Uncertainties in estimating regional methane emissions from rice paddies due to data scarcity in modeling approach

Wen Zhang[1], Qing Zhang[1], Yao Huang[2], Tingting Li[1], Junyan Bian[1,3], Pengfei 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

Abstract

Rice paddy is a major anthropogenic source of the atmospheric methane. But because of the high spatial heterogeneity, making accurate estimation of the methane emission from rice paddies is still a big challenge, even with complicated models. Data scarcity is a substantial cause of the uncertainties in estimating the methane emissions on regional scales. In the present study, we discussed how data scarcity affected the uncertainties in model estimations of rice paddy methane emissions, from site scale up to regional/national scale. The uncertainties in methane emissions from rice paddies of China was calculated with a local-scale model and the Monte Carlo simulation. The data scarcities in five of the most sensitive model variables, field irrigation, organic matter application, soil properties, rice variety and production were included in the analysis. The result showed that in each individual county, the within-cell standard deviation of methane flux, as calculated via Monte Carlo methods, was

13.5%–89.3% of the statistical mean. After spatial aggregation, the national total methane emissions were estimated 6.44–7.32 Tg, depending on the base scale of the modeling and the reliability of the input data. And with the given data availability, 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. The 95% confidence interval of the estimation was 4.5–8.7 Tg by assuming a Gamma distribution. Improving the data availability of the model input variables is expected to reduce the uncertainties significantly, especially of those factors with high model sensitivities.

Keywords: model, upscaling, uncertainty aggregation, data scarcity, methane emissions, rice paddy

1 Introduction

Methane is not only an important greenhouse gas in the atmosphere, but also an active reactor in many atmospheric chemistry processes. Rice cultivation has been recognized the major anthropogenic activity that accounted for the rapid increase of the atmospheric methane concentration. But because of the high spatial heterogeneity in methane emissions from rice paddies, huge uncertainty has long been the big problem in making reliable estimations, even after complicated models were developed and applied (Li, et al., 2002; Zhang et al., 2011; 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 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 for 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).

In model upscaling, the first problem modelers face is how to make the spatial divisions (each division was call a cell, hereafter). 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 the partition of large cells, leading to a reduced level of spatial details. 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, a finer spatial resolution requires sufficient model input data; otherwise, data must be shared among cells for at least some, if not all, the model inputs. This type 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 estimate regional/national methane emissions from rice paddies, it is critical to obtain detailed information on organic matter amendments, soil properties, rice varieties 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 at a regional scale (Zhang et al., 2011).

To analyze the uncertainty due to errors in model inputs in each cell, 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). Based on the probability distribution functions (PDFs) derived from measurements and/or a priori knowledge of the model inputs, the Monte Carlo method involves randomly and repeatedly drawing values from the PDFs to drive the model and produce varying model estimates. After the Monte Carlo simulation is performed for a within-cell uncertainty analysis in each division, we face the problem of uncertainty upscaling. In the case of "independent" partitioning of the entire subject region, an independent random variable is assigned to depict variations in the model estimate for each division (IPCC, 2000; Ogle et al., 2010), the uncertainty upscaling can be quite simple, as explained by the statistical "Law of Large Numbers". As previously noted, however, a paucity of data for some of the model variables and a small cell size 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 appropriately with this type of "dependency".

The objective of the present study is to evaluate the impacts of data scarcity on the uncertainty in regional estimations of rice paddy methane emissions, and discuss how different spatial resolutions affect the regional estimation uncertainties, given the same data availability for different spatial division schema.

2 Methods

2.1 Uncertainty assessment in model upscaling

Fig. 1 presents a flowchart of model upscaling in the case study. The solid arrows in Fig. 1 represent procedures for 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 primary focus of the present study is the aggregation of the uncertainties in model estimations due to data scarcity.

2.1.1 Within-cell variation in model estimates

When partitioning the large region under consideration into spatially adjacent divisions, the within-cell variation must be accounted for first (King 1991; van Bodegom et al., 2000; Ogle et al., 2003, 2010). The baseline model estimate is usually established 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 estimate will therefore be lost after averaging, which 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). 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 the 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 $V_0=V_1+V_2+...+V_N$. The aggregation of uncertainty, represented by the statistical variance or standard deviation, is generalized as $Var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N}\sum_{j=1}^{N} Cov(X_i, X_j)$ (Ross, 2006), and it can also be transformed into quadratic summation of the elementary variances via the standardized variance-covariance matrix:

$$\sigma_0^2 = \sum_{i,j} \sigma_i \times C_{ij} \times \sigma_j \text{, (i=1...N, j=1...N)} \qquad (1)$$

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 **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 vary coincidently because they share common data and modeled processes for the model inputs. If the estimation in cell *i* is over-/under-estimated, the estimation in cell *j* will most likely be over-/under-estimated as well 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 **C** will be an identity matrix in which the diagonal elements will be 1 and all the off-diagonal elements will be 0. The aggregation in equation (1) will thereafter indicate the arithmetic sum of the within-cell 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 via numerical experiments. For different levels of data sharing between two cells (Table 1), the model estimations for the two cells were iteratively calculated with CH4MOD. The model inputs were randomly selected from the ranges of the variables (Table B1). When there was data sharing between the two cells for a variable in Table 1, the value of the variable was selected once for both cells. And for variables 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 number, 1000 iterations in the present study, of paired model estimations for 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:

$$I_{ds} = \begin{cases} \dfrac{1}{n}\sum_{i \neq j} C_{ij}, & n > 0 \\ 0, & n = 0 \end{cases} \tag{2}$$

$$I_R = \frac{N}{\sqrt[m]{\prod_{k=1}^{m} N_k}} \tag{3}$$

where C_{ij} is the element of the DS (data sharing) matrix defined in equation (1) and n is the total number of off-diagonal, non-zero elements of the DS matrix. In equation (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 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 abundance of data 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 a 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. However, 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, adopt 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 the 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 the 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).

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 indicates the importance of these five factors in methane emissions (Table B1 in Appendix B). Fig. 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 constructed 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 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 location throughout a county, we have no more detailed data on grain yield that would allow us to make PDFs of the *GY* variable.

The data on the other two variables, *OM* and W_{ptn}, were collected and statistically analyzed to produce PDFs (Table 2 and Table 3) at provincial and grand region scales (Fig. 2b). Rice paddy methane emissions vary notably with rice variety (Singh et al., 1997). The variety index (*VI*), which accounts for the methane emission differences between rice varieties (Huang et al., 1998a, 2004), ranges from 0.5 to 1.5, and it typically has a value close to 1.0 for most rice varieties (Huang et al., 1997, 2004). We assumed that the 95% confidence interval (CI) for *VI* was 0.5 to 1.5 and that it 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 of China were used as the spatial divisions in the three scenarios, respectively. In S2 and S3, PDFs of the rice grain yield were calculated based on a 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 the correlation coefficients (C_{ij}) in Table 1 based on the state of data sharing illustrated in Fig. 2b. With the within-cell variations in methane emissions calculated via the Monte Carlo approach, the aggregation of the model estimates was then performed via equation (1) for early, late and middle rice. When combining the estimation results for the three rice ecosystems, equation (1) was again utilized for the OM and VI data shared 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).

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 M ha. 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 being present in the rice harvest area, 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 estimations for 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 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, calculated via 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 the 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 results of other studies (not listed in Table 5), e.g., Ren et al. (2010), Li et al. (2002) and Yao et al. (1996), also fell within the ranges listed 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 Tg yr^{-1} (1.7−7.9 Tg yr^{-1} C) for mid-season drainage irrigation and 8.5−16.0 Tg yr^{-1} (6.4−12.0 Tg yr^{-1} C) when continuous flooding was applied.

Uncertainties of regional estimations come from many sources, including the model imperfection due to inaccuracy of parameters and structural fallacy of the model (e.g., Kennedy and O'Hagan, 2001), as well as the data errors and poor availability of the model inputs. 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). 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 (Kennedy and O'Hagan, 2001). Incorporation of correlations between the input parameters will improve the estimation of the within-cell variances. However, facing the difficulty of data scarcity, it is necessary to parameterize the within-cell variance of each input parameter separately at present. Apart from data scarcity, model imperfections due to a poor understanding of the complexity of the ecosystem are also a primary source of estimation bias. A model comprises functions and equations that describe the physical processes of interest, but it cannot include every detail. Model inaccuracies 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 in the uncertainty assessment.

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

The uncertainty in regional methane emissions in Table 4 is primarily caused by errors and a scarcity of model input data (Fig. 2). 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). We made the model estimations at three scales (S1, S2 and S3 in Table 4). 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, 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). In Table 4, the 95% CI was 3.4—12.3 Tg when modeling was performed at a coarser resolution (S3). At 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 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; a higher I_R indicates more redundant processing.

In Table 1, sharing data for the higher-sensitivity variable, e.g., *SAND* vs. *Yield* in Table B1, may result in a larger correlation coefficient C_{ij}. Although C_{ij} in Table 1 is computation intensive, needing a large number of modeling iterations, a rough estimation (Eqn. 4) of C_{ij} may be meaningful in finding the proper spatial resolution before the model upscaling is conducted:

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

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 a value of 1 or 0. If cells *i* and *j* share data for the model input variable *k*, $I_{ij,k}$ is assigned a value of 1; otherwise, it is 0. The sensitivity index s_k

reflects the difference in the importance of the model input variables to the model output. Fig. 3 presents the comparison of the correlation coefficients calculated in two ways. Though the rough estimation of C_{ij} via Eqn. 4 differs to some extent from those in Table 1, the values exhibit the same trend in reflecting the impacts of data sharing on correlations of the model outputs between cells.

4 Conclusions

Data scarcity is a significant challenge in making regional estimates of greenhouse gas emissions. We developed a data sharing matrix to estimate the 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 the five most sensitive factors introduced an aggregated uncertainty to the estimates ranging from 4.5 to 8.7 Tg with a 95% confidence interval. Aggregated uncertainty may vary with the spatial resolution for a given dataset, and the indicator I_{ds} is useful for identifying an appropriate spatial resolution. An appropriate spatial resolution corresponds to a value between 0 and 1 for the I_{ds}, which represents a compromise between the data scarcity of different model variables. Improving the 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.

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.

Reference:

Aumann, G. Ebner, H. Tang, L. 1991. Automatic derivation of skeleton lines from digitized contours. J. Photogr. Remote Sens. 46(5) 259–268

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

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

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

Frolking, S., Qiu, J., Boles, S., Xiao, X., Liu, J., Zhuang, Y., Li, C., Qin, X., 2002. Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China. Global Biogeochem. Cy. 16(4), 1091, doi: 101029/2001GB001425

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

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil sciences.

Geoderma, 103(1－2), 3－26

Han, X., Liu, X., Gao, L., Li, L., 1987. Climatic zonification of cropping system in China (in Chinese). In: Han, X. and X. Liu (Eds) Multiple cropping systems in China. China Agriculture University Press, Beijing, pp 28－46

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

Huang, Y., Sass, R.L., Fisher, F.M., 1997. Methane emission from Texas rice paddy soils1 Quantitative multi-year dependence of CH4 emission on soil, cultivar and grain yield. Glob. Change Biol. 3, 479–489

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

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

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

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

Huang, Y., Y., Yu, Zhang, W., Sun, W., Liu, S., Jiang, J., Wu, J., Yu, W., Wang, Y., Yang, Z., 2009. Agro-C: A biogeophysical model for simulating the carbon budget of agroecosystems. Agri. Forest Meteorol. 149(1), 106–129

IPCC (Intergovernmental Panel on Climate Change), 2000. IPCC Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. Penman, J., Kruger, D., Galbally, I., etc. IPCC/OECD/IEA/IGES, Hayama, Japan

Jenkinson, D.S., Rayner, J.H., 1977. The turnover of soil organic matter in some of the Rothamsted classical experiments. Soil Sci. 123(5), 298–305

Kennedy M.C., O'Hagan A., 2011. Bayesian calibration of computer models. J. R. STAT. SOC. B. 63(3), 425–464.

Khalil, M.A.K., Butenhoff, C.L., 2008. Spatial variability of methane emission from rice fields and implications for experimental design. J. Geophys. Res. 113, G00A09, doi:101029/2007JG000517

King, A.W., Johnson, A.R., O’Neill, R.V., 1991. Transmutation and functional representation of heterogeneous landscapes. Landscape Ecol. 5(4), 239–253

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

Li, C., Frolking, S., Frolking, T., 1992. A model of nitrous oxide evolution from soil driven by rainfall events: 1 Model structure and sensitivity. J. Geophys. Res. 97 (D9), 9759–9776

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

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

Liang, G., 1983. Rice ecology(In Chinese). Agricultural Press, Beijing, China, pp. 250

Ma, X., Zhu, B., Du, D., Zheng, X., 2005. CH_4, CO_2 and N_20 emissions from the year-round flooded paddy field at fallow season(In Chinese with English abstract). J. Agro-Environ. Sci. 24, 1199–1202

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

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

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

Oberthür, T., Goovaerts, P., Dobermann, A., 1999. 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

Ogle, S., Breidt, F., Eve, M., Paustian, K., 2003. 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(11), 1521–1542

Ogle, S., Breidt, F., Easter, M., Williams, S., Killian, K., Paustian, K., 2010. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. Glob. Change Biol. 16(2), 810–822

Parton, W., Scurlock, J., Ojima, D., Gilmanov, T., Scholes, R., Schimel, D., Kirchner, T., Menaut, J.C., Seastedt, T., Moya, E., Kamnalrut, A., Kinyamario, J., 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. Global Biogeochem. Cy. 7(4) 785–809

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

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

Ross, S.M., 2006. A First Course in Probability (7th Edition). Pearson Education Inc. Prentice Hall, pp. 355–358

Shi, X., Yu, D., Warner, E.D., Pan, X., Peterson, G.W., Gong, Z., Weindorf, D.C., 2004. Soil database of 1:1,000,000 digital soil survey and reference system of the Chinese genetic soil classification system. Soil Survey Horizons, 45(4), 129‒136

Shi, X., Yu, D., Yang, G., Wang, H., Sun, W., Guo, H., Gong, Z., 2006. Crossreference benchmarks for translating the genetic soil classification of China into the Chinese soil taxonomy. Pedosphere 16(2), 147−153

Singh, S., Kumar, S., Jain, M.C., 1997. Methane emission from two India soil planted with different rice cultivars. Biology and Fertility of Soils 25, 285−289

Tian, H., Xu, X., Lu, C., Liu, M., Ren, W., Chen, G., Melillo, J., Liu, J., 2011. 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

van Bodegom, P.M., Leffelaar, P.A., Stams, A.J.M., Wassmann, R., 2000. Modeling methane emissions from rice fields: Variability, uncertainty, and sensitivity analysis of processes involved. Nutrient Cycling in Agroecosys. 58(1–3), 231–248

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

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

Wang, B., Neue, H.U., Sanmote, H.P., 1997. Effect of cultivar difference,‘IR72’, ‘IR69958’ and ‘Dular’) on methane emission. Agri. Ecosys. Environ. 62, 31−40

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

42, 163−173

Xie, B., Zhou, Z., Zheng, X., Zhang, W., Zhu, J., 2010. Modeling methane emissions from paddy rice fields under elevated atmospheric carbon dioxide conditions. Adv. Atmos. Sci. 27(1), 100–114

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

Yan, X., Cai, Z., Ohara, T., Akimoto, H., 2003. Methane emission from rice fields in mainland China: Amount and seasonal and spatial distribution. J. Geophys. Res. 108(D16), 4505, doi:10.1029/2002JD003182

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

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

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

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

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

Appendix A: Description of CH4MOD and the compilation of model inputs

CH4MOD is an 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). This model calculates the production of methanogenic substrates from rice plant root exudates and added organic matter (OM) decomposition. Both OM decomposition and rice-plant-induced substrate production are significantly influenced by environmental factors, including soil texture and temperature. Soil moisture controls the fraction of the substrates transformed into methane. There are two major paths by which the methane produced in rice paddy soils is emitted into the atmosphere. One path 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 time 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 indicated that 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 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. The available data regarding 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 comprises 10 km×10 km raster datasets of soil properties at 10 cm depth intervals from the surface down. The spatial resolution of the soil data is the finest among the CH4MOD input parameters (Fig. 2).

A4 Organic matter amendment in rice paddies

The organic matter amended into rice fields includes various types of farm manure (green manure, animal manure etc.) and crop straw as well as dead roots and stubble from previous crops. Roots remaining in the soil can be accounted for using the root/shoot ratio (Huang et al., 2007). Stubble was assumed to represent 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 at a provincial level with the census data (Table 2). The straw application in Table 2 is not rice-specific but, rather, incorporates all the crops in each province. The bias may not be significant in provinces where rice dominates 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 concerning the application of manure in rice cultivation. In this study, 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). The amount of farmyard manure application in each province (Table 2) was part of the investigation results.

Appendix B: Sensitivity analysis of CH4MOD

Data on an environmental factor are usually expressed as M ± e, where M represents the measurement and e represents the error. When used as model inputs, imprecise data 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 represents the variability of the model output in response to variations in model inputs. Usually, an individual variable sensitivity analysis is performed by "varying one variable at a time". 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(lnx). The purpose of the model sensitivity analysis in the present study is to explore the modeled methane flux variability to variations of the model input parameters as in formula (b 1):

$$\frac{dy}{y} \propto s_k \times \frac{dx_k}{x_k} \qquad \text{or} \qquad d(\ln y) \propto s_k \times d(\ln x_k) \qquad \text{(b 1)}$$

where k is used to identify each model parameter and y represents the seasonal methane emissions flux (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 for the relationship between methane flux and the model input variables in terms of fractal variation.

The Monte Carlo approach was adopted as the first step to randomly select values of the model input parameters from their value domains (Table B1), at which point 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, a 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:

$$s_w = \frac{1}{N} \times \sum_{k \neq l} \frac{\left| \overline{y_l} - \overline{y_k} \right|}{y_0}, \qquad k, l \in W \qquad \text{(b 2)}$$

where W = (1, 2, 3, 4, 5) in the formula (b 2) 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_o 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 the dates of transplanting to the harvest. In the model sensitivity analysis, the temperature data are virtually created by the following equations:

$$T_{air}^{(t)} = \overline{T}_{\max} - \left| t - S_{\max} \right| \times D_T + R(-0.5, 0.5) \qquad \text{(b 3)}$$

$$D_T = \begin{cases} (\overline{T}_{\max} - \overline{T}_{\min})/(S_{\max} - S_s), & T \le S_{\max} \\ (\overline{T}_{\max} - \overline{T}_{\min})/(S_e - S_{\max}), & T > S_{\max} \end{cases} \quad \text{(b 4)}$$

$$S_{max} = R(S_s, S_e) \quad \text{(b 5)}$$

$$\overline{T}_{\max} = R(25.0, 35.0) \quad \text{(b 6)}$$

$$\overline{T}_{\min} = R(10.0, 20.0) \quad \text{(b 7)}$$

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 for the rice season. The time variable t ($S_s \le t \le S_e$) represents days after transplanting.

The results indicated 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.