Answers to reviewer # 1

We thank the reviewer for his/her comments. We answer them in the following. The comments are in bold and our answers in normal font, the new text being in blue.

This manuscript investigates the spatio-temporal resolution of global methane inversions using different observing systems. This information is useful for a proper interpretation of inversion results obtained using existing observing systems, and for the design of new systems. An expected outcome is that larger regions are better resolved than smaller regions. Less expected is the finding that smaller regions are better resolved at the seasonal than at the inter-annual time scale. While trying to understand this, a couple of questions came up, as explained below, which I found have not been dealt with adequately yet. To make this study acceptable for publication this will have to be repaired, and/or explained more clearly.

GENERAL COMMENTS

Figure 3 and the tables depend on a detection criterion, like threshold SNR value which should be exceeded to declare a region as detected or not. This criterion should be defined explicitly in the text. From one of the figure captions I found out that the criterion corresponds to SNR=1. This sounds like a rather loose criterion. Wouldn't something like a 95% confidence criterion be more appropriate? Whatever choice is made it should be stated and motivated clearly.

We have modified Section 2.4 and added the missing information:

"Our criterion consists in evaluating the ability of the observing systems to detect CH_4 anomalies of a given amplitude, defined by the reference inversion. For this, we define a signal-to-noise ratio:

- the inversion with surface measurements is chosen to provide the signal as the data covers a long time window (2000-2011) as compared to the two other observing systems. This longer window makes it possible to sample the CH_4 IAV more robustly than a 2-3 year inversion. We assume that the fluxes inferred by this inversion are representatitve of state-of-the art inversions currently published. The signal is actually the CH_4 anomalies for the various time scales derived from REFSURF.

- for the three observing systems (SURF, IASI and GOSAT), the Bayesian posterior errors of the year-to-year changes of CH_4 fluxes, computed

from the Monte-Carlo ensemble as described in Section 2.3, constitute the noise associated to each observing system.

Finally, the criterion for detecting CH_4 anomalies is that the signal-to-noise ratio is larger than 1 (≈68% confidence). "

It would be more robust to use SNR=2 i.e. 95% but then almost nothing would be considered as detected so that would prevent any constructive comments. We also have stressed this point in the conclusion "Our criterion is based on a 68% confidence interval (1 sigma). At almost all regional time-space scales (except in NorthAmbor, AfrEquat at the longer time-scales and a few cases in India, Indonesia, EastEurRussia and FarEastSib), the three observing systems would fail the test at 2 sigmas (95%), a more stringent criterion commonly used in other scientific communities."

In this study the REFSURF inversion is used as reference, representing what the true variability would be like. As long as we don't know the true flux the results of an inversion may seem a defensible choice. However, this is only true as long as the validity of this approximation doesn't interfere with the conclusions that are derived from it in the end. Since this reference set of fluxes comes from an inversion of surface data itself, it suffers from the same flux detection limitations as the SURF inversion. Suppose that the setups of REFSURF and SURF were statistically equivalent, wouldn't you expect SNR=1? I mean if their posterior uncertainties are the same then REFSURF would be like a random instance of the posterior uncertainty of SURF. In this case what remains is equivalent to a comparison of the posterior uncertainties of the SURF, IASI, and GOSAT inversions.

The comment above has implications for the conclusions regarding the scale dependency of flux detection. For example, if REFSURF is not capable of resolving small-scale variability, it will generate noise (depending on the a priori constraints). If the use of GOSAT leads to better-constrained small-scale fluxes it may end up 'detecting' the noise of the REFSURF inversion, rather than the variability of the true fluxes at that scale. All we learn in the end is that GOSAT is better able to resolve small-scale fluxes than the surface network. That is something we could have concluded already looking only at their posterior uncertainties. Then what is the added value of the method that is used here?

REFSURF and SURF couldn't be statistically equivalent: REFSURF is driven by the observation vector whereas SURF is driven by the statistics

in the covariance matrices (of observations and the state vector) and by the (transport-)model.
In this paper we assume that REFSURF reasonably represents the discussed scale as other state-of-the art inversions. We agree with the reviewer's comment on the implications if this assumption was wrong, and we tried to clarify even more this assumption in the text e.g. in Section 2.4 "We assume that the fluxes inferred by this inversion are representatitve of state-of-the art inversions currently published." and "However, the quality of the chosen signal remains debatable and our diagnostic for GOSAT and IASI may be pessimistic in areas where SURF signal-to-noise ratio is low." We could have compared the noises of the three systems. But positionning the noises with the signal allows to go further than comparing the three observing systems. Indeed, the largest noise may nevertheless be small enough compared to the signal we want to detect (or the smallest one be too large). This is why the signal computed from the 12-year REFSURF inversion is used to normalize the noises. The aim is to state whether the inter-annual variability at various time and spatial scales can be detected by one or several of the observing system. This is now stated explicitly in Section 2.4:
"Comparing signal-to-noise ratios amounts to comparing noises normalized by the expected signals. The normalization provides an absolute criterion to assess the time scales and regions at which the CH_4 anomalies are reliable."

The choice of reference period for calculating the signal should be explained better. It is chosen because 'it corresponds to a period of minimum atmospheric growth rate'. I guess this means that it has a minimal contribution from the long-term trend. However, shouldn't it be representative of the entire period also? The fact that you get anomalies that are predominantly positive suggests it is not. As a result your IAV signal will inevitably contain signal from a time scale outside the 3-year IAV window. It becomes even worse for the annual and seasonal time scale. I don't see how the method avoids signals from longer time-scales affecting the seasonal time scale. Wouldn't I have been better to take out variations on longer time-scales before computing seasonal anomalies?
We agree that the way we defined flux anomalies had to be made clearer. Our aim is not to isolate the various frequencies in the signal due to the fluxes (at seasonal, annual, year-to-year scales and long-term trend). Since our noise refers to the year to year changes in methane emissions at various scales, we defined the signal i.e. "anomalies", as the difference between one given season/year/3year-period and a reference (seasonal/yearly/3yearly mean) to

be consistent with the noise definition. Doing so, a given time scale contains contributions from other time scales. The signal is here mostly to give an order of magnitude, at various time scales, of what we want to detect, considering the year-to-year noises defined for the three observing systems.
The choice of 2004-2005 is mostly arbitrary and we agree that, as it corresponds to a minimum in methane growth rate and internannual variations it leads to more positive anomalies for the longer time scales. This is less true for smaller spatial and temporal scales. We have modified Section 2.4:
"CH_4 regional flux anomalies are defined here as the deviation from a reference of the CH_4 inferred fluxes for various time periods, from the monthly to the 3-yearly scale. The reference is the 2004-2005 mean over the same time-period. The aim of this definition is to get the order of magnitude of the year-to-year changes at various time scales. As the 2004-2005 reference corresponds to a period of minimum atmospheric methane growth rate (Dlugokencky et al., 2011), it leads to more positive anomalies for the longer time scales. "
Taking out the longer time scales to get at the amplitude of seasonal cycles would be another study.

The role of the prior flux uncertainty should be explained better. I wonder if some regions get already 'detected' without using any data, just because the prior uncertainty is already small enough to satisfy the detection criterion. It would explain why some regions are detected without the observing contributing any significant constraint (for example IASI detecting fluxes from NorthAmericanBoreal, when only data between 30S and 30N are used).
Section 2.3 now states more clearly how the prior uncertainty is used to compute the posterior uncertainty (used as noise):
"The posterior error statistics (the "noise" for our study) are estimated as follows:

- we estimate the ratio of posterior to prior standard deviations of the annual flux errors $r = \frac{\sigma_a}{\sigma_b}$ from the ensemble, a quantity which is more robust than σ_a and σ_b individually for small ensembles (because some of the underspread affects the prior and the posterior in a similar way); the number of members in the ensemble depends on the time scale e.g. 10 members for the yearly time scale (10 inversions, each one covering 1 year), 120 members for the monthly time scale

- we estimate the posterior standard deviations of the annual flux errors by multiplying r to the known value of σ_b i.e. the one implied by our

error covariance matrix (computed from the above assumptions)

- the posterior standard deviations of the pluri-annual flux errors errors for n years is obtained by applying a factor of $\frac{1}{\sqrt{n}}$ to the previous result, assuming that the errors are uncorrelated from one year to the next

- the posterior standard deviations of the difference between fluxes from one year to the next (i.e. the error on the IAV for two consecutive years) is computed by applying an inflation factor of $\sqrt{2}$ to the previous result, still assuming that the errors are uncorrelated from one year to the next. We assume this apporach to be a conservative hypothesis since in reality some of the transport and retrieval errors are recurrent, thereby inducing positive correlations and reducing the inflation factor."

Our prior fluxes have nearly no any interannual variability and are therefore not supposed to detect anomalies in general. However, we checked that our prior error statistics reflect this property by confronting them with the anomalies in the same way as what we present for the posterior. The property is verified for all regions and time scales (i.e. the detection rate is marginal or null) except for NorthAmBor and BorN at the seasonal time scale, where detection rates of 58% and 37%, respectively, are obtained. This result was correctly intuited by the reviewer and may suggest that our prior error statistics in NorthAmBor are underestimated. We have added a note of caution in the main text: "The detection rate is above 50% for the three observing systems in this region (Table 1), but, in contrast to the other regions and to the other time scales, the prior error statistics already lead to detection rates of 58% for the prior. This shows that the Tropical IASI soundings do not add information for this region and at this time scale, as expected. "

Another factor influencing the scale dependency of flux detection is the accuracy at which posterior fluxes are approximated. Give that only 10 Mont Carlo ensemble members are used this accuracy cannot be that high (see for example the appendix of Pandey et al, 2016 for a formula to compute the uncertainty of a Monte Carlo derived uncertainty for a given number of iterations). Although the limited ensemble size should not introduce a scale dependency, the number of iterations per inversion could do that, because, depending on the search algorithm used, the large scales may be solved first being the dominant eigenvectors of the

optimization problem.
As now stated in Section 2.3, "we estimate the ratio of posterior to prior standard deviations of the annual flux errors $r = \frac{\sigma_a}{\sigma_b}$ from the ensemble, a quantity which is more robust than σ_a and σ_b individually for small ensembles (because some of the underspread affects the prior and the posterior in a similar way)."
No information is given about the number of iterations that is used, but in our experience the M1QN3 could converge slowly. Therefore additional information about the convergence of small-scale fluxes is needed.
The convergence of the inversions with M1QN3 was stopped on the criterion of the ratio final/initial gradient norm: it must be less than 0.01. This information has been added in Section 2.1.

Somewhere in the discussion a note of caution is required that the posterior uncertainties are derived without proper accounting for systematic errors in the satellite retrieval and transport model. Because of this, despite the use of real data, the detection statistics probably end up being rather optimistic. The use of Desroziers recipe for error tuning does not account for the neglect of off diagonals in the R matrix.
A "note of caution" has been added in the discussion in Section 5, Conclusions: "We also have neglected the impact of likely state-dependent systematic errors in current satellite retrievals and transport models that further reduce the inversion performance to an unknown extent."

SPECIFIC COMMENTS

Page 3, line 5: How appropriate is it here to use prior fluxes without IAV? It means that the prior is biased with regard to IAV, and as a result the posterior IAV will be low biased too (assuming that all other statistical assumptions are satisfied).
In inversion, the choice of no prior IAV is generally made to minimize the influence of prior emissions on the inferred signals and let observations generate the IAV part of it. We choose this assumption following the same spirit although we agree it may smooth the generated IAV. Indeed, Bergamaschi *et al.* (2013) tested such "flat" priors (their S3 scenarios) and showed that the IAV was derived from the assimilated data. We added a sentence about this in Section 2.1: "It is important here to recall that the prior fluxes (fires excepted) have no inter-annual variability (IAV). This choice is made for IAV to be generated by atmospheric observations and atmospheric transport and chemistry and not by prior IAVs of emissions (and sinks) which are still uncertain or even controversial (e.g. Schaefer et al. (2016); Hausmann

et al. (2016); Nisbet et al. (2014))"

Page 4, line 1: I'm trying to understand the logic of the sqrt(2). How do you define inter-annual variability? Wouldn't it be the difference from one year to another? A sqrt(2) inflation rather suggests the variability of the 2-yearly flux in Tg/(2 year). How does that fit with the 3 year time windows? Apart from this I don't see why the assumption of uncorrelated errors would lead to a conservative estimate. I would rather think of posterior uncertainties as being negatively correlated because of limitations in independently resolving the yearly fluxes. Because of these complications I wonder how appropriate it is to address the 3 yearly time scale using a one year inversion anyway.

The explanation in Section 2.3 and 2.4 have been made clearer; regarding the $\sqrt{2}$ factor: "the posterior standard deviations of the difference between fluxes from one year to the next (i.e. the error on the IAV for two consecutive years) is computed by applying an inflation factor of $\sqrt{2}$ to the previous result, still assuming that the errors are uncorrelated from one year to the next. We assume this approach to be a conservative hypothesis since in reality some of the transport and retrieval errors are recurrent, thereby inducing positive correlations and reducing the inflation factor." One issue with this hypothesis is that the limitation induced by observing systems to solve independently the methane fluxes also leads to negatively correlated errors. Remember that we use standard deviations, so that the standard deviation for a difference is the standard deviation of the sum of the two terms, here one year and the next one; since the two terms are the same (because we have only one year in the ensembles), we get that the standard deviation of the difference from one year to the next is $\sqrt{2}$ $\times$ the standard deviation of the available year.

Page 4, line 6: 'The uncertainty in OH (5% after optimization)' Which scale does this refer to? If it is the global scale, then how about the uncertainty per latitude band?

5% refers to the global scale and we did not compute the uncertainty for the different latitudinal bands. However, as prior error are already low for OH, we expect very little changes compared to the prior values.

Page 9, line 7: 'PBSURF signal is twice as often detected' Why is this? I guess it depends on the size of the regions in PBSURF compared to the scale of the regions that are evaluated. If the latter are smaller then wouldn't you rather expect that the large-region inversion suppresses the within-region variability? In that case they would become harder, rather than easier to detect.

Some further discussion at this point would be helpful.
This is a good point. Large-region based inversions relying on sparse surface networks generally generate larger IAVs than pixel-based inversions when aggregating the results of the latter at the same regional scale. One illustration of this can be found in Pison *et al.*, 2013. This can be understood as a large-region based inversion scales up or down a whole region and changes can be large, especially when the atmospheric constraints are far from the region (e.g. South America in Pison *et al.*, 2013). In pixel based inversions, as we only prescribe loose spatial correlation, changes appear to be more located around the stations. We added this in Section 4.2: "The large-region-scale inversion means that the spatial variability of the prior is kept within each region and is only scaled (contrary to REFSURF, which is performed at the pixel scale i.e. is able to vary only a few pixels to match the data). This difference in the methods may lead to very different spatial variability in each of the regions of interest (Figure 4), a larger variability allowing a better detection rate with our criterion. Indeed, the large-region-scale inversion may lead to larger variability than pixel-based inversions in some regions (e.g. Pison et al., 2013) because of the homothetic scaling of the pixels composing each region in PBSURF (correlations between pixels of 1) as opposed to the individual scaling of model pixels with soft constraints in REFSURF (spatial correlations less than 1)."

TECHNICAL COMMENTS

Page 9, line 7: 'REFSURF' i.o. 'SURF'. Please check if there are other instances where REFSURF was meant.
OK

Page 11, line 11: 'acknowledge' i.o. 'aknowledge'
OK.

Figure 5: What happens to the whisker-boxes at the 2Y time scale? I guess they become too compressed to see. If so, then please mention this somewhere (it shows up in other figures also).
There are no whisker-boxes for the 3Y scale since there is only one window of three years in each of the 3-year periods (the whisker-boxes for Y are made with 3 values, 6 values for 6M, etc).

Figure 5, 6 & 7, caption: 'aggregation' i.o. 'agregation'
OK.

Answers to reviewer #2

We thank the reviewer for his/her comments. We answer them in the following. The comments are in bold and our answers in normal font, the new text being in blue.

This paper shows results from a set of CH4 inversions using three different observation sets (in situ, IASI, and GOSAT) to test whether anomalies in flux can be detected across a range of time and spatial scales. The ultimate goal is to determine whether such inversions can be used to attribute methane flux signals, like the change in global growth rate through the 2000s, to a particular region or regions and, perhaps, biogeochemical processes. The authors have done a lot of work to make the results statistically meaningful, the approach is generally sound, figures and tables are informative, and the discussion is accurate, if perhaps not fully satisfying. In my opinion the material is clearly worthy of publication in ACP after satisfying the concerns of the reviewers.

The paper suffers at times from lack of clarity and some inverse methodological issues exist, which are well-characterized by Anonymous Reviewer #1 and the comment from T. G. Nuñez Ramirez. I did not find the tables too difficult, but they do take some focus.

We have particularly re-written Sections 2.3 and 2.4 to make them clearer and answered the comments of reviewer #1 and T.G. Nuñez Ramirez.

These issues aside, the question that remains to me is: so what? What are the implications of the findings for using CH4 measurements and inverse models to understand the underlying processes? What is the message for carbon cycle science? To my reading the answer to the title question is: NO, except on the broadest of scales and strongest of signals (seasonality), which doesn't really require very extensive measurements or sophisticated mathematical techniques and holds little useful information. This is a serious problem for understanding the current CH4 budget, for projecting future interactions of CH4 and climate, and for designing mitigation policies to reduce the radiative forcing of CH4. The paper alludes to some of the most egregious short-comings, but never really comes out and says our data and techniques are inadequate and what should be done about it.

We modified the end of the paper (Section 5 Conclusions) to acknowledge that the situation of methane atmospheric inversions may not be as optimistic as found in current papers, although our study has limitations that

we also acknowledge more clearly (e.g. underestimation of the signal, overestimation of the noises).

I fully agree with other comments that setting the detection criterion to SNR =1 is a very low bar for attributing anomalies to specific locations and processes.
It would be more robust to use SNR=2 i.e. 95% but then almost nothing would be considered as detected so that would prevent any constructive comments. We also have stressed this point in the conclusion "Our criterion is based on a 68% confidence interval (1 sigma). At almost all regional time-space scales (except in NorthAmbor, AfrEquat at the longer time-scales and a few cases in India, Indonesia, EastEurRussia and FarEastSib), the three observing systems would fail the test at 2 sigmas (95%), a more stringent criterion commonly used in other scientific communities."

The paper is also sometimes seemingly overly optimistic about the model ability to capture signals, e.g., Conclusions line 8-11, where having any detectable anomalies (≈25% on average) is called 'fair to good' and Abstract, where regional scale signals are said to be 'properly detected.'
Taking into account this comment, we found that we had chosen a too optimistic way of computing the noises. We have therefore corrected the problem and modified the results and the text accordingly.

Clearly something much better than current observations and/or existing model formulations is needed. I think the paper should not shy away from such a statement and point out specifically where the problems reside in the analysis. The fact that the detectability depends on the underlying (modeled) signal configuration is further indictment of the overall flux analysis method. The statements that inversions 'should always include an uncertainty assessment', 'attribution... needs more attention', and 'more observations and ... improved transport' are platitudes that don't require a detailed analysis like the one produced in this paper. Go ahead and give the discussion some punch.
We rephrased the end of the paper (Section 5: conclusions) in this sense, although one has to consider that our work has limitations and is not positioned as the ultimate "killer" of atmospheric inversions. We now acknowledge more clearly the challenges for atmospheric inversions and for the work presented here. We still think that the message to deliver systematically an uncertainty assessment in inversion papers is useful to mention as it is not always reported in current papers. Indeed, the cost of running Monte-Carlo ensembles or the explicit computation of posterior errors with the Hessian

matrix or the difficulty in designing relevant sensitivity studies often limit the uncertainty analysis proposed in papers. Our work also shows the critical importance to do so systematically to have the proper level of significance of the inferred fluxes.

Minor Comments:

The analysis does not address transport issues at all, although perhaps it could. Such analysis could include impact of transport uncertainty on inference of fluxes in unobserved regions (e.g., satellite data in dark or high latitudes) and resulting 'noise.' Expand discussion or delete from Conclusions lines 32-33.

OK, this part has been deleted.

Not clear that detection of anomalies at grid scale in Amazon is robust. Depends on signal, which may not be realistic from sparse data constraint. Maybe examine more closely or moderate expectations.

We moderated our statements here.

Can we detect regional methane anomalies? A comparison between three observing systems.

Cindy Cressot[1], Isabelle Pison[1], Peter J. Rayner[2], Philippe Bousquet[1], Audrey Fortems-Cheiney[1], and Frédéric Chevallier[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA/CNRS/UVSQ, Gif-sur-Yvette, France.
[2]School of Earth Sciences, University of Melbourne, Melbourne, Australia

Correspondence to: I. Pison (isabelle.pison@lsce.ipsl.fr)

Abstract. A Bayesian inversion system is used to evaluate the capability of the current global surface network and of the space-borne GOSAT/TANSO-FTS and IASI instruments to quantify surface flux anomalies of methane at various spatial (global, semi-hemispheric and regional) and time (seasonal, yearly, 3-yearly) scales. The evaluation is based on a signal-to-noise ratio analysis, the signal being the methane fluxes inferred from the surface-based inversion from 2000 to 2011 and the noise (i.e. precision) of each of the three observing systems being computed from the Bayesian equation~~for each of the three inversions using either surface or satellite data~~. At the global and semi-hemispheric scales, all observing systems ~~properly~~ detect flux anomalies at ~~all the tested time-scales~~most of the tested time scales. At the regional scale, ~~seasonal flux signals are properly detected by all~~ some seasonal flux anomalies are detected by the three observing systems, but year-to-year ~~changes~~ anomalies and longer-term trends are only poorly detected. Moreover, reliably detected regions depend on the reference surface-based inversion used as ~~a~~ the signal. Indeed, tropical flux inter-annual variability, for instance, can be attributed mostly to Africa in the reference inversion or spread between tropical regions ~~and China~~in Africa and America. Our results show that inter-annual analyses of methane emissions inferred by atmospheric inversions should always include an uncertainty assessment and that the attribution of ~~the atmospheric methane increase since 2007 to a particular region still needs more attention i.e.~~ current trends in atmospheric methane to particular regions needs increased effort, for instance gathering more observations ~~for the futureand using improved~~ (in the future) and improving transport models. At all scales, GOSAT generally ~~obtains the best results~~ shows the best performance of the three observing systems.

1 Introduction

As the second most important anthropogenic greenhouse gas after carbon dioxide in terms of radiative forcing, methane (CH_4) is an important climate driver. Monitoring atmospheric CH_4 concentrations and their driving emissions are therefore primary research objectives for Earth observation science. These two objectives are combined in atmospheric inversion systems. Such systems infer the space-time variations of the global or regional emissions from the assimilation of observations of atmospheric mole fractions into chemistry-transport models (CTMs) (Houweling et al., 1999; Bergamaschi et al., 2007; Bousquet et al., 2011; Pison et al., 2013). For these systems, explaining the trends of CH_4 concentrations, such as their stability between

2000 and 2006 and their later increase (Kirschke et al., 2013), is a major scientific objective. Despite considerable efforts in developing observing systems at the Earth's surface, in the atmosphere and from space, the ~~inverted~~ inferred fluxes are associated with large uncertainties. This still allows diverging interpretations of the trends, depending on which CTM is used or on how the inversion set-up is defined (Bousquet et al., 2006, 2011; Rigby et al., 2008; Dlugokencky et al., 2009; Bergamaschi et al., 2013). In principle, the Bayesian framework should reconcile all well-tuned inversion systems because it characterizes the uncertainty of each inversion product at all space-time scales, thereby weighting each scenario suggested by the inversion approach. In practice, posterior uncertainties are often difficult to compute and are also affected by mis-specified prior or observation uncertainties (Berchet et al., 2015). In a previous study, Cressot et al. (2014) applied objective tuning methods imported from Numerical Weather Prediction (Desroziers et al., 2005) within a robust Monte-Carlo approach to optimize the input error covariance matrices of a global CH_4 inversion system. Here, we use their results as a starting point to characterize the uncertainty of the year-to-year variations of the ~~inverted~~ inferred fluxes at various temporal ~~and spatial scales,~~ (e.g. seasonal, annual, 3-yearly, monthly) and spatial (global, latitudinal bands, large regions) scales in order to document which anomaly signals from the inversions are reliable and which are not within our framework. To do so, three different global CH_4 observation systems are considered: surface sites from various global networks (flasks and continuous), the space-borne ~~Infrared~~ Infra-red Atmospheric Sounding Interferometer (IASI) that provides a mid-to-upper-tropospheric column and the Thermal And Near ~~infrared~~ infra-red Sensor for carbon Observation - Fourier Transform Spectrometer (TANSO-FTS), that observes the total column from space. Using the flux anomalies of the surface inversion as the signal, signal-to-noise ratios for different temporal and spatial scales are computed, the noise being the uncertainty (precision) of the year-to-year changes of the ~~inverted~~ inferred fluxes for each observing system. Signal-to-noise ratios are then considered as a statistical criterion to evaluate the ability of an observing system to retrieve the CH_4 flux inter-annual variability.

The paper is structured as follows. The theoretical framework and the different data sets are presented in Section 2. The signal-to-noise ratios are presented in Section 3 and further discussed in Section 4.

2 Method

2.1 Inversion Framework

Our inversion system is based on a variational formulation of Bayes' theorem, as detailed by Chevallier et al. (2005), which has been adapted to the inversion of CH_4 fluxes by Pison et al. (2009). It allows ~~inverting~~ inferring grid-point-scale fluxes, thereby avoiding gross aggregation errors (Kaminski et al., 2001), while assimilating the large flow of satellite data at appropriate observation times and locations. It ingests observations of CH_4 mole fractions and prior information about the variables that are to be optimized, with associated error covariance matrices. Bayesian error statistics of the ~~inverted~~ inferred variables are computed from a Monte-Carlo ensemble of inversions which is consistent with the assigned prior and observation errors (Chevallier et al., 2007). The inversion system includes the LMDz transport model of Hourdin et al. (2006) at resolution $3.75° \times 2.5°$ (longitude x latitude) for 19 vertical levels ~~in a nudged and offline mode, which we couple~~ nudged to ECMWF analysed winds in its on-line mode. We use here its off-line mode that exploits the output variables of the on-line version.

We couple it to a simplified chemistry module ~~(SACS)~~ to represent the interactions between CH_4 and the hydroxyl radical (OH), its main sink in the atmosphere, and between methyl chloroform (MCF) and OH. Note that the loss due to chlorine in the marine boundary layer is not implemented yet in this model. When it assimilates both CH_4 and MCF mole fractions, as is done here, it synergistically optimizes both CH_4 surface sources at weekly and model grid resolution and OH at weekly resolution over 4 latitude bands (-90/-30, -30/0, 0/30, 30/90)~~, therefore dynamically distinguishing~~ . This set-up therefore dynamically distinguishes between CH_4 ~~emission and~~ net surface emissions (soil uptake included) and atmospheric loss. The system iteratively minimizes the Bayesian cost function (made non-quadratic by the non-linear chemistry) using the M1QN3 algorithm (Gilbert and Lemaréchal, 1989).

This system is applied here to assimilate data from each one of three CH_4 observing systems ~~and one~~ together with data from a MCF observing system (to constrain OH concentrations), in the configuration used by Cressot et al. (2014). The reader is referred to Cressot et al. (2014) for a detailed description of this configuration. It is ~~enough~~ important here to recall that the prior fluxes (fires excepted) have no inter-annual variability (IAV). ~~Therefore, IAV is generated from~~ This choice is made for IAV to be generated by atmospheric observations and atmospheric transport and chemistry ~~.~~ and not by prior IAVs of emissions (and sinks) which are still uncertain or even controversial (e.g. Schaefer et al. (2016); Hausmann et al. (2016); Nisbet et al. (2014)).

Two types of inversions are presented in this study:

- a reference inversion (hereafter called REFSURF) using CH_4 and MCF surface measurements from December 1999 to December 2011 ~~; and~~
- three ensembles of inversions (see Section 2.3 for the use of these), one using surface measurements only (called SURF hereafter), one using IASI ~~observations~~ data and MCF observations only (called IASI hereafter) and one using TANSO-FTS ~~observations~~ data and MCF observations only (called GOSAT hereafter~~), each ensemble consisting of ten~~ , from the name of the platform, Greenhouse gases Observing SATellite); each ensemble consists of ten one-year inversions from 10/2009 to 09/~~2010.~~ 2010, with respective inversion set-ups tuned according to an objective analysis described in Cressot et al. (2014).

For all inversions, the minimization of the non-quadratic cost function is stopped when the ratio of the final to the initial norm of the gradient is less than 0.01.

2.2 Data sets

In order to have continuous and homogeneous surface data throughout the extended assimilation window of REFSURF, we restrict the methane site list to 36 instead of 49 as used in Cressot et al. (2014). They come from the National Oceanic and Atmospheric Administration (NOAA) global cooperative air sampling network (Dlugokencky et al., 1994, 2009), the Commonwealth Scientific and Industrial Research Organisation (CSIRO) (Francey et al., 1999) and the National Institute of Water and Atmospheric Research (NIWA) (Lowe et al., 1991). We also use ~~the~~ station Alert (ALT) from Environment Canada

(EC) (Worthy et al., 2009). MCF measurements are provided by 11 NOAA surface sites (Montzka et al., 2011) and are used to constrain OH concentrations (Pison et al., 2009). The surface sites used in our inversions are presented in Figure 1.

We use observations of the mid-to-upper tropospheric CH_4 column made by IASI, a thermal interferometer on-board the Meteorological Operational (MetOp) satellites. This quantity is retrieved based on a non-linear inference scheme (Crevoisier et al., 2009) within 30 degrees of the Equator over both land and ocean at about 09:30 a.m./p.m. local time, with an accuracy of 1.2% (≈20 ppb).

Last, we use observations of the CH_4 atmospheric total column over land from TANSO-FTS, a ~~near-infrared~~ near-infra-red spectrometer on-board ~~the Greenhouse gases Observing SATellite (GOSAT)~~GOSAT. Total columns are retrieved by optimal estimation using the algorithm of Parker et al. (2011) and with a precision of 0.6% (≈10 ppb).

The averaging kernel or weighting function and the prior profile (when available) of each IASI or TANSO-FTS retrieval are directly accounted for in the inversion system following Connor et al. (2008).

2.3 Error statistics

The error statistics are described in detail in Cressot et al. (2014). For the fluxes, the spatial correlations are defined by e-folding lengths of 500 km over land and 1000 km over ocean (no correlation between land and ocean); time correlations are defined by an e-folding length of 2 weeks: it was checked that these choices lead to a budget uncertainty which is consistent with the uncertainty of bottom-up inventories as described in Kirschke et al. (2013).

The input error statistics for the prior and the observations are tuned using objective diagnostics as described by Cressot et al. (2014). This means that they exhibit some objectivity that is seen to translate into realistic Bayesian posterior error statistics, which in particular make all present inversions statistically consistent at the annual and global or regional scales (Cressot et al., 2014).

In order to keep the computational burden to a reasonable level, we compute the posterior error statistics from a Monte-Carlo inversion ensemble of 10 times one year ~~(~~for each of the three observing systems (ensembles GOSAT, IASI and SURF as described in Section 2.1).

The posterior error statistics (the "noise" for our study) are estimated as follows:

- we estimate the ratio of posterior to prior standard deviations of the annual flux errors $r = \frac{\sigma_a}{\sigma_b}$ from the ensemble, a quantity which is more robust than σ_a and σ_b individually for small ensembles (because some of the underspread affects the prior and the posterior in a similar way); the number of members in the ensemble depends on the time scale e.g. 10 ~~/2009 to 09/2010). Therefore, posterior error statistics of inter-annual emissions are computed from~~ members for the yearly time scale (10 inversions, each one covering 1 year), 120 members for the monthly time scale
- we estimate the posterior standard deviations of the annual flux errors by multiplying r to the known value of σ_b i.e. the one implied by our error covariance matrix (computed from the above assumptions)
- the posterior standard deviations of the ~~ones of annual emissions~~ pluri-annual flux errors errors for n years is obtained by applying a factor of $\frac{1}{\sqrt{n}}$ to the previous result, assuming that the errors are uncorrelated from one year to the next

– the posterior standard deviations of the difference between fluxes from one year to the next (i.e. the error on the IAV for two consecutive years) is computed by applying an inflation factor of $\sqrt{2}$ ~~. This means that we consider that~~ to the previous result, still assuming that the errors are uncorrelated from one year to the next. ~~This is~~ We assume this approach to be a conservative hypothesis since in reality some of the transport and retrieval errors are recurrent, thereby inducing positive correlations and reducing the inflation factor.

The variability of CH_4 concentrations depends on the oxidizing capacity of the atmosphere, which is largely controlled by OH concentrations. Since OH concentrations are constrained through MCF data in our multi-species inversion system (Section 2.1), the uncertainty on OH (≈5% after optimization) is accounted for in the uncertainty of the ~~inverted~~ inferred CH_4 emissions and of their inter-annual variations.

At a given space-time scale, the differences between the posterior errors of the three observing systems are mainly due to the constraints that each observing system brings on the flux estimates. This in turn is linked to the number of data, to their distribution in time and space, and also to their sensitivity to methane surface fluxes and to their uncertainty. It may also depends on the ability of the transport model to properly represent the various data.

2.4 Evaluation criterion

CH_4 regional flux anomalies are defined here as the deviation from ~~the 2004-2005 mean~~ a reference of the CH_4 ~~inverted fluxes~~ . inferred fluxes for various time periods, from the monthly to the 3-yearly scale. The reference is the 2004-2005 ~~has been chosen as a reference because it~~ mean over the same time-period. The aim of this definition is to get the order of magnitude of the year-to-year changes at various time scales. As the 2004-2005 reference corresponds to a period of minimum atmospheric methane growth rate (Dlugokencky et al., 2011), it leads to more positive anomalies for the longer time scales. The regional scale is based on the regions shown in Figure 2 and large latitudinal bands are defined as BorN for latitudes higher than 60 degrees North, MidN between 30 and 60 degrees North, TropN between 0 and 30 degrees North, TropS between 0 and 30 degrees South, MidS between 30 and 60 degrees South and BorS higher than 60 degrees South. We study various ~~timescales from the week to 3 years~~spatial and temporal scales of inferred flux anomalies.

Our criterion consists in evaluating the ability of the observing systems to detect CH_4 anomalies of a given amplitude, defined by the reference inversion. ~~The inversion of~~ For this, we define a signal-to-noise ratio:

– the inversion with surface measurements is chosen to provide the signal as the data ~~cover~~ covers a long time window (2000-2011) as compared to the two other observing systems. This longer window makes it possible to sample the CH_4 IAV more robustly than a 2-3 year inversion. We ~~compare the~~ assume that the fluxes inferred by this inversion are representative of state-of-the art inversions currently published. The signal is actually the CH_4 anomalies for the various time scales derived from REFSURF~~to the error variances computed for each observing system (from~~ .

– for the three observing systems (SURF, IASI and GOSAT)~~. The Bayesian posterior error variances associated with the IAV~~ , the Bayesian posterior errors of the year-to-year changes of CH_4 fluxes~~are~~ , computed from the Monte-Carlo

ensemble as described in Section 2.3and , constitute the noise associated to each observing system. To evaluate the space-time scales at which the anomalies are larger than the detection limit of each observing system, the

Finally, the criterion for detecting CH_4 anomalies is that the signal-to-noise ratio is larger than 1 (≈68% confidence).

Comparing signal-to-noise ratios are computed at the same spatial scales for weekly to 3-yearly time scales. This statistical criterion estimates for which amounts to comparing noises normalized by the expected signals. The normalization provides an absolute criterion to assess the time scales and regions at which the CH_4 anomalies are reliablefor each observing system. . However, the quality of the chosen signal remains debatable and our diagnostic for GOSAT and IASI may be pessimistic in areas where SURF signal-to-noise ratio is low.

In the following, the presentation of the results is done for three timescales time scales (seasonal, yearly, and 3-yearly trends) before assessing their sensitivity to temporal and spatial aggregations.

3 Results: signal-to-noise ratios

3.1 Seasonal-scale detection

The signal-to-noise ratios are computed over three-month periods (JFM, AMJ, JAS and OND, hereafter referred to as "seasons" for simplicity) from 2000 to 2011 i.e. 48 occurrences (12 years of 4 seasons). JFM, 12 AMJ, 12 JAS and 12 OND).

The three observing systems are able to detect almost all the anomalies at the global scale (Table 1). As expected, the fraction of detected anomalies decrease decreases with the spatial scale. At the global scale, 93 to 9791 to 93% of the flux anomalies are detected depending on the observing system (Table 1). At semi-hemispheric scales (excluding MidS and BorS areas), this range is of 10-910-87% (median = 5249.5%), GOSAT having the best range (25-918-87%) compared to IASI (22-7912-60%) and SURF (10-810-66%). The lack of detection in MidS and BorS is not significant considering the small methane fluxes involved. At the regional scale, the detection range is 0-970-79% (median = 104%), with large contrasts. Again the range is more favorable favourable for GOSAT (0-970-79%, median = 207%) than for SURF (0-870-75%, median = 103%) and IASI (0-520-72%, median = 60%). Only anomalies in Central America Anomalies in the USA, Central America [CentralAm], temperate Africa [SouthernAfr], Middle East and Australia&New-Zealand [AustrNZ] are not detected by any of the three observing systems. GOSAT is the only one of the three observing systems to detect any anomalies in the USA, anomaly in temperate South America [SouthSAm] and temperate northern Africa [SouthernAfrNorthAfrWest, NorthAfrEast].

At the seasonal time-scaletime scale, large signals are due to various causescaused by various processes, depending on the emitting area. At high Northern northern latitudes, a large seasonal cycle is expected for wetland emissionsemission areas, with mostly no emissions during winter and maximum emissions during summer: this leads to four seasons very different from their average and therefore to large anomalies. This is illustrated on the detection in NorthAmBor The detection rate is above 50% for the three observing systems in this region (Table 1): GOSAT is able to detect almost all the anomalies, half , but, in contrast to the other regions and to the other time scales, the prior error statistics already lead to detection rates of 58% for the prior. This shows that the Tropical IASI soundings do not add information for this region and at this time scale, as expected. GOSAT

performs better by detecting more than three quarters of the anomalies, about one third of which are ~~positive (Table 1/~~in winter (Figure 3, due to ~~maximum emissions in spring and summer) and half negative (Table 1/~~almost null emissions when the surface is snow-covered), one third in summer and one third in fall (Figure 3, due to ~~almost null emissions when the surface is snow-covered~~maximum emissions in summer). Due to a larger noise (≈~~1.4~~1.5 Tg vs ≈~~1~~1.2 Tg for GOSAT, Figure 4 [a]), SURF misses ~~some~~ all springs (Figure 3~~); and IASI, with the largest noise (≈1.9 Tg, Figure 4 a), mostly detects winter and summer (Figure 3)~~. In the larger BorN area, only winter and summer are detected (Figure 3).

In the Tropics, some areas also have large seasonal variations, mainly due to biomass burning or rice-paddies. In AfrEquat, some of the AMJ positive signals generated are ~~almost all~~ detected by GOSAT and IASI (Figure 4 [a]). Note that SURF performs poorly in this area (Table 1), due to the lack of stations which leads to large noise (≈3.3 Tg, Figure 4 [a]). In India and China, ~~the~~ rice-paddy practices lead to a seasonal cycle of methane emissions with a maximum in JAS and a minimum in JFM (Matthews et al., 1991). The three systems detect anomalies in JFM and JAS (Figure 3) with consistent signs (≈ half positive, half negative anomalies~~for GOSAT and IASI, positive~~) for GOSAT, negative anomalies preferentially detected by IASI and SURF (Table 1).

3.2 Yearly-scale detection

The signal-to-noise ratios are computed ~~over the years~~ for each year from 2000 to 2011 i.e. 12 occurrences. At the yearly scale, detection rates are smaller than at the seasonal scale, at all spatial scales. Note that most anomalies are positive since the reference for computing the signal is 2004-2005 i.e. the period of global minimum over 2000-2011. At the global scale, detection rates range from 58% to ~~83~~75% (Table 2). The Boreal zone [BorN] is ~~only poorly detected (8%)~~ not detected whereas the Tropics [TropN and TropS] remain the best detected zone (~~16-58~~16-50%). At the regional scale, the detection rates range between 0 and ~~58~~33% with a median of 0%: the only regions above 25% of detection are tropical Africa [~~NorthAfrWest, NorthAfrEast,~~ AfrEquat] ~~, Middle East for GOSATand Eastern Siberia FarEastSib~~and NorthAfrWest for GOSAT. No detection is obtained in key regions for methane emissions such as Amazonia ~~, India, China (except SURF at 16%)~~(except GOSAT at 8%), India, China and North America [NorthAmBor, USA].

The differences between the three observing systems are larger at the yearly scale than at the seasonal scale: GOSAT and IASI detect ~~more than~~ 75% of the 12 possible global occurrences versus 58% for SURF (Table 2). At the regional scale, GOSAT detects more anomalies than the two other systems~~, IASI and SURF being comparable in their detection rates~~. Indeed, GOSAT noises are smaller than the two other systems (<~~1.5~~3.5 Tg in ~~MiddleEast~~ AfrEquat for GOSAT against >~~4.5~~3.5 Tg for ~~SURF and IASI~~ IASI and >5.8 for SURF; <~~1.8~~2.5 Tg in ~~NorthAfrEast~~ NorthAfrWest for GOSAT against > ~~2.9~~4.7 Tg for ~~SURF and IASI~~ IASI and SURF). This is partly due to the large number of data available in these two regions (Table 4): with ~~NorthAfrWest, CentralAsia and AustrNZ, they have~~ NorthAfrEast, NorthAfrWest has the largest number of GOSAT and IASI data, mainly because ~~they are~~ it is among the driest areas i.e. with the lowest cloud cover. In agreement with the intuition of Bergamaschi et al. (2013) that performing gross averages makes it possible to extract a signal from the inversion, the detection is enhanced in the latitudinal bands e.g. detection rates ~~>50% in MidN and~~ ≥25% in TropN for GOSAT ~~, TropN for~~ and SURF.

But ~~, at the regional scale,~~ it remains difficult to robustly extract yearly flux anomalies. Therefore, we now focus our analysis on longer time scales, with a longer time aggregation of three years, to get hints at the longer trends in methane emissions.

3.3 Trend detection over 2000-2011

To study the detection of flux long-term trend over 12 years, a compromise has to be found between the rather short length of this time window and the time aggregation of fluxes, which needs to filter out year-to-year changes. Aggregating through time while still retaining a small enough resolution to discuss ~~tendencies~~ trends over 2000-2011, we define four ~~time-windows~~ time windows of three years each: 2000-2002, 2003-2005, 2006-2008 and 2009-2011. The reference period for the definition of the anomalies of each of these four periods is still 2004-2005 (Section 2.4).

At the global scale, the emissions have slowly decreased from 2000 to 2005, with a global minimum in 2004-2005, then increased at a larger rate after 2006 (Kirschke et al., 2013). The three observing systems are able to detect the large positive anomalies after 2006 and ~~consistently detect nothing or small positive anomalies~~ detect nothing before (Table 3). The three observing systems are able to detect the same ~~time-evolution~~ temporal evolution of the signal in TropN and TropS. Only GOSAT and SURF detect MidN anomalies; the lower detection by IASI at these latitudes is expected since the data used here are only within ~~+/-30~~ ±30 degrees of the Equator (Table 4: no IASI data in MidN). The signal in BorN is never detected. This is consistent with the recent increase of methane global emissions coming mostly from the Tropics and to a lesser extent from the northern mid-latitudes, as suggested by Bergamaschi et al. (2013) and Nisbet et al. (2014).

Being able to detect anomalies at a smaller spatial scale could help attributing the changes in methane emissions to particular processes. Unfortunately, even when aggregating 3 years together (instead of one as in Section 3.2), it is still difficult to detect regional anomalies. ~~On top of the regions already detected at the yearly time scale, a positive change in Chinese emissions is detected with the three-year aggregation, but only by IASI and SURF. The lack of detection by GOSAT stems from the small number of GOSAT data compared to IASI over India and China (Table 4), which is due to cloud cover and aerosol column content. On the contrary, GOSAT alone suggests detectable negative anomalies in NorthAmBor in 2000-2002 and 2009-2011. This is consistent with the lack of surface sites in this area (e.g. the Canadian stations from Environment Canada were not used here) and the lack of data by IASI North of 30 degrees (Table 4).~~

~~At high northern latitudes, positive anomalies in FarEastSib are detected by all three systems in 2000-2005 and again in 2009-2011 by GOSAT and SURF, even though the emissions in this area are small (1 Tg in 2004-2005, Table 3). This is due to the very small noises, mainly due to the small prior errors, which are built proportional to the fluxes. Moreover, for SURF, 3 stations are available downwind of this region.~~

In TropN, among the regions with a good detection rate are NorthAfrWest and NorthAfrEast plus ~~part~~ some of AfrEquat, the remainder of this region being in TropS. In these regions, all three observing systems detect anomalies, even though GOSAT has the largest signal-to-noise ratios. Note that SURF seems to ~~be able to make use of~~ benefit from the stations located mostly on the coasts (only ASK is actually ~~in the land-mass~~inland). GOSAT is also able to detect ~~large~~ negative (2000-2003) and positive (2006-2011) anomalies in the MiddleEast; SURF is under the detection threshold because the available station in the region, WIS, is upwind ~~of~~ the area and no other station is available close enough downwind; the anomalies are not detected by IASI

either because IASI ~~'s weight-function~~ weighting function peaks in the mid-troposphere. In a region dominated by subsidence, like the MiddleEast, the altitude concentrations seen by IASI are not directly connected to the surface. The detection of surface variations in the fluxes is therefore poor, contrary to regions dominated by convection like Indonesia, where IASI has the best detection rates. In ~~China, the three systems agree on the detectable negative anomalies in 2000-2002 and do not detect any signal in 2003-2005. After 2006, SURF detects the positive anomalies, because its noise is the smallest (≈12 Tg) with about 3 stations providing relatively direct constraints in the region. The two satellites, for which noises are 15 to 40% larger (≈14 and ≈17 Tg), do not detect this signal.~~

~~In~~ Indonesia, IASI and GOSAT agree on detectable positive anomalies in 2000-2002 and ~~2006-2008 and~~ nothing detectable for ~~the other two periods~~2003-2005 and 2009-2011. Indeed, no large El Niño occurred during the first decade of the 21st century with the associated large fires such as those experienced ~~late in 2015~~ in 1997-1998 or more recently in 2015-2016 for instance (National Weather Service - Climate Prediction Center, 2016).

Among the key-areas for methane emissions, signals in Amazonia (dominated by tropical wetlands) and in BorN, particularly in SiberianLowlands (dominated by boreal wetlands in summer), remain undetectable by the three systems. In SiberianLowlands, the noises of the three systems are small (between ~~3 and 6~~3.8 and 7.8 Tg [not shown]); in Amazonia, the noises of the satellites are relatively small (≈~~8~~ 6 and ≈~~6~~7 Tg resp. for GOSAT and IASI), whereas the noise of SURF, for which no stations are available closer than ASC in the Atlantic, is ≈~~19~~24 Tg (Figure 7, 3Y case). Nevertheless, all these anomalies remain smaller than the smaller noise, and are therefore not detectable in our framework. This is because the signal variability ~~remain~~ remains small after inversion (less than 20% of the average mass over 2004-2005). ~~As there is no~~ Possible reasons for this are an actual low variability in these regions for this period and the fact that the choice to limit IAV in the prior emissions ~~(except biomass burning),~~ to biomass burning together with the lack of constraints from the atmosphere ~~leads some~~ lead the inferred fluxes to stick to the low-IAV prior~~, leading to small anomalies~~.

3.4 Detection at other ~~timescales~~time scales

As shown previously, the temporal scale at which the signal and noise are computed has an impact on the detection~~. Section 3.1 deals with the~~ : 3-monthly ~~time-scale~~ ("seasonal", Section 3.1) and yearly (year-to-year changes, Section 3.3) time scales over a 12-year ~~time-window; Section 3.3 deals with the~~ time window; 3-yearly ~~time-scale~~ time scale in 3-year ~~time-windows. The impact of temporal aggregation~~ time windows (trend, Section 3.3). We investigate the impact of the time scale of flux aggregation within the 3-year time windows of Section 3.3 on the noise and the signal ~~in these time-windows is displayed in Figure 5, Figure 6 and Figure 7~~ for three areas: ~~Global, hemispheric with the example of BorN, and regional with the example of Amazonia .~~ global (Figure 5), the northern Tropics (Figure 6) and Amazonia (Figure 7). For each area, we perform time aggregations from 3-yearly to monthly scales, the 3-yearly case corresponds to the results commented in Section 3.3 about trend. At all spatial scales, the noises and signals are smaller when the ~~time-scale~~ time scale is smaller (from 3-yearly to ~~weekly~~monthly). As expected for emissions with "seasonal" cycles, the seasonal scale (4- or 3-monthly) is particularly detected (Figure 5, Figure 6) in our relatively large areas.

~~The minimum time-resolution of one week could be relevant in regions where the signal is mostly from wetland emissions and/or biomass burning; it would be useful to be able to detect the beginning of the emitting season for wetlands and the short-lived fires. In NorthAmBor, where both these sources are found, about 55% of weekly anomalies are detected by GOSAT and SURF. In all other regions, the detection rates at this time-scale are small (<≈25%, not shown).~~

In key-region Amazonia (Figure 7), no signal is detected at the 3-yearly ~~time-scale~~ time scale nor at the ~~weekly time-scale~~ monthly time scale by any of the three systems; only GOSAT detects about ~~15~~8% of the anomalies at the yearly ~~time-scale~~time scale. Actually, the ~~time-scale~~ time scale at which the best detection rates are found depends on the region and varies from the largest possible (~~12-year~~ 3-year scale) to the 2-month scale. In ~~most of~~ Africa [~~NorhtAfrWest~~NorthAfrWest, NorthAfrEast, AfrEquat, SouthernAfr], the ~~signal at the 12-year scale is detected~~ best detection rates are obtained at the 3-year scale by all three systems~~(it is detected by GOSAT only in SouthernAfr). In India and China,~~ , as in Europe, Indonesia and Australia&New-Zealand [AustrNZ]. In the North of Eurasia [EastEurRussia, SiberianLowlands, SiberianHighlands, FarEastSib], the best detection rates ~~are obtained at the 2- or~~ range from the 3-yearly to the 3-monthly ~~time-scale for GOSAT (77 and 43% respectively) but at the 4-monthly time scalefor IASI (66 and 30% respectively) and SURF (61 and 44% respectively)~~time scales. In Central Asia, IASI obtains the best detection rates at the 2-monthly time scale. At high latitudes [~~BorN~~NorthAmBor], the best detection rates are found at the 2-monthly ~~time-scale (between 70% for IASI and 86% for GOSAT)~~(SURF), 3-monthly (IASI) and 4-monthly (GOSAT) time scales (with 88 to 100% for GOSAT, up to 75% for IASI (but which is not better than the prior detection rate, see Section 3.1) and up to 77% for SURF), which is consistent with seasonal cycles with a large magnitude over a short period of time in this region.

In order to further understand the various levels of detection described above, we investigate the sensitivity of our results to two main parameters of our set-up: spatial aggregation and signal used.

4 Sensitivity analysis

4.1 Impact of spatial aggregation on trend detection

Our inversion ~~systems solves~~ system solves for methane fluxes at ~~the~~ model resolution (3.75°x2.5°) worldwide. Although spatial and temporal correlations are prescribed (see Section 2.3), flux anomalies of different signs may still be obtained. These anomalies may be either the realistic result of the constraints or due to the optimization taking an easy path when too few constraints are available. The definition of larger areas may lead to summing up anomalies of opposite signs and hide (realistic or not) spatial variations. We ~~then try~~ try here to investigate the impact of the spatial aggregation of model pixels in the case of one illustrative region, Amazonia, which is a key-area for methane emissions and remains poorly detected by all the studied observing systems at all ~~time-scales~~ time scales (see Section 3.4). In the region as defined on our model grid, the signal at the pixel scale is indeed patchy (Figure 8). Dipoles of negative/positive signal are summed up when aggregating at ~~the region 's~~ region scale. The impact of the progressive aggregation of rings of pixels from the center of Amazonia is displayed in Figure 9 ~~:~~ for the 3-yearly ~~signal could be~~ time scale: the signal is detected by all systems for the four 3-year periods up to the 3rd ring i.e. for a region covering 25 pixels instead of 66. It would then be possible to define the regions based on the spatial

aggregation that allows the best detection rates for the chosen observing system. ~~This would nevertheless lead to the issue of the user~~ Nevertheless, this may be inconsistent with users' needs e.g. ~~whether the regions are actually relevant for country~~ if they are expressed in terms of country-based budgets.

4.2 Impact of the signal on seasonal and yearly detection

Since the signal is obtained from one inversion only~~i.e. depends on numerous~~ , it depends on a series of assumptions (error statistics, ~~set of assimilated data~~ data selection, etc) and ~~has potentially~~ may have large uncertainties in various areas (e.g. far from the observing stations)~~, another signal definition has been tested. It must cover enough years of analysis to be representative of the variability of methane fluxes. We therefore chose~~ . Another signal definition is therefore tested. We choose an inversion by Bousquet et al. (2011), (called PBSURF hereafter) instead of the ~~SURF~~ REFSURF inversion described above. ~~PBSURF solves~~ Like REFSURF, PBSURF covers enough years of analysis to be representative of the variability of methane fluxes. The main differences between PBSURF and REFSURF are:

- PBSURF uses an analytical inversion whereas REFSURF is variational,
- because of this, PBSURF solves for methane fluxes for large regions ~~and~~ whereas REFSURF works at the pixel scale,
- PBSURF retrieves monthly fluxes whereas REFSURF retrieves fluxes at a weekly resolution,
- PBSURF solves for methane fluxes for several processes in each region ~~, using observations from a set of surface stations different from SURF.~~ whereas REFSURF solves for net emissions,
- as a consequence of the three previous points, the **B** matrices of the two inversions are quite different,
- PBSURF uses monthly means of the surface observations as constraints whereas REFSURF uses hourly data,
- because of this, the sets of surface stations used by PBSURF and REFSURF are different.

The large-region-scale inversion means that the spatial variability of the prior is kept within each region and is only scaled (contrary to ~~SURF~~REFSURF, which is performed at the pixel scale i.e. is able to vary only a few pixels to match the data). This difference in the methods may lead to very different spatial variability in each of the regions of interest (Figure 4), a larger variability ~~ensuring~~ allowing a better detection rate with our criterion. Indeed, the large-region-scale inversion may lead to larger variability than pixel-based inversions in some regions (e.g. (Pison et al., 2013)) because of the homothetic scaling of the pixels composing each region in PBSURF (correlations between pixels of 1) as opposed to the individual scaling of model pixels with soft constraints in REFSURF (spatial correlations less than 1).

We first focus on the seasonal (3-monthly) scale, which is the ~~time-scale~~ time scale at which the detection is the most ~~favorable for SURF~~ favourable in the largest areas (Section 3.4) while being relevant for methane emissions at the regional scale defined here. The issue here is not whether the two inversions agree on the retrieved fluxes but whether the detection rates differ. Europe illustrates how the detection rates of two signals can differ: for ~~all three observing systems, PBSURF signal~~

GOSAT, signal PBSURF is more than twice as often detected as ~~SURF~~ REFSURF and the signs of the detected anomalies are opposite (positive for ~~SURF,~~ REFSURF, mostly negative with PBSURF, Table 1 and Table 5: less positive anomalies are detected for a larger total number of detected anomalies).

~~The signal by~~ Signal PBSURF contains more negative anomalies than ~~SURF~~ REFSURF at the global scale and in ~~BorN and MidN (for GOSAT and SURF)~~MidN and TropN. This is due to the fact that the two years of global minimum in ~~SURFPB~~ PBSURF are not 2004 and 2005 but 2004 and 2006, so that using 2004-2005 as the reference period does not lead to mainly positive anomalies. For the three observing systems, detection is better with ~~PBSURF signal~~ signal PBSURF in the Southern hemisphere ~~(TropS, MidS~~Tropics (TropS). In the Northern hemisphere, at the regional scale, the detection rate is shifted in longitude. NorthAmBor seasons are about ~~half as~~ 25% less often detected whereas up to ~~5~~ 30 times more occurrences are detected in SiberianLowlands, SiberianHighlands and FarEastSib. In SiberianLowlands and FarEastSib, the larger number is due to negative signals for GOSAT and SURF. The same pattern is seen in the mid-latitudes where MiddleEast, India and China, which are almost never detected with ~~PBSURF signal~~ signal PBSURF (only India for GOSAT), versus NorthAfrWest and NorthAfrEast, in which mainly positive anomalies are detected (IASI and SURF) or both positive and negative anomalies (GOSAT). The regional scale in the Southern hemisphere confirms the better detection with ~~PBSURF signal~~ signal PBSURF (Amazonia, SouthSAm, SouthernAfr~~, Indonesia, AustrNZ~~). In Amazonia, the (mainly positive) signals are detectable by GOSAT and IASI, but China (resp. India) is not ~~anymore~~ any more (resp. poorly) detectable using PBSURF.

At the yearly scale (Table 6), the detection rates are shifted to the ~~North in the Northern hemisphere and to the South in the Southern hemisphere~~ South (from TropN and MidN to ~~BorN and~~ TropS). Detection rates higher than 50% are found in Amazonia for GOSAT and IASI; in Europe for GOSAT~~and SURF; in Indonesia for GOSAT and IASI~~.

One important outcome of this sensitivity test to the signal is that some regional or hemispheric flux anomalies are detected ~~at all latitudes at most time-scales~~ but the localization of the detected signal varies depending on the inversion characteristics (including the observations used). This is of course one important limitation in attributing the observed atmospheric changes to particular regions and to the underlying emission processes.

The impact of the signal on the detection of anomalies has also been tested by using a variational inversion at the pixel scale assimilating both surface and IASI data. With this signal, the detection rates are higher in the Tropics (particularly in India and China) and in the Southern hemisphere at mid-latitudes [not shown]. This suggests that the joint assimilation of surface and satellite data may lead to a better localization of the anomalies of the surface methane fluxes. Nevertheless, this requires that the consistency between the two types of data (surface and remote-sensed) be improved (Locatelli et al., 2015; Monteil et al., 2013).

5 Conclusions

~~The aim of this study was to investigate which~~ This study aimed at investigating the spatial and temporal scales ~~current atmospheric inversions may detect in terms~~ of methane surface flux anomalies that current atmospheric inversions can detect. To do so, we have proposed a signal-to-noise ratio analysis, the signal being the methane fluxes inferred from a reference surface-

based inversion from 2000 to 2011 and the noise being computed from three inversion systems using surface or satellite data (GOSAT and IASI). At the global and semi-hemispheric scales, all observing systems detect flux anomalies at ~~all time-scales~~ various time scales from seasonal (3-month average) to long-term trend (3-year average). At all scales, GOSAT generally shows the best results among the different systems, as could be expected from the density of the data and their sensitivity to surface emissions.

At the regional scale, the results are more variable. ~~The seasonal changes are all detected with fair to good rates by at least one network (GOSAT), and more than 50% of the regions are detected by~~ In 8 regions out of 20, anomalies are detected by the ~~three networks~~three networks; in 5 regions, no anomaly is detected by any of the three systems. The year-to-year changes ~~and longer term trends (three year averages) are detected for up to 50% of the regions (by GOSAT) with detection rates mostly lower than 50%. Anomalies~~ are detected in 9 regions by GOSAT but with poor detection rates (lower than 40%). Longer term trends (three year averages) in African regions ~~(all), Middle East (GOSAT), Eastern Siberia and Europe (all)~~ are detected with variable rates by the three systems. In some key regions for the methane cycle, anomalies are hardly detected, both in the case of dominant anthropogenic emissions (North America) or natural emissions (Amazonia, Siberian lowlands). A sensitivity test to the spatial scale through aggregation shows that dipole effects in the retrieved flux anomalies prevent anomalies in Amazonia (as defined in this study) to be detected. Flux anomalies in India and China, two areas with large and mixed (natural and anthropogenic) methane emissions, are generally poorly detected. ~~Only a long-term trend over China is detected, with larger emissions after 2006 for IASI and the surface network but not for GOSAT, which has a lower number of observations over these regions because of cloud cover and aerosol layers. North-American emission changes are not detected, with the exception of the long term trend of boreal North America (negative after 2006 for GOSAT). Overall, the~~ A sensitivity test with a second signal, also obtained from an inversion with surface constraints, shows that overall, the detection at a yearly scale ~~is generally~~ remains poor to fair ~~for the two signals tested, both obtained with surface constraints~~(>50% in Amazonia for the test signal). These tests point at the importance of properly determining the spatial aggregation at which the inferred fluxes are used, with the issue that such an aggregation depends on the inversion system used. This suggests that the ability of the inversions to retrieve significant inter-annual variations in the methane fluxes is not evident and should be evaluated against uncertainties, which are not always computed and/or provided with the inversion products.

The use of another signal (which is from a different surface-based inversion) does not change the main conclusion that anomalies at the regional scale are ~~only fairly well~~ not well detected but shows that the regions which are not seen may be different: some yearly changes in Amazonia ~~and India~~ can be detected but tropical Africa is much less detected with the second signal. Therefore, the precise identification of flux anomalies in the Tropics appears not to be robust with ~~regards~~ regard to changes in the inversion used for the signal. This is of course an issue when attributing the increase observed in atmospheric methane since 2006 to a particular region, as already noticed by Locatelli et al. (2015).

~~To increase the detection rates, the number of constraints (i. e. of assimilated data, either from satellite or from surface sites) should be~~ Our criterion is based on a 68% confidence interval (1 sigma). At almost all regional time-space scales (except in NorthAmBor, AfrEquat at the longer time-scales and a few cases in India, Indonesia, EastEurRussia and FarEastSib), the three observing systems would fail the test at 2 sigmas (95%), a more stringent criterion commonly used in other scientific

communities. We also have neglected the impact of likely state-dependent systematic errors in current satellite retrievals and transport models that further reduce the inversion performance to an unknown extent.

Overall, our study may appear to be pessimistic about the skill of current inversions at the regional scale. However, at least two elements put this view into perspective.

First, we focussed on the first decade of the XXIst century, a time period with relatively flat methane signals. Neither a strong El Niño, nor a large volcanic eruption occurred, contrary to the previous decade (1990-1999). As an illustration, the methane atmospheric growth rate fluctuates from 2 to 16 ppb/yr in the 1990s (standard deviation of yearly annual increase of ±4.5 ppb/yr) as compared to -4 to +7 ppb/yr (standard deviation of yearly annual increase of ±3.5 ppb/yr) in the 2000s (Dlugokencky et al., 2011). This reduces methane flux anomalies and their detectability for a given noise. A time period with larger year-to-year changes in the methane cycle could lead to an improved detectability.

Second, as mentioned in Section 2, we have been relatively conservative to estimate the noise, possibly leading to its overestimation, therefore also limiting the detectability of methane flux anomalies.

Our work has several implications for methane inversions.

First, inversion results should never be presented without an extensive uncertainty analysis to distinguish between robust and more hypothetical results. This may seem obvious but such an analysis is not always provided, or only partially, in inversion papers, mostly because of its computational cost.

Second, to increase the detection robustness, the information amount from the satellite data and from the surface sites should be dramatically increased, as shown by the regional differences between the two surface-based inversions (e.g. Africa versus Tropical regions and China) and between the satellite based inversions~~(more IASI observations over China and India than GOSAT ones). To increase~~. Defining smaller regions, as tested here in Amazonia, may also improve the detection of anomalies in small key areas with intense methane emissions. An increase in the robustness of the attribution of flux anomalies to a particular region ~~, transport models should be improved together with the~~ goes with the improvement of the consistency of error statistics prescribed for ~~flux~~ fluxes and observations (Berchet et al., 2015). ~~Defining smaller regions, as tested here in Amazonia, may also improve the detection of anomalies in small key areas with intense methane emissions~~

Third, as the regions robustly inferred depend on the assimilated datasets, but also on the transport model and inversion set-up, it seems important to push for regular comparisons and syntheses of the various transport models and inversion systems, which is at present the only way to approach the full range of uncertainty.

With time, the increase of observations in density, precision, and accuracy, if sustained by long-term funding of surface networks and development of satellite instruments, together with the necessary improvement of transport models, should allow to reduce uncertainties in methane flux estimates. The joint assimilation of surface and satellite observations could be a solution to ~~better constrain the surface methane~~ further improve the constraint on methane surface fluxes, if the consistency between surface and remote sensed data can be improved (Locatelli et al., 2015; Monteil et al., 2013; Cressot et al., 2014). Cloud cover and aerosol layers ~~may~~ limit the observability of key regions such as China and India ~~.~~ or even Amazonia and induce systematic errors in passive satellite instruments (e.g. Buchwitz et al. (2016)). Solar based satellite instruments also provide limited data at high latitudes. The future space mission MERLIN, based on a differential active LIDAR measurement with a very small spot

on the ground, ~~is less sensitive to cloud cover and does not need light to provide data~~ should overcome these issues and provide data at all latitudes and all seasons (Kiemle et al., 2014). In this context, MERLIN seems a promising mission to improve some of the limitations raised in this paper.

Acknowledgements. The authors are very grateful to the many people involved in the surface and satellite measurement and in the archiving of these data. The authors particularly thank E.J. Dlugokencky (NOAA), S.A. Montzka (NOAA), C. Crevoisier (LMD), H. Boesch (University of Leicester), R. Parker (University of Leicester), P.B. Krummel (CSIRO), L.P. Steele (CSIRO), R.L Langenfelds (CSIRO), S. Nichol (NIWA) and D. Worthy (EC). We ~~aknowledge~~ acknowledge the contributors to the World Data Center for Greenhouse Gases for providing their data of methane and methyl-chloroform atmospheric mole fractions. The first author is funded by CNES and CEA. P. J. Rayner is in receipt of an Australian Professorial Fellowship (DP1096309). This work was performed using HPC resources from CCRT under the allocation 2014-t2014012201 made by GENCI (Grand Equipement National de Calcul Intensif) and a DSM allocation. We also thank the computing support team of the LSCE led by F. Marabelle.

References

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Bonne, J.-L., and Paris, J.-D.: Objectified quantification of uncertainties in Bayesian atmospheric inversions, Geoscientific Model Development, 8, 1525–1546, doi:10.5194/gmd-8-1525-2015, http://www.geosci-model-dev.net/8/1525/2015/, 2015.

Bergamaschi, P., Frankenberg, C., Meirink, J., Krol, M., Dentener, F., Wagner, T., Platt, U., Kaplan, J., K "orner, S., Heimann, M., et al.: Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations, Journal of Geophysical Research-Atmospheres, 112, D02 304, 2007.

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R., Dlugokencky, E., Wofsy, S., Kort, E., Sweeney, C., et al.: Atmospheric CH4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, Journal of Geophysical Research: Atmospheres, 118, 7350–7369, 2013.

Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E., Brunke, E., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D., et al.: Source attribution of the changes in atmospheric methane for, Atmos. Chem. Phys, 11, 3689–3700, 2011.

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Van der Werf, G. R., Peylin, P., Brunke, E. G., Carouge, C., Langenfelds, R. L., Lathiere, J., Papa, F., Ramonet, M., Schmidt, M., Steele, L. P., Tyler, S. C., and White, J.: Contribution of anthropogenic and natural sources to atmospheric methane variability, Nature, 443, 439–443, doi:10.1038/nature05132, 2006.

Buchwitz, M., Dils, B., Boesch, H., Crevoisier, C.and Detmers, R., Frankenberg, C., Hasekamp, O., Hewson, W., Laeng, A., Noel, S., Nothold, J., Parker, R., Reuter, M., and Schneising, O.: Product Validation and Intercomparison Report (PVIR) for the Essential Climate Variable (ECV) Greenhouse Gases (GHG), Tech. Rep. report version 4, ESA Climate Change Initiative (CCI), http://www.esa-ghg-cci.org/?q=webfm_send/300, 2016.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Breon, F., Chédin, A., and Ciais, P. F.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res, 110, 2005.

Chevallier, F., Bréon, F.-M., and Rayner, P.: The contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research, 112, doi:10.1029/2006JD007375, 2007.

Connor, B. J., Boesch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, Journal of Geophysical Research: Atmospheres, 113, 2008.

Cressot, C., Chevallier, F., Bousquet, P., Crevoisier, C., Dlugokencky, E., Fortems-Cheiney, A., Frankenberg, C., Parker, R., Pison, I., Scheepmaker, R., et al.: On the consistency between global and regional methane emissions inferred from SCIAMACHY, TANSO-FTS, IASI and surface measurements, Atmospheric Chemistry and Physics, 14, 577–592, 2014.

Crevoisier, C., Nobileau, D., Fiore, A., Armante, R., Chédin, A., and Scott, N.: A new insight on tropospheric methane in the Tropics–first year from IASI hyperspectral infrared observations, Atmospheric Chemistry and Physics Discussions, 9, 6855–6887, 2009.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, 2005.

Dlugokencky, E., Steele, L., Lang, P., and Masarie, K.: The growth rate and distribution of atmospheric methane, Journal of Geophysical Research, 99, 1994.

Dlugokencky, E., Bruhwiler, L., White, J., Emmons, L., Novelli, P., Montzka, S., Masarie, K., Crotwell, A., Miller, J., and Gatti, L.: Observational constraints on recent increases in the atmospheric CH_4 burden, Geophys. Res. Lett., doi:10.1029/2009GL039780, in press, 2009.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369, 2058–2072, doi:10.1098/rsta.2010.0341, http://rsta.royalsocietypublishing.org/content/369/1943/2058, 2011.

Francey, R., Steele, L., Langenfelds, R., and Pak, B.: High Precision Long-Term Monitoring of Radiatively Active and Related Trace Gases at Surface Sites and from Aircraft in the Southern Hemisphere Atmosphere, Journal of the Atmospheric Sciences, 56, 279–285, 1999.

Gilbert, J.-C. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, Mathematical programming, 45, 407–435, 1989.

Hausmann, P., Sussmann, R., and Smale, D.: Contribution of oil and natural gas production to renewed increase of atmospheric methane (2007–2014): top-down estimate from ethane and methane column observations, Atmospheric Chemistry and Physics, 16, 3227–3244, doi:10.5194/acp-16-3227-2016, 2016.

Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J., Fairhead, L., Filiberti, M., Friedlingstein, P., Grandpeix, J., et al.: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, Climate Dynamics, 27, 787–813, 2006.

Houweling, S., Kaminski, T., Dentener, F., Lelieveld, J., and Heimann, M.: Inverse modeling of methane sources and sinks using the adjoint of a global transport model, Journal of Geophysical Research, 104, 26–137, 1999.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, Journal of Geophysical Research, 105, 4703–4715, 2001.

Kiemle, C., Kawa, S. R., Quatrevalet, M., and Browell, E. V.: Performance simulations for a spaceborne methane lidar mission, Journal of Geophysical Research: Atmospheres, 119, 4365–4379, 2014.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., et al.: Three decades of global methane sources and sinks, Nature Geoscience, 6, 813–823, 2013.

Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreux, F., Grandpeix, J.-Y., Lefebvre, M.-P., Rio, C., Bergamaschi, P., et al.: Atmospheric transport and chemistry of trace gases in LMDz5B: evaluation and implications for inverse modelling, Geoscientific Model Development, 8, 129–150, 2015.

Lowe, D., Brenninkmeijer, C., Tyler, S., and Dlugkencky, E.: Determination of the Isotopic Composition of Atmospheric Methane and its Application in the Antarctic, J. Geophys. Res., 96, 15 455–15 467, 1991.

Matthews, E., Fung, I., and Lerner, J.: Methane emission from rice cultivation: Geographic and seasonal distribution of cultivated areas and emissions, Global Biogeochemical Cycles, 5, 3–24, 1991.

Monteil, G., Houweling, S., Butz, A., Guerlet, S., Schepers, D., Hasekamp, O., Frankenberg, C., Scheepmaker, R., Aben, I., and Röckmann, T.: Comparison of CH4 inversions based on 15 months of GOSAT and SCIAMACHY observations, Journal of Geophysical Research: Atmospheres, 118, 2013.

Montzka, S., Krol, M., Dlugokencky, E., Hall, B., Jöckel, P., and Lelieveld, J.: Small interannual variability of global atmospheric hydroxyl, Science, 331, 67, 2011.

National Weather Service - Climate Prediction Center: Cold & Warm Episodes by Season, http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml, 2016.

Nisbet, E. G., Dlugokencky, E. J., Bousquet, P., et al.: Methane on the rise-again, Science, 343, 493–495, 2014.

Parker, R., Boesch, H., Cogan, A., Fraser, A., Feng, L., Palmer, P. I., Messerschmidt, J., Deutscher, N., Griffith, D. W., Notholt, J., et al.: Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations, Geophysical Research Letters, 38, 2011.

Pison, I., Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D.: Multi-species inversion of CH_4, CO and H_2 emissions from surface measurements, Atmospheric Chemistry and Physics, 9, 5281–5297, http://www.atmos-chem-phys.net/9/5281/2009/, 2009.

Pison, I., Ringeval, B., Bousquet, P., Prigent, C., and Papa, F.: Stable atmospheric methane in the 2000s: key-role of emissions from natural wetlands, Atmospheric Chemistry and Physics Discussions, 13, 9017–9049, doi:10.5194/acpd-13-9017-2013, 2013.

Rigby, M., Prinn, R., Fraser, P., Simmonds, P., Langenfelds, R., Huang, J., Cunnold, D., Steele, L., Krummel, P., Weiss, R., O'Doherty, S., Salameh, P., Wang, H., Harth, C., Mülhe, J., and Porter, L.: Renewed growth of atmospheric methane, Geophys. Res. Lett, 35, L22 805, 2008.

Schaefer, H., Fletcher, S. E. M., Veidt, C., Lassey, K. R., Brailsford, G. W., Bromley, T. M., Dlugokencky, E. J., Michel, S. E., Miller, J. B., Levin, I., Lowe, D. C., Martin, R. J., Vaughn, B. H., and White, J. W. C.: A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by 13CH4, Science, 352, 80–84, doi:0.1126/science.aad2705, 2016.

Worthy, D. E., Chan, E., Ishizawa, M., Chan, D., Poss, C., Dlugokencky, E. J., Maksyutov, S., and Levin, I.: Decreasing anthropogenic methane emissions in Europe and Siberia inferred from continuous carbon dioxide and methane observations at Alert, Canada, Journal of Geophysical Research: Atmospheres, 114, 2009.

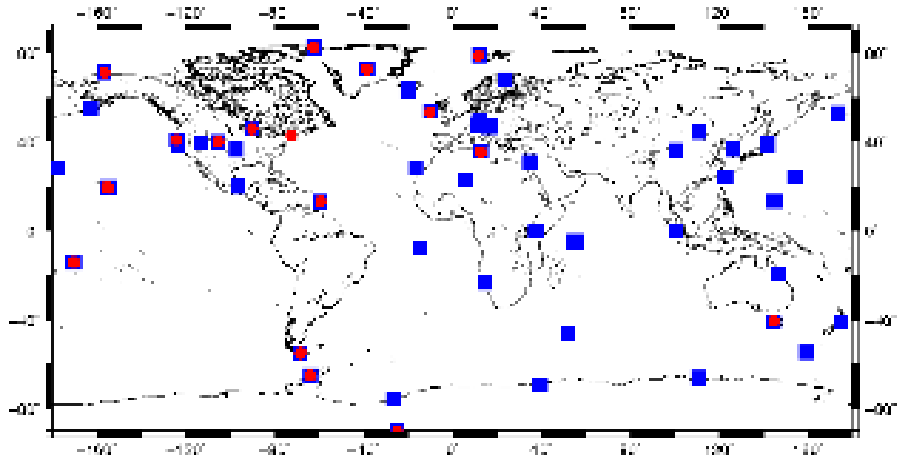


Figure 1. *Surface sites from the NOAA, CSIRO, NIWA and EC networks used in this study with red circles for surface sites observing MCF dry air mole fractions and blue squares for surface sites observing CH_4 dry air mole fractions.*

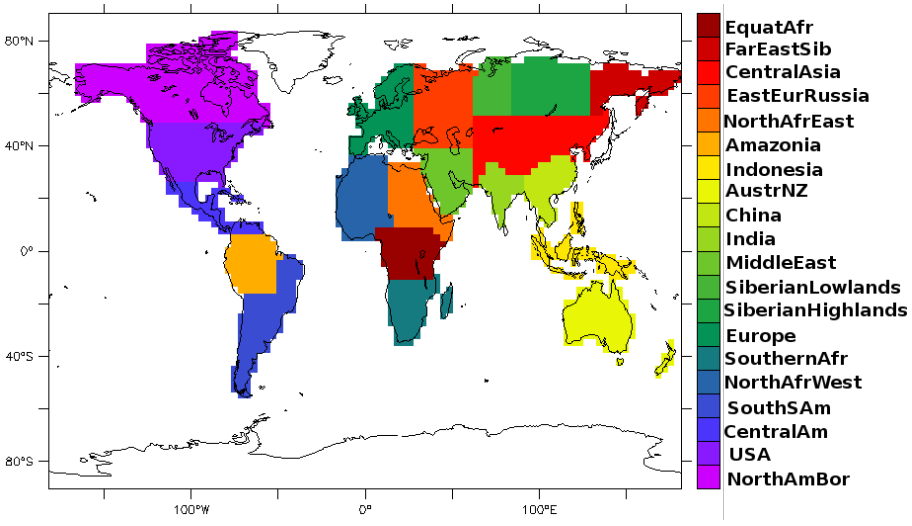


Figure 2. *Regions on the model grid, adapted to key-area for methane fluxes.*

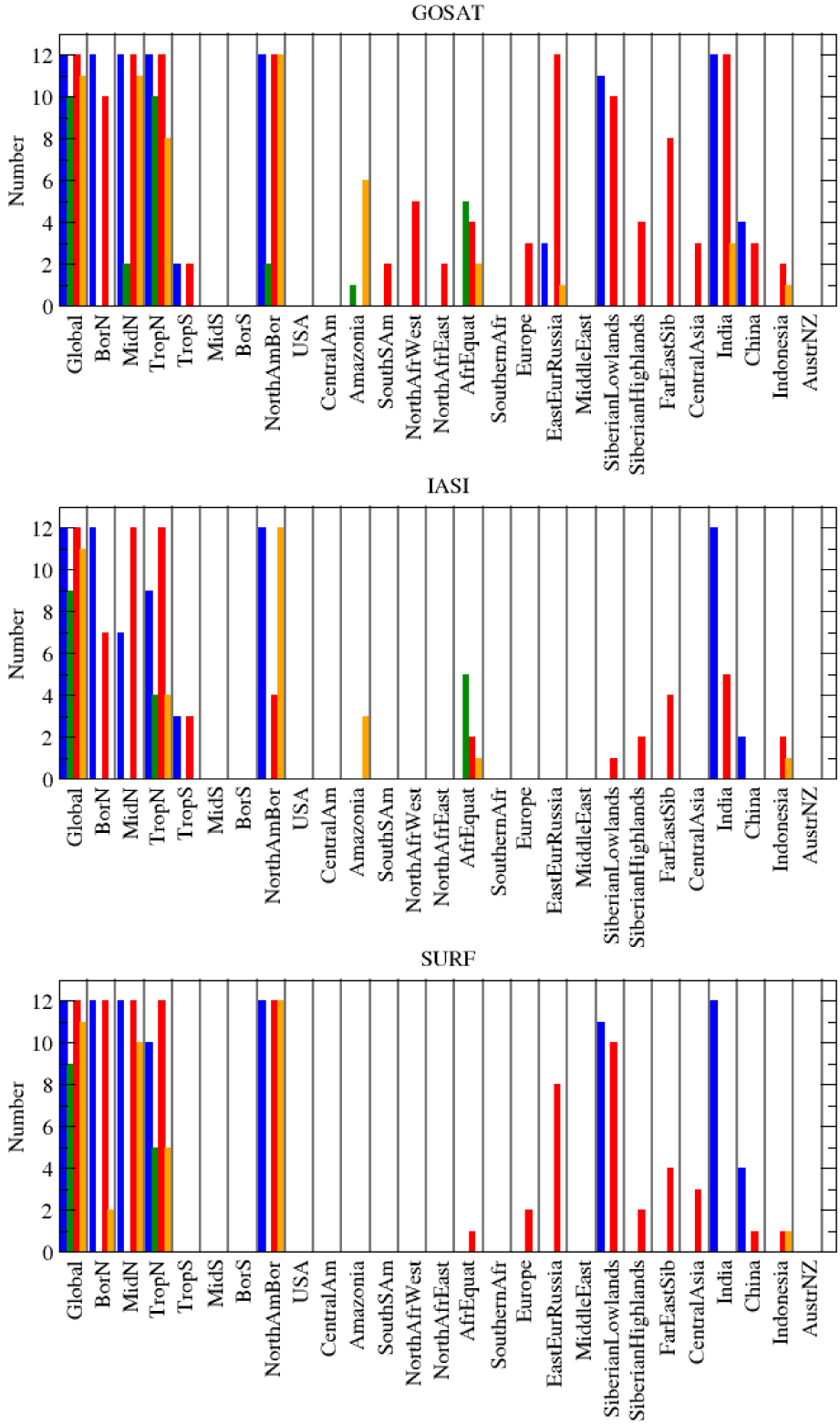
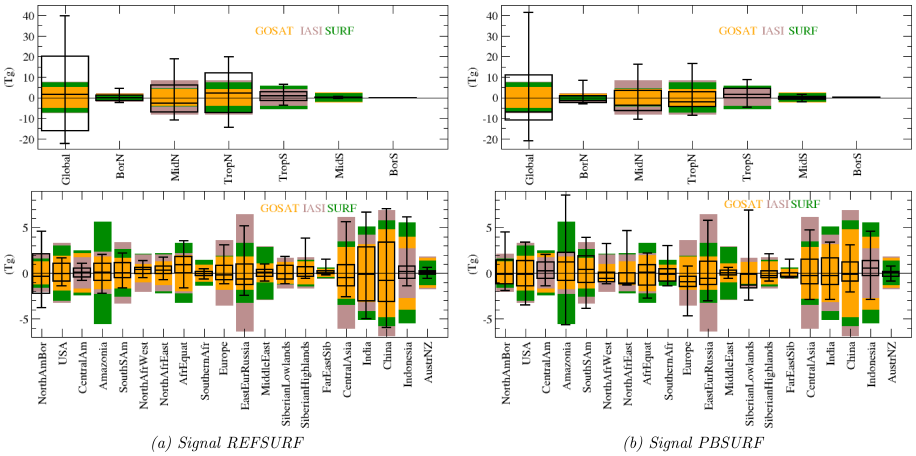


Figure 3. *Number of detected seasons over the 12 possible for winter (JFM, blue), spring (AMJ, green), summer (JAS, red) and fall (OND, orange) in the various regions.*



(a) Signal REFSURF *(b) Signal PBSURF*

Figure 4. *Noise at the seasonal time scale by the three observing systems (bars) and box plots (median, 25 and 75%) for the signal in various areas (latitudinal bands and regions). Detection is achieved when the signal is larger than the noise i.e. for all the occurrences in each box plot which lay outside the matching ~~colored~~ coloured bar.*

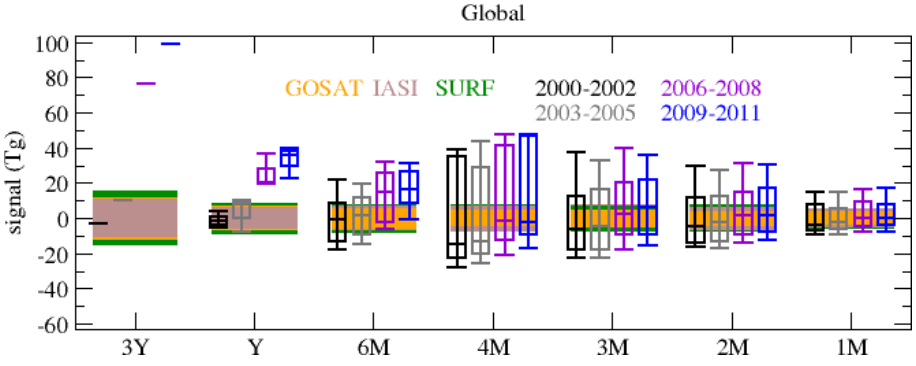


Figure 5. *Impact of temporal ~~agregation~~ aggregation on noise (bars) and signal (box plots with median, 25 and 75%) over 3-year ~~time-windows~~time windows. Detection is achieved when the signal is larger than the noise i.e. for all the occurrences in each box plot which lay outside the matching coloured bar. Link to Table 3: the Global lines of the Table corresponds to the 3Y bars here.*

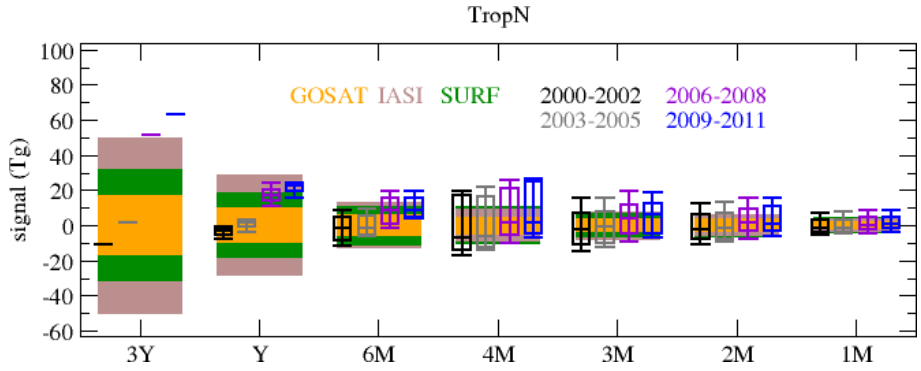


Figure 6. *Impact of temporal ~~agregation~~ aggregation on noise (bars) and signal (box plots with median, 25 and 75%) over 3-year ~~time-windows~~time windows. Detection is achieved when the signal is larger than the noise i.e. for all the occurrences in each box plot which lay outside the matching coloured bar. Link to Table 3: the TropN lines of the Table corresponds to the 3Y bars here.*

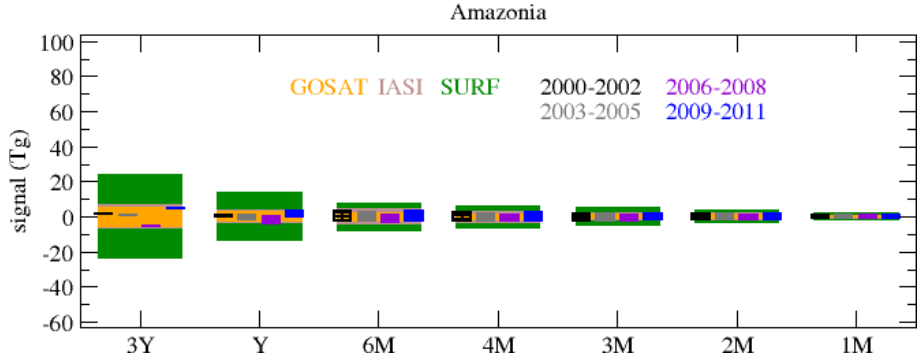


Figure 7. *Impact of temporal ~~agregation~~ aggregation on noise (bars) and signal (box plots with median, 25 and 75%) over 3-year ~~time-windows~~time windows. Detection is achieved when the signal is larger than the noise i.e. for all the occurrences in each box plot which lay outside the matching coloured bar. Link to Table 3: the Amazonia lines of the Table corresponds to the 3Y bars here.*

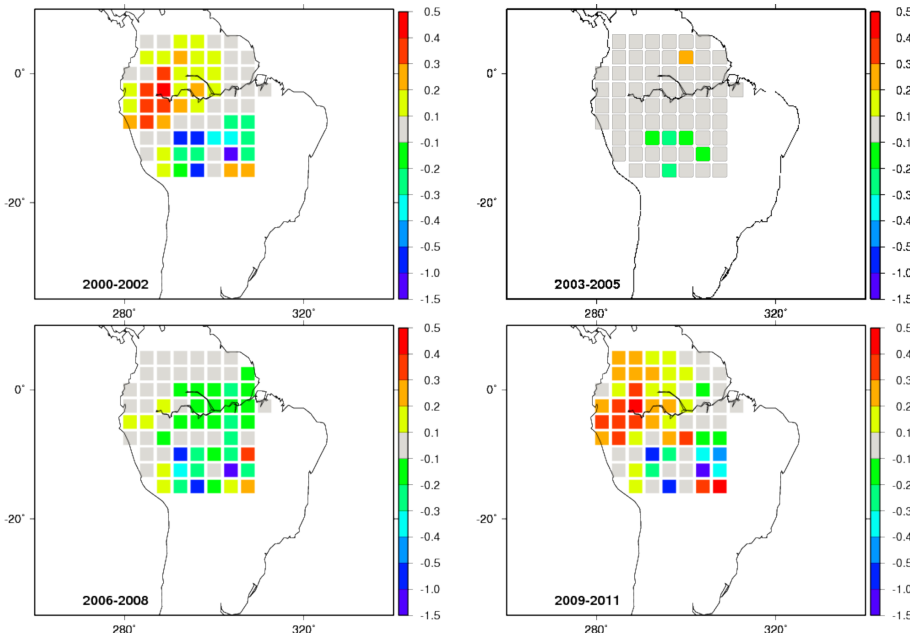


Figure 8. *Signal (Tg) for the four 3-year ~~time-windows~~ time windows at the pixel scale.*

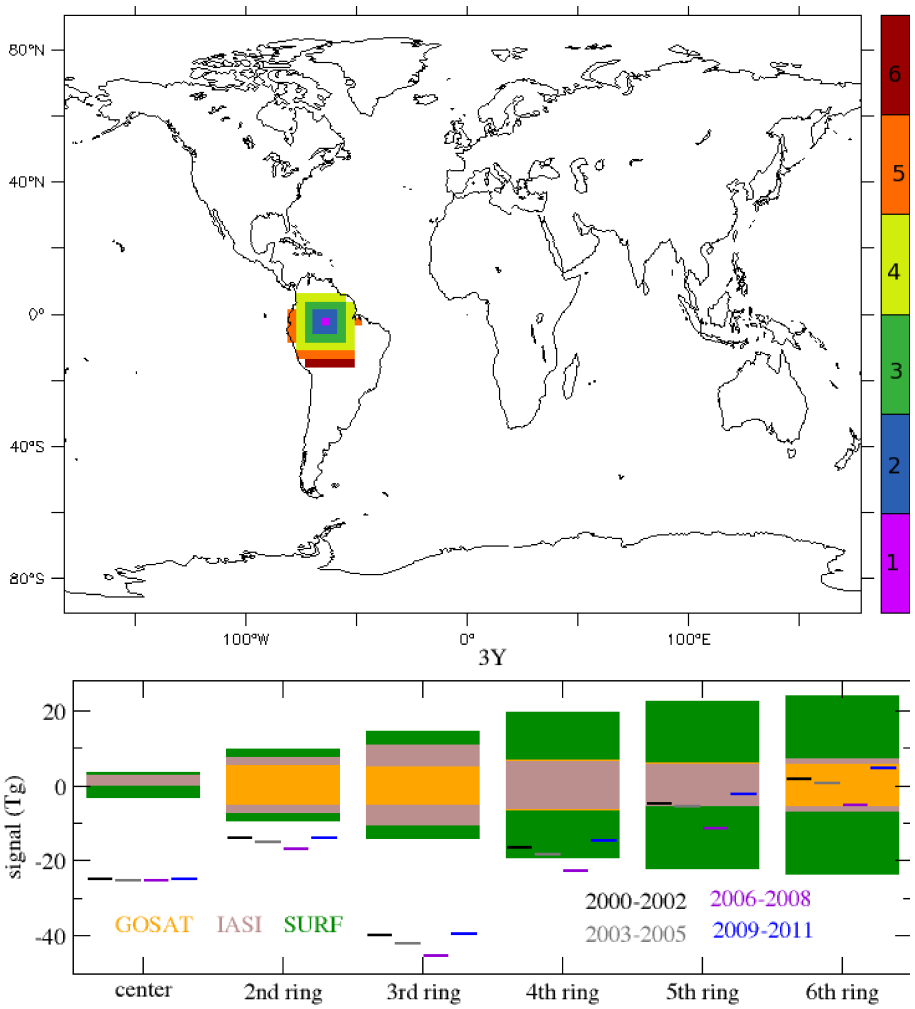


Figure 9. *Impact of spatial aggregation in Amazonia on noise (bars) and signal (box plots with median, 25 and 75%) over 3-year time windows: from a unique pixel to larger rings around it. Detection is achieved when the signal is larger than the noise i.e. for all the occurrences which lay outside the matching coloured bar.*

Table 1: Detection of the signal consisting in the anomalies at the "seasonal" ~~time-scale~~ time scale i.e. quarters of the year (JFM, AMJ, JAS, OND). The signal is the difference between each quarter in the 2000-2011 period (i.e. 48 occurrences) and the 2004-2005 average from REFSURF. The noise is computed at the quarter ~~time-scale~~ time scale from each of the three observation systems, GOSAT, IASI and SURF. See Section 2.4 and Section 2.3 for details. In each cell of the Table, we show X%(YY/ZZ) where X% is the percentage of quarterly anomalies detected (among 48 possible), YY is the number of positive anomalies detected among the ZZ detected anomalies. Column "Ave. mass" indicates the average emitted mass of CH_4 over 2004-2005 in the area.

Region	Ave. mass (Tg)	GOSAT	IASI	SURF
Global	517	~~97%(24/47)~~ 93%(22/45)	~~93%(22~~91%(21/~~45~~44)	91%(21/44)
BorN	18	~~50%(12~~45%(10/~~24~~22)	~~50%(12~~39%(07/~~24~~19)	~~52%(13~~54%(12/~~25~~26)
MidN	177	~~87%(18~~77%(14/~~42~~37)	~~54~~39%(12/~~26~~19)	~~81%(16~~70%(12/~~39~~34)
TropN	194	~~91%(24~~87%(22/~~44~~42)	~~79%(20~~60%(16/~~38~~29)	~~81%(21~~66%(17/~~39~~32)
TropS	115	~~25%(10~~08%(03/~~12~~04)	~~22%(10~~12%(04/~~11~~06)	~~10%(05/05)~~ ∅
MidS	12	∅	∅	∅
BorS	1	∅	∅	∅
NorthAmBor	20	~~97%(23~~79%(14/~~47~~38)	~~52%(13~~58%(04/~~25~~28)	~~87%(18~~75%(12/~~42~~36)
USA	37	~~20%(09/10)~~ ∅	∅	∅
CentralAm	17	∅	∅	∅
Amazonia	38	~~08~~14%(01/~~04~~07)	~~02~~06%(00/~~01~~03)	∅
SouthSAm	30	~~06%(03~~04%(00/~~03~~02)	∅	∅
NorthAfrWest	13	~~16%(08~~10%(05/~~08~~05)	∅	~~04%(02/02)~~ ∅
NorthAfrEast	11	~~20%(10~~04%(02/~~10~~02)	~~06%(03/03)~~ ∅	~~02%(01/01)~~ ∅
AfrEquat	32	~~35%(17~~22%(11/~~17~~11)	~~20%(10~~16%(08/~~10~~08)	~~04%(02~~02%(01/~~02~~01)
SouthernAfr	10	~~04%(00/02)~~ ∅	∅	∅
Europe	33	~~14%(07~~06%(03/~~07~~03)	∅	04%(02/02) ~~14%(07/07)~~
EastEurRussia	30	~~45~~33%(12/~~22~~16)	∅	16%(08~~%(04~~/~~04~~08) ~~22%(11/11)~~

Table 1: (continued) Detection of the signal consisting in the anomalies at the "seasonal" ~~time-scale~~time scale.

Region	Ave. mass (Tg)	GOSAT	IASI	SURF
MiddleEast	16	~~14%(04/07)~~ ∅	∅	∅
SiberianLowlands	8	~~47%(12~~43%(10/~~23~~21)	~~12%(06~~02%(01/~~06~~01)	~~35%(12~~43%(10/~~17~~21)
SiberianHighlands	5	~~22%(11~~08%(04/~~11~~04)	~~06%(03~~04%(02/~~03~~02)	~~22%(11~~04%(02/~~11~~02)
FarEastSib	1	16%(08/08)	~~16%(08~~08%(04/~~08~~04)	~~16%(08~~08%(04/~~08~~04)
CentralAsia	28	~~33%(09~~06%(03/~~16~~03)	~~02%(01/01)~~ ∅	~~18%(06~~06%(03/~~09~~03)
India	50	~~62%(13/30) 50%(~~56%(12/~~24~~27)	35%(~~12~~05/17)	25%(00/12)
China	64	~~43%(11~~14%(03/~~21~~07)	~~10%(03~~04%(00/~~05~~02)	~~31%(09~~10%(01/~~15~~05)
Indonesia	36	06%(03/03)	~~16%(07/08)~~ 06%(03/03)	04%(02/02)
AustrNZ	6	~~02%(01/01)~~ ∅	∅	~~02%(01/01)~~ ∅

Table 2: Detection of the signal consisting in the anomalies at the yearly ~~time-scale~~time scale. The signal is the difference between each year in the 2000-2011 period (i.e. 12 occurrences) and the 2004-2005 average from REFSURF. The noise is computed at the yearly ~~time-scale~~ time scale from each of the three observation systems, GOSAT, IASI and SURF. See Section 2.4 and Section 2.3 for details. In each cell of the Table, we show X%(YY/ZZ) where X% is the percentage of yearly anomalies detected (among 12 possible), YY is the number of positive anomalies detected among the ZZ detected anomalies. Column "Ave. mass" indicates the average emitted mass of CH_4 over 2004-2005 in the area.

Region	Ave. mass (Tg)	~~Gosat~~ GOSAT	~~Iasi~~ IASI	~~Surf~~S
Global	517	~~83~~75%(08/~~10~~09)	75%(08/09)	58%(
BorN	18	∅	∅	∅
MidN	177	08%(01/01)	∅	08%(
~~MidN 177 66%(07/08) ∅ ∅~~ TropN	194	~~58~~50%(06/~~07~~06)	~~41%(05/05)~~ ∅	~~50%(06~~25%
TropS	115	~~25%(03~~16%(02/~~03~~02)	~~33%(04~~16%(02/~~04~~02)	16%(
MidS	12	∅	∅	∅
BorS	1	∅	∅	∅
NorthAmBor	20	∅	∅	∅
USA	37	∅	∅	∅
CentralAm	17	∅	∅	∅
Amazonia	38	~~∅~~ 08%(00/01)	∅	∅
SouthSAm	30	~~08%(01/01)~~ ∅	∅	∅
NorthAfrWest	13	~~41%(05/05)~~ 33%(04/04)	~~41%(05/05)~~ ∅	∅
NorthAfrEast	11	~~50%(06/06) 25%(03/03)~~ 08%(01/01)	∅	∅
AfrEquat	32	~~41%(05~~33%(04/~~05~~04)	33%(04/04)	~~33%(04~~25%
SouthernAfr	10	∅	∅	∅
Europe	33	16%(02/02)	~~08%(01/01)~~ ∅	~~16%(02~~
EastEurRussia	30	∅	∅	∅
MiddleEast	16	~~58%(04/07)~~ ∅	∅	∅

Table 2: (continued) Detection of the signal consisting in the anomalies at the yearly ~~time-scale~~time scale.

Region	Ave. mass (Tg)	Gosat	Iasi	Su
SiberianLowlands	8	∅	∅	∅
SiberianHighlands	5	08%(01/01)	~~08%(01/01)~~ ∅	~~08%(01~~
FarEastSib	1	~~25%(03~~08%(01/~~03~~01)	~~25%(03/03)~~ ∅	~~25%(03~~
CentralAsia	28	08%(00/01)	∅	∅
India	50	∅	∅	∅
China	64	∅	∅	~~16%(00~~
Indonesia	36	16%(02/02)	~~25%(02/03)~~ 16%(02/02)	∅
AustrNZ	6	~~16%(02/02)~~ ∅	∅	∅

Table 3: Detection of the signal consisting in the anomalies at the 3-yearly ~~time-scale~~time scale. The signal is the difference between each 3-year ~~time-window~~ time window in the 2000-2011 period (2000-2002, 2003-2005, 2006-2008, 2009-2011) and the 2004-2005 average from REFSURF. The noise is computed at the 3-yearly ~~time-scale~~ time scale from each of the three observation systems, GOSAT, IASI and SURF. See Section 2.4 and Section 2.3 for details.

In each cell of the Table, we show whether a positive anomaly, a negative anomaly or no anomaly is detected and with which signal-to-noise ratio: positive anomaly detected: +++ = with stn ratio > 3, ++= stn ratio > 2 and + = stn ratio > 1; negative anomaly detected with - -= stn ratio <-2, - = stn ratio <-2,∅ = no anomaly detected.

The number below the name of the area is the average emitted mass of CH_4 over 2004-2005 in the area.

Region	System	2000-2002	2003-2005	2006-2008	2009-2011
Global	Gosat	∅	~~+~~∅	+++	+++
517	Iasi	∅	~~+~~∅	+++	+++
	Surf	∅	∅	+++	+++
BorN	Gosat	∅	∅	∅	∅
18	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
MidN	Gosat	~~-~~∅	∅	+ ~~+~~	+ ~~+~~
177	Iasi	∅	∅	∅	∅
	Surf	∅	∅	+	+
TropN	Gosat	∅	∅	++ ~~+~~	+++
194	Iasi	∅	∅	+ ~~+~~	+ ~~+~~
	Surf	∅	∅	+ ~~++~~	+ ~~++~~
TropS	Gosat	~~+~~∅	∅	∅	+
115	Iasi	+	∅	∅	+
	Surf	~~+~~∅	∅	∅	+
MidS	Gosat	∅	∅	∅	∅
12	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
BorS	Gosat	∅	∅	∅	∅
1	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅

Table 3: (continued) Detection of the signal consisting in the anomalies at the 3-yearly ~~time-scale~~time scale.

Region	System	2000-2002	2003-2005	2006-2008	2009-2011
NorthAmBor	Gosat	~~-~~∅	∅	∅	~~-~~∅
20	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
USA	Gosat	∅	∅	∅	∅
37	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
CentralAm	Gosat	∅	∅	∅	∅
17	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
Amazonia	Gosat	∅	∅	∅	∅
38	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
SouthSAm	Gosat	∅	∅	∅	+
30	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	~~+~~∅
NorthAfrWest	Gosat	∅	∅	+	++
13	Iasi	∅	∅	~~+~~∅	+ ~~+~~
	Surf	∅	∅	~~+~~∅	+ ~~+~~
NorthAfrEast	Gosat	∅	∅	+ ~~++~~	+ ~~++~~
11	Iasi	∅	∅	+	+
	Surf	∅	∅	+	+
AfrEquat	Gosat	~~+~~	∅	++ ~~+~~	+++
32	Iasi	∅	∅	++	+++
	Surf	∅	∅	+	++
SouthernAfr	Gosat	∅	∅	∅	∅
10	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
Europe	Gosat	+	∅	∅	∅

Table 3: (continued) Detection of the signal consisting in the anomalies at the 3-yearly ~~time-scale~~time scale.

Region	System	2000-2002	2003-2005	2006-2008	2009-2011
33	Iasi	+∅	∅	∅	∅
	Surf	+	∅	∅	∅
EastEurRussia	Gosat	∅	∅	∅	∅
30	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
MiddleEast	Gosat	- −	∅	+∅	+ +
16	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
SiberianLowlands	Gosat	∅	∅	∅	∅
8	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
SiberianHighlands	Gosat	+∅	∅	∅	∅
5	Iasi	∅	∅	∅	∅
	Surf	+∅	∅	∅	∅
FarEastSib	Gosat	+ +	+∅	∅	+∅
1	Iasi	+∅	+∅	∅	∅
	Surf	++∅	+∅	∅	+∅
CentralAsia	Gosat	-	∅	∅	∅
28	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
India	Gosat	∅	∅	∅	∅
50	Iasi	∅	∅	∅	∅
	Surf	∅	∅	∅	∅
China	Gosat	−∅	∅	∅	∅
64	Iasi	-	∅	+∅	∅
	Surf	−∅	∅	+∅	+∅
Indonesia	Gosat	+	∅	+∅	∅
36	Iasi	+ +	∅	+ +	∅

Table 3: (continued) Detection of the signal consisting in the anomalies at the 3-yearly ~~time-scale~~time scale.

Region	System	2000-2002	2003-2005	2006-2008	2009-2011
	Surf	∅	∅	∅	∅
AustrNZ	Gosat	+∅	∅	∅	∅
6	Iasi	+∅	∅	∅	∅
	Surf	+∅	∅	∅	∅

Appendix A: Supplementary tables

Table 4: Yearly mean number of observations over the period used for the Monte-Carlo noise computation (10/2009-09/2010) in the various regions for the three observing systems.

Region	Area (x10^6km^2)	GOSAT	IASI	SURF
Global	510	32348	240084	1722
BorN	31	92	00	172
MidN	91	9060	00	556
TropN	126	14934	121756	602
TropS	128	6118	107148	156
MidS	95	2132	9078	140
BorS	37	00	00	96
NorthAmBor	14	194	00	00
USA	11	2516	2218	124
CentralAm	05	608	6328	24
Amazonia	07	802	3366	00
SouthSAm	10	1780	3068	24
NorthAfrWest	10	4986	4564	94
NorthAfrEast	07	3756	5148	00
AfrEquat	07	1394	3572	14
SouthernAfr	07	1488	3246	28
Europe	06	572	00	94
EastEurRussia	07	896	00	00
MiddleEast	06	2456	3748	26
SiberianLowlands	02	170	00	00

Table 4: (continued) Yearly mean number of observations.

Region	Area (x10^6km^2)	GOSAT	IASI	SURF
SiberianHighlands	05	126	00	00
FarEastSib	03	54	00	00
CentralAsia	12	3864	694	74
India	03	1180	4190	00
China	05	1164	4574	00
Indonesia	07	312	3324	26
AustrNZ	10	3308	4362	50

Table 5: Detection of the signal consisting in the anomalies at the "seasonal" ~~time-scale~~ time scale (JFM, AMJ, JAS, OND). The signal is the difference between each quarter in the 2000-2011 period (i.e. 48 occurrences) and the 2004-2005 average from PBSURF. The noise is computed at the quarter ~~time-scale~~ time scale from each of the three observation systems, GOSAT, IASI and SURF. See Section 2.4 and Section 2.3 for details. In each cell of the Table, we show X% [±TT] (±YY/±ZZ) where X% is the percentage of quarterly anomalies detected, [±TT] is the difference with REFSURF (Table 1), ±YY is the difference in the number of positive anomalies detected compared to REFSURF and ±ZZ is the difference in the total number of detected anomalies compared to REFSURF. Ave. mass= average emitted mass of CH_4 over 2004-2005.

Region Ave. mass (Tg) REFSURF/PBSURF	Gosat	Iasi	Surf
Global 517/499	~~93~~ 87% [~~-4~~-6] (~~-11~~-10/~~-2~~-3)	~~75~~ 72% [~~-18~~-19] (~~-10~~-9/-9)	~~85~~ 72% [~~-8~~-19] (~~-10~~-9/~~-4~~-9)
BorN 18/17	~~87~~ 75% [+~~37~~30] (~~0~~+2/+~~18~~14)	~~81~~ 75% [+~~31~~36] (~~0~~+5/+~~15~~17)	~~87~~ 77% [+~~35~~23] (~~-1~~0/+~~17~~11)
MidN 177/172	~~83~~ 66% [~~-4~~-11] (~~-6~~/-2/~~-5~~)	~~50~~ 35% [-4] (0/-2)	~~77~~ 62% [~~-4~~-8] (~~-4~~0/~~-2~~-4)
TropN 194/165	~~64~~ 47% [~~-27~~-40] (~~-12~~-10/~~-13~~-19)	~~37~~ 27% [~~-42~~-33] (~~-8~~-5/~~-20~~-16)	~~39~~ 29% [~~-42~~-37] (~~-9~~-6/~~-20~~-18)
TropS 115/120	~~43~~ 29% [+~~18~~21] (+~~7~~9/+~~9~~10)	~~43~~ 31% [+~~21~~19] (+~~7~~9/+~~10~~9)	~~27~~ 10% [+~~17~~10] (+~~7~~5/+~~8~~5)
MidS 12/25	~~10 +10(+2/+5)~~ ∅	~~10 +10(+2/+5)~~ ∅	~~14~~ 04% [+~~14~~4] (+~~3~~0/+~~7~~2)
BorS 1/0	~~97 +97(+23/+47)~~ ∅	~~10 +10(0/0)~~ ∅	~~91 +91(+20/+44)~~ ∅
NorthAmBor 20/8	~~54~~ 64% [~~-43~~-15] (~~-11~~-2/~~-21~~-7)	~~27~~ 43% [~~-25~~-15] (~~-1~~+3/~~-12~~-7)	~~39~~ 58% [~~-48~~-17] (~~-6~~0/~~-23~~-8)
USA 37/54	~~58~~ 3%1 [+~~38~~31] (+~~4~~8/+~~18~~15)	~~14~~ 06% [+~~14~~6] (+~~3~~1/+~~7~~3)	~~08~~ 10% [+~~8~~10] (+~~2~~3/+~~4~~5)
CentralAm 17/13	~~12 +12(+6/+6)~~ ∅	~~25~~ 02% [+~~25~~2] (+~~11~~1/+~~12~~1)	~~10 +10(+5/+5)~~ ∅
Amazonia 38/31	~~47~~ 45% [+~~39~~31] (+~~17~~19/+~~19~~15)	35% [+~~33~~29] (+~~14~~15/+~~16~~14)	~~08~~ 04% [+~~8~~4] (+~~3~~2/+~~4~~2)
SouthSAm 30/45	~~47~~ 45% [+41] (+~~12~~15/+20)	~~25~~ 08% [+~~25~~8] (+~~5~~3/+~~12~~4)	~~25~~ 20% [+~~25~~20] (+~~5~~6/+~~12~~10)
NorthAfrWest 13/13	~~58~~ 41% [+~~42~~31] (+~~4~~7/+~~20~~15)	~~25~~ 16% [+~~25~~16] (+~~12~~8/+~~12~~8)	~~25~~ 16% [+~~21~~16] (+~~10~~8/+~~10~~8)
NorthAfrEast 11/12	~~97~~ 39% [+~~77~~35] (+~~2~~10/+~~37~~17)	~~29~~ 25% [+~~23~~25] (+~~9~~12/+~~11~~12)	25% [+~~23~~25] (+~~11~~12/+~~11~~12)
AfrEquat 32/33	~~25~~ 18% [~~-10~~-4] (~~-14~~-10/~~-5~~-2)	~~14~~ 10% [-6] (-8/-3)	00% [~~-4~~-2] (~~-2~~-1/~~-2~~-1)
SouthernAfr 10/14	~~52~~ 43% [+~~48~~43] (+~~10~~7/+~~23~~21)	~~37~~ 14% [+~~37~~14] (+~~9~~5/+~~18~~7)	~~25~~ 14% [+~~25~~14] (+~~7~~5/+~~12~~7)
Europe 33/33	~~31~~ 14% [+~~17~~8] (~~-7~~-3/+~~8~~4)	~~08~~ 04% [+4] (~~-2~~0/+2)	~~37~~ 12% [+~~23~~8] (~~-7~~-2/+~~11~~4)

Table 5: (continued) Detection of the signal consisting in the anomalies at the "seasonal" ~~time-scale~~time scale.

Region Ave. mass (Tg) REFSURF/PBSURF	GOSAT	IASI	SURF
EastEurRussia 30/27	~~47~~ 33% [~~+2~~0] (~~-1/+1) 04 -4(~~-2/~~-2~~0)	~~18~~ ∅	10% [~~-4~~-6] (-3/~~-2~~-3)
MiddleEast 16/14	~~00 -14(-4/-7)~~ ∅	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅
SiberianLowlands 8/14	~~97~~ 89% [+~~50~~46] (~~0~~+2/+~~24~~22)	~~64~~ 60% [+~~52~~58] (+~~6~~11/+~~25~~28)	~~91~~ 85% [+~~56~~42] (~~0~~+2/+~~27~~20)
SiberianHighlands 5/4	~~25~~ 22% [+~~3~~14] (~~0~~+7/+~~1~~7)	~~22~~ 12% [+~~16~~8] (+~~8~~4/+~~8~~4)	~~25~~ 20% [+~~3~~16] (~~0~~+8/+~~1~~8)
FarEastSib 1/2	~~87~~ 52% [+~~71~~36] (+4/+~~34~~17)	~~72~~ 50% [+~~56~~42] (+~~4~~8/+~~27~~20)	~~83~~ 50% [+~~67~~42] (+~~4~~8/+~~32~~20)
CentralAsia 28/32	~~37~~ 20% [+~~4~~14] (+~~3~~5/+~~2~~7)	~~02 0(0/0)~~ ∅	~~31~~ 08% [+~~1~~32] (+~~5~~1/+~~6~~1)
India 50/45	~~22~~ 10% [~~-40~~-46] (~~-7~~-11/~~-19~~-22)	02% [~~-48~~-33] (~~-11~~-5/~~-23~~-16)	00% [~~-35~~-25] (~~-12~~0/~~-17~~-12)
China 64/46	00% [~~-43~~-14] (~~-11~~-3/~~-21~~-7)	00% [~~-10~~-4] (~~-3~~0/~~-5~~-2)	00% [~~-31~~-10] (~~-9~~-1/~~-15~~-5)
Indonesia 36/33	~~20~~ 06% [~~+14~~0] (~~+4~~0/~~+7~~0)	~~31~~ 12% [+~~15~~6] (+~~5~~2/+~~7~~3)	~~08~~ 00% [~~+2~~-4] (~~+1~~-2/~~+1~~-2)
AustrNZ 6/6	~~06 +4(+1/+2)~~ ∅	~~00 0(0/0)~~ ∅	~~06 +4(+1/+2)~~ ∅

Table 6: Detection of the signal consisting in the anomalies at the yearly ~~time-scale~~time scale. The signal is the difference between each year in the 2000-2011 period (i.e. 12 occurrences) and the 2004-2005 average from PBSURF. The noise is computed at the yearly ~~time-scale~~ time scale from each of the three observation systems, GOSAT, IASI and SURF. See Section 2.4 and Section 2.3 for details. In each cell of the Table, we show X% [±TT] (±YY/±ZZ) where X% is the percentage of yearly anomalies detected, [±TT] is the difference with REFSURF (Table 2), ±YY is the difference in the number of positive anomalies detected compared to REFSURF and ±ZZ is the difference in the total number of detected anomalies compared to REFSURF. Ave. mass= average emitted mass of CH_4 over 2004-2005.

Region Ave. mass (Tg) REFSURF/PBSURF	GOSAT	IASI	SURF
Global 517/499	~~75~~ 58% [~~-8~~-17] (~~0/~~-1/~~-2~~)	66% [-9] (0/-1)	~~50~~ 4%1 [~~-8~~-17] (~~-1~~-2/~~-1~~
BorN 18/17	~~41 +33(+4/+4)~~ ∅	~~00 0(0/0)~~ ∅	~~41 +33(+4/+4)~~ ∅
MidN 177/172	~~25 -41(-6/-5)~~ 00% [~~0~~-8] (~~0~~-1/~~0~~-1)	∅	00% [~~0~~-8] (~~0~~-1/~~0~~-1)
TropN 194/165	~~16~~ 00% [~~-42~~-50] (~~-4~~-6/~~-5~~-6)	∅	00% [~~-41~~-25] (~~-5~~-3/~~-5) 08 -42~~
TropS 115/120	~~50~~ 41% [+25] (+3/+3)	~~66~~ 41% [+~~33~~25] (+3/+~~43~~)	41% [+25] (+3/+3)
MidS 12/25	~~08 +8(+1/+1)~~ ∅	~~33 +33(+1/+4)~~ ∅	~~00 0(0/0)~~ ∅
BorS 1/0	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅
NorthAmBor 20/8	~~41~~ 25% [+~~41~~25] (+~~5~~3/+~~53~~)	~~08 +8(+1/+1)~~ ∅	~~16 +16(+2/+2)~~ ∅
USA 37/54	~~16 +16(0/+2)~~ ∅	~~16 +16(0/+2)~~ ∅	~~00 0(0/0)~~ ∅
CentralAm 17/13	~~00 0(0/0)~~ ∅	~~25 +25(+3/+3)~~ ∅	~~00 0(0/0)~~ ∅
Amazonia 38/31	~~50~~ 58% [+50] (+~~6~~7/+6)	58% [+58] (+7/+7)	08% [+8] (+1/+1)
SouthSAm 30/45	33% [+~~25~~33] (+~~2~~3/+~~34~~)	~~25~~ 16% [+~~25~~16] (+2/+~~32~~)	~~25~~ 16% [+~~25~~16] (+2/+
NorthAfrWest 13/13	00% [~~-41(-5/-5) 00~~ -33] (-4/-4)	~~00 -41(-5/-5)~~ ∅	∅
NorthAfrEast 11/12	00% [~~-50(-6/-6) 00 -25(-3/-3) 00~~ -8] (-1/-1)	∅	∅
AfrEquat 32/33	~~08~~ 00% [-33] (~~-5~~-4/-4)	00% [-33] (-4/-4)	00% [~~-33~~-25] (~~-4~~-3/~~-4~~
SouthernAfr 10/14	~~25 +25(+2/+3)~~ 16% [+16] (+1/+2)	~~00 0(0/0)~~ ∅	∅
Europe 33/33	50% [+34] (-2/+4)	~~41~~ 08% [+~~33~~8] (~~-1~~0/~~4~~+1)	~~50~~ 33% [+~~34~~33] (~~-2~~0/~~

Table 6: (continued) Detection of the signal consisting in the anomalies at the yearly ~~time-scale~~time scale.

Region Ave. mass (Tg) REFSURF/PBSURF	Gosat	Iasi	Surf
EastEurRussia 30/27	16% [+16] (+1/+2)	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅
MiddleEast 16/14	~~00 -58(-4/-7)~~ ∅	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅
SiberianLowlands 8/+14	~~25 +25(+2/+3) 00 0(0/0)~~ 16% [+16] (+1/+2)	∅	∅
SiberianHighlands 5/4	00% [-8] (-1/-1)	~~00 -8(-1/-1)~~ ∅	~~00 -8(-1/-1)~~ ∅
FarEastSib 1/2	~~08 -17(-2/-2)~~ 00% [~~-25~~-8] (~~-3~~-1/~~-3~~-1)	~~08 -17(-2/-2)~~ ∅	∅
CentralAsia 28/32	00% [-8] (0/-1)	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅
India 50/45	~~08 +8(0/+1)~~ ∅	~~08 +8(0/+1)~~ ∅	~~00 0(0/0)~~ ∅
China 64/46	~~00 0(0/0)~~ ∅	~~00 0(0/0)~~ ∅	~~00 -16(0/-2)~~ ∅
Indonesia 36/33	~~50~~ 33% [+~~34~~17] (+~~3~~2/+~~4~~2)	~~66~~ 33% [+~~41~~17] (+~~4~~2/+~~5~~2)	~~25 +9(+1/+1)~~ ∅
AustrNZ 6/6	~~16 0(-1/0)~~ ∅	~~08 +8(0/+1)~~ ∅	~~08 +8(0/+1)~~ ∅