Geosci. Model Dev. Discuss., 7, 181–216, 2014
www.geosci-model-dev-discuss.net/7/181/2014/
doi:10.5194/gmdd-7-181-2014

Geoscientific Model Development
Discussions
Open Access

Estimation of uncertainties due to data scarcity in model upscaling: a case study of methane emissions from rice paddies in China

W. Zhang[1], T. Li[1], Y. Huang[2], Q. Zhang[1], J. Bian[1,3], and P. Han[1]

[1]LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China
[2]LVEC, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China
[3]Faculty of Resources and Environmental Science, Hubei University, Wuhan, 430062, China

Received: 22 October 2013 – Accepted: 30 November 2013 – Published: 10 January 2014

Correspondence to: T. Li (litingting@mail.iap.ac.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Data scarcity is a major cause of substantial uncertainties in regional estimations conducted with model upscaling. To evaluate the impact of data scarcity on model upscaling, we introduce an approach for aggregating uncertainties in model estimations. A data sharing matrix was developed to aggregate the modeled uncertainties in divisions of a subject region. In a case study, the uncertainty in methane emissions from rice paddies on mainland China was calculated with a local-scale model CH4MOD. The data scarcities in five of the most sensitive model variables were included in the analysis. The national total methane emissions were 6.44–7.32 Tg, depending on the spatial resolution used for modeling, with a 95 % confidence interval of 4.5–8.7 Tg. Based on the data sharing matrix, two numeral indices, I_R and I_{ds}, were also introduced to suggest the proper spatial resolution in model upscaling.

1 Introduction

Global change, including climate change due to the accumulation of greenhouse gases (GHG) in the atmosphere, is an important issue for scientists and policy-makers around the world. Because of the high spatial heterogeneities in global changes, models are widely used to delineate spatial variations and make regional estimations (Harvey, 2000). The models used in regional or global studies differ widely in terms of their spatial scales. Many of these models are site-specific, describing processes at local scales, including Century (Parton et al., 1993), RothC (Jenkinson and Rayner, 1977), DNDC (Li et al., 1992) and Agro-C (Huang et al., 2009; Yu et al., 2012). Extrapolating a site-specific model to a regional or global scale is usually referred to as "model upscaling" (King, 1991; van Bodegom et al., 2000). A common framework of this upscaling involves partitioning a large region into smaller, individual areas and running the model for each area (Matthews et al., 2000; Li et al., 2004; Yu et al., 2012).

Here, we use the term "cell" to refer to each small division of a large region. The cells may be of equal size (such as pixels of an image or grids of a raster dataset), but they can also be spatially irregular (e.g., counties or provinces of a nation). In model upscaling, the first problem modelers face is how to make the divisions. It is preferable to partition the region so that the model inputs in the cells are as statistically independent of each other as possible (King, 1991; Ogle et al., 2003, 2010). When data are scarce, however, the criterion of inter-cell independence may result in a partitions of large cells, leading to a reduced level of detail. An additional challenge is the great variability in the availability of data for the model inputs, which complicates the selection of an appropriate cell size. A properly partitioned subject region should balance the differences in spatial data abundance among model inputs. If the cell size is too large, substantial spatial variation in the model input variables will be lost after within-cell averaging (van Bodgegom et al., 2002; Verburg et al., 2006). Scientists tend to use the finest spatial resolution possible to express details in spatial variation in their modeling results. However, finer spatial resolution requires sufficient model input data; otherwise, data must be shared among cells for at least some, if not all, of the model inputs. This kind of inter-cell non-independence among the cells (resulting from data scarcity and requiring data sharing) complicates the uncertainty analysis (Ogle et al., 2003) when finer spatial resolutions are adopted.

To address as many details of spatial variation as possible, sufficient data are needed for model inputs. However, the data are usually far from abundant, and they differ greatly from one model input variable to the other. For example, to estimate national methane emissions from rice paddies, it is critical to obtain detailed information on organic matter amendments and field irrigation in rice cultivation (Khalil et al., 2008; Peng et al., 2007; van Bodegom et al., 2000; Wassmann et al., 1996). Such data, however, are seldom available on a regional scale (Zhang et al., 2011).

No matter which cell size is selected, the problem of estimating uncertainty in model estimates is always about estimating within-cell variance and, thereafter, the aggregation of the within-cell variances. The uncertainty analysis over a large region thus has two steps: (i) properly accounting for the uncertainty in each small division, i.e., within-cell variance, and (ii) aggregating the within-cell uncertainty correctly to produce the overall regional uncertainty (King, 1991; van Bodegom et al., 2000). The uncertainty in model estimates is a consequence of imprecision in model performance, errors in model inputs and validation data etc. (Klepper, 1997; van Bodegom et al., 2000). To analyze uncertainties due to errors in model inputs, representative errors and measurement errors, the Monte Carlo simulation has been recognized as an effective method (IPCC, 2000), and it has been applied in many studies (Ogle et al., 2003, 2010; Yu et al., 2012). By constructing probability density functions (PDFs) for model input variables and coefficients, or joint/Bayesian PDF if possible, variations of the model inputs over a certain area or at a specific site can be statistically depicted. Based on the PDFs derived from measurement data and/or a priori knowledge, the Monte Carlo method involves randomly and repeatedly drawing values from the PDFs to drive the model running and produce varying model estimates.

After the Monte Carlo simulation is performed for within-cell uncertainty analysis in each division of a large subject region, we face the problem of uncertainty upscaling. In the case of "independent" partitioning of the entire subject region, the uncertainty upscaling can be quite simple, as explained by the statistical "Law of Large Numbers". An independent, random variable is assigned to depict variations in the model estimation in each division (IPCC, 2000; Ogle et al., 2010). As previously noted, however, a paucity of data for some of the model variables and a small cell size division of the entire subject area may result in data sharing among divisions, which is problematic for the model variables that lack sufficient data to support fine-resolution partitioning. Upscaling the uncertainties in the model outputs must deal with this kind of "dependency" appropriately.

The objective of the present study is to find a way of evaluating impacts of data scarcity on regional estimation uncertainties, apart from the other causes of uncertainties. We will develop a data sharing matrix as the kernel of the uncertainty aggregation in model upscaling and discuss how different spatial resolutions affect the regional

GMDD
7, 181–216, 2014

estimation uncertainties, given the same data availability for different spatial division schema. As a case study, we performed the uncertainty analysis of the national rice paddy methane emissions inventory of mainland China with CH4MOD, a model developed to simulate methane emissions from rice paddies (Huang et al., 1998a, 2004).

2 Methods

2.1 Uncertainty assessment in model upscaling

Figure 1 shows a flowchart of model upscaling in the case study. The solid arrows in Fig. 1 represent procedures of estimating national methane emissions, and the hollow arrows describe the uncertainty assessments accompanying the model upscaling. Although many studies have demonstrated how to upscale a model to make regional estimations from various baseline scenarios (Matthews et al., 2000; Li et al., 2004; Ogle et al., 2010), the main focus of the present study is on aggregating the uncertainties in model estimations due to data scarcity.

2.1.1 Within-cell variation in model estimation

When partitioning the large region under consideration into spatially adjacent divisions, the within-cell variation must be accounted for at first (King, 1991; van Bodegom et al., 2000; Ogle et al., 2003, 2010). The baseline model estimation is usually performed by running the model once in a cell. Each model input variable will have one datum or one time series of data, e.g., daily weather observations. If there are multiple data available for a model input variable in a cell, they are averaged before modeling. The within-cell heterogeneity of the model estimation will therefore be lost after averaging and will cause errors in the model's estimation. This type of error is referred to as the "fallacy of average" (Verburg et al., 2006). In contrast, the within-cell PDF of the variation in the model variable can also be established by statistical analysis of the data and/or expert estimation (Ogle et al., 2010; IPCC, 2000). The Monte Carlo simulation is considered an effective approach to evaluate within-cell variation or uncertainty in model estimates due to errors in model input variables and their interactions, and it is thus used in the present study (Fig. 1).

2.1.2 Spatial uncertainty aggregation in case of data scarcity

In each cell, the model estimation via Monte Carlo iteration produces a numeric depiction of a random variable $V_i(m_i, \sigma_i)$, where m_i and σ_i are the statistical mean and standard deviation, respectively, of the random variable V_i. Thereafter, the model upscaling involves the summation of the random variables $V_0 = V_1 + V_2 + \ldots + V_N$. The aggregation of uncertainty, represented by the statistical variance or standard deviation, is generalized as $\mathrm{Var}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}(X_i, X_j)$ (Ross, 2006), and it can also be transformed into quadratic summation of the elementary variances via the standardized variance-covariance matrix:

$$\sigma_0^2 = \sum_{i,j} \sigma_i \times C_{ij} \times \sigma_j, \quad (i = 1, \ldots, N, \quad j = 1, \ldots, N) \tag{1}$$

where σ_0^2 is the aggregated variance of the regional estimation and σ_i and σ_j are the standard deviations of the within-cell variations in cells i and j, respectively. The matrix $\mathbf{C}$ is comprised of coefficients C_{ij}, which stand for "correlations" between individual cells. Here the "correlation" is a measure of how the model outputs in two cells varying concurrently because they may share common data of the model inputs. If the estimation in cell i is over-/under-estimated, the estimation in cell j will probably be over-/under-estimated also because they share common data, and vice versa. The aggregation of the model outputs can be quite simple if the model estimate is made with independent data in each cell. In this case, the matrix $\mathbf{C}$ will be an identity matrix in which the diagonal elements will be 1 and all of the off-diagonal elements will be 0. The aggregation in Eq. (1) will thereafter indicate the arithmetic sum of the within-cell

GMDD
7, 181–216, 2014

variances, as addressed by the *Law of Large Numbers*. However, when there are not sufficient data to support independent calculation among cells, the off-diagonal elements, C_{ij}, of the matrix **C** will no longer be zero.

In the present study, C_{ij} was empirically calculated by the numerical experiments. For different level of data sharing between two cells (Table 1), the model estimations in the two cells were iteratively calculated with CH4MOD. The model inputs were randomly selected from the ranges of the variables (Table B1), respectively. When there was data sharing between the two cells for a variable in Table 1, the value of the variable was selected once for the two cells. And for the variable with no data sharing, the value of the variable was selected separately for the two cells. The correlation coefficients (C_{ij}) of the model estimations in the two cells was statistically calculated with a large amount, 1000 iterations in the present study, of the paired model estimations in the two cells.

2.1.3 Indicators of data scarcity in model estimation

A common problem in making a model estimation for a large region is that the available data for the model input variables differ greatly. To evaluate the overall data scarcity of the model input variables, two indicators are defined:

$$I_{\mathrm{ds}} = \begin{cases} \frac{1}{n}\sum\limits_{i \neq j} C_{ij}, & n > 0 \\ 0, & n = 0 \end{cases} \tag{2}$$

$$I_{\mathrm{R}} = \frac{N}{\sqrt[m]{\prod\limits_{k=1}^{m} N_k}} \tag{3}$$

where C_{ij} is the element of the DS matrix defined in Eq. (2) and n is the total number of off-diagonal, non-zero elements of the DS matrix. In Eq. (3), N is the total number of cells (divisions) that partition the entire region under consideration and N_k is the number of data points for the model variable k. When the off-diagonal elements of the sharing matrix are all 0, indicating abundant data (no sharing) among the cells for all of the model input variables, $I_{ds} = 0$ and $I_R = 1$. The other extreme, when the off-diagonal elements of the DS matrix are all 1, indicates a severe data scarcity and complete data sharing among the cells for every model input variable, $I_{ds} = 1$ and $I_R = N$.

Data scarcity refers to the data abundance relative to the spatial resolution, i.e., spatial details we intend to depict via the model simulation. With all the model input data on hand, we may expect more data scarcity, and larger I_{ds}, when we choose a smaller cell size and vice versa. An I_{ds} of 0 indicates a "perfect" data abundance for the chosen spatial resolution. But this "perfection" may, conversely, imply that we have chosen too large of a cell size and that some spatially varying details in the model inputs were lost, a severe "fallacy of average." The regional partitioning should, in this case, take a finer spatial resolution to show more heterogeneous details in the model estimation.

2.2 Uncertainty assessment of estimated methane emissions from rice paddies in China

2.2.1 CH4MOD and input variables

In this case study, we used the model CH4MOD to estimate methane emissions from rice paddies in China. CH4MOD is a semi-empirical model that simulates methane production and emissions from rice paddies under various environmental conditions and agricultural practices (Huang et al., 1998a, 2004; Xie et al., 2010).

The CH4MOD model runs with a daily step and is driven by air temperature. The main input variables include soil sand percentage (SAND), organic matter amendments (OM), rice grain yield (GY), water management pattern (W_{ptn}) and rice cultivar index (VI). Appendix A describes CH4MOD and the compilation of model inputs. More detailed information regarding the model development, validation and application has been provided elsewhere by the authors (Huang et al., 2004, 2006; Zhang et al., 2011).

GMDD
7, 181–216, 2014

2.2.2 PDFs of the model input variables

Many studies (Khalil and Butenhoff, 2008; Li et al., 2004; Matthews et al., 2000; Van Bodegom et al., 2002) have suggested that a significant proportion of the uncertainty in regional rice paddy methane emissions arises from data scarcity, especially with regard to the soil sand content (SAND), organic matter amendments (OM), rice grain yield (GY), water management (W_{ptn}) and rice cultivar index (VI). The CH4MOD sensitivity analysis similarly shows the importance of these five factors in methane emissions (Table B1 in Appendix B). Figure 2 illustrates the data abundance of the five model variables. The data for soil sand content is a 10 km by 10 km raster dataset made from soil profiles via spatial interpolation (Oberthür et al., 1999; Shi et al., 2004, 2006). Although a certain proportion of the immense spatial variation in soil properties may be lost after spatial interpolation (Goovaerts, 2001; van Bodegom et al., 2002), the gridded soil data are still the most detailed data of the five model inputs. In descending order of data abundance, the other four factors are GY, OM, W_{ptn} and VI. Assuming a normal distribution, the PDFs of four factors (all except W_{ptn}) were parameterized by statistical analysis of their data.

With a specific spatial resolution, e.g., using administrative counties as divisions, the PDF of SAND in a division was calculated with the grid data within the division. Because every county has only one datum for GY, no PDF was assumed for GY when counties were adopted as divisions. Although the yield of rice grain is not the same at every place throughout a county, we have no more detailed data of the grain yield that would allow us to make PDFs of the GY variable.

The data on other two variables, OM and W_{ptn}, were collected and statistically analyzed to produce PDFs (Tables 2 and 3) on provincial and regional scales (Fig. 2b). The rice paddy methane emissions vary notably with rice varieties (Singh et al., 1997). The variety index (VI), which accounts for the methane emissions differences between rice varieties (Huang et al., 1998a, 2004), ranged from 0.5 to 1.5, and it typically has the value around 1.0 for most rice varieties (Huang et al., 1997, 2004). We assumed the 95 % CI in VI was 0.5 to 1.5 and exhibited a normal distribution. In the case of partitioning the entire nation into counties, the counties included within a province and/or grand region must share data and PDFs for the variables OM, W_{ptn} and VI.

The PDFs in the case study of rice paddy methane emissions did not encompass all sources of uncertainties for the five variables. Careful planning in building PDFs of the model variables will improve the reliability of the uncertainty assessment. At present, we are focused on uncertainty aggregation in model upscaling when facing data scarcity.

2.2.3 Uncertainty calculation and aggregation

To evaluate how the adoption of cell sizes influences the uncertainty of regional estimations, we used three partitioning schema, S1, S2 and S3, to estimate the methane emissions in China with the same previously described datasets. The counties, provinces and grand regions were used as the spatial divisions of China in the three scenarios, respectively. In S2 and S3, the PDFs of rice grain yield were calculated with statistical analysis of census data. The Monte Carlo iteration was performed 500 times in each cell to calculate the within-cell uncertainty.

For each of the three scenarios, the elements of the DS matrix were valued by referencing to the correlation coefficients (C_{ij}) in Table 1 based on the state of data sharing illustrated in Fig. 2b. With the within-cell variations of methane emissions calculated via the Monte Carlo approach, the aggregation of the model estimates were then fulfilled via Eq. (1) for early, late and middle rice. When combing the estimation results for the three rice ecosystems, Eq. (1) was again utilized for the OM and VI data sharing by the three rice ecosystems.

After aggregation, the confidence interval, e.g., 95 % CI of the national methane emission, was derived via the parameterized PDF of the aggregated estimate. Assuming a Gamma distribution (Fig. B1 in Appendix B), the two parameters of the PDF, shape (α) and scale (β) were calculated by the momentum method, where β = variance/average and α = average/β (Ross, 2006).

GMDD
7, 181–216, 2014

3 Results and discussion

3.1 Methane emissions from rice paddies in China and their uncertainties

In 2010, the total rice harvest area of China was 29.9 Mha (1 M = 10^6). The national total methane emissions were 6.44–7.32 Tg, depending on the spatial resolution used for modeling (Table 4). In each individual county, the within-cell standard deviation of methane flux, seasonal methane emissions per unit area, as calculated via Monte Carlo methods, was 13.5–89.3 % of the statistical mean. Because no errors were considered in the area from which rice was harvested, the relative uncertainty for methane emissions was the same as in the methane flux estimation. In the case of errors in the rice harvest area being included, the uncertainty of methane emissions in each cell can be calculated with Rule B of IPCC (2000) before aggregation.

When data sharing between counties was not accounted for, the falsely aggregated standard deviation was approximately 1.7–2.2 % of the national emissions according to the *Law of Large Numbers*. However, when the correlation of the model estimation in cells was considered (Table 1), the overall aggregated standard deviation was 16.3 % of the total emissions, ranging from 18.3–28.0 % for early, late and middle rice ecosystems (Table 4). This finding implies that intensifying data quantities significantly reduces uncertainties in regional estimations by reducing the data sharing and the correlations in the DS matrix. Assuming a Gamma distribution (Fig. B1 in Appendix B), the 95 % confidence interval (CI) of the national total methane emissions, made by the moment matching approach with m_0 and σ_0, was 4.5–8.7 Tg at the S1 spatial resolution (Table 4).

The national methane emissions from rice paddies in China have been estimated in many previous studies. Table 5 lists those studies that included uncertainty assessments. With the exception of results from Huang et al. (1998), in which higher emissions were produced because of the continuous flooding used for rice cultivation in the study, the uncertainties in all other studies largely overlapped with those of the present study, although significance levels for the uncertainties were not explicitly provided. The

GMDD
7, 181–216, 2014

results in other studies (not list in Table 5), e.g., Ren et al. (2010), Li et al. (2002) and Yao et al. (1996), also fell within the ranges shown in Table 4. Most of these previous studies focused on organic matter application and water regimes in their estimations of uncertainty (Table 5) because of data scarcity in these two factors. Taking into consideration the tremendous spatial heterogeneity of soil characteristics, Li et al. (2004) believed that these were the most sensitive factors accounting for uncertainties, and the uncertainty was between 2.3–10.5 $\mathrm{Tg\,yr^{-1}}$ (1.7–7.9 $\mathrm{Tg\,yr^{-1}}$ C) for mid-season drainage irrigation and 8.5–16.0 $\mathrm{Tg\,yr^{-1}}$ (6.4–12.0 $\mathrm{Tg\,yr^{-1}}$ C) when continuous flooding was applied.

Uncertainty in regional rice paddy methane emission models comes from multiple sources. A comprehensive uncertainty analysis should synthetically include all major uncertainty sources (IPCC, 2000; van Bodegom et al., 2002). In the present study, the within-cell variances of the five most sensitive factors, i.e., SAND, GR, OM, W_{ptn} and VI, were parameterized and included in the Monte Carlo simulations, but there are also other factors that may contribute to uncertainties (van Bodegom et al., 2002). And moreover, there may be covariance between the input parameters. For example, the rice variety (VI) and/or soil texture (SAND) may have impacts on the irrigation applied (W_{ptn}). With sufficient data, we may quantify the correlations between the input parameters, and then build a joint/Bayesian PDF of the input parameters. Incorporation of correlations of the input parameters will improve the estimation of the within-cell variances. But facing the difficulty of data scarcity, we have to parameterize the within-cell variance of each input parameter separately at present. Apart from the data scarcity, model imperfections due to a poor understanding of the complexity of the ecosystem is also a primary source of estimation bias. A model encompasses functions and equations that describe the physical processes of interest, but it cannot involve every detail. The model inaccuracy may bias the estimation away from the true value, which is usually evaluated by model validation (Huang et al., 2004). In the present study, however, we did not incorporate the error of model inaccuracy into the uncertainty assessment.

GMDD
7, 181–216, 2014

3.2 Data scarcity, spatial resolution and the uncertainties in regional estimation

The uncertainty of regional methane emissions in Table 4 is mainly caused by errors and a scarcity of model input data (Fig. 2). When the abundance of data for model variable differs significantly (Fig. 2), modeling at a finer spatial resolution does help to reduce the estimation uncertainty. In Table 4, the 95 % CI was 3.4–12.3 Tg when modeling was performed at coarse resolution (S3). On the provincial scale (Scenario S2), however, the 95 % CI narrowed to 4.8–10.4 Tg, and the aggregated standard deviation was 19.5 % of the national total emissions. However, without sufficient data support (Fig. 2), upscaling a model at an over-fine resolution makes no substantial difference, as that in Table 4 for S1. Although the uncertainty was reduced further when the spatial resolution was at the county level, this approach is not cost-effective, and the indicator I_R rises rapidly from up to 3 at the provincial scale to more than 27 at the county scale (Table 4). The I_R indicates the redundant cost, higher I_R means more redundant processing. Beyond the processing cost, the spatial resolution of 10 km × 10 km (Fig. 2b) might reduce the estimation uncertainty a bit more.

In Table 1, sharing data of the higher sensitivity variable, e.g., SAND vs. Yield in Table B1, may result in a larger correlation coefficient C_{ij}. While C_{ij} in Table 1 is computation intensive which needs a large amount of modeling iterations, a rough estimation (Eq. 4) of C_{ij} may be meaningful in finding the proper spatial resolution before the model up scaling is carried out.

$$C_{ij} = \frac{\sum_{k=1}^{m} I_{ij,k} \times s_k}{\sum_{k=1}^{m} s_k} \tag{4}$$

where s_k is the sensitivity index of the model parameter k (e.g., Table B1 in the Appendix) and m is the number of model input variables under consideration. $I_{ij,k}$ is a binary variable taking the value of 1 or 0. If cells i and j share data for the model input

GMDD
7, 181–216, 2014

variable k, $I_{ij,k}$ is assigned the value of 1; otherwise, it is 0. The sensitivity index s_k is to reflect the difference in importance of the model input variables to the model output. Figure 3 shows the comparison of the correlation coefficients calculated by the two ways. Though the rough estimation of the C_{ij} by Eq. (4) differs to some extent from those in Table 1, they show the same trend in reflecting impacts of data sharing on correlations of the model outputs in two cells.

3.3 Model upscaling and spatial aggregation in a general view

To produce regional estimations in model upscaling, the aggregation procedure can be divided into two parts: the pre-positive aggregation and the post-positive aggregation (Fig. 4). An example of the pre-positive aggregation in the present study is the averaging of the soil data included in each division. When no pre-positive aggregation is applied (Fig. 4a, I_{ds} is close to 1 and I_R is far larger than 1), nearly every detail in the model input parameters is kept before the model is run. The pathway in Fig. 4a is obviously computation-heavy, but it may not be effective in reducing the estimation uncertainty, e.g., S1 vs. S2 in Table 4. At the other extreme (Fig. 4c, $I_{ds} = 0$), all aggregations are performed before the modeling, and no spatial variation of the model output is depicted. A finer spatial resolution is therefore necessary to explore some spatial variation details in the model outputs. When facing remarkably diverse data abundance in model input variables, as in the case of rice paddy methane emissions (Fig. 2), determining where to place the modeling on the pathway in Fig. 4 is a balance between the model variables with respect to data scarcity and the corresponding sensitivities.

The indicators I_{ds} and I_R may be referred to in finding the right position at which the pre-positive aggregation stops and the modeling is actually carried out (Fig. 4b). As previously mentioned, a value of 0 or 1 for I_{ds}, all pre-positive or post-positive aggregation, is not a good choice. I_{ds} should be a value between 0 and 1 to indicate a compromise between data scarcity of model variables. From S3 to S2, both the I_{ds} and I_R increase and the estimation uncertainty is reduced significantly (Table 4). However, from S2 to S1, the estimation uncertainty and I_{ds} change little, whereas I_R increases markedly.

GMDD
7, 181–216, 2014

This pattern implies that when I_R increases much faster than I_{ds}, the pre-positive aggregation should stop, the modeling should take place, and the other aggregation should be performed in the post-positive stage.

4 Conclusions

Data scarcity is a big challenge in making regional estimates of greenhouse gas emissions. We developed a data sharing matrix to estimate aggregated uncertainties in China's rice paddy methane emission introduced by data scarcity. Based on the data sharing matrix, we estimated that data scarcity in five most sensitive factors introduced an aggregated uncertainty to the estimates, ranging from 4.5 to 8.7 Tg with 95 % confidence interval. Aggregated uncertainty may vary with the spatial resolution for a given dataset, while the indicator of I_{ds} is useful for identifying a proper spatial resolution. A proper spatial resolution corresponds to a value between 0 and 1 for the I_{ds}, which represents the compromise of the data scarcity between model variables. Improving data abundance of model inputs is expected to reduce the uncertainties in estimating terrestrial greenhouse gas emission, in which the sensitivity of the model inputs also plays a key role.

Appendix A

Description of CH4MOD and the compilation of model inputs

CH4MOD is an empirical model that simulates methane production and emissions from rice paddies under various environmental conditions and agricultural practices (Huang et al., 1998a, 2004; Xie et al., 2010). This model calculates methanogenic substrates' production from rice plant root exudates and added organic matter (OM) decomposition. Both the OM decomposition and rice-plant-induced substrate production are significantly influenced by environmental factors, including soil texture and temperature.

GMDD
7, 181–216, 2014

The soil moisture controls the transformation fraction of the substrates into methane. There are two major routes by which methane produced in rice paddy soils is emitted into the atmosphere. One way is the aerenchyma system of the rice plants, and the other is methane bubbles. Both pathways of methane emissions are formulated in the model.

CH4MOD runs on a daily step, and it is driven by daily air temperature. Its input parameters include soil sand percentage (SAND), organic matter amendment (OM), rice grain yield (GY), water management pattern (W_{ptn}) and rice cultivar index (VI).

A1 Rice harvest area and grain production

Data on rice production and the harvest area of each province in 2010 were extracted for early, late and middle rice from the nation's statistical yearbook (EBCAY, 2011). The county-level rice production census was obtained from the Chinese Academy of Agricultural Sciences. Although the county-level data do not record fractions of early, late or single rice cultivation, the rotation type in each county was represented using the approach of Frolking et al. (2002) by referring to the climatic zonification of the cropping system in China (Han et al., 1987).

Many studies have shown that the methane emissions differ notably among rice varieties (Singh et al., 1997; Wang et al., 1997). In CH4MOD, the impact of rice variety on methane emissions was parameterized as the variety index (VI) (Huang et al., 1998a, 2004). The VI ranges from 0.5 to 1.5 and typically has a value of approximately 1.0 for most rice varieties (Huang et al., 1997, 2004).

A2 Climate data and rice phenology

Daily mean air temperature is the only meteorological data required to drive the CH4MOD model. Observations of the air temperature at 678 Chinese meteorological stations in 2010 were acquired from the National Meteorological Information Center (NMIC), China Meteorological Administration (CMA) (http://cdc.cma.gov.cn/). For counties without a meteorological station, the air temperatures at the nearest neighboring station was used.

The rice phenology, including transplanting and harvesting dates, controls the start and end of CH4MOD's run in simulating methane emissions. Data regarding the rice phenology were originally iso-line maps, edited by Zhang et al. (1987), in the Atlas of Agricultural Climate in China. The transplanting and harvesting dates for each grid were spatially interpolated from the iso-lines via the TIN (Triangular Irregular Network) technique (Aumann et al., 1991) and assigned to each county.

A3 Soil properties

The spatial database of soil sand content (SAND) is one of the databases developed by the Institute of Soil Sciences, Chinese Academy of Sciences, from the samples of soil profiles obtained during the Program of the Second Soil Survey of China and subsequent surveys. The database is comprised of 10 km × 10 km raster datasets of soil properties at depths of every 10 cm from the surface down. The spatial resolution of the soil is the finest among the CH4MOD input parameters (Fig. 2).

A4 Organic matter amendment in rice paddies

The organic matter amended into rice fields include various types of farm manure (green manure and animal manure, etc.) and crop straw as well as dead roots and stubble from the previous crops. Roots remaining in the soil can be calculated with the root/shoot ratio (Huang et al., 2007). Stubble was assumed to be one-tenth of the aboveground straw biomass. The fraction of straw incorporation and farm manure application, however, is not well known, and limited data are available. In the First National Census of Pollution Sources conducted by the Ministry of Environmental Protection of China (EPFNCPS, 2011), straw application in croplands was summarized on a provincial level with the census data (Table 2). The straw application in Table 2 is not rice-specific but, rather, incorporates all of the crops in each province. The bias may not be significant in provinces where rice dominates the crop cultivation. In addition to crop straw, the incorporated crop residues include dead crop roots and stubble. According to Zhao and Li (2001), stubble accounts for approximately 13 % of the total straw in dry weight.

Until now, no regular statistical data or comprehensive census data have been available about the application of manure in rice cultivation. In this study, we investigated organic matter application in rice cultivation. More than 1000 such investigation papers were collected and validated. The amount of farmyard manure application in each province (Table 2) was part of the investigation results.

Appendix B

Sensitivity analysis of CH4MOD

Data of an environmental factor is usually expressed as $M \pm e$, where M represents the measurement and e represents the error. When used as model inputs, the data imprecision can result in uncertainties in the model outputs with diverse magnitudes depending not only on the data imprecision but also on the model sensitivity. Model sensitivity addresses the variability of the model output to variations in model inputs. Usually, an individual variable sensitivity analysis is carried out by the "varying one variable at a time" approach. In contrast to the individual variable sensitivity analysis, a regional sensitivity analysis is performed in the present study, and simultaneous variations of the model inputs account for interactions of the variables in the model. The Monte Carlo method is commonly applied to simultaneously produce variations of model inputs.

To scale the model input variation, the e/M is adopted for each of the variables to make them comparable to each other, and all the CH4MOD input parameters have positive values. In differential form, the expression e/M can be expressed generally as $\frac{dx}{x}$ or $d(\ln x)$. The purpose of the model sensitivity analysis in the present study is to

GMDD
7, 181–216, 2014

explore the modeled methane flux variability to variations of the model input parameters as in formula (B1):

$$\frac{\mathrm{d}y}{y} \propto s_k \times \frac{\mathrm{d}x_k}{x_k} \quad \text{or } \mathrm{d}(\ln y) \propto s_k \times \mathrm{d}(\ln x_k) \tag{B1}$$

where k is used to identify each model parameter and y represents the seasonal methane emissions flux ($\mathrm{g\,CH_4\,m^{-2}}$) calculated by CH4MOD with x_k as input. S_k is the sensitivity index of the model variable k, and it is defined as the linear coefficient between methane flux and the model input variables in term of fractal variation.

The Monte Carlo approach was adopted as the first step to pick values of the model input parameters randomly from their value domains (Table B1), upon which the methane flux was calculated with CH4MOD. This picking-and-calculating procedure iterates for 20 000 cycles. After logarithmic transformation of the model inputs and outputs, simple variable linear regression was performed, and the sensitivity index was defined as the slope coefficient of the regression equation.

Water management in rice cultivation is a key factor that impacts methane emissions from rice paddies. In CH4MOD, the diverse water management strategies in Chinese rice cultivation are grouped into five irrigation patterns and include flooding, drainage and intermittent irrigation (Huang et al., 2004). In the case of this nominal variable, the sensitivity index was calculated as follows:

$$s_w = \frac{1}{N} \times \sum_{k \neq l} \frac{\left|\overline{y_l} - \overline{y_k}\right|}{y_0}, \quad k,l \in W \tag{B2}$$

where $w = (1,2,3,4,5)$ in the formula (B2) is the code set of the irrigation water patterns (Table B1). N is the total number of (j,k) pairs, and $\overline{y_l}$, $\overline{y_k}$ and y_0 represent the mean methane flux for irrigation water pattern l, k and all water patterns, respectively.

To run the CH4MOD simulation, daily air temperatures must be available for the duration of rice growth from dates of transplanting to the harvest. In model sensitivity

GMDD
7, 181–216, 2014

analysis, the temperature data are virtually created by the following equations from Eqs. (B3) to (B7).

$$T_{\mathrm{air}}^{(t)} = \overline{T}_{\max} - |t - S_{\max}| \times D_{\mathrm{T}} + R(-0.5, 0.5) \tag{B3}$$

$$D_{\mathrm{T}} = \begin{cases} \left(\overline{T}_{\max} - \overline{T}_{\min}\right) / (S_{\max} - S_{\mathrm{s}}), & T \leq S_{\max} \\ \left(\overline{T}_{\max} - \overline{T}_{\min}\right) / (S_{\mathrm{e}} - S_{\max}), & T > S_{\max} \end{cases} \tag{B4}$$

$$S_{\max} = R(S_{\mathrm{s}}, S_{\mathrm{e}}) \tag{B5}$$

$$\overline{T}_{\max} = R(25.0, 35.0) \tag{B6}$$

$$\overline{T}_{\min} = R(10.0, 20.0) \tag{B7}$$

where the function $R(v_1, v_2)$ returns a random number between v_1 and v_2. S_{s} and S_{e} represent the transplanting and harvesting dates, respectively, and $S_{\max}$ is the day on which the air temperature reaches its maximum of the rice season. Time variable t ($S_{\mathrm{s}} \leq t \leq S_{\mathrm{e}}$) represents days after transplanting.

The results showed that methane emissions are most sensitive to field irrigation, with a sensitivity index of 0.67 (Table B1). The soil texture, rice variety and organic matter application rank lower, with sensitivity indices of 0.63, 0.51 and 0.47, respectively.

Acknowledgements. The study was jointly supported by the National Natural Science Foundation of China (Grant No. 41175132, 41075107, 41021004) and the National Key Basic Research Development Foundation of China (Grant No. 2010CB950603). We thank the Resources and Environmental Scientific Data Center of the Chinese Academy of Sciences and the National Meteorological Information Center of the Chinese Meteorological Administration for their support in providing the data. We also thank Heikki Järvinen and Timo Vesala of University of Helsinki for their suggestions and help to this manuscript. We are very grateful to the reviewers for their help in improving the manuscript.

GMDD
7, 181–216, 2014

References

Aumann, G., Ebner, H., and Tang, L.: Automatic derivation of skeleton lines from digitized contours, J. Photogr. Remote Sens., 46, 259–268, 1991.

Cai, Z. C.: A category for estimate of CH_4 emission from rice paddy fields in China, Nutr. Cycl. Agroecosys. 49, 171–179, 1997.

Commission of The First National Pollution Source Census Data Compilation of China (CFPC): Datasets of China Pollution Source Census, China Environmental Science Press, Beijing, China, 2011.

Editorial Board of China Agriculture Yearbook (EBCAY): China Agriculture Yearbook, China Agriculture Press, Beijing, China, 2011.

Frolking, S., Qiu, J., Boles, S., Xiao, X., Liu, J., Zhuang, Y., Li, C., and Qin, X.: Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China, Global Biogeochem. Cy., 16, 1091, doi:10.1029/2001GB001425, 2002.

Gao, L. and Li, L.: Rice Meteorology and Ecology, China Agriculture Press, Beijing, 1992 (in Chinese).

Goovaerts, P.: Geostatistical modelling of uncertainty in soil sciences, Geoderma, 103, 3–26, 2001.

Han, X., Liu, X., Gao, L., and Li, L.: Climatic zonification of cropping system in China, in: Multiple Cropping Systems in China, edited by: Han, X. and Liu, X., China Agriculture University Press, Beijing, 28–46, 1987 (in Chinese).

Harvey, L. D. D.: Up scaling in global change research, Climatic Change, 44, 225–263, 2000.

Huang, Y., Sass, R. L., and Fisher, F. M.: Methane emission from Texas rice paddy soils1 Quantitative multi-year dependence of CH_4 emission on soil, cultivar and grain yield, Glob. Change Biol., 3, 479–489, 1997.

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

Huang, Y., Sass, R. L., and Fisher, F. M.: Model estimates of methane emission from irrigated rice cultivation of China, Glob. Change Biol., 4, 809–821, 1998b.

Huang, Y., Zhang, W., Zheng, X., Li, J., and Yu, Y.: Modeling methane emission from rice paddies with various agricultural practices, J. Geophys. Res., 109, D08113, doi:10.1029/2003JD004401, 2004.

Huang, Y., Zhang, W., Zheng, X., Han, S., and Yu, Y.: Estimates of methane emissions from Chinese rice paddies by linking a model to GIS database, Acta Ecol. Sinica, 26, 980–987, 2006.

Huang, Y., Zhang, W., Sun, W., and Zheng, X.: Net primary production of Chinese croplands from 1955 to 1999, Ecol. Appl., 17, 692–701, 2007.

Huang, Y., Yu, Y., Zhang, W., Sun, W., Liu, S., Jiang, J., Wu, J., Yu, W., Wang, Y., and Yang, Z.: Agro-C: a biogeophysical model for simulating the carbon budget of agroecosystems, Agr. Forest Meteorol., 149, 106–129, 2009.

IPCC (Intergovernmental Panel on Climate Change): IPCC Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, edited by: Penman, J., Kruger, D., Galbally, I. E., Hiraishi, T., Nyenzi, B., Emmanuel, S., Buendia, L., Hoppaus, R., Martinsen, T., Meijer, J., Miwa, K., and Tanabe, K., IPCC/OECD/IEA/IGES, Hayama, Japan, 2000.

Jenkinson, D. S. and Rayner, J. H.: The turnover of soil organic matter in some of the Rothamsted classical experiments, Soil Sci., 123, 298–305, 1977.

Khalil, M. A. K. and Butenhoff, C. L.: Spatial variability of methane emission from rice fields and implications for experimental design, J. Geophys. Res., 113, G00A09, doi:10.1029/2007JG000517, 2008.

King, A. W., Johnson, A. R., and O'Neill, R. V.: Transmutation and functional representation of heterogeneous landscapes, Landscape Ecol., 5, 239–253, 1991.

Klepper, O.: Multivariate aspects of model uncertainty analysis: tools for sensitivity analysis and calibration, Ecol. Model., 101, 1–13, 1997.

Li, C., Frolking, S., and Frolking, T.: A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity, J. Geophys. Res., 97, 9759–9776, 1992.

Li, C., Mosier, A., Wassmann, R., Cai, Z., Zheng, X., Huang, Y., Tsuruta, H., Boonjawat, J., and Lantin, R.: Modeling greenhouse gas emissions from rice-based production systems: sensitivity and up scaling, Global Biogeochem. Cy., 18, GB1043, doi:10.1029/2003GB002045, 2004.

Li, Y.: Research and practice of water-saving irrigation for rice in China, in: Proceedings of an International Workshop Held in Wuhan, China: Water-Saving Irrigation for Rice, edited by: Barker, R., Loeve, R., Li, Y., and Tuong, T., 23–25 March 2001, Colombo, Sri Lanka, International Water Management Institute, 1–10, 2001.

Liang, G.: Rice Ecology, Agricultural Press, Beijing, China, 250 pp., 1983 (in Chinese).

Ma, X., Zhu, B., Du, D., and Zheng, X.: CH_4, CO_2 and N_2O emissions from the year-round flooded paddy field at fallow season, J. Agro-Environ. Sci., 24, 1199–1202, 2005 (in Chinese with English abstract).

Mao, L.: Irrigation in Rice Paddies, China Agricultural Press, Beijing, 1981 (in Chinese).

Matthews, R. B., Wassmann, R., Knox, J. W., and Buendia, L. V.: 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, 2000.

Ministry of Water Resources and Utilization of China (MWRUC), Division of Rural Water Resources and Utilization: National Program for Conservation of Irrigation Water in Chinese Agriculture, Chinese Agricultural Press, Beijing, China, 1996 (in Chinese).

Oberthür, T., Goovaerts, P., and Dobermann, A.: mapping soil texture classes using field texturing, particle size distribution and local knowledge by both conventional and geostatistical methods, Eur. J. Soil Sci., 50, 457–479, 1999.

Ogle, S., Breidt, F., Eve, M., and Paustian, K.: Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agricultural lands between 1982 and 1997, Glob. Change Biol., 9, 1521–1542, 2003.

Ogle, S., Breidt, F., Easter, M., Williams, S., Killian, K., and Paustian, K.: Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model, Glob. Change Biol., 16, 810–822, 2010.

Parton, W., Scurlock, J., Ojima, D., Gilmanov, T., Scholes, R., Schimel, D., Kirchner, T., Menaut, J. C., Seastedt, T., Moya, E., Kamnalrut, A., and Kinyamario, J.: Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide, Global Biogeochem. Cy., 7, 785–809, 1993.

Peng, S., Li, D., Xu, J., Ding, J., He, Y., and Yu, J.: Effect of water-saving irrigation on the law of CH_4 emission from paddy field, Environ. Sci., 28, 9–13, 2007.

Ren, W., Tian, H., Xu, X., Liu, M., Lu, C., Chen, G., Melillo, J., Reilly, J., and Liu, J.: Spatial and temporal patterns of CO_2 and CH_4 fluxes in China's croplands in response to multifactor environmental changes, Tellus B, 63, 222–240, doi:10.1111/j.1600-0889.2010.00522.x, 2010.

Ross, S. M.: A First Course in Probability, 7th Edn., Pearson Education Inc., Prentice Hall, 355–358, 2006.

GMDD
7, 181–216, 2014

Shi, X., Yu, D., Warner, E. D., Pan, X., Peterson, G. W., Gong, Z., and Weindorf, D. C.: Soil database of 1 : 1 000 000 digital soil survey and reference system of the Chinese genetic soil classification system, Soil Survey Horizons, 45, 129–136, 2004.

Shi, X., Yu, D., Yang, G., Wang, H., Sun, W., Guo, H., and Gong, Z.: Cross-reference benchmarks for translating the genetic soil classification of China into the Chinese soil taxonomy, Pedosphere, 16, 147–153, 2006.

Singh, S., Kumar, S., and Jain, M. C.: Methane emission from two India soil planted with different rice cultivars, Biol. Fert. Soils, 25, 285–289, 1997.

Tian, H., Xu, X., Lu, C., Liu, M., Ren, W., Chen, G., Melillo, J., and Liu, J.: Net exchanges of CO_2, CH_4, and N_2O between China's terrestrial ecosystems and the atmosphere and their contributions to global climate warming, J. Geophys. Res., 116, G02011, doi:10.1029/2010JG001393, 2011.

van Bodegom, P. M., Leffelaar, P. A., Stams, A. J. M., and Wassmann, R.: Modeling methane emissions from rice fields: variability, uncertainty, and sensitivity analysis of processes involved, Nutr. Cycl. Agroecosys., 58, 231–248, 2000.

van Bodegom, P. M., Verburg, P. H., Stein, A., Adiningsih, S., and Denier van der Gon, H. A. C.: Effects of interpolation and data resolution on methane emission estimation from rice paddies, Environ. Ecol. Stat., 9, 5–26, 2002.

Verburg, P. H., van Bodegom, P. M., Denier van der Gon, H. A. C., Bergsma, A., and van Breemen, N.: Upscaling regional emissions of greenhouse gases from rice cultivation: methods and sources of uncertainty, Plant Ecol., 182, 89–106, 2006.

Wang, B., Neue, H. U., and Sanmote, H. P.: Effect of cultivar difference, "IR72", "IR69958" and "Dular") on methane emission, Agr. Ecosys. Environ., 62, 31–40, 1997.

Wassmann, R., Neue, H. U., Alberto, M. C. R., Lantin, R. S., Bueno, C., Llenaresas, D., Arah, J. R. M., Papen, H., Rennenberg, H., and Seiler, W.: Flux and pools of methane in wetland rice soils with varying organic inputs, Environ. Monit. Assess., 42, 163–173, 1996.

Xie, B., Zhou, Z., Zheng, X., Zhang, W., and Zhu, J.: Modeling methane emissions from paddy rice fields under elevated atmospheric carbon dioxide conditions, Adv. Atmos. Sci., 27, 100–114, 2010.

Xiong, Z., Cai, H., Min, S., and Li, B.: Rice in China, China Agricultural Science and Technology Press, Beijing, 197–606, 1992 (in Chinese).

Yan, X., Cai, Z., Ohara, T., and Akimoto, H.: Methane emission from rice fields in mainland China: amount and seasonal and spatial distribution, J. Geophys. Res., 108, 4505, doi:10.1029/2002JD003182, 2003.

Yao, H., Huang, Y., and Chen, Z.: Estimation of methane from rice paddies in mainland China, Global Biogeochem. Cy., 10, 641–649, 1996.

Yu, Y., Huang, Y., and Zhang, W.: Modelling soil organic carbon change in croplands of China, 1980–2009, Global Planet. Change, 82–83, 115–128, 2012.

Zhang, F., Wang, D., and Qiu, B.: Map of Agro-Climate in China, Science Press, Beijing, China, 1987 (in Chinese).

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

Zhao, Q. and Li, Q.: Presents and prospects of crop straw application in southern China, in: Mechanism and Techniques in Straw Application, edited by: Liu, X., Gao, W., and Zhu, W. China Agricultural Science and Technology Press, Beijing, China, 138–146, 2001 (in Chinese).

Table 1. Lookup table of correlation coefficients of the model outputs in two cells due to data sharing.

Data sharing between cell i and j						Data sharing between cell i and j					
Yield	OM	Sand	W_{Ptn}	VI	C_{ij}	Yield	OM	Sand	W_{Ptn}	VI	C_{ij}
0*	0	0	0	1	0.069	1	0	0	0	1	0.136
0	0	0	1	0	0.347	1	0	0	1	0	0.430
0	0	0	1	1	0.413	1	0	0	1	1	0.520
0	0	1	0	0	0.295	1	0	1	0	0	0.343
0	0	1	0	1	0.375	1	0	1	0	1	0.478
0	0	1	1	0	0.674	1	0	1	1	0	0.776
0	0	1	1	1	0.796	1	0	1	1	1	0.900
0	1	0	0	0	0.082	1	1	0	0	0	0.170
0	1	0	0	1	0.167	1	1	0	0	1	0.225
0	1	0	1	0	0.436	1	1	0	1	0	0.481
0	1	0	1	1	0.519	1	1	0	1	1	0.616
0	1	1	0	0	0.396	1	1	1	0	0	0.458
0	1	1	0	1	0.499	1	1	1	0	1	0.575
0	1	1	1	0	0.760	1	1	1	1	0	0.849
0	1	1	1	1	0.878	1	1	1	1	1	1.000
1	0	0	0	0	0.066						

* 1 means the two cells share data for the variable and 0 means they do not share data for the variable.

GMDD
7, 181–216, 2014

Table 2. Fraction of straw incorporation and farm manure application in rice cultivation[a].

Province	Fraction of straw incorporation[b]	Farm manure ($kg\,OM\,ha^{-1}$)[c]		Province	Fraction of straw incorporation	Farm manure ($kg\,OM\,ha^{-1}$)	
		Mean	Range			Mean	Range
Beijing	0.41	821.6	321.6–1321.6	Henan	0.56	1539.2	776.2–2302.1
Tianjin	0.29	927.4	123.1–1731.6	Hubei	0.20	2101.3	981.1–3221.6
Hebei	0.62	1519.3	959.5–2079.1	Hunan	0.34	1836.9	846.7–2827.2
Shanxi	0.44	1824.8	1195.5–2454.2	Guangdong	0.23	1243.2	634.5–1851.8
Inner Mon.	0.12	1837.5	1042.4–2632.7	Guangxi	0.27	1384.7	645.4–2124.1
Liaoning	0.03	1108.5	657.8–1559.3	Hainan	0.22	1408.5	964.8–1852.1
Jilin	0.03	1308.4	421.5–2195.4	Chongqing	0.17	1608.7	801.5–2415.8
Heilongjiang	0.23	1800.8	836.0–2765.6	Sichuan	0.18	1922.7	940.7–2904.7
Jiangsu	0.23	1263.5	605.6–1921.4	Guizhou	0.09	1793.2	740.2–2546.1
Zhejiang	0.35	1276.2	734.1–1818.3	Yunnan	0.10	1802.3	853.1–2751.5
Anhui	0.19	1507.5	424.3–2590.7	Shaanxi	0.34	1769.6	555.3–2983.9
Fujian	0.32	1123.1	852.6–1393.6	Gansu	0.03	1923.0	375.9–3470.1
Jiangxi	0.38	1612.2	842.3–2382.1	Ningxia	0.15	1448.6	515.5–2381.7
Shandong	0.55	1032.8	530.8–1534.7	Xinjiang	0.45	1612.0	407.7–2816.3

[a] No data of farm manure application is available for Shanghai and Tibet. The data of Jiangsu and Guizhou was adopted for them, respectively.
[b] Statistics of the first national pollution source census conducted by the Ministry of Environmental Protection of China (CFPC, 2011); but no variation range provided in the publication.
[c] Statistics of the investigation data of the organic application in crop cultivation made by the Institute of Atmospheric Physics, Chinese Academy of Sciences. Green manure was not included in the present study because it accounts for a minor proportion in total organic matter application in rice cultivation.

Table 3. Proportions of different water irrigation patterns[a] in each grand region.

Grand Region	Baseline fraction	Uncertainty fraction
I	3: 0.92; 4: 0.08[b]	1: 0.31; 2: 0.31; 3: 0.30; 4: 0.08
II	2: 0.95; 4: 0.05	1: 0.32; 2: 0.32; 3: 0.31; 4: 0.05
III	2: 0.82; 4: 0.18	1: 0.27; 2: 0.28; 3: 0.27; 4: 0.18
IV	1: 1.0	1: 0.34; 2: 0.33; 3: 0.33
V	1: 1.0	1: 0.34; 2: 0.33; 3: 0.33

[a] Refer to Huang et al. (2004) for the definition of water irrigation patterns.
[b] Means the water irrigation pattern 3 was applied in 92 % of the rice cultivation area in the Grand Region I (Fig. 2a), and the rest 8 % of rice area was continuously flooded (water irrigation pattern 4).

GMDD
7, 181–216, 2014

Table 4. Estimated methane emissions from rice paddies of China and their uncertainties.

Scenario	Spatial Resolution	CH_4 Emission (Tg)	SD of the Estimation (Tg)	95 % CI[a] (Tg)	I_{ds}	I_R
Middle rice						
S1	County	4.03	0.74 (18.3 %)	2.7–5.6	0.147	30.3
S2	Province	4.37	0.94 (21.4 %)	2.7–6.4	0.153	2.7
S3	GR[b]	4.13	1.53 (37.1 %)	1.7–7.6	0.069	1.4
Early rice						
S1	County	1.02	0.28 (28.0 %)	0.5–1.6	0.157	27.6
S2	Province	1.40	0.44 (31.4 %)	0.7–2.4	0.117	1.9
S3	GR	1.34	0.60 (44.6 %)	0.4–2.7	0.069	1.2
Late rice						
S1	County	1.39	0.30 (21.6 %)	0.9–2.0	0.157	28.4
S2	Province	1.56	0.45 (28.7 %)	0.8–2.5	0.118	2.0
S3	GR	1.73	0.79 (45.3 %)	0.6–3.6	0.069	1.2
All rice						
S1	County	6.44	1.05 (16.3 %)	4.5–8.7	–	–
S2	Province	7.32	1.43 (19.5 %)	4.8–10.4	–	–
S3	GR	7.20	2.29 (31.8 %)	3.4–12.3	–	–

[a] 95 % CI (Confidence Interval) of the estimation was calculated from Gamma distribution. The shape and scale parameters of the Gamma distribution were estimated by the emission estimation and the corresponding SD.
[b] GR: Grand Regions in Fig. 2a.

GMDD
7, 181–216, 2014

Table 5. Uncertainties in methane emission from rice paddies of China via various methods.

Method	Uncertainty Range	Reference
IPCC Tier 2/Statistical analysis on measured methane fluxes[a]	8.1 ± 3.7 (1993)[b]	Cai et al. (1997)
Simplified CH4MOD model/Model input scenarios	7.2–13.6 (1993)[c]	Huang et al. (1998b)
MERES model/Organic matter scenarios	3.4–8.6 (1993)	Matthews et al. (2000)
IPCC Tier 2/Organic matter amendment and irrigation scenarios	5.8–9.6 (1995)	Yan et al. (2003)
DNDC model/Most Sensitive Factors	2.3–10.5 (1990)	Li et al. (2004)
CH4MOD model/Monte Carlo	4.2–9.1 (2010)	This study

[a] Method the uncertainty assessment was made;
[b] the number in parentheses indicates the year when the estimation of methane emission was made;
[c] assuming continuous flooding in rice cultivation.

Table B1. Sensitivity indexes of CH4MOD parameters.

Parameters	Value range Min.	Max.	Sensitivity Index
Grain yield ($kg\,ha^{-1}$)	1000	9000	0.35
Soil sand content (%)	6	90	0.63
OM amendment ($kg\,ha^{-1}$)*	200	6500	0.47
Rice cultivar index	0.5	1.5	0.51
Water regime	1, 2, 3, 4, 5		0.67

* The fraction of OMN and OMS in the amended organic matters varies harmoniously between 0.45 and 0.55 to reflect differences in OM types.

GMDD
7, 181–216, 2014

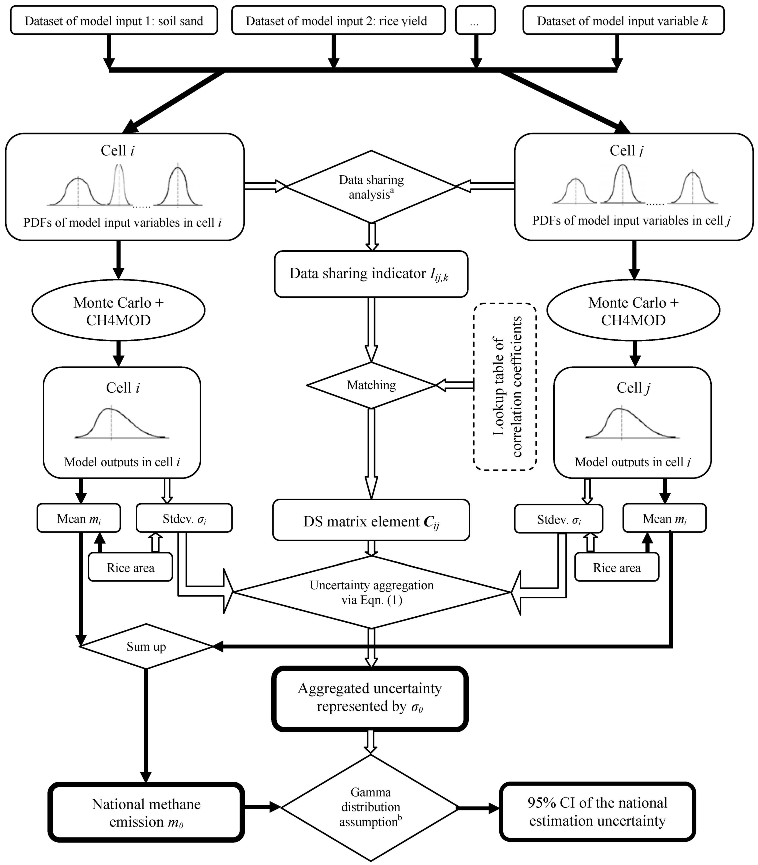

Fig. 1. Flowchart of upscaling CH4MOD to estimate methane emission from rice paddies of China and the uncertainty aggregation. Notes: (a) if cell i and j sharing data of the model input variable k, then $I_{ij,k} = 1$, otherwise $I_{ij,k} = 0$; (b) the assumption of Gamma distribution of the national methane emission was based on the results in model sensitivity analysis in Appendix B.

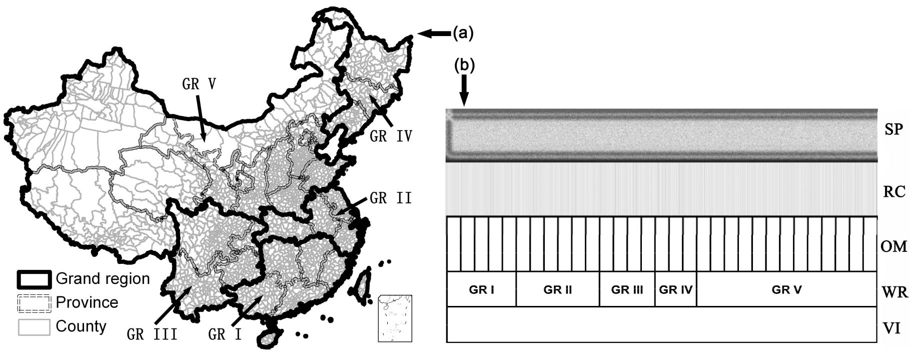

Fig. 2. Administration boundaries of China on different scales and data abundance of the CH4MOD input variables on different spatial resolutions. A grand region (GR) is a cluster of provinces that are similar in rice cultivation: GR I (Guangdong, Guangxi, Hainan, Hunan and Jiangxi); GR II (Shanghai, Jiangsu, Zhejiang, Anhui, Fujian and Hubei); GR III (Chongqing, Sichuan, Guizhou and Yunnan); GR IV (Heilongjiang, Liaoning and Jilin) and GR V (Other provinces).

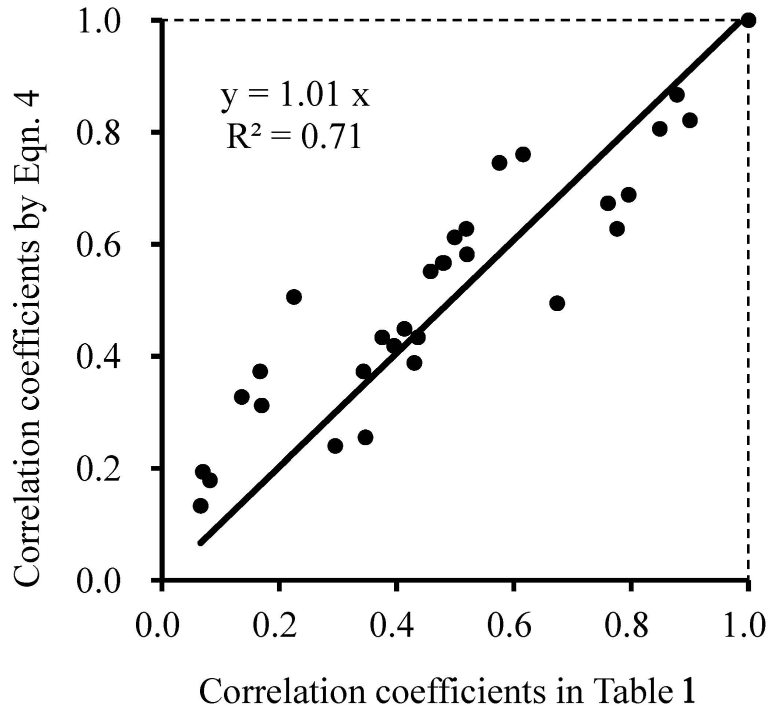

Fig. 3. Comparison of the correlation coefficients C_{ij} calculated by two methods.

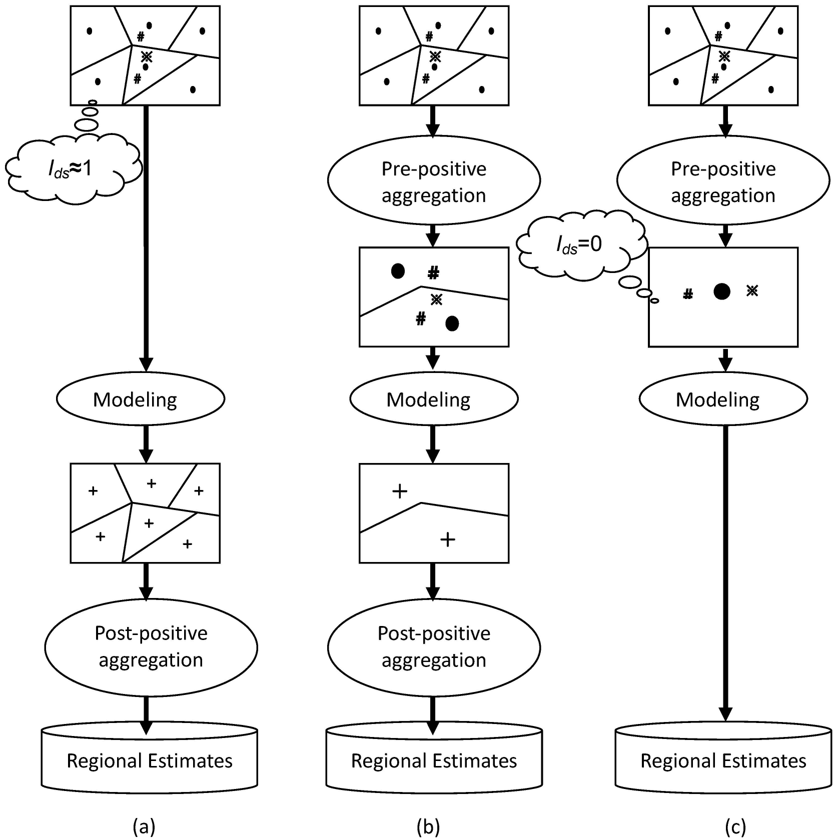

Fig. 4. Aggregation procedures in making regional estimation via modeling. Aggregation can be carried out before/after the modeling or on both sides depending on data scarcity of the model input variables. (• low sensitive variable; # medium sensitive variable; ※ high sensitive variable).

GMDD
7, 181–216, 2014

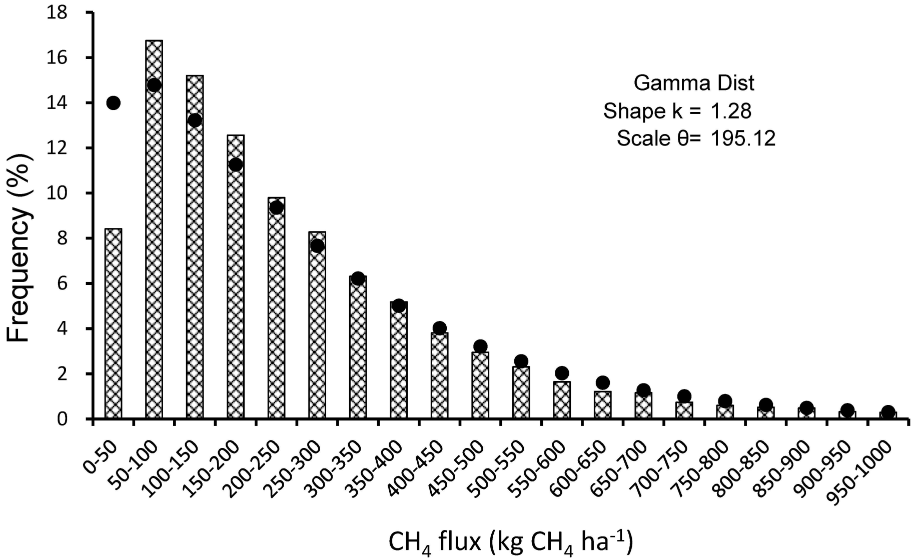

Fig. B1. Frequency distribution of the modeled methane fluxes in sensitivity analysis. The filled bars are the CH4MOD outputs, and the filled circles are outputs of Gamma distribution. The shape and scale parameters of the Gamma distribution were calculated with the statistical average and standard deviation of the CH4MOD outputs: $\beta = (\text{std})^2/(\text{avg})$ and $\alpha = (\text{avg})/\beta$.

GMDD
7, 181–216, 2014

Estimation of uncertainties due to data scarcity

W. Zhang et al.