This study presents the methodology to upscale the uncertainty in model outputs associated with data sharing and spatial resolution taking a methane emission model for paddy rice applied to China as the example. Model upscaling and the quantification of model uncertainty due to upscaling themselves are commendable aims and of interest for broader modeling communities. However, after reading twice, the following concerns came out:

1. The concept of data sharing is in question. Fig. 1 itself is clear and I could roughly understand what the authors proceeded. However, what is the difference among the following three: (1) input at a site is shared by that at a neighboring site; (2) the identical value is used for multiple sites because only coarser-resolution data are available; (3) and there is less spatial variability of a variable of input. In my understanding, (1) and (2) are the same. If so, the simplest approach is to create inputs at difference grid sizes from the finer input and run the model at each scale. Otherwise the impacts of decreased variability of input (coarser-resolution input in general have less variability than finer-resolution input) on model output have never quantified. I don't understand why the authors go such a tricky way that aggregates model outputs using correlation between two cells.

Re: We agree with the reviewer of what the data scarcity is and where the data-sharing rises because of data shortage [(1), (2) and (3)]. As mentioned by the reviewer, because only coarser-resolution data are available, the identical value is used for multiple neighboring sites (or grid cells) when making model estimations. Also analog to the "simplest approach" recommended by the reviewer, we ran the model at three scales (S1, S2 and S3 in Table 4), though not all of the five scales represented in Fig. 2. At each scale, S1 for instance, the finer input (data of *SAND*, 10km×10km raster dataset) was aggregated to create input of *SAND* at the scale of S1. But to run the model at a specific scale, the data of the other model variables, i.e., *OM*, *Wptn* and *VI*, must be shared between neighboring grid cells because they are coarser than the specific grid size of S1. Table 4 shows the scale effects of the model estimations, in other words, the impacts of decreased variability of input on the model output. At each of the specific scales (S1, S2 or S3), the direct model output is of the variation in each of the grid cells (in a county at S1, a province at S2 or a GR at S3). Up to the entire region from the individual grid cells, aggregation of the model outputs in all the grid cells of the region must be performed in a proper way. And this is what the manuscript do, to introduce an approach of uncertainty aggregation in model up-scaling.

2. Theoretically, a model should be calibrated at any scale before its application. As far as I read Haung et al. (1997) describing the modeling of CH4MOD, the structure of this model is very simple and is almost a multiplicative product of a few (five or so) variables. Why don't the authors directly calibrate empirical parameters of the model using coarser-resolution data? Also even for a complex process-based model (though the model is not for methane emission), the methodology to incorporate the information on subgrid-scale heterogeneity into a coarse-resolution model is feasible (e.g., Iizumi et al., 2013).

Re: We agree that a model, simple or complex, must be calibrated and validated before its application. The model (CH4MOD) in the present study, though not a complex model, has every essential element of a model. We had made the model calibration and validation in a previous study (Huang et al., 2004). In that study, we thoroughly calibrated and validated the

model with field measurements of rice paddy methane emissions and the associated environmental and agronomic conditions. Because the measurements of rice paddy methane emissions were only available at site scale, the model calibration and validation were, therefore, not performed at scales other than the site scale. This is why we call the CH4MOD a site scale model and why up-scaling is necessary when making regional estimations with the site scale model.

Spatial heterogeneity makes the regional estimation a hard work, and even harder when we lack sufficient data to describe the spatial variations in details. Because of data scarcity, we have little information of the spatial variation of methane emissions and also the factors that influence the methane emissions (Fig. 2). In the present study, we showed, with a case example, how to quantify the impacts of data scarcity on the regional estimation uncertainty via model up-scaling. It is not the purpose of the study to delineate the spatial heterogeneity at coarse resolution with finer "subgrid-scale" data via a reliable approach, such as that in Iizumi et al., 2013, owing to the fact that we usually lack the "subgrid-scale" data of the model input variables. Instead, we assumed probability distributions of the inner-cell variations of the model inputs with the poorly available information of them (Section 2.2.2 of the manuscript).

3. There is no mentioning of the importance of parameter values of a model as an important source of uncertainty of model output (e.g., Kennedy and O'Hagan, 2001; Makowski et al., 2002). This is especially true for empirical model like CH4MOD.
Re: Uncertainties of regional estimations come from many sources, including the model imperfection due to inaccuracy of parameters and structural fallacy of the model, as well as the data errors and poor availability of the model inputs. The inaccuracy of model parameters is no doubt an important source of uncertainties of the model outputs. Calibrating the model with field measurements helps to improve the model performance. The imprecision of CH4MOD was evaluated in a previous study (Huang et al., 2004). In the present study, we paid our attention on how to quantify the impacts of data scarcity on the uncertainty of regional estimations, instead of comprehensively analyzing the contribution of every possible source. In Section 3.1 of the manuscript, we briefly discussed, without many details, other possible sources of the uncertainty besides the data availability (please see Line 285-307 in the revised MS).

Specific comments:
4. P182L8. "five of the most sensitive model variables" is misleading. As long as I read Haung et al. (1997), this CH4MOD model has a few (five or so) variables as a whole.
Re: Apart from the "five of the most sensitive model variables", the CH4MOD has other inputs, including daily temperatures and rice calendar etc. that were mentioned by the reviews in the subsequent Specific Comment 5 and 18. The reviewer might not take the right paper of the model description. "Huang et al. (1997)" was not for describing the entire model structure. The bulk model description was first addressed in Huang et al. (1998) and further development and comprehensive validation of the model was provided in Huang et al. (2004).

5. P183L23-25. I suggest adding the number of harvesting of rice crop in a year and historical crop calendar here. These variables are important in estimating methane emission from rice

paddy and have been changing with time (e.g., Tao et al., 2003).

Re: In China, the rice-based crop rotations changed with time and in different climatic zones from the temperate to the tropical. The typical crop rotations in a year include rice-rice, rice-winter wheat and single rice etc. In CH4MOD, we consider the type and amount of the previous crop residues amended into the field (the OM variable). But because the present study is not about the impacts of long term changes of land use, crop rotation and crop calendar etc. (in fact, the issue of decadal changes of methane emission from rice paddies in China had been discussed in one of our previously published paper, Zhang et al, (2011)), we only made model estimation in a single year. We are aware that the imprecision in rice calendar is a source of the uncertainty, but in the sensitivity analysis of the model, we found it was not so sensitive as the "five of the most sensitive model variables".

6. P186L6. Does V_i (M_i, sigma_i) indicate a normal distribution?

Re: A random variable V_i usually has several parameters, e.g. mathematic expectation (mean, M_i) and the variance or its square root (standard deviation, sigma_i), whether the random variable is of normal distribution or other probability distribution. In this study, the V_i is severely skew and was fitted by Gamma distribution (Fig. B1).

7. P186L17-18. Could you elaborate how you can distinguish high correlation due to similar geographical condition across sites and that due to the sharing of an identical data value associated with the limited availability of finer-resolution data?

Re: The rice paddy methane emissions may show high spatial correlations due to similar geographical conditions. The correlation might be analyzed via approaches such as geostatistical methods, had we got spatially sufficient measurements of rice paddy methane emissions. Unfortunately, we don't have enough direct measurements of rice paddy methane emissions and this is why the modeling is necessary. To model methane emissions, we need to collect data of geographical condition and agronomic activities to drive the model. Because of the similarity of the geographical condition across neighboring sites, the estimated methane emissions inevitably showed similarity. But this kind of similarity, or correlation, is conceptually different from the correlation owing to data scarcity. The difference is analogous to that between *how two random phenomena correlate* and *how the sampling of the two random phenomena correlates*. The outputs of the model in each grid cell were parameterized into m_i and sigma_i. The variations of the m_i among all the grid cells delineate the spatial variations and similarity of methane emissions due to similar geographical conditions and agronomic activities. What we discussed in the present study is the uncertainty of the model estimation (represented by sigma_i) and the correlation of the estimation due to data sharing.

8. P186L20. "share common data" would be more precise to say "share common data and modeled processes".

Re: Revised.

9. P187L21. This "DS matrix" is the first time to appear. It is more appropriate to write "data-sharing (DS) matrix" here if my understanding is correct.

Re: Revised.

10. P188L19. "semi-empirical" Please be consistent throughout the manuscript (P195L19 says "empirical").
Re: Revised.

11. P189L24. It is more readable to say "grand region scales" than "regional scales".
Re: Revised.

12. P190L1. The "CI" is the first time to appear as far as I read. Please replace this by "confidence interval (CI)" if my understanding is correct.
Re: Revised.

13. P190L25-28. I have no specific criticism to use a Gunma distribution; however, I wonder if a use of bootstrap method is more useful to avoid a fitting error of a Gamma distribution because a Gamma distribution fails to capture the frequency of low methane emission values (Fig. B1) and this study focuses on the uncertainty of mode output (you have a lot of model output data, indicating that a fitting is not necessarily essential).
Re: In each grid cell, we made a lot of model outputs via the Monte Carlo approach and the bootstrap methods would be efficient in exploring the within-cell uncertainty, as pointed out by the reviewer. But when aggregating the within-cell uncertainties of all the grid cells to produce the uncertainty of the whole region, the bootstrap is questionable because there are "correlations" among the with-cell uncertainties owing to data sharing among them. After fitting a parametric Gamma distribution to the model outputs in each grid cell, we took the probability theorem of the summation of random variables as the start point (Equation 1). Thereafter, we discussed the problem of data scarcity and how it acts on the aggregation of uncertainties in model up-scaling.

14. P191L3. You can delete "1M=10^6".
Re: Deleted.

15. P193L3-5. I don't understand which result support this statement. Figure 2 doesn't work for this. This seems to be contradictory with the results presented in Table 4. The same comment can be applied to P193L14-15.
Re: We revised the sentence to clarify what it intends to. Fig. 2 visually shows how the model input differed in data availability. And with the model inputs showed in Fig. 2, Table 4 shows the changes of estimation uncertainty when different spatial scales (S1, S2 and S3) were adopted to make the regional estimation via modeling. Comparing the modeling results for S1, S2 and S3, we came to the inference that adopting finer resolution reduced the estimation uncertainty. The sentence is revised as "Even if the data abundance of the model variables differ significantly (Fig. 2), modeling at a finer spatial resolution does help to reduce the estimation uncertainty (Table 4)". The statement in P193L14-15 is an extending inference of the result in Table 4. We deleted it in the revision.

16. P194L26-P195L2. It would be nice if the author present this in a figure.

Re: The discussion between P194L26 and P195L2 is based on Table 4 that lists the value of i_ds and i_R at each of the three scales. We presented the value of i_ds and i_R in a table instead of a figure because we think it is better to list the value of i_ds/i_R along with the corresponding uncertainty ranges in Table 4, and also there were already 5 figures in the manuscript.

17. P194L22-23. However, both I_DS and I_R require finer-resolution data to calculate.
Re: According to their definition (Eq.2 and 3), I_DS and I_R can be calculated at any specific resolution (Table 4). With a given database (As shown in Fig. 2), however, finer-resolution yields higher I_DS and I_R values that indicate less estimation uncertainties (Table 4).

18. P197L1-2. I think the spatially interpolated temperature is an important source of uncertainty of model output. Why don't the authors include this into the analysis?
Re: Temperature is no doubt an important factor that influences methane emissions from rice paddies. But in regions where rice is cultivated, we have meteorological observations of daily air temperatures, and no spatial interpolation was made for S1, S2 and S3 scenarios. In other previous studies (Li et al., 2004; Matthews et al., 2000; Peng et al., 2008), the most important contributors to the uncertainties of methane emission were soil properties, OM amendment and field irrigations. Comparing to those factors, the error in temperature contributed much less to the uncertainty. In the paper, we therefore didn't include the contribution of the error and scarcity of air temperature in the analysis.

19. P197L5. It is well-documented that crop phenology including rice in China changed with time (e.g., Tao et al., 2006). Is this factor considered for the estimation of national level methane emission? The estimated value appeared in P182L9 is tied to the reliability of input data. It is recommended the authors mention this point clearly in Abstract.
Re: The crop phenology is one of the CH4MOD inputs that controls when the model starts to run and at what time it stops. But because we didn't focus on the long term change of the national rice paddy methane emission, the changes of crop phenology and climate were not considered in the present study. We agree that the estimated national methane emission relied on all the input of the model variables and we revised the abstract to mention the point explicitly.

20. P198L7-9. Please elaborate how you select these papers from which sources.
Re: Because no regular or comprehensive census data of the OM amendment is available, the investigation of how much OM amended into rice cultivation was made during the compilation of the national inventory of methane emission from rice cultivation of China. We delivered investigation papers to farmers in all the typical rice cultivation regions of China and summarized the returned data. The details of the data collection and the quality control can be found in the Supporting Information to a previously published paper (Zhang et al., 2011). In the revision, we add a reference pointing to the description of the data source.

21. Fig. 3. As far as I read, Fig. 3 is not cited in the main text.

Re: Fig.3 is cited in the manuscript in Page 194 Line 3. But it was showed as "Figure 3". We corrected it as "Fig. 3".

Reference:

Huang, Y., Sass, R.L., Fisher, F.M., 1998a. A semi−empirical model of methane emission from flooded rice paddy soils. Glob. Change Biol. 4, 247−268

Huang, Y., Zhang, W., Zheng, X., Li, J., Yu, Y., 2004. Modeling methane emission from rice paddies with various agricultural practices. J. Geophys. Res. 109: D08113, doi:101029/2003JD004401

Li, C., Mosier, A., Wassmann, R., Cai, Z., Zheng, X., Huang, Y., Tsuruta, H., Boonjawat, J., Lantin, R., 2004. Modeling greenhouse gas emissions from rice-based production systems: Sensitivity and up scaling. Global Biogeochem. Cy. 18, doi:101029/2003GB002045

Matthews, R.B., Wassmann, R., Knox, J.W., Buendia, L.V., 2000. Using a Crop/Soil Simulation Model and GIS Techniques to Assess Methane Emissions from Rice Fields in Asia. IV. Upscaling to National Levels. Nutr. Cycl. Agroecosys. 58, 210–217

Peng, S., Li, D., Xu, J., Ding, J., He, Y., Yu, J., 2007. Effect of Water-Saving Irrigation on the Law of CH_4 Emission from Paddy Field. Environ. Sci. 28, 9–13

Zhang, W., Yu, Y., Huang, Y., Li, T., Wang, P., 2011. Modeling methane emissions from irrigated rice cultivation in China from 1960 to 2050. Glob. Change Biol. 17, 3511–3523

<u>Response to reviewer2:</u>

General comments: This study presents the methodology to evaluate the impact of data scarcity on model upscaling. They developed a data sharing matrix to aggregate the modeled uncertainties in divisions of a subject region. The study is a valuable contribution to the literature on estimation of uncertainties due to data scarcity in model upscaling. There are a number of issues that the authors will need to address.

Specific comments:

1. As far as I can see, besides data sharing, there are other reasons causing correlation among different cells (e.g. similar geographical condition), which means that the uncertainties calculated in this paper may be underestimated in aggregation. It is better to mention the other sources that cause additional correlation and their effects on estimating the uncertainties in aggregation.

Re: We agree with the reviewer that the uncertainty of the methane emission from rice paddies of China was underestimated owing to the exclusion of the reasons other than the data scarcity and errors considered in the present study. Uncertainties of regional estimations come from many sources, including the model imperfection due to inaccuracy of parameters and structural fallacy of the model, as well as the data errors and poor availability of the model inputs. In Section 3.1 of the manuscript, we briefly discussed, without many details, other possible sources of the uncertainty besides the data availability (please see Line 285-307 in the revised MS).

2. P187 L4-L12. The most valuable contribution of this paper is that they developed a data sharing matrix. More details should be provided about how to calculate Cij.

Re: In the present study, we described two way of the Cij calculation. The first one is to calculate it by numeric experiments (P187L4-13, Table 1), and thereafter assigns value to the specific data sharing by looking up Table 1. The second way is quite simple and presented by the Equation 4 (P193). The Equation 4 can be used as the preliminary evaluation of the data sharing and the impacts on uncertainty aggregation.

3. P187 L21. What is DS matrix? Should "Eq.(2)" be changed to "Eq.(1)".

Re: DS matrix is "data sharing" matrix. Literal errors corrected.

4. Fig. 2. GR II contains "Fujian" according to the legend which disagrees with the figure. How did the authors divide China into five GRs? The criteria should be provided more clearly.

Re: Fujian should be in GR I and we had fixed the problem. The GRs was defined by the dominant crop rotations in rice paddies and the climatic conditions. In GR I, double rice is the primary crop rotation and in rice paddies of GR II, rice usually rotates with upland crops in a year round. We add brief description in the figure caption to explain the division.

Revision lists:

The revisions to this manuscript are highlighted in the file. They are listed bellowed:

1, Page 1, Line 4: We have changed the authors' sequence according to the contribution of each author.

2, Page 2, Line 24-25.

3, Page 7, Line 133.

4, Page 8, Line 159.

5, Page 11, Line 214.

6, Page 11, Line 219.

7, Page 14, Line 285-288.

8, Page 15, Line 310-321.

9, Page 18, Line 365-367.

10, Page 28, Line 583-589.