL), a GC-FID was applied until January 2009, and then replaced by a cavity ring-down spectrometer (CRDS). CRDS measurements (which are superior in precision compared to GC-FID) also started at other measurement sites, but here we used the GC measurements wherever available for the sake of time-series consistency while CRDS measurements were included for quality control and error assessment.

The InGOS measurements are calibrated against the NOAA-2004 standard scale (which is equivalent to the World Meteorological Organization Global Atmosphere Watch WMO-CH₄-X2004 CH₄ mole fraction scale) [Dlugokencky *et al.*, 2005], except the InGOS measurements at Mace Head (MHD), for which the Tohoku University (TU) CH₄ standard scale has been used [Aoki *et al.*, 1992; Prinn *et al.*, 2000]. The two calibration scales are in close agreement. Based on parallel measurements by NOAA and Advanced Global Atmospheric Gases Experiment (AGAGE) at five globally distributed stations over more than 20 years an average difference of 0.3 ± 1.2 ppb between the two scales has been found. This difference is considered as not significant, and therefore no scale correction has been applied. In this study, we use the InGOS "release 2014" data set.

Six InGOS stations are equipped with tall towers, with uppermost sampling heights of 97-300 m above the surface, eight sites are surface stations (at low altitudes) with sampling heights of 6-60 m, and four sites are mountain stations (at altitudes between 1205 m and 3575 m asl).

The in-situ measurements at the InGOS stations are complemented by discrete air samples from the NOAA Earth System Research Laboratory (ESRL) global cooperative air sampling network at 11 European sites (and additional global NOAA sites used for the global inverse models) [Dlugokencky *et al.*, 1994; Dlugokencky *et al.*, 2009] and at five sites from the French RAMCES (Réseau Atmosphérique de Mesure des Composés à Effet de Serre) network [Schmidt *et al.*, 2006]. The discrete air measurements are taken from samples which are usually collected weekly.

For validation of the inverse models, we use CH₄ measurements of discrete air samples from four European aircraft profile sites at Griffin, Scotland (GRI), Orléans, France (ORL), Hegyhátsál, Hungary (HNG) and Bialystok, Poland (BIK) (see Figure 1). The analyses of the samples from GRI, ORL and HNG were performed at the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) with the same GC used for RAMCES sites, those from BIK at the Max Planck Institute for Biogeochemistry (MPI).

Furthermore, we use airborne in-situ measurements from a campaign over Europe, which was performed in September / October 2009 as part of the "Infrastructure for Measurement of the European Carbon Cycle" (IMECC) project [Geibel *et al.*,

2012]. All measurements of the discrete air samples (from the NOAA and RAMCES surfaces sites and LSCE and MPI aircraft profile sites) and from the IMECC aircraft campaign are calibrated against the WMO-CH4-X2004 scale.

3 Modelling

3.1 Inversions

5 Four inversions were performed, investigating the impact of different sets of stations and the use of 'a priori' information on emissions (see Table 2). Inversion S1 covers 2006-2012 using a base set of observations (including only stations with maximum data gaps of 1 year), while inversions S2, S3, and S4 were performed for the years 2010-2012 and include additional stations, for which not all data are available before 2010. In S1, S2, and S3 the InGOS data set is used along with the discrete air samples from NOAA and RAMCES surfaces sites, while in S4 only the InGOS data are used. The exact sets of stations
10 applied in the different inversion experiments are indicated in Table 1. Inversion S1, S2, and S4 use 'a priori' information of CH₄ emissions from gridded inventories. For the anthropogenic CH₄ emissions, the "EDGARv4.2FT-InGOS" inventory is used, which integrates information on major point sources from the European Pollutant Release and Transfer Register (E-PRTR) into the EDGARv4.2FastTrack CH₄ inventory (<http://edgar.jrc.ec.europa.eu/overview.php?v=ingos>) [*Janssens-Maenhout et al.*, 2014]. Since EDGARv4.2FT-InGOS covers only the period 2000-2010, the inventory of 2010 has been
15 applied as 'a priori' also for 2011 and 2012. For the natural CH₄ emissions from wetlands, most models used the wetland inventory of J. Kaplan [*Bergamaschi et al.*, 2007] as 'a priori', except TM5-CTE, which applied LPX- Bern v1.0 [*Spahni et al.*, 2013] instead. Inversion S3 was performed without using detailed bottom-up inventories as 'a priori', in order to analyse the constraints of observed atmospheric CH₄ on emissions independent of 'a priori' information (using a homogeneous distribution of emissions over land and over the ocean, respectively, as starting point for the inversions in a similar manner as
20 in *Bergamaschi et al.* [2015]; [for further details see section 1 of supplementary material](#)).

3.2 Atmospheric models

The atmospheric models used in this study are listed in Table 3. The models include global Eulerian models with zoom over Europe (TM5-4DVAR, TM5-CTE, LMDZ), regional Eulerian models (CHIMERE) and Lagrangian dispersion models (STILT, NAME, COMET). The horizontal resolutions over Europe are ~ 1.0 - 1.2° (longitude) \times ~ 0.8 - 1.0° (latitude) for the
25 global models (zoom), and ~ 0.17 - 0.56° (longitude) \times ~ 0.17 - 0.5° (longitude) for the regional models. The regional models use boundary conditions ([background CH₄ mole fractions](#)) from inversions of the global models (STILT from TM3, COMET from TM5-4DVAR, CHIMERE from LMDZ, or estimate the boundary conditions in the inversions (NAME), using baseline observations at Mace Head as 'a priori' estimates. [In case of NAME and CHIMERE, the boundary conditions are further optimized in the inversion.](#)
30 All models used the same observational data set described in section 2 (except the stations ZEP and ICE, that are outside the domain of some regional models and except the mountain stations JFJ, PDM and KAS, which were not used in the NAME inversions). For the stations with in-situ measurements in the boundary layer, most models assimilated only measurements in the early afternoon (between 12:00 and 15:00 LT), and for mountain stations only night-time measurements (between 00:00 and 03:00 LT) [*Bergamaschi et al.*, 2015]. However, NAME and COMET used observations at all times. The individual
35 inverse models are described in more detail in the supplementary material ([section 1](#)).

4 Results and discussion

4.1 European CH₄ emissions

Figure 2 shows the maps of the European CH₄ emissions (average 2010-2012) derived from the seven inverse models for inversion S4. The corresponding maps for inversions S1-S3 (available from five models) are shown in the supplementary material (Figures 1S-3S). In S1, S2, and S4, which are guided by the 'a priori' information from the emission inventories, the 'a posteriori' spatial distributions are usually close to the prior patterns on smaller scales (determined by the chosen spatial correlation scale lengths). The NAME inversion groups together grid cells for which the observational constraints are weak, i.e., it averages over increasingly larger areas at larger distances from the observations. Consequently, in the NAME inversion the 'fine structure' of the 'a priori' inventories disappears in areas which are not well constrained (e.g., Spain).

10 Comparing inversions S1, S2, and S4 shows overall very similar spatial patterns for all inverse models, indicating only moderate differences in the observational constraints of the three different sets of stations. In particular, addition of NOAA and RAMCES discrete air samples (inversion S2 vs. S4) results in only minor differences in the derived emissions. When the larger set of InGOS stations (S2 vs. S1) is used, most models yield higher CH₄ emissions from Northern Italy. This is most likely mainly due to the observations from Ispra (IPR), at the north-western edge of the Po valley, while this area is not well
15 constrained in S1.

The information content of the observations is further examined in inversion S3, which does not use detailed emission inventories (Figure 3S), similar to a previous sensitivity experiment in *Bergamaschi et al.* [2015]. Especially TM5-4DVAR and TM3-STILT yield similar spatial distributions with elevated CH₄ emissions from the BENELUX area and northwestern Germany, from the coastal area of northwestern France, Ireland, UK, and the Po valley. Most of these patterns are visible also
20 in inversion S3 of NAME, however with more variability on smaller scales (while TM5-4DVAR and TM3-STILT show much smoother distributions). These regional hotspots are broadly consistent with the bottom-up inventories, which illustrates the principal capability of inverse modelling to derive emissions that are independent of detailed 'a priori' inventories in the vicinity of observations. LMDZ and TM5-CTE also show elevated emissions over western and central Europe, but in contrast to the other three inverse models no regional hotspots. For TM5-CTE this is related to the applied inversion technique (adjusting
25 emissions uniformly over large predefined regions), which effectively limits the number of degrees of freedom, and does not allow retrieval of regional hotspots, if such patterns are not 'a priori' present within the predefined regions. For LMDZ, the lack of regional hotspots is probably related to the specific settings for this scenario, with a spatial correlation scale length of 500 km uncertainties per grid cell of only 200%, significantly larger than in TM5-4DVAR (50 km) and TM3-STILT (60 km).

Figure 3a displays the annual total European CH₄ emissions derived by the models for 2006-2012 in inversion S1, and for
30 2010-2012 in S2-S4. The figure shows the total emissions from all EU-28 countries, and separately from northern Europe (Norway, Sweden, Finland, Baltic countries, and Denmark), western Europe (UK, Ireland, Netherlands, Belgium, Luxembourg, France, Germany, Switzerland, and Austria), eastern Europe (Poland, Czech Republic, Slovakia, and Hungary), and southern Europe (Portugal, Spain, Italy, Slovenia, Croatia, Greece, Romania, and Bulgaria). The non-EU-28 countries Norway and Switzerland are included here in 'northern Europe' and 'western Europe', respectively, but not in EU-28. Six of
35 the seven models yield considerably higher total CH₄ emissions from the EU-28 compared to the anthropogenic CH₄ emissions reported to UNFCCC (submission 2016), while NAME is very close to the UNFCCC emissions. This behaviour is apparent also for the European subregions western, eastern and southern Europe, while for northern Europe (where natural CH₄ emissions play a large role) also NAME yields higher total CH₄ emissions compared to UNFCCC (except for S3 in 2011 and 2012).

40 Figure 3a also shows the results from the previous study of *Bergamaschi et al.* [2015], which used four inverse models (previous versions of those applied in this study) and a set of 10 European stations with continuous measurements

(complemented by discrete air samples) to estimate CH₄ emissions in 2006-2007. For TM5-4DVAR, TM3-STILT, and LMDZ the results are relatively similar (within ~10% for EU-28) to this study, while the CH₄ emissions from NAME were ~20% lower (EU-28). Despite the significantly larger number of European monitoring stations in the present study, however, we emphasize that the available stations do not very well cover the whole EU-28 area. Consequently, the emissions especially from Southern Europe remain poorly constrained.

For comparison of total emissions derived by the inverse models and anthropogenic emissions from emission inventories it is essential to account for natural emissions, especially from wetlands, peatlands and wet soils. As an estimate of these emissions and their uncertainties, we use an ensemble of seven wetland inventories from the "Wetland and Wetland CH₄ Inter-comparison of Models Project" (WETCHIMP) [Melton *et al.*, 2013; Wania *et al.*, 2013] (the spatial distribution of European CH₄ emissions from the different individual WETCHIMP inventories is shown in Figure 4S). Figure 3a shows the mean, median, minimum and maximum CH₄ emissions from this ensemble for EU-28 and the different European subregions. These quantities are evaluated after integrating over the corresponding areas, using the multi-annual mean (1993-2004) of the WETCHIMP inventories. For northern Europe, in particular, the estimated wetland emissions are high (2.5 (1.7-4.3) Tg CH₄ yr⁻¹ (mean, minimum, maximum)) and exceed the anthropogenic CH₄ emissions (UNFCCC: 1.3 Tg CH₄ yr⁻¹; mean 2006-2012). Substantial wetland emissions are also estimated for western Europe (1.6 (0.4-3.1) Tg CH₄ yr⁻¹), but wetland emissions are also non-negligible for eastern Europe (0.3 (0.03-0.9) Tg CH₄ yr⁻¹) and southern Europe (0.6 (0.01-1.1) Tg CH₄ yr⁻¹), especially when considering the upper range of these estimates. For EU-28, wetland emissions of 4.3 (2.3-8.2) Tg CH₄ yr⁻¹ are estimated, corresponding to 22% (11%-41%) of reported anthropogenic CH₄ emissions.

Taking into account the estimates of the WETCHIMP ensemble brings the results of the six inverse models that derive high emissions into the upper uncertainty range of the sum of anthropogenic emissions (reported to UNFCCC) and wetland emissions, while the emissions derived by NAME fall in the lower range (Figure 3b). This analysis suggests broad consistency between bottom-up and top-down emission estimates, albeit with a clear tendency (6 of 7 models) towards the upper range of the bottom-up inventories for the total CH₄ emissions from EU-28. This behaviour is apparent also for western and southern Europe, while for eastern Europe several models are close to or above the upper uncertainty bound (NAME is very close to the mean), and for northern Europe several models are rather in the lower range (or below the lower uncertainty bound) of the combined UNFCCC and WETCHIMP inventory.

Critical to the assessment of consistency between the different approaches, is the analysis of their uncertainties. Inverse models usually propagate estimated observation and model errors to the estimated emissions, however in particular the model errors are generally based on simplified assumptions. Furthermore, the error estimates of the inverse models take usually only random errors into account, and are based on the assumption that observation and model errors are unbiased. Estimated 2 σ uncertainties for EU-28 top-down emissions range between ~7% and ~33% (except for inversion S3 of NAME, for which uncertainties are larger than 50%). For the subregions 'northern Europe' and 'southern Europe', which are poorly constrained by measurements, the model estimates of the relative uncertainties are significantly larger, ranging between ~20% and more than ~100%.

The (2 σ) uncertainties of the UNFCCC inventories shown in Figure 3a are based on the uncertainties of major CH₄ source categories reported by the countries in their national inventory reports. To calculate the uncertainties of total emissions per country (or group or countries), the reported uncertainties per category were aggregated as described in Bergamaschi *et al.* [2015]. We note, however, that uncertainties reported for the same category by different countries exhibit large differences (e.g., for coal between 9 and 300%, for oil and natural gas between 5 and 460%, for enteric fermentation between 7 and 50%, for manure management between 5 and 100%, and for solid waste disposal between 22 and 126%), with the lower uncertainty estimates appearing unrealistically low. Furthermore, the estimates of the total uncertainties consider only the major categories

(EU-28: 93% of reported emissions) and do not take into account potential additional emissions (and their uncertainties) that are not covered by the inventories.

Figure 3a includes also the anthropogenic CH₄ emissions from EDGARv4.2FT-InGOS (for 2006-2010), which are at the upper uncertainty bound of the UNFCCC inventories for EU-28. The difference between UNFCCC and EDGAR is mainly due to significant differences in CH₄ emissions from fossil fuels (coal, oil, and natural gas), which, however, might be overestimated in some cases in EDGAR [Bergamaschi *et al.*, 2015].

For wetlands, very large differences between the different inventories of the WETCHIMP ensemble are apparent regarding the spatial emission distribution (see Figure 4S) and the magnitude of the emissions, illustrating the very high uncertainties in the current estimates. Comparing the different wetland inventories, a striking pattern is visible for LPJ-WHyMe, with very high CH₄ emissions for the British Isles. The climate of this region has mild winters that allow simulated wetland CH₄ emissions to continue year-round, yielding high annual emissions intensity for LPJ-WHyMe [Melton *et al.*, 2013].

In the previous analysis of Bergamaschi *et al.* [2015] the contribution from natural sources in western and eastern Europe was considered to be very small, based on the wetland inventory of J. Kaplan [Bergamaschi *et al.*, 2007]. However, that inventory is close to the lower estimates of the WETCHIMP ensemble. Unfortunately, direct comparisons of CH₄ emissions simulated by the different wetland inventories with local or regional CH₄ flux measurements in European wetland areas are lacking. Therefore, no conclusions can be drawn as to which of the inventories is most realistic.

To further investigate the contribution of wetland emissions we analyse the seasonal variations. Figure 4 illustrates that four inverse models (TM5-4DVAR, TM5-CTE, TM3-STILT, and LMDZ) calculate pronounced seasonal variations in total emissions. For EU-28 the derived seasonality is largely consistent with the seasonality of the wetland emissions from the WETCHIMP ensemble (both regarding the amplitude, and the phase with maxima in summer). For northern Europe the seasonal variations derived by the four inverse models are somewhat smaller compared to the mean of the WETCHIMP ensemble, while for western and eastern Europe they are somewhat larger, but still broadly within the minimum-maximum range of the WETCHIMP inventories. For southern Europe, the seasonality of the four inverse models is more irregular, and the maximum emissions for the wetland ensemble show a clear peak in winter, which however is not apparent in the mean or median of the ensemble. This is probably due to the important role of precipitation for the wetland emissions in southern Europe, while for temperate and boreal regions the seasonal variation of wetland emissions is mainly driven by temperature (e.g., [Christensen *et al.*, 2003; Hodson *et al.*, 2011]). In contrast to the discussed four models, NAME derives much smaller seasonal variations, and for western Europe, eastern Europe, and EU-28 with opposite phase (small maximum in winter). Only for northern Europe, also NAME estimates maximum emissions in summer, however the amplitude is much smaller compared to the other models and the WETCHIMP wetland inventories. One reason contributing to the smaller amplitude is that NAME provides only 3-monthly emissions (compared to monthly resolution of the other four inverse models), but the lower temporal resolution of NAME clearly explains only a smaller part of the different seasonal cycles. Figure 5S shows that also in inversion S3 (which is not using any detailed a priori inventory nor any a priori seasonal cycle) significant seasonal cycles of CH₄ emissions are derived by TM5-4DVAR, TM3-STILT, LMDZ, and TM5-CTE, which demonstrates that the derived seasonal cycles are mainly driven by the observations, and not by the a priori.

Apart from the different behaviour of NAME, the finding that four inverse models derive seasonal cycles that are broadly consistent with the seasonal cycles calculated by the WETCHIMP ensemble supports a significant contribution of wetlands to the total CH₄ emissions. Commonly, anthropogenic CH₄ emissions are assumed to have no significant seasonal variations, except CH₄ emissions from rice and biomass burning (which however play only a minor role in Europe). Unfortunately, only very limited information is available about potential seasonal variations of anthropogenic CH₄ sources (other than rice and biomass burning). Ulyatt *et al.* [2010] reported significant seasonal variations of CH₄ emissions from dairy cows, mainly

related to the lactation periods of cows. *VanderZaag et al.* [2014], estimating total CH₄ emissions from two dairy farms, found higher CH₄ emissions in fall compared to spring, mainly due to varying CH₄ emissions from manure management. Beside agricultural CH₄ sources, CH₄ from landfills [*Spokas et al.*, 2011] and waste water may also exhibit seasonal variations, while only small seasonal variations were found for natural gas distribution systems [*McKain et al.*, 2015; *Wennberg et al.*, 2012; *Wong et al.*, 2016] (and further references therein). Quantitative estimates of potential seasonal variations of anthropogenic sources cannot be made due to the limited number of studies, but the relative variability of the total anthropogenic sources is expected to be much smaller compared to wetlands.

Model simulations and bottom-up inventories for individual countries (or group of countries) are shown in the supplementary material (Figure [5S6S](#)), illustrating further that wetland emissions are important, particularly in northern European countries, but may also contribute significantly for many other countries.

Finally, we analyse the trends in CH₄ emissions (Figure 7S). Anthropogenic CH₄ emissions reported to UNFCCC for EU-28 decreased by -0.44 ± 0.02 Tg CH₄ yr⁻² during 2006-2012. Also all 5 inversions which are available for this period (inversion S1) derive negative CH₄ emission trends ranging between -0.19 and -0.58 Tg CH₄ yr⁻². The uncertainties given for the trends of the individual inversions (and the reported CH₄ emissions), however, include only the uncertainty of the linear regression (i.e. reflecting the scatter of the annual values around the linear trend), but do not take into account the uncertainties of the annual mean values and the error correlations between different years. In particular the latter remain very difficult to estimate, which currently limits clear conclusions about the significance of the trends.

4.2 Evaluation of inverse models

First we evaluate the performance of model simulations at the atmospheric monitoring stations. Figure [6S-8S](#) shows the correlation coefficients, bias, ~~and~~ root mean square (RMS) difference, and the ratio between modelled and observed standard deviation for inversion S4, including stations that were assimilated and stations that were used for validation only. For the evaluation of the statistics for the in-situ measurements, we use only early afternoon data (between 12:00 and 15:00 LT). Averaging over all stations, the correlation coefficients are between 0.65 and 0.79 for 6 models, and 0.5 for COMET. The ranking of models in terms of correlation coefficients is closely reflected in the achieved average RMS values, ranging between 33 and 70 ppb (with models with higher correlation coefficients typically achieving lower average RMS). At several tall towers a clear tendency of decreasing RMS with increasing sampling height is visible, demonstrating the benefit of higher sampling heights, which allow more representative measurements that are less affected by local sources and that can be better reproduced by the models.

While the evaluation of the model simulations at the monitoring stations provides a measure of the quality of the inversions and the atmospheric transport models applied (e.g., with the correlation coefficients describing how much of the observed variability can be explained by the models), the analysis of the station statistics cannot quantify how realistic the derived emissions are, but gives only some qualitative indications about potential biases of the emissions. The inverse models optimize model emissions to achieve an optimal agreement between simulated and observed atmospheric CH₄ mole fractions (taking into account the a priori constraints). This implies that potential biases of the model (or the observations) may be compensated in the inversions by introducing biases in the derived emissions. In particular, vertical mixing of the models is very critical in this context. For example, too strong vertical mixing of the transport models may be compensated in the inversion by enhancing the model emissions (i.e. deriving model emissions that are higher than real emissions) such that a good agreement between simulated and observed mole fractions at the surface can still be achieved. An important diagnostic to identify such potential systematic errors is the analysis of vertical profiles (including the boundary layer and the free troposphere). For this purpose we compare our model simulations with regular aircraft profiles at four European sites (Figure 5). At Griffin (GRI), observed

and simulated mole fractions show only small vertical gradients, while at Orléans (ORL), Hegyhátsál (HNG), and Bialystok (BIK) large vertical gradients are visible, with increasing values towards the surface. The figure also includes the background mole fractions in the absence of model emissions over Europe calculated by TM5-4DVAR (based on the scheme of Rödenbeck *et al.* [2009]). At GRI, the measurements are in general very close to the background mole fractions, illustrating that the impact of European emission is rather limited at this site. In contrast, pronounced enhancements in measured and simulated CH₄ compared to the background are apparent at the other three sites, especially in the lower ~2 km due to regional emissions. These enhancements show some seasonal variation, with largest vertical extension during summer (~2 km), while they are confined to the lower ~1 km during winter, due to the seasonal variations in the average boundary layer height [Koffi *et al.*, 2016]. Please note that the differences of the background mole fractions which are visible in Figure 5 between some sites, are partly due to the different temporal sampling at the different sites (compare Figure 6).

To analyse potential model biases more quantitatively, we evaluate in the following the enhancement of observations and model simulations compared to background CH₄ values (1) integrated over the entire boundary layer, and (2) integrated over the lower troposphere up to ~3-4 km. The rationale behind this approach is that emissions initially mainly accumulate within the boundary layer. Therefore, potential biases in model emissions should be reflected in differences between the observed and modelled integrated enhancement within the boundary layer. For the overall budget, however, mixing between boundary layer and free troposphere plays an important role. Thus, the enhancement integrated over the entire lower troposphere provides additional diagnostics for potential model biases.

The integration of the enhancements is shown for the individual profiles at ORL, HNG and BIK in the supplementary material (Figures 7S9S, 8S10S, 9S11S). In addition, we use also aircraft measurements from the IMECC campaign in September / October 2009 (Figure 10S12S). These include profile measurements at Orléans and Bialystok, but also at Karlsruhe, Jena, and Bremen, hence extending the spatial coverage of the sites with regular profiles (ORL, HNG and BIK). To calculate the enhancements for the individual profiles, we apply the background mole fractions calculated for the TM5-4DVAR zoom domain as the common reference for the observations and the model simulations for all global models (i.e. TM5-4DVAR, TM5-CTE, and LMDZ). For STILT and NAME, the background CH₄ is calculated calculated for the STILT and NAME domains, illustrate that but the dependence of the background mole fractions (calculated by TM5-4DVAR) on the exact extension of the domain is generally rather small. However, the CH₄ background mole fractions used in the inversions of the regional models (for NAME based on baseline observations at Mace Head and for TM3-STILT based on the TM3 model) shows significant differences compared to the TM5-4DVAR background, with typically ~10 ppb higher values at the three continental aircraft sites (ORL, HNG, BIK; see Figure 5). In order to investigate which background mole fractions are more realistic we compared the model simulations with the aircraft observations for events with very low simulated contribution (< 3 ppb) from European CH₄ emissions (Figure 14S). This analysis shows that TM5-4DVAR simulations are close to the observations (average bias between -1.1 and 3.5 ppb), which indicates that the TM5-4DVAR background is relatively realistic, while NAME and TM3-STILT are consistently higher at the continental aircraft sites with average biases of 12-13 ppb for NAME and 9-12 ppb for TM3-STILT. This supports the use of the background calculated with TM5-4DVAR as reference for the measurements. For the evaluation of the simulated CH₄ enhancements of the regional models, however, we use the actual background used in NAME and TM3-STILT.

For the integration over the boundary layer, we use the boundary layer height (BLH) diagnosed by TM5. A recent comparison of the TM5 BLH with observations from the NOAA Integrated Global Radiosonde Archive (IGRA) [Koffi *et al.*, 2016] showed that TM5 reproduces the daytime BLH relatively well (within ~10–20%), but larger deviations were found for the nocturnal BLH, especially during summer, when very low BLHs (< 100 m) are observed. Here, we use only profiles for which the (TM5 diagnosed) BLH is not lower than 500 m. The average enhancement of the measurements and model simulations in the

boundary layer compared to background is denoted as $\Delta c_{MOD, BL}$ and $\Delta c_{OBS, BL}$, respectively (further details about the evaluation of the enhancements are given in the supplementary material). Figure 6 shows the derived 'relative bias', defined as:

$$rb_{BL} = (\Delta c_{MOD, BL} - \Delta c_{OBS, BL}) / \Delta c_{OBS, BL}$$

for ORL, HNG, BIK for the entire target period 2006-2012 (inversion S1). ~~At ORL, The~~ three global inverse models (i.e. TM5-

5 4DVAR, TM5-CTE, and LMDZ) show in general only a small average positive relative bias (rb_{BL} between -7% and 10%) at the three aircraft sites (TM5-4DVAR: 20%; TM5-CTE: 22%; LMDZ: 30%), indicating that these models likely overestimate the regional emissions. In contrast, TM3-STILT and NAME have significant negative relative biases (TM3-STILT: rb_{BL} between -13% and -24% for the three sites; NAME $rb_{BL} = -30\%$ for ORL and HNG), very small biases of only 1% and 3%, respectively, at this profile site. Also at HNG, three models show some positive bias on average (TM5-4DVAR: 14%, LMDZ: 16%, TM3-STILT: 11%), while the bias of NAME and TM5-CTE is close to zero. In contrast, at BIK all models show small relative bias (TM5-4DVAR: -2%; TM3-STILT: 5%; TM5-CTE: 6%; LMDZ: 6%; NAME: not available).

15 These negative biases are likely related to the positive bias in the background CH_4 used for NAME and TM3-STILT (see above), since the regional models invert the difference between the observations and the assumed background. In fact, also at most continental atmospheric monitoring stations, the background used for NAME and TM3-STILT is significantly higher (~10 ppb) compared to the TM5-4DVAR background (Figure 15S).

The 'relative bias' is also extracted separately for different seasons (right panel of Figure 6). ~~At ORL, all models have relative biases close to zero in spring (March—May), while larger relative biases are visible during other seasons. Apart from this feature, there is no clear seasonal cycle in the relative bias apparent and the variability between the different seasons is generally small at HNG and BIK (data points at BIK for DJF are considered not significant as they are from one single profile only). From this analysis there is no evidence that the seasonal cycle of emissions derived by four inverse models (TM5-4DVAR, TM5-CTE, TM3-STILT, and LMDZ; see section 4.1) with clear maxima in summer could be due to a seasonal bias in the transport models. At the same time, however, NAME, which calculates much smaller seasonal variations of emissions, also shows no only small seasonal variations of the average bias at ORL and HNG, NAME has ~20-40% lower average bias between December and May compared to June to November, which seems to contrast with the seasonality of the emissions derived by NAME; however~~ However, especially at this site HNG the total number of profiles is rather small (n=22), which limits the analysis of potential seasonal transport biases.

25 Figure ~~11S-13S~~ shows the relative bias of the CH_4 enhancements integrated over the lower troposphere, defined as:

$$rb_{COL} = (\Delta c_{MOD, COL} - \Delta c_{OBS, COL}) / \Delta c_{OBS, COL}$$

30 The three global inverse models (i.e. TM5-4DVAR, TM5-CTE, and LMDZ) have a relative bias between of -4% and 20% at the three aircraft sites, indicating a small tendency to overestimate the European CH_4 emissions, while the regional models show a negative relative bias (TM3-STILT: between -9% and -20% for the three sites; NAME -31% for ORL and -40% HNG). At ORL all models exhibit a significant positive bias, ranging between 31% and 41%. At HNG, the relative biases are between 7% and 27% for the five models and at BIK between 6% and 25% for TM5-4DVAR, TM3-STILT, and TM5-CTE and LMDZ.

35 Figure 7 presents an overview of the derived relative biases for the enhancement integrated over the boundary layer (rb_{BL} , top panel of figure) and in the lower troposphere (rb_{COL} , lower panel). The ~~finding that differences of~~ the relative bias integrated over the lower troposphere is somewhat higher than compared to that integrated only over the boundary layer (e.g., $rb_{COL} > rb_{BL}$ for TM5-4DVAR and TM5-CTE at ORL and BIK) ~~in several cases~~ suggests that shortcomings of the models to simulate the exchange between the boundary layer and the free troposphere may contribute significantly to the bias in the derived emissions. An illustrative example of the shortcomings of the models to simulate the free troposphere are the IMECC profiles at Bialystok on 30 September 2009 (Figure ~~10S12S~~).

40 The measurements show a considerable CH_4 enhancement (~25 ppb) at

around 3.5 to 4 km, which is not reproduced by the models. This could indicate that cloud convective transport was missed by the models.

A general limitation of the analysis of the enhancements integrated over the lower troposphere, however, is that this analysis is more sensitive to potential errors in the simulated background mole fractions in the free troposphere compared to the boundary layer, because of the generally much lower enhancements in the free troposphere.

Finally, we analyse the correlation between the relative bias of the integrated CH₄ enhancements and the regional model emissions. Figure 16S shows the relationship between rb_{BL} and the average model emissions around the aircraft site, integrating all model grid cells with a maximum distance of 400 km (hereafter referred to as integration radius) from the aircraft site. At all three sites clear correlations between rb_{BL} and the regional model emissions are found, which confirms that rb_{BL}, derived from the aircraft profiles, can be used to diagnose biases in the regional model emissions.

The derived correlations depend on the chosen area, over which model emissions are integrated. For ORL and HNG, significant correlations were found for integration radii between 200 and 800 km, while for BIK different integration radii resulted in poorer correlations (now shown), probably related to significant differences in the spatial emission patterns derived by the different models around this site. To further improve the analysis, the 'footprints' (i.e. sensitivities of atmospheric concentrations to surface emissions) of the individual aircraft profiles should be taken into account in the future. Furthermore, it would be useful, to calculate for all global models individually the background mole fractions using the scheme of Rödénbeck *et al.* [2009]. This would allow to derive the modelled CH₄ enhancements more accurately.

5 Conclusions

We have presented estimates of European CH₄ emissions for 2006-2012 using the new InGOS data set of in-situ measurements from 18 European monitoring stations (and additional discrete air sampling sites) and an ensemble of seven different inverse models. For the EU-28, total CH₄ emissions of 26.7-8 (20.2-29.7) Tg CH₄ yr⁻¹ are derived (mean, 10% percentile, and 90% percentile from all inversions), compared to total anthropogenic CH₄ emissions of 21.3 Tg CH₄ yr⁻¹ (2006) to 18.8 Tg CH₄ yr⁻¹ (2012) reported to UNFCCC. Our analysis highlights the potential significant contribution ~~suggests that~~ of natural emissions from wetlands (including peatlands and wet soils) ~~contribute significantly~~ to the total European emissions, with total wetland emissions of 4.3 (2.3-8.2) Tg CH₄ yr⁻¹ (EU-28) estimated from the WETCHIMP ensemble of seven different wetland inventories [Melton *et al.*, 2013; Wania *et al.*, 2013]. The hypothesis of a significant contribution from natural emissions is supported by the finding that four inverse models (TM5-4DVAR, TM5-CTE, TM3-STILT, LMDZ) derive significant seasonal variations of CH₄ emissions with maxima in summer. However, the NAME model calculates only a weak seasonal cycle, with small maximum (of EU-28 total CH₄ emissions) in winter. Furthermore, it needs to be emphasized that wetland inventories have large uncertainties and show large differences in the spatial distribution of CH₄ emissions.

Taking into account the estimates of the WETCHIMP ensemble, the bottom-up and top-down estimates of total EU-28 CH₄ emissions are broadly consistent within the estimated uncertainties. However, the results from six inverse models are in the upper uncertainty range of the sum of anthropogenic emissions (reported to UNFCCC) and wetland emissions, while the emissions derived by NAME are in the lower range. Furthermore, the comparison of bottom-up and top-down estimates shows some differences for the different European subregions. For northern Europe (including Norway) several models are rather in the lower range (or below the lower uncertainty bound) of the combined UNFCCC and WETCHIMP inventory, while for eastern Europe several models are close to the upper uncertainty bound or above (NAME is very close to the mean). Considering the estimated uncertainties of the inverse models, however, the uncertainty ranges of bottom-up and top-down estimates generally overlap for the different European subregions.

To estimate potential biases of the emissions derived by the inverse models, we analysed the enhancements of CH₄ mole fractions compared to the background, integrated over the entire boundary layer and over the lower troposphere, using regular aircraft profiles at four European sites and the IMECC aircraft campaign.

This analysis showed for the three global inverse models (TM5-4DVAR, TM5-CTE, and LMDZ) a relatively small average relative bias (rb_{BL} between -7% and 10%, rb_{COL} -4% and 20% for ORL, HNG and BIK). The regional models revealed a significant negative bias (TM3-STILT: rb_{BL} between -13% and -24%, rb_{COL} between -9% and -20% for ORL, HNG and BIK; NAME $rb_{BL} = -30%$, rb_{COL} between -31% and -40% at ORL and HNG). A potential cause for the negative relative bias of TM3-STILT and NAME is the significant positive bias of the background used in TM3-STILT (from global TM3 inversion) and NAME (based on measurements at baseline conditions at Mace Head).

~~suggests that several models have a significant positive bias over Orléans (France), with 3 models (TM5-4DVAR, TM5-CTE, LMDZ) calculating 20%–30% larger enhancements integrated over the boundary layer, while TM3-STILT and NAME showed low biases in the boundary layer (1%–3%), but also significant biases of the enhancement integrated over the total column of the lower ~3 km. Smaller biases were found for the aircraft profile sites in Hegyhátsál (Hungary) and Białystok (Poland). For the latter, several models even show a small negative bias.~~

~~The relative bias rb_B shows clear correlations with regional model emissions around the aircraft profile sites, which confirms that rb_{BL} can be used to diagnose biases in the regional model emissions. The analysis of the integrated enhancements compared to the background allows us, in principle, to relate the observed biases in the mole fractions to biases in the model emissions. The accuracy of the estimated relative biases, A limitation of this approach, however, depends on the quality of is potential errors in the simulated background mole fractions.~~

In particular the enhancements derived for the lower troposphere above the boundary layer (which are usually much smaller than the enhancements within the boundary layer) are very sensitive to the background mole fractions. Therefore, potential model errors in the exchange between the boundary layer and the free troposphere (and their impact on the derived emissions) remain difficult to quantify.

Our study highlights the challenge to verify anthropogenic bottom-up emission inventories with the small uncertainties desirable for the international climate agreements. To reduce the uncertainties of the top-down estimates (1) the natural emissions need to be better quantified, (2) transport models need to be further improved, including their spatial resolution and in particular the simulation of vertical mixing, and (3) the network of atmospheric monitoring stations should be further extended, especially in southern Europe, which is currently clearly under-sampled. Furthermore, the uncertainty estimates of bottom-up inventories (including both the anthropogenic and natural emissions) and atmospheric inversions need to be further improved.

Acknowledgements

This work has been supported by the European Commission Seventh Framework Programme (FP7/2007–2013) project InGOS under grant agreement 284274. We thank Joe Melton for providing the WETCHIMP data set and for discussion about the wetland emission. We are grateful to Maarten Krol and Frank Dentener for helpful comments on the manuscript, and to Ernest Koffi for support of the further analyses of the boundary layer heights. ECMWF meteorological data have been preprocessed by Philippe Le Sager into the TM5 input format. We thank Arjo Segers for support of the TM5 modelling. Furthermore, we thank Johannes Burgstaller for the compilation of UNFCCC emission uncertainties. We are grateful to ECMWF for providing computing resources under the special project "Global and Regional Inverse Modeling of Atmospheric CH₄ and N₂O (2012–2014)" and "Improve estimates of global and regional CH₄ and N₂O emissions based on inverse modelling using in situ and satellite measurements (2015–2017)". We thank François Marabelle and his team for computing support at LSCE.

References

- Alexe, M., et al., Inverse modelling of CH₄ emissions for 2010-2011 using different satellite retrieval products from GOSAT and SCIAMACHY, *Atmos. Chem. Phys.*, *15*, 113–133, doi: 10.5194/acp-15-113-2015, 2015.
- 5 Andrews, A. E., et al., CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos Meas Tech*, *7*(2), 647-687, doi: 10.5194/amt-7-647-2014, 2014.
- Aoki, S., T. Nakazawa, S. Murayama, and S. Kawaguchi, Measurements of atmospheric methane at the Japanese Antarctic Station, Syowa, *Tellus B*, *44*, 273–281, 1992.
- 10 Bergamaschi, P., et al., Satellite cartography of atmospheric methane from SCIAMACHY onboard ENVISAT: 2. Evaluation based on inverse model simulations, *J. Geophys. Res.*, *112*(D02304), doi: 10.1029/2006JD007268, 2007.
- Bergamaschi, P., et al., Inverse modeling of global and regional CH₄ emissions using SCIAMACHY satellite retrievals, *J. Geophys. Res.*, *114*, doi:10.1029/2009JD012287, 2009.
- Bergamaschi, P., et al., Atmospheric CH₄ in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, *J Geophys Res-Atmos*, *118*(13), 7350-7369, doi: 10.1002/jgrd.50480, 2013.
- 15 Bergamaschi, P., et al., Inverse modeling of European CH₄ emissions 2001-2006, *J. Geophys. Res.*, *115*(D22309), doi:10.1029/2010JD014180, 2010.
- Bergamaschi, P., et al., Top-down estimates of European CH₄ and N₂O emissions based on four different inverse models, *Atmos. Chem. Phys.*, *15*, 715-736, doi: 10.5194/acp-15-715-2015, 2015.
- 20 Blake, D., and F. Rowland, Continuing worldwide increase in tropospheric methane, *Science*, *239*, 1129–1131, 1988.
- Bousquet, P., et al., Contribution of anthropogenic and natural sources to atmospheric methane variability, *Nature*, *443*, doi:10.1038/nature05132, 2006.
- Brandt, A. R., et al., Methane Leaks from North American Natural Gas Systems, *Science*, *343*, doi: 10.1126/science.1247045, 2014.
- 25 Butler, J. H., and S. A. Montzka, The NOAA annual greenhouse gas index (AGGI), <https://www.esrl.noaa.gov/gmd/aggi/aggi.html>, last access: 19 September 2017, 2017.
- Christensen, T. R., A. Ekberg, L. Ström, M. Mastepanov, N. Panikov, M. Öquist, B. H. Svensson, H. Nykänen, P. J. Martikainen, and H. Oskarsson, Factors controlling large scale variations in methane emissions from wetlands, *Geophys Res Lett*, *30*(7), doi: 10.1029/2002gl016848, 2003.
- 30 Cunnold, D. M., et al., In situ measurements of atmospheric methane at GAGE/AGAGE sites during 1985-2000 and resulting inferences, *J. Geophys. Res.*, *107*(D14), 2002.
- Dlugokencky, E., Trends in Atmospheric Methane, www.esrl.noaa.gov/gmd/ccgg/trends_ch4/, last access: 19 September 2017, 2017.
- Dlugokencky, E. J., L. P. Steele, P. M. Lang, and K. A. Masarie, The growth rate and distribution of atmospheric methane, *J. Geophys. Res.*, *99*, 17021-17043, 1994.
- 35 Dlugokencky, E. J., E. G. Nisbet, R. Fisher, and D. Lowry, Global atmospheric methane: budget, changes and dangers, *Phil. Trans. R. Soc. A* *369*(1943), 2058-2072, doi: 10.1098/rsta.2010.0341, 2011.
- Dlugokencky, E. J., R. C. Myers, P. M. Lang, K. A. Masarie, A. M. Crotwell, K. W. Thoning, B. D. Hall, J. W. Elkins, and L. P. Steele, Conversion of NOAA atmospheric dry air CH₄ mole fractions to a gravimetrically prepared standard scale, *J. Geophys. Res.*, *110*(D18306), doi: 10.1029/2005JD006035, 2005.
- 40 Dlugokencky, E. J., et al., Observational constraints on recent increases in the atmospheric CH₄ burden, *Geophys. Res. Lett.*, *36*(L18803), doi: 10.1029/2009GL039780, 2009.
- Ganesan, A. L., A. J. Manning, A. Grant, D. Young, D. E. Oram, W. T. Sturges, J. B. Moncrieff, and S. O'Doherty, Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, *Atmos. Chem. Phys.*, *15*, 6393–6406, doi: 10.5194/acp-15-6393-2015, 2015.
- 45 Geibel, M. C., et al., Calibration of column-averaged CH₄ over European TCCON FTS sites with airborne in-situ measurements, *Atmos. Chem. Phys.*, *12*, 8763–8775, doi: 10.5194/acp-12-8763-2012, 2012.
- Hammer, S., D. W. T. Griffith, G. Konrad, S. Vardag, C. Caldow, and I. Levin, Assessment of a multi-species in situ FTIR for precise atmospheric greenhouse gas observations, *Atmos Meas Tech*, *6*(5), 1153-1170, doi: 10.5194/amt-6-1153-2013, 2013.
- 50 Henne, S., D. Brunner, B. Oney, M. Leuenberger, W. Eugster, I. Bamberger, F. Meinhardt, M. Steinbacher, and L. Emmenegger, Validation of the Swiss methane emission inventory by atmospheric observations and inverse modelling, *Atmos Chem Phys*, *16*(6), 3683-3710, doi: 10.5194/acp-16-3683-2016, 2016.
- Hodson, E. L., B. Poulter, N. E. Zimmermann, C. Prigent, and J. O. Kaplan, The El Niño-Southern Oscillation and wetland methane interannual variability, *Geophys Res Lett*, *38*(8), doi: 10.1029/2011gl046861, 2011.
- 55 Houweling, S., P. Bergamaschi, F. Chevallier, M. Heimann, T. Kaminski, M. Krol, A. M. Michalak, and P. Patra, Global inverse modeling of CH₄ sources and sinks: an overview of methods, *Atmos Chem Phys*, *17*(1), 235-256, doi: 10.5194/acp-17-235-2017, 2017.
- IPCC, *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme* Institute for Global Environmental Strategies, Japan, 2006.
- 60

- Janssens-Maenhout, G., D. Guizzardi, P. Bergamaschi, and M. Muntean, On the CH₄ and N₂O emission inventory compiled by EDGAR and improved with EPTR data for the INGOS project, JRC technical report, EUR 26633 EN, 2014.
- Kirschke, S., et al., Three decades of global methane sources and sinks, *Nature Geoscience*, 6(10), 813-823, doi: 10.1038/NGEO1955, 2013.
- 5 Koffi, E. N., et al., Evaluation of the boundary layer dynamics of the TM5 model over Europe, *Geosci Model Dev*, 9, 3137-3160, doi: 10.5194/gmd-9-3137-2016, 2016.
- Kort, E. A., J. Eluszkiewicz, B. B. Stephens, J. B. Miller, C. Gerbig, T. Nehrkorn, D. B.C., J. O. Kaplan, S. Houweling, and S. C. Wofsy, Emissions of CH₄ and N₂O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, *Geophys. Res. Lett.*, 35(L18808), doi:10.1029/2008GL034031, 10 2008.
- Levin, I., S. Hammer, E. Eichelmann, and F. Vogel, Verification of greenhouse gas emission reductions: The prospect of atmospheric monitoring in polluted areas, *Philosophical Transactions*, 369, 1906-1924, doi: 10.1098/rsta.2010.0249, 2011.
- Levin, I., H. Glatzel-Mattheier, T. Marik, M. Cuntz, and M. Schmidt, Verification of German methane emission inventories and their recent changes based on atmospheric observations, *J. Geophys. Res.*, 104(D3), 3447-3456, doi: 15 10.1029/1998JD100064, 1999.
- Locatelli, R., et al., Atmospheric transport and chemistry of trace gases in LMDz5B: evaluation and implications for inverse modelling, *Geosci. Model Dev.*, 8, 129-150, doi: 10.5194/gmd-8-129-2015, 2015.
- Lopez, M., M. Schmidt, M. Ramonet, J. L. Bonne, A. Colomb, V. Kazan, P. Laj, and J. M. Pichon, Three years of semicontinuous greenhouse gas measurements at the Puy de Dôme station (central France), *Atmos Meas Tech*, 8(9), 3941- 20 3958, doi: 10.5194/amt-8-3941-2015, 2015.
- Manning, A. J., S. O'Doherty, A. R. Jones, P. G. Simmonds, and R. G. Derwent, Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, *J. Geophys. Res.*, 116(D02305), doi:10.1029/2010JD014763, 2011.
- McKain, K., et al., Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts, *Proceedings of the National Academy of Sciences*, 112, 1941-1946, doi: 10.1073/pnas.1416261112, 2015.
- 25 Melton, J. R., et al., Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), *Biogeosciences*, 10(2), 753-788, doi: 10.5194/bg-10-753-2013, 2013.
- Mikaloff Fletcher, S. E., P. P. Tans, L. M. Bruhwiler, J. B. Miller, and M. Heimann, CH₄ sources estimated from atmospheric observations of CH₄ and its ¹³C/¹²C isotopic ratios: 1. Inverse modelling of source processes, *Global Biogeochem. Cycles*, 30 18, doi:10.1029/2004GB002223, 2004a.
- Mikaloff Fletcher, S. E., P. P. Tans, L. M. Bruhwiler, J. B. Miller, and M. Heimann, CH₄ sources estimated from atmospheric observations of CH₄ and its ¹³C/¹²C isotopic ratios: 2. Inverse modelling of CH₄ fluxes from geographical regions, *Global Biogeochem. Cycles*, 18, doi:10.1029/2004GB002224, 2004b.
- Miller, S. M., et al., Anthropogenic emissions of methane in the United States, *Proceedings of the National Academy of Sciences of the United States of America*, 110, 20018-20022, doi: 10.1073/pnas.1314392110, 2013.
- 35 National Academy of Science, Verifying Greenhouse Gas Emissions: Methods to Support International Climate Agreements, The National Academies Press, Washington D.C., doi: 10.17226/12883, 2010.
- Nisbet, E., and R. Weiss, Top-Down Versus Bottom-Up, *Science*, 328(5983), 1241-1243, doi: 10.1126/science.1189936 2010.
- Nisbet, E. G., E. J. Dlugokencky, and P. Bousquet, Methane on the Rise-Again, *Science*, 343(6170), 493-495, doi: 40 10.1126/science.1247828, 2014.
- Popa, M. E., M. Gloor, A. C. Manning, A. Jordan, U. Schultz, F. Haensel, T. Seifert, and M. Heimann, Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, *Atmos. Meas. Tech.*, 3, 407-427, doi: 10.5194/amt-3-407-2010, 2010.
- Prinn, R. G., et al., A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, *J. Geophys. Res.*, 115, 17751-17792, 2000.
- 45 Rigby, M., et al., Renewed growth of atmospheric methane, *Geophys. Res. Lett.*, 35(L22805), doi:10.1029/2008GL036037, 2008.
- Rödenbeck, C., C. Gerbig, K. Trusilova, and M. Heimann, A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, *Atmos. Chem. Phys.*, 9, 5331-5342, doi: 10.5194/acp-9-5331-2009, 2009.
- 50 Saunio, M., et al., The global methane budget 2000-2012, *Earth System Science Data*, 8(2), 697-751, doi: 10.5194/essd-8-697-2016, 2016.
- Schmidt, M., et al., RAMCES: The French Network of Atmospheric Greenhouse Gas Monitoring, *13th WMO/IAEA Meeting of Experts on Carbon Dioxide Concentration and Related Tracers Measurement Techniques*, WMO report 168, 165-174, 2006.
- 55 Schmidt, M., et al., High-precision quasi-continuous atmospheric greenhouse gas measurements at Trainou tower (Orléans forest, France), *Atmos Meas Tech*, 7(7), 2283-2296, doi: 10.5194/amt-7-2283-2014, 2014.
- Spahni, R., F. Joos, B. D. Stocker, M. Steinacher, and Z. C. Yu, Transient simulations of the carbon and nitrogen dynamics in northern peatlands: from the Last Glacial Maximum to the 21st century, *Clim. Past*, 9(3), 1287-1308, doi: 10.5194/cp-9-1287-2013, 2013.
- 60 Spokas, K., J. Bogner, and J. Chanton, A process-based inventory model for landfill CH₄ emissions inclusive of seasonal soil microclimate and CH₄ oxidation, *Journal of Geophysical Research*, 116(G4), doi: 10.1029/2011jg001741, 2011.

- Ulyatt, M. J., K. R. Lassey, I. D. Shelton, and C. F. Walker, Seasonal variation in methane emission from dairy cows and breeding ewes grazing ryegrass/white clover pasture in New Zealand, *New Zealand Journal of Agricultural Research*, 45(4), 217-226, doi: 10.1080/00288233.2002.9513512, 2010.
- 5 VanderZaag, A. C., T. K. Flesch, R. L. Desjardins, H. Baldé, and T. Wright, Measuring methane emissions from two dairy farms: Seasonal and manure-management effects, *Agricultural and Forest Meteorology*, 194, 259-267, doi: 10.1016/j.agrformet.2014.02.003, 2014.
- Vermeulen, A., InGOS Integrated non-CO₂ Greenhouse gas Observing System, Final Report, Grant Agreement Number 284274, http://www.ingos-infrastructure.eu/?wpfb_dl=547, 2016.
- 10 Vermeulen, A. T., A. Hensen, M. E. Popp, W. C. M. van den Bulk, and P. A. C. Jongejan, Greenhouse gas observations from Cabauw Tall Tower (1992–2010), *Atmos. Meas. Tech.*, 4, 617–644, doi: 10.5194/amt-4-617-2011, 2011.
- Wania, R., et al., Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP), *Geosci Model Dev*, 6(3), 617-641, doi: 10.5194/gmd-6-617-2013, 2013.
- Wecht, K. J., D. J. Jacob, C. Frankenberg, Z. Jiang, and D. R. Blake, Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, *Journal of Geophysical Research: Atmospheres*, 119(12), 7741-7756, doi: 10.1002/2014jd021551, 2014.
- 15 Weiss, R. F., and R. G. Prinn, Quantifying greenhouse-gas emissions from atmospheric measurements: a critical reality check for climate legislation, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1943), 1925-1942, doi: 10.1098/rsta.2011.0006, 2011.
- Wennberg, P. O., et al., On the Sources of Methane to the Los Angeles Atmosphere, *Environmental Science & Technology*, 46(17), 9282-9289, doi: 10.1021/es301138y, 2012.
- 20 WMO, Report of the Seventh WMO Meeting of Experts on Carbon Dioxide Concentration and Isotopic Measurement Techniques, Rome, Italy, 7-10 September 1993, 1993.
- WMO, Integrated Global Greenhouse Gas Information System (IG3IS), <http://www.wmo.int/pages/prog/arep/gaw/ghg/IG3IS-info.html>, last access: 22 March 2017, 2016a.
- 25 WMO, WMO Greenhouse Gas Bulletin (GHG Bulletin) - No. 12: The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2015 8pp, http://library.wmo.int/opac/doc_num.php?explnum_id=3084, 2016b.
- Wong, C. K., T. J. Pongetti, T. Oda, P. Rao, K. R. Gurney, S. Newman, R. M. Duren, C. E. Miller, Y. L. Yung, and S. P. Sander, Monthly trends of methane emissions in Los Angeles from 2011 to 2015 inferred by CLARS-FTS observations, *Atmos Chem Phys*, 16(20), 13121-13130, doi: 10.5194/acp-16-13121-2016, 2016.
- 30 Zavala-Araiza, D., et al., Reconciling divergent estimates of oil and gas methane emissions, *Proceedings of the National Academy of Sciences*, 201522126, doi: 10.1073/pnas.1522126112, 2015.

Tables

Table 1: European monitoring stations used in this study. "s.h." is the sampling height (m) above ground, "ST" specifies the sampling type ("I": in-situ measurements; "D": discrete air sample measurements). The last four columns indicate the use of the corresponding station data set in the inversions S1-S4 (see section 3.1 and Table 2).

ID	station name	data provider	lat	lon	alt	s. h.	ST	S1	S2	S3	S4
ZEP	Ny-Alesund	InGOS/NILU ¹	78.91	11.88	474	15	I	•	•	•	•
		NOAA	78.91	11.88	474	5	D	•	•	•	
SUM	Summit	NOAA	72.60	-38.42	3210	5	D	•	•	•	
PAL	Pallas	InGOS/FMI ²	67.97	24.12	565	7	I	•	•	•	•
		NOAA	67.97	24.12	560	5	D	•	•	•	
ICE	Storhofdi,	NOAA	63.40	-20.29	118	9	D	•	•	•	
VKV	Voekovo	InGOS/MGO ³	59.95	30.70	70	6	I		•	•	•
TTA	Angus	InGOS/UoE ⁴	56.55	-2.98	313	222	I	•	•	•	•
BAL	Baltic Sea	NOAA	55.35	17.22	3	25	D				
LUT	Lutjewad	InGOS/CIO ⁵	53.40	6.35	1	60	I	•	•	•	•
MHD	Mace Head	InGOS/UoB ⁶	53.33	-9.90	25	15	I	•	•	•	•
		NOAA	53.33	-9.90	5	21	D	•	•	•	
BIK1	Bialystok	InGOS/MPI ⁷	53.23	23.03	183	5	I				
BIK2						30	I				
BIK3						90	I				
BIK4						180	I				
BIK5						300	I	•	•	•	•
CBW1	Cabauw	InGOS/ECN ⁸	51.97	4.93	-1	20	I				
CBW2						60	I				
CBW3						120	I				
CBW4						200	I	•	•	•	•
OXK1	Ochsenkopf	InGOS/MPI ⁷	50.03	11.82	1022	23	I				
OXK2						90	I				
OXK3						163	I	•	•	•	•
OXK		NOAA	50.03	11.82	1022	163	D				
HEI	Heidelberg	InGOS/IUP ⁹	49.42	8.67	116	30	I	•	•	•	•
KAS	Kasprowy Wierch	InGOS/AGH ¹⁰	49.23	19.98	1987	2	I		•	•	•
LPO	Ile Grande	RAMCES	48.80	-3.58	20	10	D	•	•	•	
GIF	Gif sur Yvette	InGOS/LSCE ¹¹	48.71	2.15	160	7	I	•	•	•	•
TRN1	Trainou	InGOS/LSCE ¹¹	47.96	2.11	131	5	I				
TRN2						50	I				
TRN3						100	I				
TRN4						180	I		•	•	•
SCH	Schauinsland	InGOS/UBA ¹²	47.91	7.91	1205	8	I	•	•	•	•
HPB	Hohenpeissenberg	NOAA	47.80	11.01	985	5	D	•	•	•	
HUN	Hegyhátsál	InGOS/HMS ¹³	46.95	16.65	248	96	I	•	•	•	•
HUN		NOAA	46.95	16.65	248	96	D	•	•	•	
JFJ	Jungfraujoch	InGOS/EMPA ¹⁴	46.55	7.98	3575	5	I	•	•	•	•
IPR	Ispra	InGOS/JRC ¹⁵	45.81	8.63	223	15	I		•	•	•
PUY	Puy de Dome	InGOS/LSCE ¹¹	45.77	2.97	1465	10	I		•	•	•
PUY		RAMCES	45.77	2.97	1465	10	D	•	•	•	
BSC	Black Sea	NOAA	44.17	28.68	0	5	D				
PDM	Pic du Midi	RAMCES	42.94	0.14	2877	10	D	•	•	•	
BGU	Begur	RAMCES	41.97	3.23	13	2	D	•	•	•	
LMP	Lampedusa	NOAA	35.52	12.62	45	5	D	•	•	•	
FIK	Finokalia	RAMCES	35.34	25.67	150	15	D		•	•	

¹ Norwegian Institute for Air Research, Norway

² Finnish Meteorological Institute, Helsinki, Finland

³ Main Geophysical Observatory, St. Petersburg, Russia

⁴ University of Edinburgh, Edinburgh, UK

⁵ Center for Isotope Research, Groningen, Netherlands

⁶ University of Bristol, Bristol, UK

5 ⁷ Max Planck Institute for Biogeochemistry, Jena, Germany

⁸ Energy research Centre of the Netherlands, Petten, Netherlands

⁹ Institut für Umweltphysik, Heidelberg, Germany

¹⁰ University of Science and Technology, Krakow, Poland

¹¹ Laboratoire des Sciences du Climat et de l' Environnement, Gif-sur-Yvette, France

10 ¹² Umweltbundesamt Germany, Messstelle Schauinsland, Kirchzarten, Germany

¹³ Hungarian Meteorological Service, Budapest, Hungary

¹⁴ Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland

¹⁵ European Commission Joint Research Centre, Ispra, Italy

15

20

Table 2: CH₄ inversions

25

inversion	a priori emissions	period	InGOS station	NOAA+RAMCES discrete air samples
S1	EDGARv4.2FT-InGOS	2006-2012	base	●
S2	EDGARv4.2FT-InGOS	2010-2012	extended	●
S3	no <u>detailed</u> a priori <u>inventory</u> ¹	2010-2012	extended	●
S4	EDGARv4.2FT-InGOS	2010-2012	extended	-

¹ [see section 3.1](#)

30 **Table 3:** Atmospheric models

Model	Institution	Resolution of transport model: Horizontal (lon × lat)	Vertical	Meteorology	Background CH ₄ (regional models)
TM5-4DVAR	EC JRC	Europe: 1° × 1° Global: 6° × 4°	25	ECMWF ERA-INTERIM	
TM5-CTE	FMI	Europe: 1° × 1° Global: 6° × 4°	25	ECMWF ERA-INTERIM	
TM3-STILT	MPI-BGC	Europe: 0.25° × 0.25° (STILT) Global: 5° × 4° (TM3)	61 (STILT) 26 (TM3)	ECMWF operational analysis (STILT) ECMWF ERA-INTERIM (TM3)	TM3 ³
LMDZ	LSCE	Europe: ~1.2° × 0.8° Global: ~7° × 3.6°	19	Nudged to ECMWF ERA-INTERIM	
NAME	Met Office	0.5625° × 0.375° ¹ 0.3516° × 0.2344° ²	31 ³ 59 ⁴	Met Office Unified Model (UM)	based on measurements at Mace Head ⁶
CHIMERE	LSCE	0.5° × 0.5°	29	ECMWF ERA-INTERIM	LMDZ ⁶
COMET	ECN	0.17° × 0.17°	60	ECMWF ERA-INTERIM	TM5-4DVAR

¹ for simulation period 01/2006-03/2010

35 ² for simulation period 03/2010-12/2012

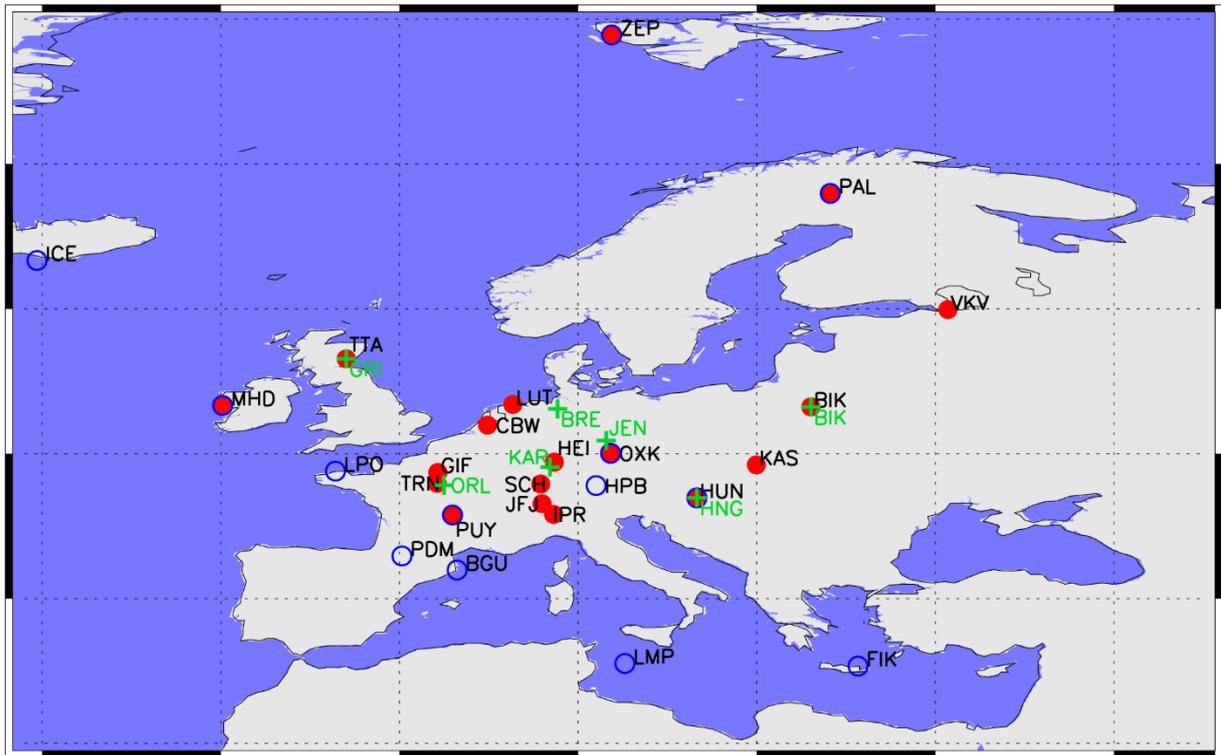
³ for simulation period 01/2006-10/2009

⁴ for simulation period 10/2009-12/2012

⁵ [coupling based on the method of Rödenbeck et al. \[2009\]](#),

⁶ [further optimized in the inversion](#)

40



- stations with in-situ measurements (InGOS)
- stations with discrete air sampling (NOAA / RAMCES)
- + aircraft profiles

35 **Figure 1:** Map showing locations of InGOS atmospheric monitoring stations with in-situ CH₄ measurements (filled red circles), additional stations with discrete air sampling (open blue circles), and the locations of the aircraft profiles (green symbols).

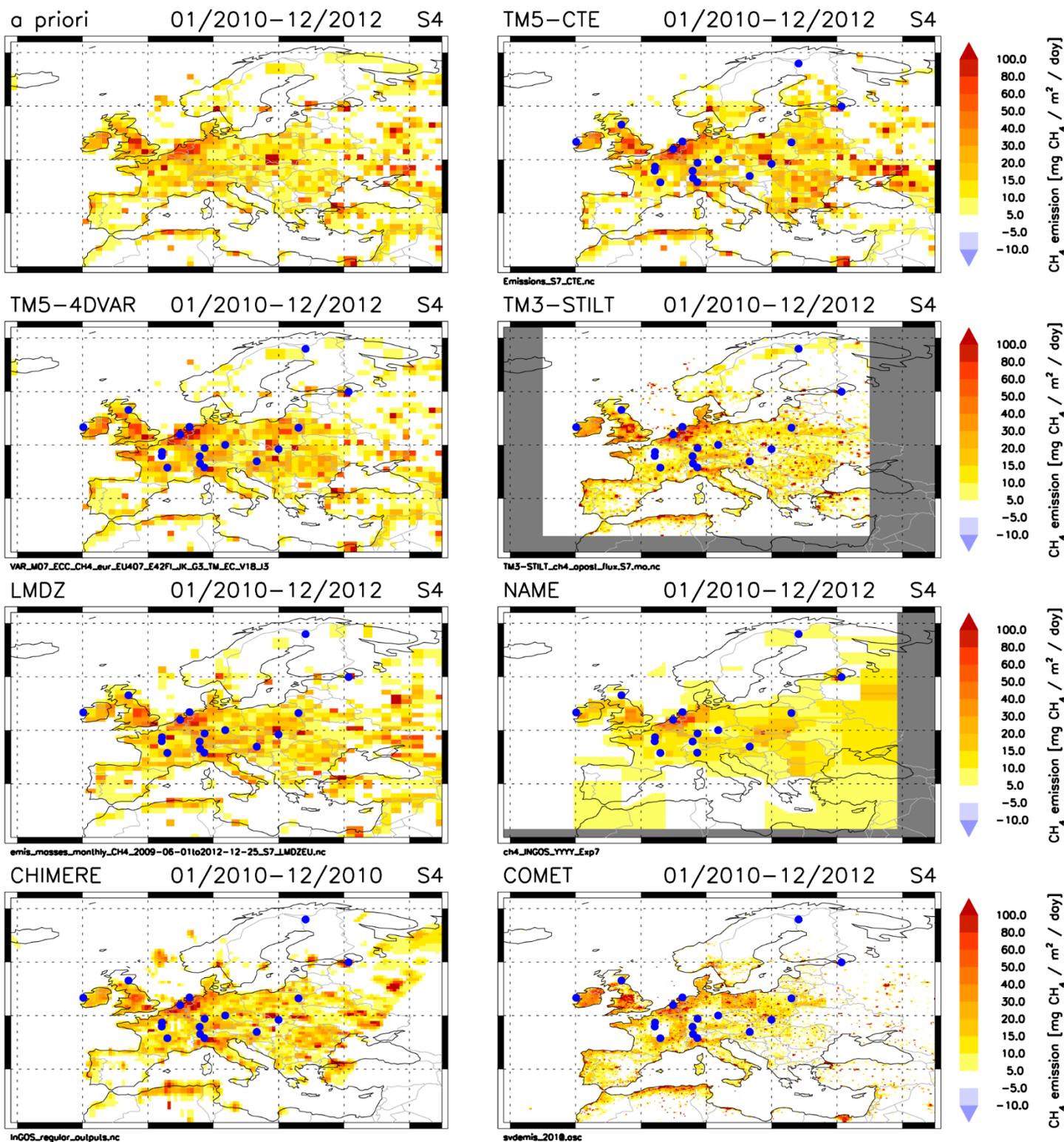


Figure 2: European CH₄ emissions derived from the seven inverse models (inversion S4; average 2010–2012; for CHIMERE only 2010). Filled blue circles are the locations of the InGOS measurement stations. Upper left panel shows a priori CH₄ emissions (as applied in TM5-4DVAR at 1°×1° resolution, while regional models use higher resolution for the a priori emissions).

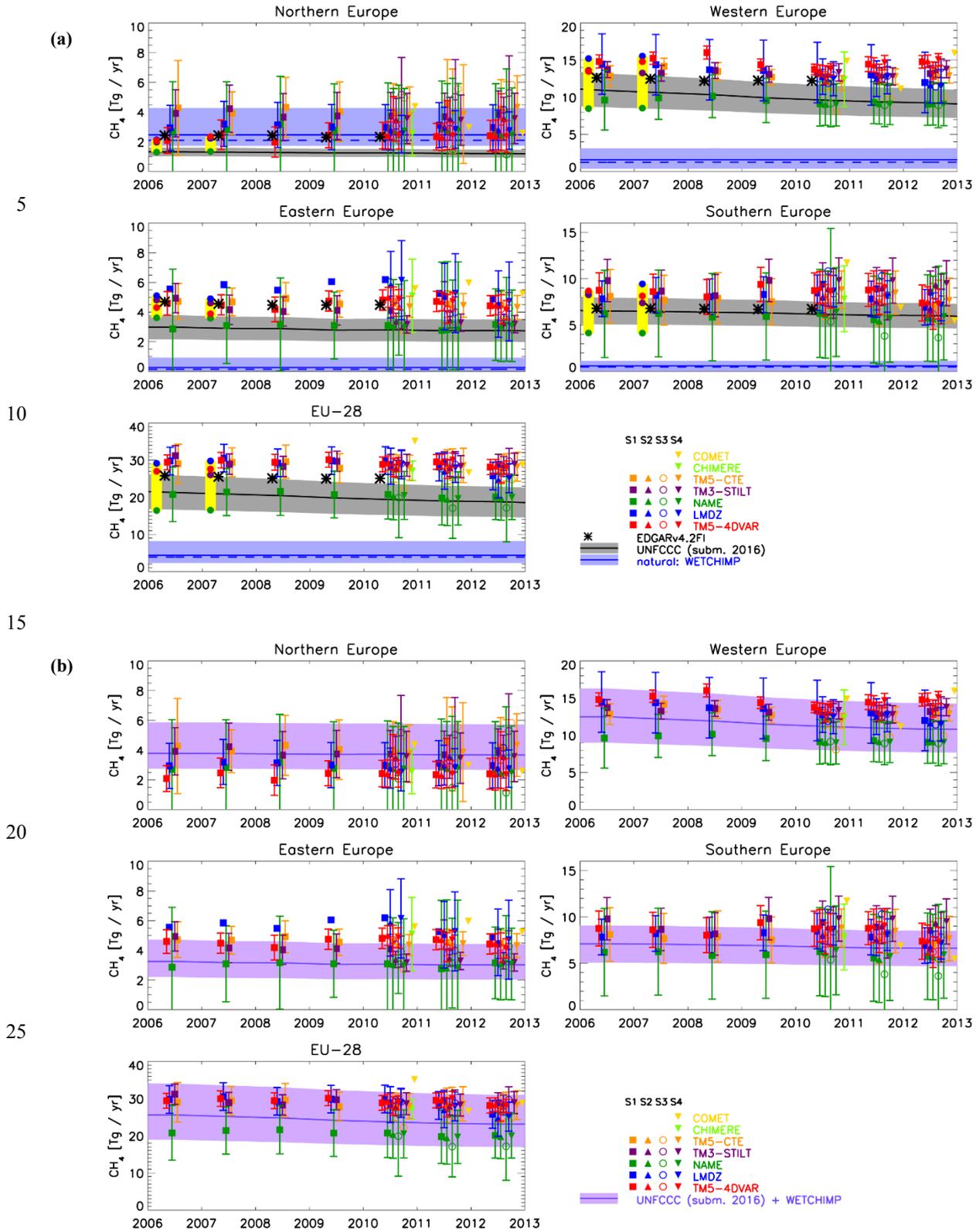
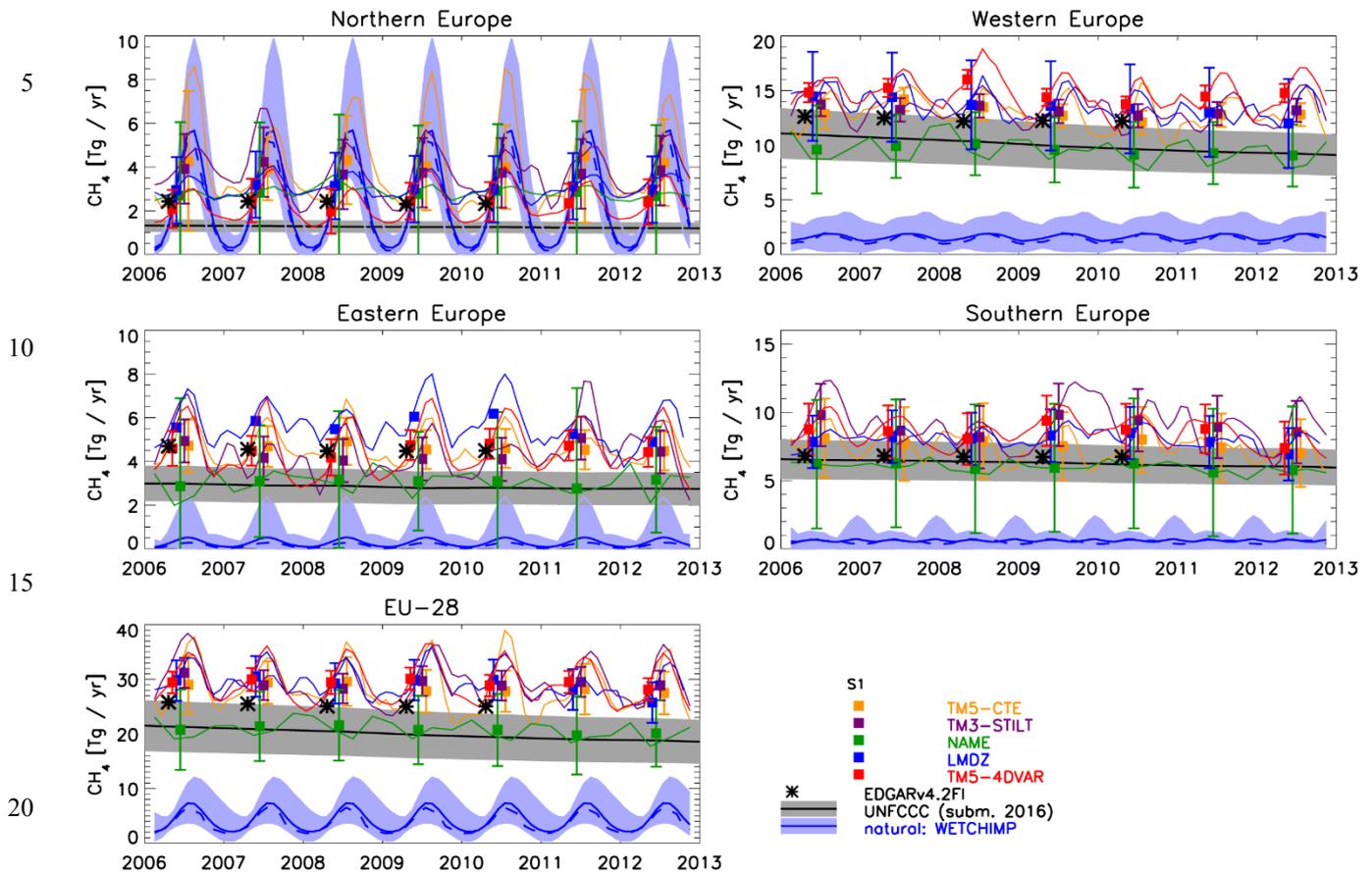


Figure 3: (a) Annual total CH₄ emissions derived from inversions for northern, western, eastern, and southern Europe, and for EU-28 (coloured symbols; bars show estimated 2σ uncertainties). For comparison, anthropogenic CH₄ emissions reported to UNFCCC (black line; grey range: 2σ uncertainty estimate based on National Inventory Reports), and from EDGARv4.2FT-InGOS (black stars) are shown. Furthermore, the blue lines show wetland CH₄ emissions from the WETCHIMP ensemble of seven models (mean (blue solid line); median (blue dashed line); minimum-maximum range (light-blue range)). The previous estimates of total CH₄ emissions from Bergamaschi et al. [2015] for 2006 and 2007 are shown within the yellow rectangles. **(b)** Comparison of annual total CH₄ emissions derived from inversions with the sum of anthropogenic CH₄ emissions reported to UNFCCC and wetland CH₄ emissions from the WETCHIMP ensemble (violet line; the light-violet range is the combined uncertainty range based on the 2σ uncertainty of UNFCCC inventories and the minimum-maximum range of the WETCHIMP ensemble).



25 **Figure 4:** Same as Fig. 3a, but including seasonal variation of CH₄ emissions derived from the inversions (S1 only; 3-monthly running mean (coloured solid lines)), and seasonal variation of wetland CH₄ emissions from the WETCHIMP ensemble of seven models (mean (blue solid line); median (blue dashed line); minimum-maximum range (light-blue range); 3-monthly running mean).

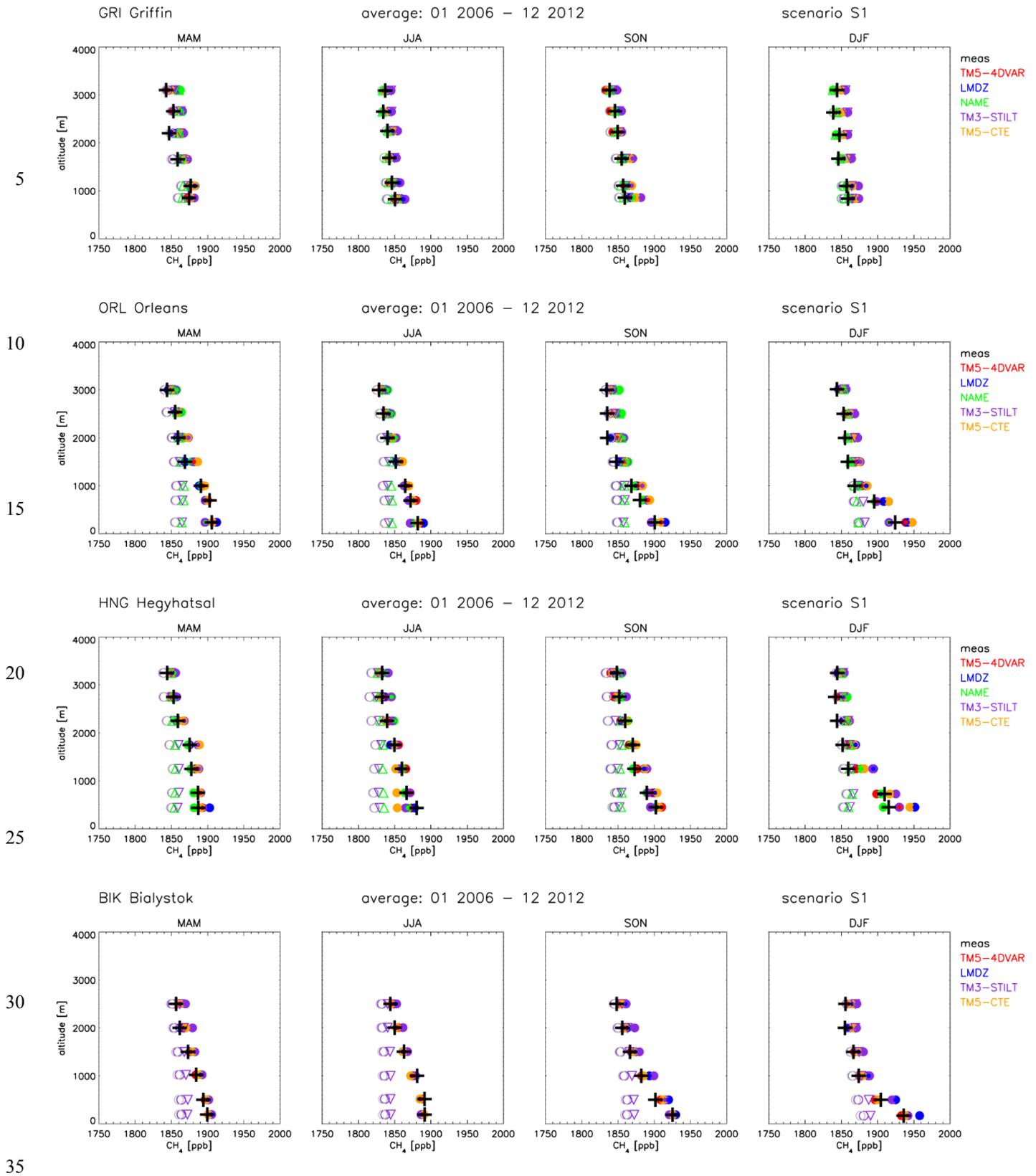


Figure 5: Seasonal averages over all available aircraft profile measurements of CH₄ at Griffon (Scotland), Orléans (France), Hegyhatsál (Hungary), and Białystok (Poland) (black crosses) during 2006–2012 and average of corresponding model simulations (filled coloured symbols). The open circles show the calculated background mole fractions, based on the method of *Rödenbeck et al.* [2009], calculated with TM5-4DVAR for the TM5-4DVAR zoom domain (grey), and for the NAME (green) and TM3-STILT (violet) domains (the latter are, however, only partially visible, since they largely overlap with the background for the TM5-4DVAR zoom domain). The open upper triangles (green) are the background mole fractions used in NAME (based on baseline observations at Mace Head), and open lower triangles (violet) are the background mole fractions used in TM3-STILT (based on TM3 model).

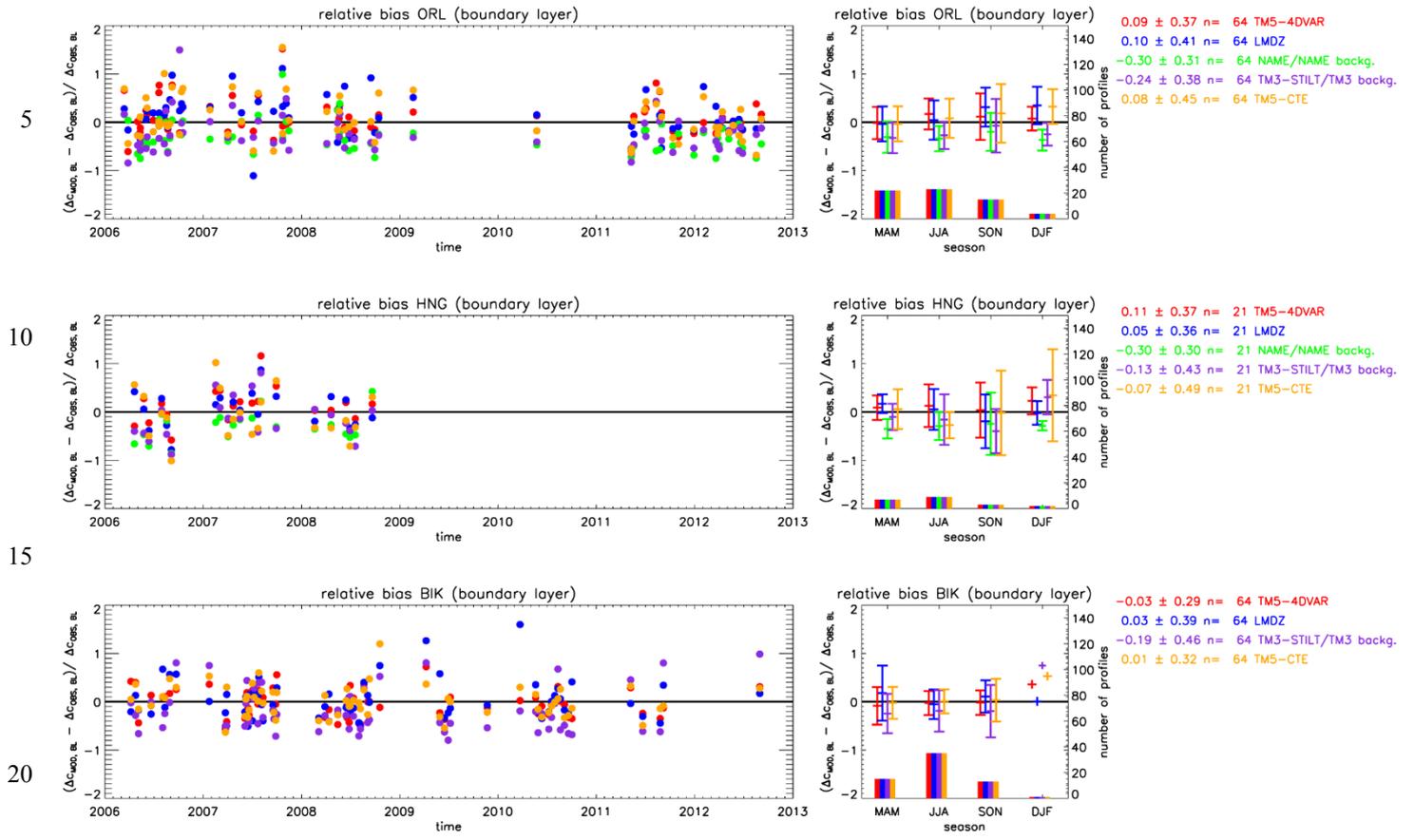


Figure 6: 'Relative' bias within the boundary layer evaluated from simulated and observed CH₄ mole fraction enhancements compared to the background ($rb_{BL} = (\Delta_{\text{MOD, BL}} - \Delta_{\text{COBS, BL}}) / \Delta_{\text{COBS, BL}}$; see section 4.2). For NAME the model enhancement has been evaluated using the NAME background, for TM3-STILT using the TM3 background, while for all other models the TM5-4DVAR background is used. Left: time series (dashed lines: linear fits over entire period); right: seasonal averages (including 1σ standard deviation) with numbers of available profiles given as bargraphs (see right axis). The numbers on the right side are the average relative bias, 1σ standard deviation, and total number of profiles over the entire period.

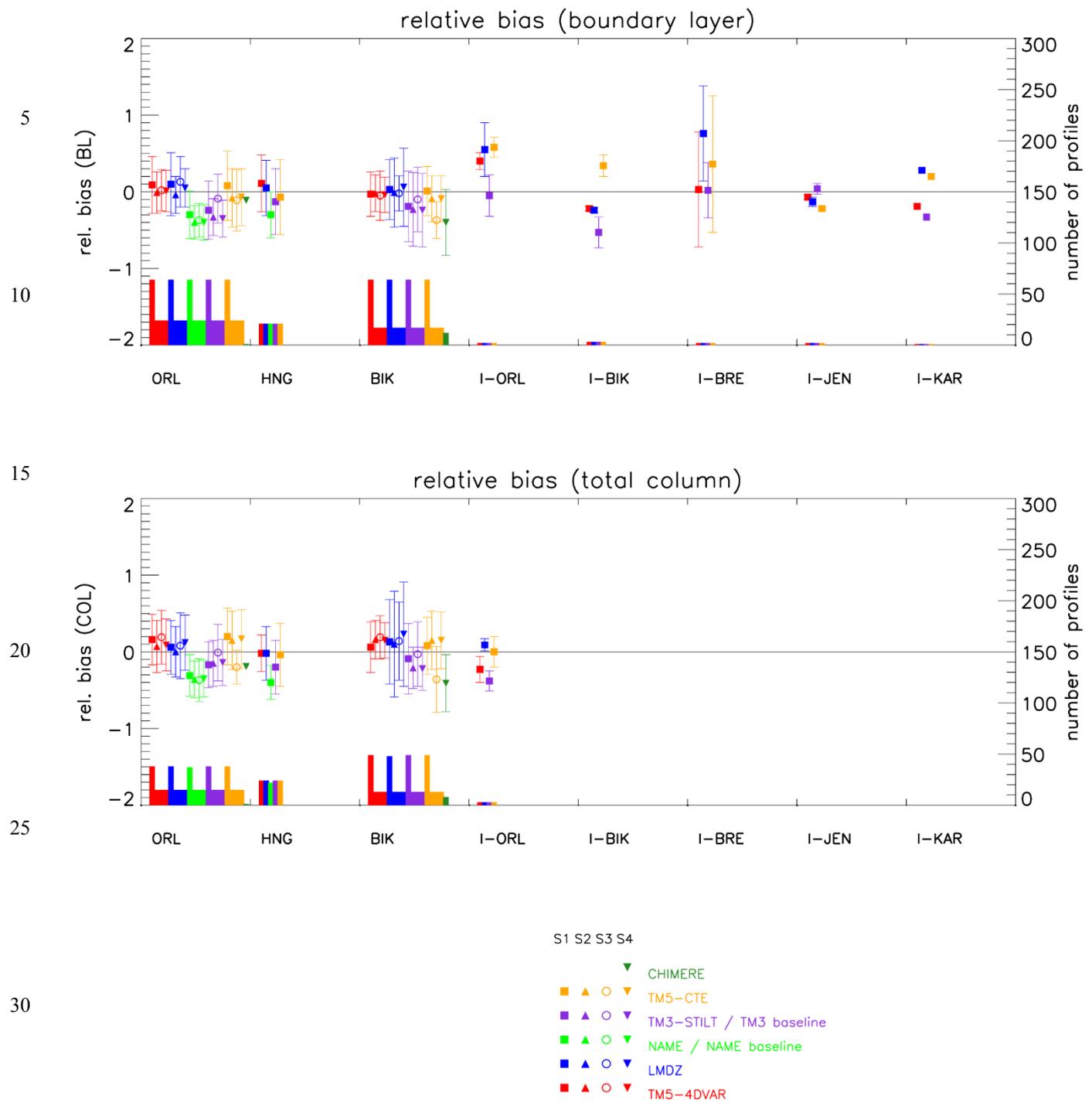


Figure 7: Overview of 'relative' bias at different aircraft sites. Top: 'relative' bias within the boundary layer (rb_{BL}). Bottom: column-averaged 'relative' bias (rb_{COL}). For NAME the relative bias has been evaluated using the NAME background, for TM3-STILT using the TM3 background, while for all other models the TM5-4DVAR background is used. Numbers of available profiles given as bargraphs (see right axis).