

- 5 This document describes the detailed point-by-point response to re-review of Anonymous Referee #1 and Co-Editor Decision letter by Jens-Uwe Grooß (13 Nov 2017)

(1) *comments from Referee / Co-Editor [in italics]*

(2) **author's response [bold]**

(3) **author's changes in manuscript [bold blue]**.

10

1. Response to re-review of Anonymous Referee #1

15 *I do not find the changes between this version and the original manuscript to be compelling enough to change my original stance on this manuscript. It is not clear to this reviewer that this manuscript contributes much to the current literature.*

We do not agree with this statement, since we had provided a point-by-point reply addressing all points raised by all 3 referees and revised significantly the manuscript.

20 We believe that our paper makes a significant contribution to the literature in the field of inverse modelling of European / regional CH₄ emissions and includes many new elements compared to the previous [Bergamaschi et al., ACP, 2015] paper:

(1) use of the new quality-controlled and harmonized InGOS CH₄ in-situ data set including 18 stations. To our knowledge this is the largest (and most consistent) data set of European atmospheric measurements used
25 in any study of top-down estimates of European CH₄ emissions so far.

(2) extension of time series (now 7-years period 2006-2012, compared to 2-years period in [Bergamaschi et al., 2015], including a short analysis of CH₄ trends.

(3) detailed analysis of potential contribution of natural sources

(4) comprehensive validation of model results using an extended set of aircraft observations, providing for
30 the first time quantitative estimates of potential biases in derived regional emissions.

Point #1 ("I find the wetland hypothesis wholly unconvincing"): The authors have added one sentence to the abstract and one sentence to the conclusions.

We did not only include the additional sentences in the abstract and the conclusions, which clearly emphasize the uncertainties of the wetland emissions, especially regarding their spatio-temporal emission distribution (furthermore these uncertainties are discussed in sections 4.1 (already in the original discussion paper)). Following the suggestion of this referee we had included an analysis of the seasonality of derived CH₄ emissions in scenario S3 (which is not using any detailed a priori inventory nor any a priori seasonal cycle; see section 4.1 and new Figure 5S, which confirms that the derived emissions are driven by the observations and not by the a priori inventories). Furthermore, we had addressed the further specific point raised by this referee on this topic, especially the potential impact of the background (see section 4.2, new Figure 14S).

Point #2 ("the methods description is poor, making it hard to gain any insight from the different inversions"): The methods description is still poor. There is still just a single paragraph in the main text describing the inversions.

As explained in our response to the reviewers, the inverse modelling system are described in the supplementary material (SM), section 1 "Atmospheric models" (summarizing the main elements of each system; this description extends over more than 4.5 pages). Furthermore, all seven inverse models are described comprehensively in separate specific papers (see references in the SM).

Since for most models used in this study only smaller updates were applied (compared to previously published applications), we think that is more appropriate to keep the summarizing description of the individual models in the SM.

However, following the request of this referee we had updated the description in the SM (e.g. the applied a priori probability density functions (pdf's) for the individual models, and the assumed uncertainties for the observations (including estimates of model errors). Furthermore, we had included the applied boundary conditions (background) in Table 3 and the information about the optimization of the background in the revised main paper.

Following also the request of the co-editor, we included further model details in the main paper (section 3.2 / table 3), as detailed below (see under "2. Response to Co-Editor Decision letter by Jens-Uwe Grooß (13 Nov 2017)")

It is left to the reader to guess at why the inverse models obtain, in some cases, radically different emissions. Figure 2 is a good example of this. The emissions from COMET look totally different from the others (e.g., why is there a source in Northern Poland that isn't in the prior or any other model?). Figure 2 seems to only be mentioned a single time in the manuscript (in the first paragraph of Section 4.1).

A significant part of the 'visual' differences in the spatial patterns between different models is related to different spatial resolutions (and, as explained in the text, for NAME related to the averaging of emissions at larger distances from the observations). Integrated over larger areas (e.g. whole EU-28), the models show a remarkable consistency (apart from the generally lower CH₄ emissions derived by NAME). Differences on smaller spatial scales are probably partly due to differences in model transport and different weighting of

the observations (i.e. different assumptions of model-data mismatch errors), but may reflect to some extent some noise of the inverse modelling systems.

In order to address this specific new comment of the reviewer, we have added the following short additional paragraph in the manuscript (in section 4.1):

5 "Apart from this specific feature of the NAME model, also some further differences in the spatial patterns derived by the different models are apparent. One example are the relatively high emissions derived by the COMET model in north-western Poland / north-eastern Germany. Such differences on smaller spatial scales are probably partly due to differences in model transport and different weighting of the observations (i.e. different assumptions of model-data mismatch errors), but may reflect to some extent also some noise of the
10 inverse modelling systems."

*In the author's response they state: "But this [understanding the differences between top-down emissions] 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
15 inverse models) and to evaluate the model performance by validation against independent observations." Was this not already done in the 2015 paper by many of the same authors (Bergamaschi et al., "Top-down estimates of European CH₄ and N₂O emissions based on four different inverse models", ACP, 2017)? The authors have just added in a few more models and a few more years of data.*

20 We assume that the referee refers here to our [Bergamaschi et al., ACP, 2015] (and not 2017) paper. As outlined above we think that the current paper provides many new elements and provides a significantly extended and improved analysis of European CH₄ emissions. What the referee calls "just added... a few more years" is a significant extension of the time period (covering now the 7-years period 2006-2012, while in [Bergamaschi et al., 2015] only the 2 years 2006 and 2007 were analyzed).

25 *In general, the conclusions drawn do not seem to be in-line with the analysis. For example, the conclusions of this manuscript state (final paragraph): (2) transport models need to be further improved, including their spatial resolution and in particular the simulation of vertical mixing, aren't some of these models finer-resolution than others and include different treatments of vertical mixing? Do these models actually perform better?*

30 We had presented detailed conclusions in-line with our analysis in section "5. Conclusions". The referee refers here to our more general conclusions at the end (which are rather recommendations to further improve top-down estimates in the future). In our study the models with higher spatial resolution do not perform better than TM5-4DVAR with resolution of 1x1 degrees (see short discussion at beginning of section
35 4.2). Nevertheless we consider the further development of high-resolution models essential to further improve the top-down estimates (e.g. [Henne et al., 2016]).

Point #3 ("it's not clear to this reviewer that their "novel" approach to estimate bias is actually an advancement"): Regarding the 3rd point, their "novel" approach to estimate bias is not particularly useful for estimating biases (as the authors claim). The difference between simulated and measured enhancements is the term that defines the model-data mismatch in the cost-function. As I showed in my previous review, $c_{obs} - c_{mod} \equiv \Delta c_{obs} - \Delta c_{mod}$. Following this, if there were no difference between the simulated and measured enhancement then the inverse model would not deviate from the prior. Stated another way, if $\Delta c_{obs} = \Delta c_{mod}$ then the top-down emissions would be equal to the prior.

As explained in our response to the reviewers we do not agree with the statement of the referee. The term mentioned by the reviewer is indeed part of the cost function, which the inverse modelling systems aim to minimize. However, this applies to the observations that are actually used (assimilated) in the inversions, while we are analyzing here independent observations that were not used in the inversion - which is a common method to validate inverse models. The concept of using independent observations for validation of inverse models is described in detail in the recent review paper by Michalak *et al.*, [2017] and has been applied in many studies (e.g. [Alexe *et al.*, 2015; Monteil *et al.*, 2013; Houweling *et al.*, 2014; Bergamaschi *et al.*, 2013]). The rationale behind this approach is to analyze, how well the inverse model perform in areas which are less constrained by the observations. As mentioned in our initial reply to the reviewer, and explained in section 4.2 of the manuscript, especially the validation of the vertical profiles (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.

Commonly, however, such comparisons against independent observations are performed to diagnose only qualitatively, if the inverse models have biases. The novel aspect of our method is that we provide for the first time quantitative estimates of the derived regional emissions. We had included in the revised version also an analysis of the correlation between the derived relative bias (rb_{BL}) and regional model emissions around the aircraft profile sites, which confirms that rb_{BL} can be used to diagnose biases in the regional model emissions (new Figure 16S).

Since we consider Point #3 as absolutely invalid, no changes have been made in the revised version regarding this point.

2. Response to Co-Editor Decision letter by Jens-Uwe Grooß (13 Nov 2017)

1. It is not clear to me, why one model that seems to be always at the low end of the deduced CH₄ emissions: the Lagrangian NAME model. Is that because of the Lagrangian formulation of the model? Or because the lack of observation stations?

The low emissions of the NAME model are likely - at least partly - related to the positive bias in the background CH₄ used for NAME (as discussed in section 4.2 and shown in the figures 14S (for the aircraft data), since the regional models invert the difference between the observations and the assumed background.

In fact, also at most continental atmospheric monitoring stations (which are applied in the inversion), the background used for NAME (and TM3-STILT) is significantly higher (~10 ppb) compared to the TM5-4DVAR background (Figure 15S).

The potential impact of the significant positive bias of the background in NAME (and TM3-STILT) on the derived CH₄ emissions is also mentioned in the conclusions of the revised version.

It would be certainly useful to perform additional tests with the NAME model using the TM5 baselines in order to quantify the impact of the baseline on the emissions. Unfortunately, however, the NAME team was not yet able to perform this additional test, since they had no further resources after the end of the InGOS project.

We don't see any obvious reason why the Lagrangian models should yield lower emissions compared to Eulerian models (however, to our knowledge, this has not been investigated in any study in a systematic way). In fact, also STILT is a Lagrangian model and yields similar emissions as the global / regional Eulerian models.

Regarding the observations: All models use the same observational data set (however, the details, how the observations are used in the inversions, differ, in particular the assumed model-data mismatch error, and hence the weighting of the individual observations in the different models).

[Short general description of the applied the assumed model-data mismatch errors has been included in section 3.2 \(see below\)](#)

2. Can you understand that the annual cycle of wetland emissions is reduced in Northern Europe with respect to the WETCHIMP study? and opposite than the annual cycle increases in the three other parts? Would you say that the WETCHIMP study is incorrect in that respect?

We can only speculate about the potential reasons. E.g. the CH₄ emissions in the wetland models are highly sensitive to the assumed assumed temperature dependence (Q₁₀ values), but also on water table and soil properties (in particular, soil organic carbon content). We note that for Northern Europe the seasonality in the a posteriori emissions derived by TM5-CTE are actually very similar compared to the WETCHIMP mean / median. Nevertheless, it is obvious that the other models derive smaller seasonal cycles in Northern Europe. For the other regions in Europe the derived seasonality is still broadly within the minimum-maximum range of WETCHIMP (even though the seasonality in the mean/median of the WETCHIMP is clearly smaller).

We would not say that WETCHIMP is 'incorrect', but clearly the uncertainties are very large, as reflected in the very different spatio-temporal emission patterns of the different individual WETCHIMP inventories.

To Point #2 of the review:

Although the model descriptions are updated in the supplement, I would likely ask you to add some more information to section 3.2. /table 2, such that in the paper the model diversity would be better understandable

without reading the supplement. That should be a few more sentences about the methods (Lagrangian/Eulerian) underlying data (if different) resolution etc.

We included further details in section 3.2 / table 3 of the main paper

- short description of the applied meteorological fields in the text

5 **- short description of the applied inverse modelling technique (added both in the text and table 3)**

- type of model (Eulerian / Lagrangian) has now been added also in table 3 (was already in the text)

- short general description of the applied uncertainties of the observations (including the measurement and model uncertainties)

(model resolutions were already described in the text and included in table 3)

10

technical corrections:

line 11 4.3 (2.3-8.2) Tg CH₄ yr⁻¹

'Tg' has been added

line 26 globally averaged tropospheric CH₄ mole fraction

15 **'tropospheric' has been added**

(although this is not exactly the average over the entire troposphere - but with the explanation following in the text ('global average from marine surface sites') it should be clear.)

20

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.

25 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., 118, 7350–7369, doi: 10.1002/jgrd.50480, 2013.

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.

30 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.

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.

35 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.

Monteil, G., et al., Comparison of CH₄ inversions based on 15 months of GOSAT and SCIAMACHY observations, J. Geophys. Res. Atmos., 118, doi:10.1002/2013JD019760, 2013.

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.

The inverse models infer total CH₄ emissions of 26.8 (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 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) Tg 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.

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 simulated and measured enhancements of CH₄ compared to the background, integrated over the entire boundary layer and over the lower troposphere. The estimated average regional biases range between -40% and 20% at the aircraft profile sites in France, Hungary and Poland.

1 Introduction

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 2016, relative to 1750 (NOAA Annual Greenhouse Gas Index (AGGI) [Butler and Montzka, 2017]). The globally averaged tropospheric CH₄ mole fraction reached a new high of 1842.8 ± 0.5 ppb in 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 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 10.1 ± 2.3 ppb yr⁻¹ during 2014-2016 [Dlugokencky, 2017].

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; Saunio 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; Saunio et al., 2016]). In addition, satellite retrievals of GHGs have also been used in a number of studies. In particular, 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 Białystok, 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° (latitude) for the regional models. Most models are driven by meteorological fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis [Dee *et al.*, 2011]. In case of STILT, the operational ECMWF analyses were used, while for NAME meteorological analyses of the Met Office Unified Model (UM) were employed. 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
30 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.

The inverse modelling systems applied in this study use different inversion techniques. TM5-4DVAR, LMDZ, and TM3-STILT use 4DVAR variational techniques, which allow to optimize emissions of individual grid cells. These 4DVAR techniques employ an adjoint model in order to iteratively minimize the cost function using a quasi-Newton [Gilbert and
35 Lemaréchal, 1989] or conjugate gradient [Rödenbeck, 2005] algorithm. The NAME model applies a simulated annealing technique, a probabilistic technique for approximating the global minimum of the cost function. In CHIMERE and COMET, the inversions are performed analytically, after reducing the number of parameters to be optimized by aggregating individual

grid cells before the inversion. TM5-CTE applies an Ensemble Kalman Filter (EnKF) [Evensen, 2003], with a fixed-lag smoother [Peters et al., 2005].

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 different models have different approaches to estimate the uncertainties of the observations (including the measurement and model uncertainties), which determine the weighting of the individual observations in the inversions. In general, the estimated model uncertainties depend on the type of station, and for some models (TM5-4DVAR and NAME) also on the specific synoptic situation. The individual inverse ~~models~~ modelling system 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). Apart from this specific feature of the NAME model, also some further differences in the spatial patterns derived by the different models are apparent. One example are the relatively high emissions derived by the COMET model in north-western Poland / north-eastern Germany. Such differences on smaller spatial scales are probably partly due to differences in model transport and different weighting of the observations (i.e. different assumptions of model-data mismatch errors), but may reflect to some extent also some noise of the inverse modelling systems

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 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 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 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, 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 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 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).

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.

- 5 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
10 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 of countries), the reported uncertainties per category were aggregated as described in *Bergamaschi et al.*
15 [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
20 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].

- 25 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].
30 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.

- 35 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
40 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 6S), 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 8S shows the correlation coefficients, bias, 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 9S, 10S, 11S). In addition, we use also aircraft measurements from the IMECC campaign in September / October 2009 (Figure 12S). 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 for the STILT and NAME domains, 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). The three global inverse models (i.e. TM5-4DVAR, TM5-CTE, and LMDZ) show in general only a small average relative bias (rb_{BL} between -7% and 10%) at the three aircraft sites. 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).

These negative biases are likely related to the positive bias in the background CH₄ 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). There is no clear seasonal cycle in the relative bias apparent and the variability between the different seasons is generally small (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 seasonal variations of the average bias at ORL and HNG. However, especially at HNG the total number of profiles is rather small ($n=22$), which limits the analysis of potential seasonal transport biases.

Figure 13S shows the relative bias of the CH₄ enhancements integrated over the lower troposphere, defined as:

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

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 CH₄ 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).

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 differences of the relative bias integrated over the lower troposphere 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) 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 12S). 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 CH_4 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 *Rödenbeck et al.* [2009]. This would allow to derive the modelled CH_4 enhancements more accurately.

5 Conclusions

We have presented estimates of European CH_4 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 CH_4 emissions of 26.8 (20.2-29.7) Tg CH_4 yr⁻¹ are derived (mean, 10% percentile, and 90% percentile from all inversions), compared to total anthropogenic CH_4 emissions of 21.3 Tg CH_4 yr⁻¹ (2006) to 18.8 Tg CH_4 yr⁻¹ (2012) reported to UNFCCC. Our analysis highlights the potential significant contribution of natural emissions from wetlands (including peatlands and wet soils) to the total European emissions, with total wetland emissions of 4.3 (2.3-8.2) Tg CH_4 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 CH_4 emissions with maxima in summer. However, the NAME model calculates only a weak seasonal cycle, with small maximum (of EU-28 total CH_4 emissions) in winter. Furthermore, it needs to be emphasized that wetland inventories have large uncertainties and show large differences in the spatial distribution of CH_4 emissions.

Taking into account the estimates of the WETCHIMP ensemble, the bottom-up and top-down estimates of total EU-28 CH_4 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).

The relative bias rb_{BL} 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 accuracy of the estimated relative biases, however, depends on the quality of 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.
- Andrews, A. E., et al., CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's
5 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.
- Bergamaschi, P., et al., Satellite cartography of atmospheric methane from SCIAMACHY onboard ENVISAT: 2. Evaluation
10 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:
15 10.1002/jgrd.50480, 2013.
- 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.
- Dee, D. P., et al., The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi: 10.1002/qj.828, 2011.
- Dlugokencky, E., Trends in Atmospheric Methane, www.esrl.noaa.gov/gmd/ccgg/trends_ch4/, last access: 19 September
35 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.
- 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.
- 40 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.
- 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.
- 45 Evensen, G., The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynam.*, *53*(4), 343–367, doi: 10.1007/s10236-003-0036-9, 2003.
- 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.
- 50 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.
- Gilbert, J. C., and C. Lemaréchal, Some numerical experiments with variable-storage quasi-Newton algorithms. Mathematical Programming, *Mathematical Programming*, *45*, 407-435, 1989.
- Hammer, S., D. W. T. Griffith, G. Konrad, S. Vardag, C. Caldow, and I. Levin, Assessment of a multi-species in situ FTIR for
55 precise atmospheric greenhouse gas observations, *Atmos Meas Tech*, *6*(5), 1153-1170, doi: 10.5194/amt-6-1153-2013, 2013.
- 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.
- 60 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.

- 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.
- Janssens-Maenhout, G., D. Guizzardi, P. Bergamaschi, and M. Muntean, On the CH₄ and N₂O emission inventory compiled by EDGAR and improved with EPRTR 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.
- 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, 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: 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), 3941-3958, 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.
- 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*, 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.
- 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: 10.1126/science.1247828, 2014.
- Peters, W., J. B. Miller, J. Whitaker, A. S. Denning, A. Hirsch, M. C. Krol, D. Zupanski, L. Bruhwiler, and P. P. Tans, An ensemble data assimilation system to estimate CO₂ surface fluxes from atmospheric trace gas observations, *J. Geophys. Res.*, 110(D24304), doi: doi:10.1029/2005JD006157, 2005.
- 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.
- Rigby, M., et al., Renewed growth of atmospheric methane, *Geophys. Res. Lett.*, 35(L22805), doi:10.1029/2008GL036037, 2008.
- Rödenbeck, C., Estimating CO₂ sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport, Tech. Rep. 6, Max-Planck-Institut für Biogeochemie, Jena, [http://www.bgc-jena.mpg.de/mpg/website/Biogeochemie/Publikationen/Technical Reports/tech_report6.pdf](http://www.bgc-jena.mpg.de/mpg/website/Biogeochemie/Publikationen/Technical%20Reports/tech_report6.pdf), http://www.bgc-jena.mpg.de/uploads/Publications/TechnicalReports/tech_report6.pdf, 2005.
- 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.

- Saunois, 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.
- 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.
- 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.
- 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.
- Vermeulen, A. T., A. Hensen, M. E. Popa, 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.
- 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.
- 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.
- 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.
- 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	Voeikovo	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 detailed a priori inventory 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	Model type	Meteorology	Background CH ₄ (regional models)	Inversion technique
TM5-4DVAR	EC JRC	Europe: 1° × 1° Global: 6° × 4°	25	Eulerian	ECMWF ERA-INTERIM		4DVAR
TM5-CTE	FMI	Europe: 1° × 1° Global: 6° × 4°	25	Eulerian	ECMWF ERA-INTERIM		EnKF
TM3-STILT	MPI-BGC	Europe: 0.25° × 0.25° (STILT) Global: 5° × 4° (TM3)	61 (STILT) 26 (TM3)	Lagrangian (STILT) Eulerian (TM3)	ECMWF operational analysis (STILT) ECMWF ERA-INTERIM (TM3)	TM3 ⁵	4DVAR
LMDZ	LSCE	Europe: ~1.2° × 0.8° Global: ~ 7° × 3.6°	19	Eulerian	Nudged to ECMWF ERA-INTERIM		4DVAR
NAME	Met Office	0.5625° × 0.375° ¹ 0.3516° × 0.2344° ²	31 ³ 59 ⁴	Lagrangian	Met Office Unified Model (UM)	based on meas. at Mace Head ⁶	simulated annealing
CHIMERE	LSCE	0.5° × 0.5°	29	Eulerian	ECMWF ERA-INTERIM	LMDZ ⁶	analytical
COMET	ECN	0.17° × 0.17°	60	Lagrangian	ECMWF ERA-INTERIM	TM5-4DVAR	analytical

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

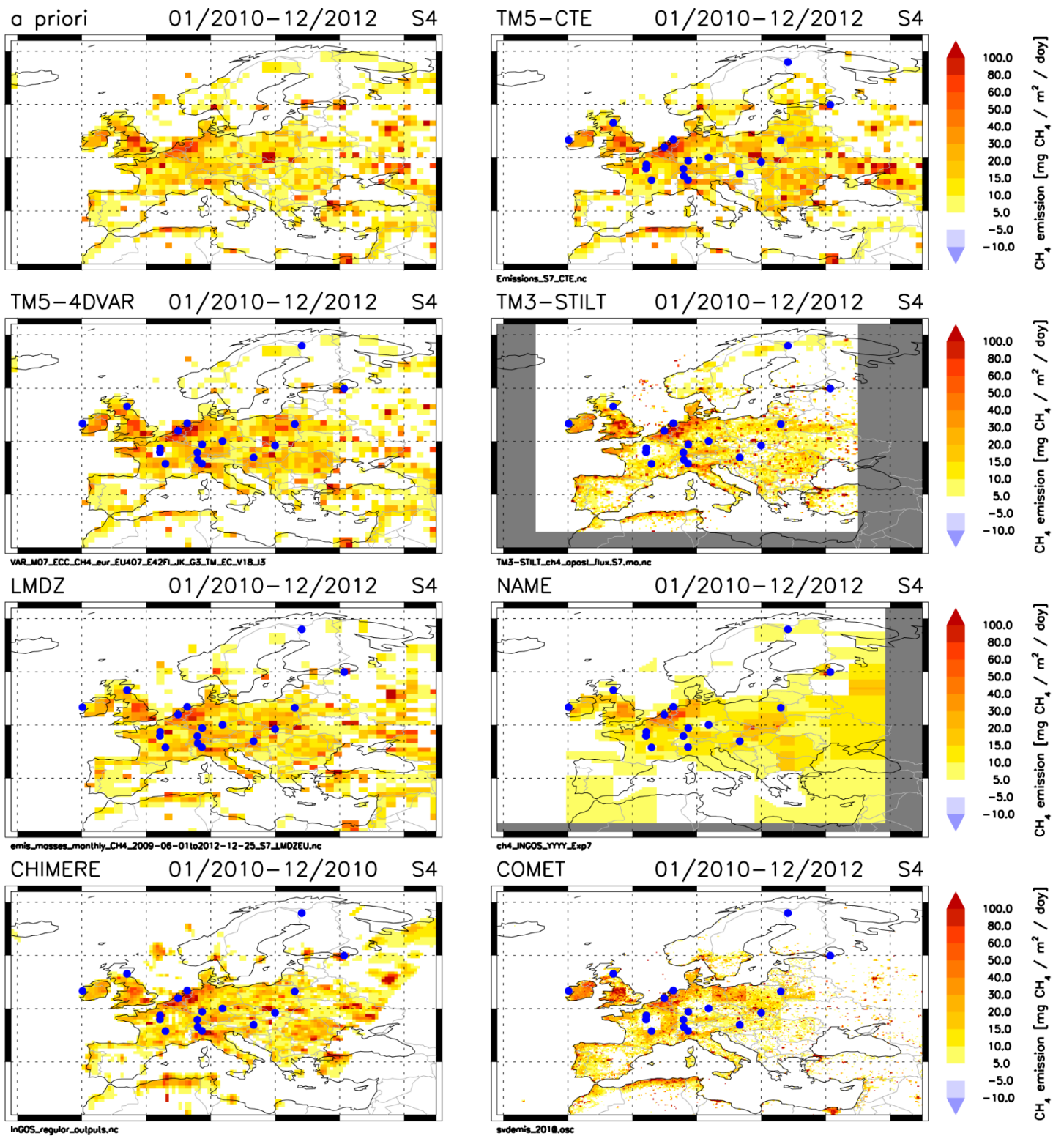


Figure 2: European CH₄ emissions derived from the seven inverse models (inversion S4; average 2010–2012; for CHIMERE only 2010). Filled blue circles are the locations of the InGOS measurement stations. Upper left panel shows a priori CH₄ emissions (as applied in TM5-4DVAR at 1°x1° resolution, while regional models use higher resolution for the a priori emissions).

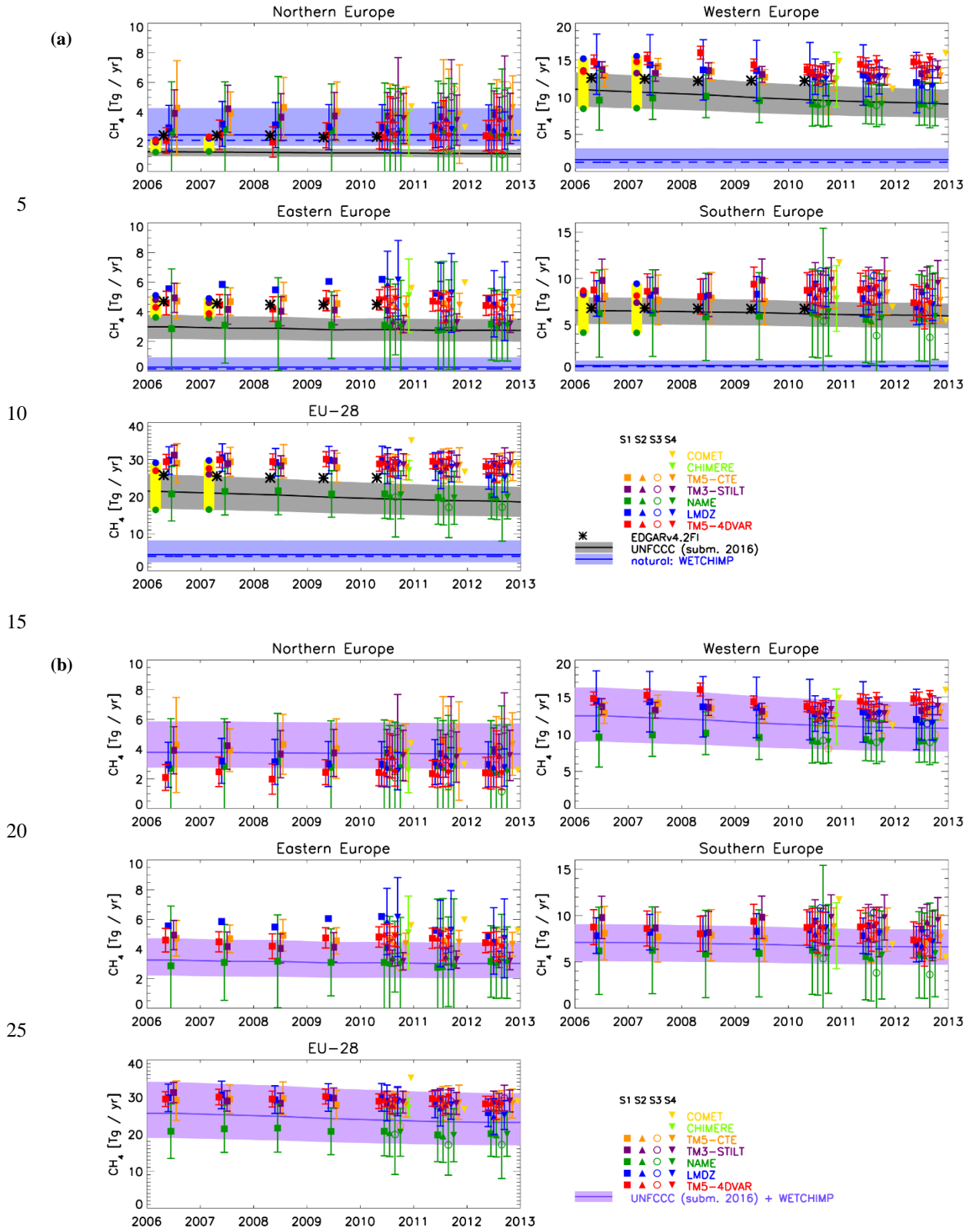
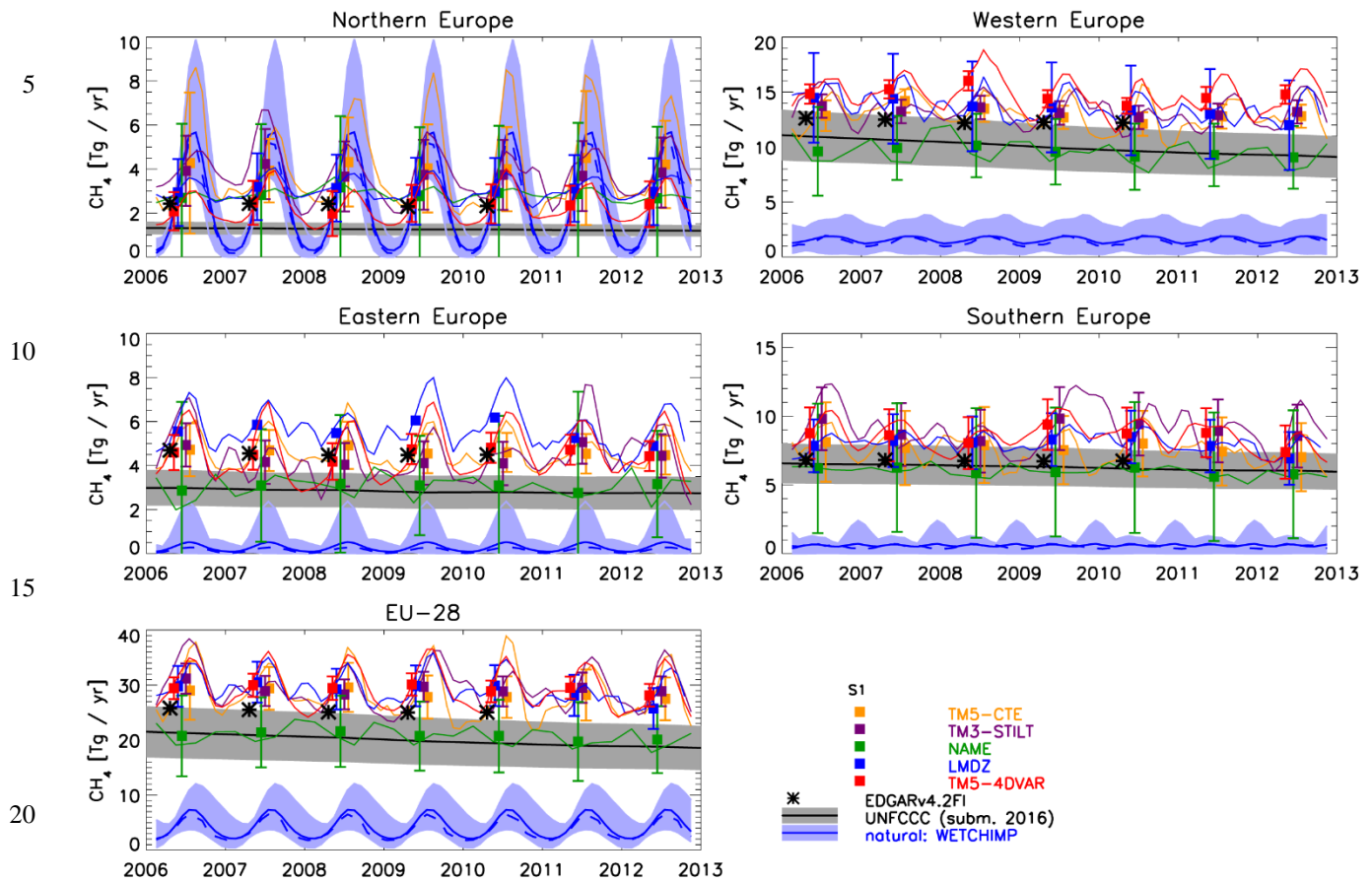


Figure 3: (a) Annual total CH_4 emissions derived from inversions for northern, western, eastern, and southern Europe, and for EU-28 (coloured symbols; bars show estimated 2σ uncertainties). For comparison, anthropogenic CH_4 emissions reported to UNFCCC (black line; grey range: 2σ uncertainty estimate based on National Inventory Reports), and from EDGARv4.2FI-InGOS (black stars) are shown. Furthermore, the blue lines show wetland CH_4 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 CH_4 emissions from Bergamaschi et al. [2015] for 2006 and 2007 are shown within the yellow rectangles. (b) Comparison of annual total CH_4 emissions derived from inversions with the sum of anthropogenic CH_4 emissions reported to UNFCCC and wetland CH_4 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 CH_4 emissions derived from the inversions (S1 only; 3-monthly running mean (coloured solid lines)), and seasonal variation of wetland CH_4 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).

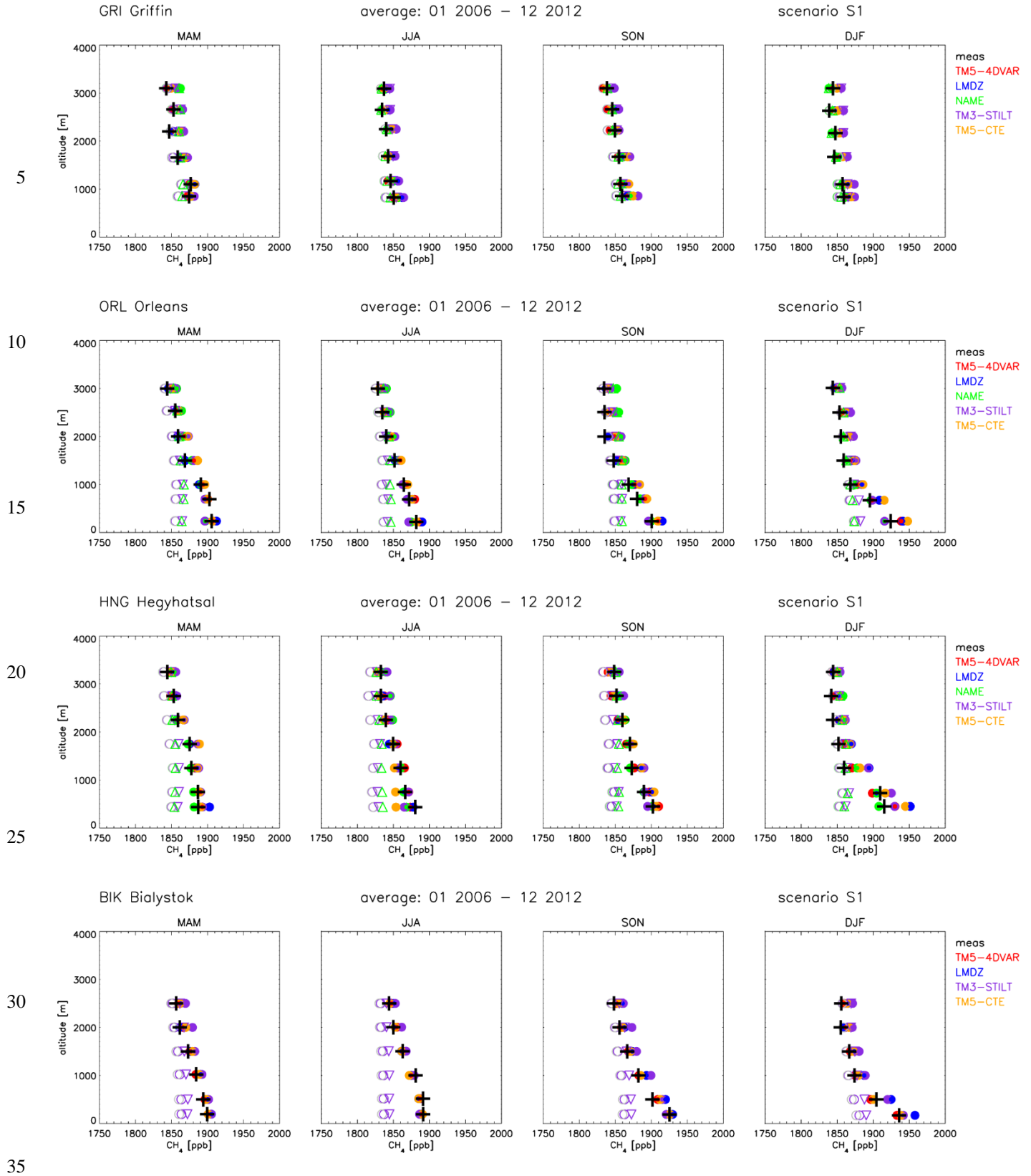


Figure 5: Seasonal averages over all available aircraft profile measurements of CH_4 at Griffin (Scotland), Orléans (France), Hegyhátsál (Hungary), and Białystok (Poland) (black crosses) during 2006–2012 and average of corresponding model simulations (filled coloured symbols). The open circles show the 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).

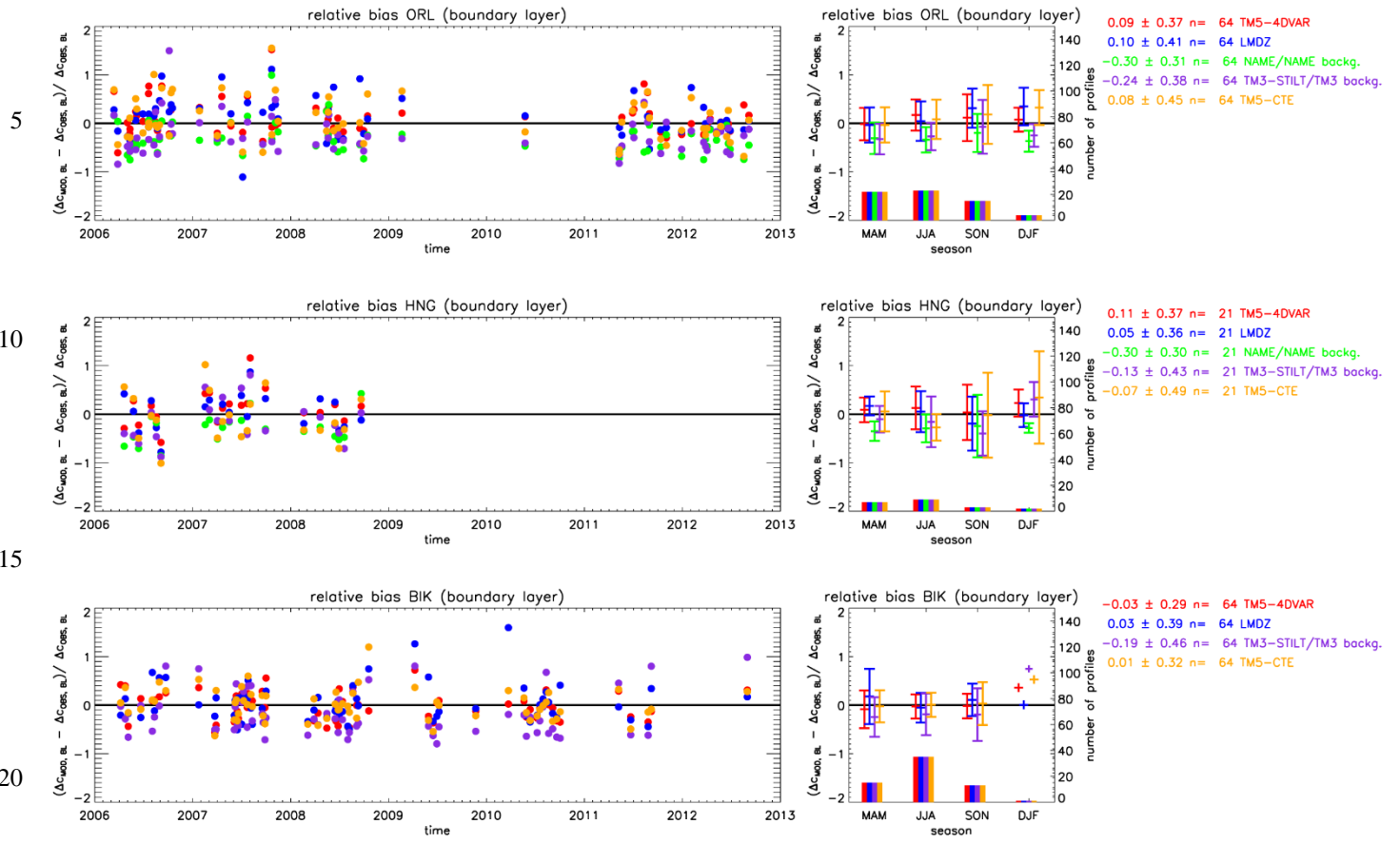


Figure 6: 'Relative' bias within the boundary layer evaluated from simulated and observed CH₄ mole fraction enhancements compared to the background ($rb_{BL} = (\Delta C_{\text{MOD, BL}} - \Delta C_{\text{OBS, BL}}) / \Delta C_{\text{OBS, BL}}$; see section 4.2). For NAME the model enhancement has been 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; right: seasonal averages (including 1 σ standard deviation) with numbers of available profiles given as bargraphs (see right axis). The numbers on the right side are the average relative bias, 1 σ standard deviation, and total number of profiles over the entire period.

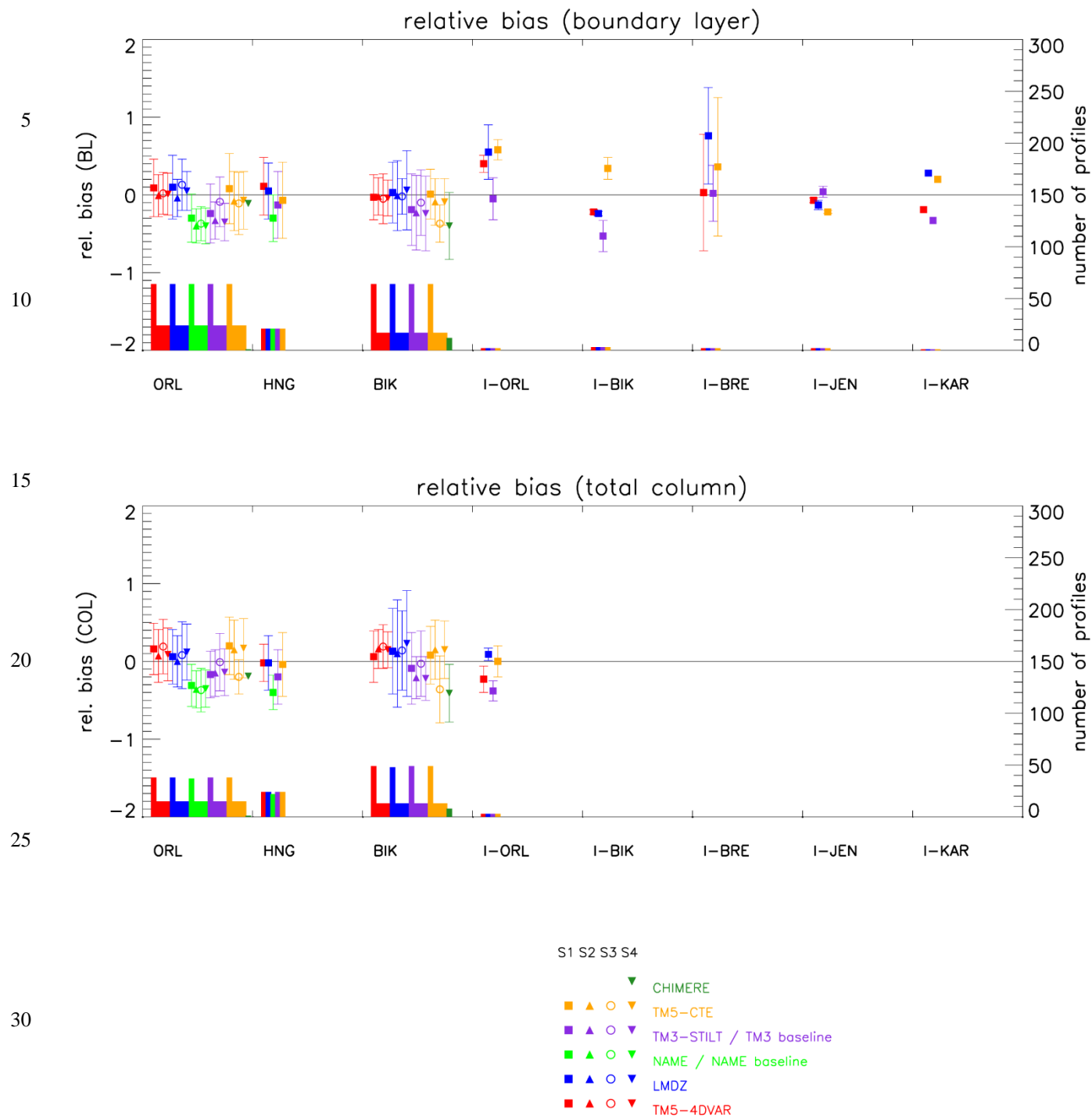


Figure 7: Overview of 'relative' bias at different aircraft sites. Top: 'relative' bias within the boundary layer (rbBL). Bottom: column-averaged 'relative' bias (rbCOL). For NAME the relative bias has been evaluated using the NAME background, for TM3-STILT using the TM3 background, while for all other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