Our responses to reviewers are in blue.

REFEREE #2
Received and published: 14 December 2011

The paper compares CH4 retrievals from the Tropospheric Emission Spectrometer (TES) with insitu aircraft profile measurements from the HIPPO program. The authors demonstrate that the new 'experimental' TES retrieval product V005 agrees better with the HIPPO observations than the current 'standard' V004 retrieval version. The paper also includes a first inverse modeling study to evaluate the usefulness of the TES retrievals.
Overall, the paper is interesting and well presented, and should be suitable for publication in ACP after consideration of the following comments:

General Comments
(1) Evaluation of bias of TES retrievals vs. HIPPO
The authors conclude that there is no systematic trend in the bias as function of latitude and no statistically significant difference in bias between TES observations over land and over ocean, based on the presentation of the bias as function of latitude in Fig. 4 (for TES V004) and Fig. 5 (for TES V005). However, any potential systematic dependence of the bias on latitude (or land vs. sea) might not be clearly visible in the plots as currently presented. In addition to presenting the single bias values ΔyR, ΔyL, ΔyU for all coinciding TES-HIPPO observations, it would be useful to present also mean and median bias values as function of latitude (e.g. in latitude bins of 5-10o degrees), to better analyze any potential latitude dependence of the bias.
Such potential systematic biases are very critical for inverse modeling. E.g., Meirink et al. [2006] demonstrated that already "Systematic measurement errors well below 1% have a dramatic impact on the quality of the derived emission". Their analysis was for SCIAMACHY, i.e. a near-IR instrument, which is sensitive also to the boundary layer, while the requirements for relative accuracy are probably even higher for thermal IR sensors such as TES, which are sensitive mainly to the middle and upper troposphere (with overall smaller XCH4 gradients). To better evaluate the usefulness of the TES retrievals, it would be essential to provide a more quantitative conclusion about any potential systematic dependence of the bias as function of latitude (and land vs. ocean). In general the requirements for relative accuracy should be discussed in more detail.
We add a paragraph to section 3.2 showing that the mean bias has no statistically significant latitudinal dependence or statistically significant dependence on land vs. ocean. We add 10-degree latitude bins of mean bias and residual standard deviation to figures 4 and 6.

(2) OSSE experiments
The OSSE inverse modeling study assigns a 50% error to emissions in the a priori error covariance matrix Sa. The authors state that this choice is 'commensurate with the perturbation made to the "true" emissions', which is also 50%. However, Sa is assumed to be diagonal (i.e. not taking into account any spatial error correlations, while the perturbation is applied on very large, continental scales (Fig. 8). Hence, on the scale of these perturbations, the aggregated a priori uncertainty is much smaller than the assumed 50% per grid cell. I would expect that adapting the settings of the a priori error covariance matrix (e.g. by taking into account some spatial error correlation) should help to retrieve the 'true' emissions at least somewhat better.

We add the following to section 4, “Although the a priori errors in figure 8 are highly spatially correlated through the use of homogeneous perturbations in large blocks, this correlation is mainly for ease of interpretation and we would not expect such correlation in actual a priori errors. We therefore do not include error covariance terms in $\mathbf{S_a}$.”

At the end of section 4 the authors state that "V005 should be more successful.", however without making any attempt to test this. The authors could easily demonstrate the impact of the reduced noise of the V005 vs V004 on the capability to better retrieve the ’true’ emissions based on their presented OSSE / inverse modeling system.
The problem is that we don’t have a global data set of averaging kernels. We say on p. 27892 line 10-11, “V005 retrievals were performed on an experimental basis along the HIPPO I and HIPPO II flight paths.”
We add a line to the end of section 4 saying, “The TES V005 data with smaller errors and higher DOFS therefore hold promise but quantitative testing must await availability of a global database of averaging kernels.”

In addition, it would be very interesting to explore the usefulness of having independent retrievals for the lower and upper troposphere for the V005 retrievals - although this might be beyond the scope of the present paper.
See response to previous comment.

Detailed Comments
Several references are missing (e.g. Rodgers (2000), Bowman et al. (2006), Payne et al. (2009), Osterman et al. (2009)....)
References added.

Introduction: The authors should include also some discussion on the requirements for relative accuracy (see general comment (1)) - currently only the precision requirements are discussed.
Added “and accuracy of at least 1% ” to the following sentence in the introduction: Space-borne observations of column methane require precision of 1-2% and accuracy of at least 1% for inverse modeling of methane sources (Meirink et al., 2006).

Section 2.2: On which CH4 scale are the HIPPO data reported (e.g. NOAA04 ?)
HIPPO data are reported on the NOAA04 scale. Added a sentence to section 2.2 explaining this.

page 27896, line 29: ’TES successfully recovers “true” sources on continental scales’:
this seems not so clearly visible in Fig. 8.
We remove the sentence in question.

Fig.1 / right panel: label at x-axis are missing.
Done.

Fig.4/5: In addition to the general comment (1), I would recommend to display the single bias values ΔyR, ΔyL, ΔyU for all coinciding TES-HIPPO observations with smaller symbols.

Furthermore it would be useful to choose colors for ocean and land, which can be better distinguished than the current blue/green.
Symbols are smaller in both figures. Land values in figure 4 are now red.

It seems that there is some latitudinal dependence of the bias (e.g. in Fig. 5b the values between 30 and 40 degrees are significantly lower than for most other latitudes).
There were discontinuities at 30N and 30S in the N2O profiles previously used to normalize the V005 CH4 profiles. We reapplied the "N2O correction" using more realistic N2O profiles from a CTM and these apparent discontinuities disappeared. See updated figure 6.
In section 3.3, we now calculate separate validation statistics for Δy_L at high and low latitudes.