WAGENINGEN UR

Uncertainty assessment in greenhouse gas emission profiles of livestock sectors in Africa, Latin America and Europe

Biqing Zhu

MSc Thesis in Environmental Sciences

July, 2014

Supervised by: Prof. Dr. Wim de Vries[1], Dr. Hans Kros[2], Dr. Jan Peter Lesschen[2] and Ir. Igor Staritsky[2]

[1]Environmental System Analysis, Wageningen University, The Netherlands

[2]Alterra Research Institute, Wageningen UR, The Netherlands

Uncertainty assessment in greenhouse gas emission profiles of livestock sectors in Africa, Latin America and Europe

Biqing Zhu

MSc Thesis in Environmental Sciences

July, 2014

Supervisor(s):

1) Prof. Dr. Wim de Vries(ESA)
wim.devries@wur.nl
2) Dr. Hans Kros(Alterra)
hans.kros@wur.nl
3) Dr. Jan Peter Lesschen
janpeter.lesschen@wur.nl
4) Ir. Igor Staritsky
igor.staritsky@wur.nl

Examiners:

1st: Prof. Dr.Wim de Vries(ESA)
wim.devries@wur.nl
2st: Prof. Dr. Rik Leemans (ESA)
rik.leemans@wur.nl

Preface

After a year of hard working, I am now happy to present you my thesis report about Uncertainty assessment in greenhouse gas emission profiles of livestock sectors in Africa, Latin America and Europe. I would like to give thanks to all of those who supported me during the implementation of my research. I would like to express my great gratitude to my supervisors Hans Kros, Jan Peter and Igor Staritsky. Thanks to Hans Kros for his excellent guidance and patience to help me starting the project and to solve the statistical problems. Thanks to Jan Peter Lesschen for his explanations and providing data related to the MITERRA-Global model. Thanks to Igor Staritsky for his great patience helping me learning R and GAMS, and helping me with modelling and data analysis. I would like to thank my supervisor Wim de Vries for his clear guidance during structuring ideas, writing and presenting. I would also like to thank Gerard Heuvelink and Dennis Walvoort for providing the R package and their help with applying the package to this project. All of these people have provided great help for conducting this research. But most importantly, I want to thank them for their kindness and patience. It has been a great honour and wonderful memory for me to conduct this research.

Biqing Zhu

Wageningen, 2014

Summary

The global animal food chain, including land use change, currently generates 14.5% of global emissions of the greenhouse gases (GHGs) carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O), as measured in CO_2 equivalents. The bulk of GHG emissions originate from CH_4 emissions from enteric fermentation, CH_4 and N_2O emissions from manure management, CO_2 and N_2O emissions from feed production, processing and transport, and CO_2 emissions from energy consumption. No research has until now been done to assess the uncertainties in GHG emission profiles of livestock. Until now, research on uncertainties has been limited to either one greenhouse gas and /or to agriculture in general. The objective of this research is to assess the uncertainty in the model inputs and parameters of GHG emission profiles of livestock sectors. This research is limited to three continents: Africa, Latin America and Europe. The GHG included are CH_4, N_2O and CO_2.

The uncertainties were estimated with MITERRA-Global, which is an environmental assessment model, calculates the global greenhouse gases (GHG) emissions as CO_2, CH_4, and N_2O, other atmospheric nitrogen emissions such as NH_3, NO_x and the N leaching to ground and surface water. It is a deterministic and static model that calculates annual emissions using emission and leaching factors. The model calculates N leaching from housing and manure storage systems and agricultural soils, and soil carbon stock changes based on IPCC Tier1 emission factors. For N_2O direct soil emissions a Tier2 approach is also available. Besides total emissions per region or hectare, the model can also express emissions per livestock sector or per product. The main input data for MITERRA-Global are: crop data, livestock data, feed data, fertilizer consumption data and spatial GIS data. The calculation is generally carried out at a sub-national level (i.e. administrative regions within a county) but the model`s output can also aggregated to national or continental levels. This research aims to assess the uncertainties in the model inputs and parameters of the simulated GHG emission profiles of livestock sectors as estimated by the MITERRA-Global model. The uncertainty assessment focuses on the probability distribution functions (pdfs) and the uncertainty propagation to the outputs of the MITERRA-Global model. This research is limited to three continents: Africa, Latin America and Europe. The GHG assessed are CH_4, N_2O and CO_2.

The model output uncertainty is caused by uncertainties in model inputs, model parameters, and model structure and model resolution. This research`s uncertainty quantification is limited to uncertainties in model inputs and model parameters, which are quantified based on the defining characteristics of their probability distribution functions, considering their spatial- and cross-correlation. The uncertainties in model inputs and parameters are derived from time series of census data, literature review or expert knowledge. These model inputs and parameters are further divided in nine groups based on data/parameter type and greenhouse gas. The final model output uncertainty and each uncertainty contribution for each nine group are quantified using Monte Carlo approach. The results are determined for each continent for the total GHG emission from the livestock sectors, the GHG emissions per livestock sector, per livestock product and per livestock production process.

Expressed as coefficient of variation (cv, as standard deviation divided by mean) results indicates that 1) The model input and parameter uncertainty ranges from 0.001 to 0.82; 2) the output uncertainty ranges from 0.12 to 0.17 for total GHG emission, 0.15 to 0.18 for CH_4 emission, 0.22 to 0.37 for N_2O emission and 0.24 to 0.31 for CO_2 emission; 3) the uncertainty in Europe is lower than in the other two continents; 4) the uncertainty ranges from 0.12 to 0.19 for beef and milk, and from 0.14 to 0.28 for pork, chicken and eggs; 5) uncertainty contribution in total GHG emissions from livestock is mainly determined by the N and CH_4 emission factors in Africa and Latin America, while in Europe, livestock parameters and other emission factors and parameters also have a substantial effect; 6) uncertainty in beef and milk products is for 80% determined by the CH_4 emission factors and for 20% by the N_2O emission factors. Uncertainty in pork is for 30% determined by the CH_4 emission factors, for 50% determined by the N_2O emission factors, and for 20% determined by other emission factors and parameters. Uncertainty in eggs and chicken is for 40% determined by the N_2O emission factors, for 40% determined by other emission factors and parameters, and for 10% determined by activity data. The uncertainty of total GHG emissions of livestock sectors is higher in Africa and in Latin America than that in Europe. The uncertainty of CH_4 emission is lower than that of N_2O and CO_2. CH_4 emission factors and N emission factors contribute the most to the model output uncertainty.

Table of Contents

List of tables

List of figures

1. Introduction

1.1 Background

The increasing ecological footprint of global animal production enhances concerns on environmental problems (Delgado, 1999; Delgado et al., 1999; Smil, 2002; Steinfeld et al., 2006; Galloway et al., 2007). Expanding livestock sectors worldwide contribute to expansion of agricultural land and associated deforestation, emissions of greenhouse gases (GHGs) (Steinfeld et al., 2006), eutrophication of surface waters (Seitzinger et al., 2005; Boyer et al., 2006) and nutrient imbalances (Smaling et al., 2008; Menzi et al., 2010). The global animal food chain, including land use change, generates 14.5% of global GHG emissions as measured in CO_2 equivalents (Gerber et al., 2013), while its contribution in Europe is near 14% (Steinfeld and Wassenaar, 2007). However, the contribution of livestock production to global anthropogenic greenhouse gas (GHG) emissions varies highly across the world (Lesschen et al., 2011b). The Food and Agriculture Organization of the United Nations (FAO) (Gerber et al., 2013) reported Latin America and the Caribbean have the highest level of emissions (almost 1.3 Gton CO_2-eq), driven by an important production of specialized beef. East Asia has the second highest level of emissions (more than 1 Gton CO_2-eq). North America and Western Europe have similar GHG emission totals (over 0.6 Gton CO_2-eq) and also fairly similar levels of protein output. South Asia's total sector emissions are at the same level as North America and Western Europe but its protein production is half what is produced in those areas (Lesschen et al., 2011b). Ruminants contribute a large share due to their high emission intensity. For the same reason, emissions in sub-Saharan Africa are large, despite a low protein output (Gerber et al., 2013).

The main GHG emission pathways related to livestock production are emissions of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) (De Boer et al., 2011). The bulk of GHG emissions originate from four main categories of processes: CH_4 emissions from enteric fermentation, CH_4 and N_2O emissions from manure management, CO_2 and N_2O emissions from feed production, processing and transport, and CO_2 emissions from energy consumption (Gerber et al., 2013). Among the GHGs mentioned above, CH_4 is estimated to be the dominant emitted GHG (about 44% of the sector's emissions). The remaining part is almost equally shared between N_2O (29%) and CO_2 (27%) (Gerber et al., 2013) .

The mandatory national reporting under the UN Climate Change Convention is usually done with the IPCC inventory approach using various default emission factors (EFs) for N_2O and CH_4 emissions from different sources. Emission factors are defined differently for N_2O and CH_4 emissions within the IPCC guidelines. For N_2O the EF is defined in terms of the mass fraction N_2O-N emissions of the different N inputs, e.g. kg N_2O-N (kg N input)$^{-1}$ (IPCC, 2006). However, for methane, the term EF is defined as the actual emission of CH_4 per animal head, e.g. the IPCC Tier 1 $EF_{(T)}$ is defined as emission factor for the defined livestock population, kg CH_4 head^{-1} yr^{-1}. For dairy and other cattle, however, a Tier 2 approach is recommended for countries with large livestock populations based on a CH_4 conversion factor (CF, kg CH_4/gross energy consumed with feed; assuming that 6.5% of the gross energy of the feed consumed is emitted as CH_4 (IPCC, 2006; Kros et al., 2012b).

There are two main types of models that have been developed to calculate the GHG emission profiles of livestock sectors: process-based dynamic models and empirically based models, which are usually emission factor (EF) based models. Kros et al. (2012b) classified the GHG models concerning livestock production as regional models, farm models and process models. Daycent (Del Grosso et al., 2005) is an example of a process based model, while CAPRI (Weiss and Leip, 2012) and MITERRA (Lesschen et al., 2011a) are EF based regional model. MITERRA-Europe is partly based on the models CAPRI (Common Agricultural Policy Regionalized Impact) (Weiss and Leip, 2012), and GAINS (Greenhouse Gas and Air Pollution Interactions and Synergies) (Winiwarter, 2005), supplemented with an N leaching module, a soil C module and a module for mitigation (Lesschen et al., 2011b). There are, however, substantial uncertainties in the base data and applied methodology such as assumptions surrounding allocation of feeds to livestock species.

Lesschen et al. (2011a) used the MITERRA model to assess regional variations in dairy, beef, pork, poultry and egg production, and related GHG emissions in the 27 Member States of the European Union (EU-27), while distinguishing enteric fermentation, manure management, direct and indirect N_2O soil

emissions, liming, fossil fuel use and fertilizer production. On a per kg product basis, they calculated GHG emissions for of 22.6 kg CO_2-eq/kg beef, 1.3 kg CO_2-eq/kg milk, 3.5 kg CO_2-eq/kg pork, 1.6 kg CO_2-eq/kg poultry, and 1.7 kg CO_2-eq/kg eggs. They reported, however, large variations in GHG emissions per unit product exist among EU countries, which are due to differences in animal production systems, feed types and nutrient use efficiencies.

No research has until now assess the uncertainties in GHG emission profiles of livestock (Kros et al., 2012a; Kros et al., 2012b). So far, research on uncertainties has been limited to either one greenhouse gas and /or to agriculture in general. Del Grosso et al. (2010) performed a nation-wide uncertainty analysis using the DAYCENT model combined with an empirically based approach to quantify uncertainties in soil N_2O emissions from croplands in the USA. Wang and Chen (2012) reviewed the state-of-the-art knowledge on the parameterization and uncertainty analysis of soil GHG emission models and presented case studies for comparing the model uncertainties of the denitrification components of four models; DAYCENT, DNDC, ECOSYS, and COMP. (Kros et al., 2012a) analysed the uncertainty propagation using the INTEGRATOR model for the emissions of N_2O for the entire EU27 and for individual countries. Karimi-Zindashty et al. (2012) used a Monte Carlo simulation to estimate the uncertainties in CH_4 emissions from livestock in Canada, using IPCC Tier 2 methodology. Sommer et al. (2009) assessed the whole-system effects of technologies for reducing GHG emissions from livestock model farms using slurry-based manure management.

This research assesses the uncertainties in GHG emission profiles of livestock sectors in Africa, Latin America and Europe. It is a follow up of research carried out on average emissions for Europe (Lesschen et al., 2011b), while further extending the research to other parts of the world (Latin America and Africa). This work is part of the FP7 EU project AnimalChange (An Integration of Mitigation and Adaptation Options for Sustainable Livestock Production under Climate Change). The integrated EU project AnimalChange (www.animalchange.eu) aims to provide scientific guidance on the integration of adaptation and mitigation objectives and design sustainable development pathways for livestock production in Northern and Sub-Saharan Africa, Latin America and Europe. One of the key objectives of this project is to quantify and reduce uncertainties in greenhouse gas (GHG) emissions and assesses climate change impacts on livestock systems (including grasslands) at regional scales, through system analyses, experiments, measurements, modelling and uncertainty analyses (Kros et al., 2012b).

1.2 Problem statement

The problem which will be addressed during this research is the uncertainty quantification of CH_4, N_2O and CO_2 emissions of livestock sectors in Africa, Latin America and Europe estimated by MITERRA-Global.

IPCC (IPCC, 2006) provides guidance in estimating and reporting uncertainties associated with annual estimates of emissions and removals. Uncertainty is defined as lack of knowledge of the true value of a variable, that can be described as a probability density function (pdf) characterizing the range and likelihood of possible values (IPCC, 2006). Among all the concepts associated with conducting an uncertainty analysis, accuracy and precision raise most attention. Accuracy refers to the agreement between the true values and the average of repeated measured observations or estimates of a variable. Precision, on the other hand, stands for the agreement among repeated measures of the same variable (IPCC, 2006). Both accuracy and precision are used to describe the model uncertainty. Thus, the lack of accuracy (known as systematic errors) and the lack of precision (known as random errors) can both be regarded as model uncertainty. Compared with random errors, systematic errors are much more difficult to be quantified. However, quantification can be made by comparing model predicted emission data with other model predictions or measured emission data.

As in many other models, the output uncertainty of the MITERRA model is determined by three categories of uncertainty sources: (1) model input and parameter uncertainty (2) model structure uncertainty, and (3) model solution uncertainty (Kros et al., 2012a). In the context of MITERRA model inputs and parameters (MIPs) refer to (1) Activity data such as animal numbers, crop yields, N fertilizer amounts (2) Model parameters, in terms of excretion and emission factors and (3) Biophysical data such as climate, land use, soil (uncertainty not considered in this study). The model input and parameter uncertainty is the main focus of uncertainty source in this study. As it has been discussed above, the uncertainty caused by model structure is relatively difficult to quantify. A possible way to assess is to compare the results of MITERRA with the results from other models. The model solution uncertainty refers to errors caused by rounding, numerical evaluation of integrals, suboptimal optimization solutions,

etc. However, it is assumed to have a marginal contribution to the output uncertainty and can therefore be ignored (Kros et al., 2012a).

In order to assess the uncertainties in model outputs, the uncertainties of the model inputs need to be quantified first. The approach taken to analyse the uncertainty will be to determine the uncertainties in activity data and emission factors, and then to combine the uncertainties to provide uncertainty estimates for the entire inventory (IPCC, 2006). The approach of this uncertainty assessment contains both uncertainty quantification (UQ) and uncertainty analysis (UA) using a Monte Carlo (MC) simulation approach.

The UQ analyses the uncertainties in model inputs and model outputs, while UA quantifies the contribution of individual sources of uncertainties to the output uncertainty. For UQ all model inputs and model parameters are considered uncertain, ranging from activity data to model constants. For UA we will select groups of model parameters for which we estimate the uncertainty contribution. The uncertainty analysis (UA) is accomplished by a Monte Carlo (MC) simulation with only one model input/parameter (group) considered uncertain for which the contribution is to be estimated. The other model inputs are considered certain by using their default values as stored in the MITERRA-Global database (Kros et al., 2012a).

Some model inputs/parameters are correlated with other model inputs/parameters, which affects the uncertainty estimation (Kros et al., 2012a). Moreover, uncertainty about spatially distributed inputs tends to be positively spatially correlated, and this influences the degree to which uncertainties cancel out by spatial aggregation (Heuvelink and Pebesma, 1999). Therefore, when quantifying the uncertainty of model inputs and parameters mentioned above, both of their cross-correlations and spatial correlations will be taken into account.

1.3 Research objective and research questions

1.3.1 Research objective

The general research objective is to assess the uncertainty in the model inputs and parameters (MIPs) of GHG emission profiles of livestock sectors in Africa, Latin America and Europe by a Monte Carlo analysis. The analysis includes probability distribution functions (pdfs) and propagation of uncertainties in the outputs of the MITERRA-Global model.

1.3.2 Research questions

The two main research questions, which will be addressed in this study, are:

1. What are the uncertainties in CH_4, N_2O and CO_2 emissions from livestock sectors in in Africa, Latin America and Europe?
2. Which model inputs and/or parameters are the main sources of uncertainty contributing to the output uncertainty in CH_4, N_2O and CO_2 emissions of livestock sectors in the three continents?

These questions will be tackled by answering the following sub-questions:

1. What are the main input data and parameters to be analysed in the uncertainty assessment?
2. How can a statistical model (pdf) be built that fully characterizes the uncertainty of the main selected input data for MITERRA-Global for the various continents, described as probability distributions, including cross correlation for certain pairs of model inputs and/or spatial correlation in these uncertainties where relevant?
3. How can model inputs and parameters realizations be sampled efficiently from their pdfs by using stochastic simulation techniques?
4. How can the uncertainty propagation analysis be carried out in batch mode and its results be stored automatically?
5. What is the uncertainty in the CH_4, N_2O and CO_2 emissions of livestock sectors in the three continents and which model inputs and/or parameters contribute the most to this uncertainty?
6. How can the results of the uncertainty contribution analysis be summarized and visualized and thus efficiently communicated to end-users in the three continents?

1.3.3 Structure of the report

In addition to the introduction chapter, this report contains another four Chapters. Chapter 2 describes the Methodology for uncertainty quantification (UQ) and uncertainty analysis (UA) and how this methodology is applied to the MITERRA-Global model. Chapter 3 provides the results for the uncertainty quantification (UQ) and uncertainty analysis (UA). Chapter 4 provides the discussion and recommendations for the methodology and for the results. Chapter 5 provides the general conclusions of the research.

2. Methodology

A Monte Carlo (MC) analysis was used to analyse uncertainty of GHG emissions of livestock sectors in Africa, Latin America and Europe, using the MITERRA-Global model. The analysis includes uncertainty quantification (UQ) and uncertainty analysis (UA). UQ and UA are important subjects of AnimalChange project (Kros et al., 2012b). This methodology is descript by Kros et al. (2012b), adapted from Heuvelink et al. (2009). The purpose of UQ is to quantify the model input uncertainty and the model output uncertainty. Whereas UA is used to determine how much the individual (group of) model input uncertainty contribute to the model output uncertainty. Eight steps are identified in this methodology (Kros et al., 2012b):

1. Defining the model, its inputs and outputs.
2. Selection of uncertainty sources and the uncertain model inputs.
3. Uncertainty quantification in model inputs.
4. Method to combine uncertainties: Monte Carlo simulation.
5. Selection of model outputs for which the uncertainty is assessed.
6. Uncertainty quantification in model outputs.
7. Uncertainty analysis of model inputs.
8. Communicate the outcomes of the UQ/UA.

Step 1 to step 7 were executed sequentially. Step 8 was applied to interpreted the UQ/UA results in the relevant steps.

2.1 Estimation of GHG emission from livestock with MITERRA-Global

MITERRA-Global is an environmental assessment model, which calculates emissions greenhouse gases as CO_2, CH_4, N_2O and atmospheric nitrogen emissions as N_2O, NH_3, NO_x and NO_3 leaching to ground and surface water on a deterministic and annual basis using emission and leaching factors (Lesschen et al., 2011a). The main emission pathways related to livestock production are shown in Figure 1.Those pathways are all included in the MITERRA-Global model except for food and feed processing.

The modelling concept of MITERRA-Global is based on the MITERRA-Europe model (Velthof et al., 2009; Lesschen et al., 2011). MITERRA-Europe was partly based on the models CAPRI (Common Agricultural Policy Regionalised Impact), and GAINS (Greenhouse Gas and Air Pollution Interactions and Synergies), supplemented with an N leaching module, a soil carbon module and a module for mitigation measures. Input data consists of activity data (e.g., livestock numbers, crop areas, animal production from Eurostat and FAO), spatial environmental data (e.g., soil and climate data) and emission factors (IPCC and GAINS). The model includes measures to mitigate GHG and NH_3 emissions and N leaching and runoff.

MITERRA-Global follows the same model structure as MITERRA-Europe based on relatively simple and transparent calculations using emission factors and statistical data. The main inputs and outputs of the MITERRA-Global model are given in Figure 2. The main input data for MITERRA-Global are crop data, livestock data, feed data, fertilizer consumption data and spatial data (GIS data) on land cover, soil and climate. Outputs include N and P budgets and GHG emissions, the latter being the focus of this study. Besides total emissions per region or hectare based emissions, the model can also express emissions on a per product basis, following the top-down LCA based approach as described in Lesschen et al. (2011a).

The calculation is carried out on a sub-national level. which is the level for which statistical input data are available for the main activities. The output can be provided at sub-national, national or continental level.

MITERRA-Global accounts for the following GHG sources: CH_4 from enteric fermentation, CH_4 and N_2O from manure management, direct and indirect N_2O soil emissions, CO_2 and N_2O from organic soils, CO_2 from liming and urea application, and GHG from fertilizer production and fossil fuel use. All emissions are converted to CO_2-eq using most recent estimates of 100 years global warming potential (GWP) values (IPCC, 2007), which are for CH_4 and N_2O 25 and 298 times the GWP of CO_2, respectively (Lesschen et al., 2014).

CH_4 emission from enteric fermentation were calculated using Tier1 emission factors (EF) derived from IPCC (2006). These Tier1 EFs are animal and continent specific. For EU27 country specific emission

factors are used for cattle. Tier2 approach is applied for enteric fermentation once the feed intake procedure is fully implemented in the model. This approach accounts for the amount and quality of feed intake and will therefore be country specific.

The N_2O emissions from agricultural soils consist of direct and indirect soil emissions. Direct N_2O emissions are from application of N fertilizer and animal manure, crop residues and cultivation of organic soils, urine and faeces produced during grazing. Indirect N_2O emissions are from N leaching, runoff and from atmospheric deposition of N volatilised from managed soils. The N_2O emissions were calculated from IPCC (2006) EFs.

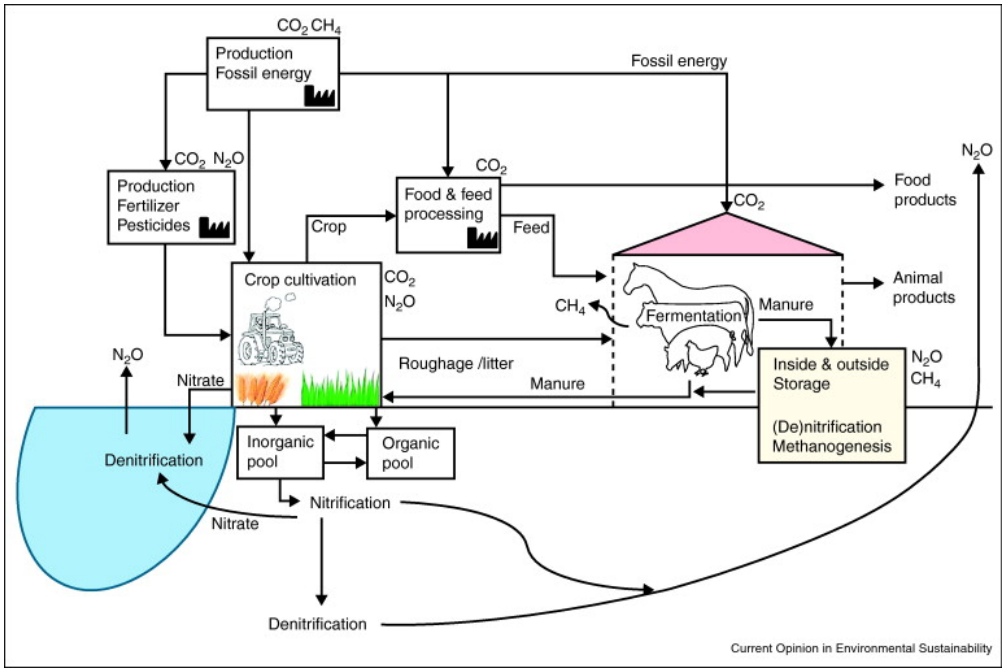


Figure 1 Main emission pathways of CO_2, CH_4 and N_2O related to livestock production (De Boer et al., 2011) as included in the MITERRA model (except for food and feed processing).

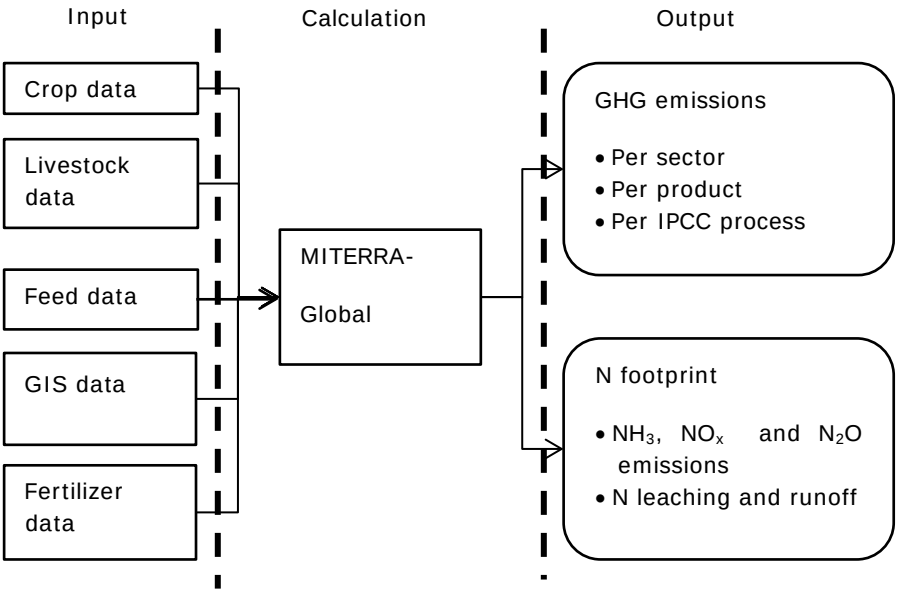


Figure 2 Inputs and outputs of the MITERRA model (adapted from Lesschen et al. (2011a))

Anaerobic decomposition of organic material in flooded rice fields produces methane, which escapes to the atmosphere primarily by transport through the rice plants. These emissions from rice cultivation are calculated according to the IPCC (2006) guidelines as a function of emission factors, cultivation period of rice and annual harvested are of rice.

Drainage and tillage of organic soils leads to loss of C due to accelerated organic matter decomposition. The CO_2 emissions from organic soils are calculated using IPCC (2006) EFs which distinguish arable land from grassland. In addition, CO_2 emissions from liming and urea application are included based on the carbon content and IPCC (2006) EFs for these soil additives.

For manure management the IPCC (2006) emission factors were used. For CH_4 the EFs depend on animal type, average annual temperature and manure system. These EFs are region specific on the basis of the average annual temperature.

2.2 Defining the uncertain model inputs and parameters

As stated before, the uncertain sources include (1) model input and parameter uncertainty (2) model structure uncertainty, and (3) model solution uncertainty (Kros et al., 2012a). The model input and parameter uncertainty is the focus of this study. The model inputs and parameters (MIPs) that are considered in this uncertainty assessment are delivered by the group of MITERRA-Global developers. The MIPs, which can directly and/or indirectly influence the GHGs emissions (CH_4, N_2O and CO_2 emissions) are selected.

Although the spatial support of MITERRA-Global is sub-national regions, not all MIPs were available for each sub-national level. Based on the data sources, four different spatial levels are identified, i.e. sub-national level, national level, continental level and generic level. The sub-national level is the lowest spatial level used in MITERRA-Global. In Europe the NUTS-2 regions are used. In Africa and Latin America, provinces or groups of provinces are used. The units of national level are equivalent to the FAO countries. The continental level uses IPCC continents. Inputs and parameters at generic level remain the same among continents. There were 89 MIPs selected and they are provided in Table 1 with their code, spatial levels and their meaning.

Table 1 Selected model inputs

Nr.	Code	Spatial level	Meaning
1	AgroMapsCA_barley	sub-national	Crop(barley) area at sub-national level from Agromaps[1]
2	AgroMapsCA_maize	sub-national	Crop(maize) area at sub-national level from Agromaps
3	AgroMapsCA_other	sub-national	Crop(other) area at sub-national level from Agromaps
4	AgroMapsCA_soybean	sub-national	Crop(soybean) area at sub-national level from Agromaps
5	AgroMapsCA_wheat	sub-national	Crop(wheat) area at sub-national level from Agromaps
6	AgroMapsCropProd	sub-national	Crop production at sub-national level from Agromaps
7	BMFac_Gras	generic	fraction of N of bruto mineralisation available on grass
8	BMFac_OtherAreable	generic	fraction of N of bruto mineralisation on arable land
9	bNumAniRAINS	national	Animal numbers at country level for EU countries from GAINS
10	CAPRI_NumAni	sub-national	Livestock numbers at $NUTS_2$ level for EU
11	CH4_EF_EntFer_cs	sub-national	Enteric fermentation (kg CH4 per animal per year) for cattle and sheep
12	CH4_EF_EntFer_other	sub-national	Enteric fermentation (kg CH_4 per animal per year) for other animals
13	CH4_EF_ManManage	sub-national	Manure management methane EF by temperature (kg CH_4 per head per year)
14	CO2_L_gasoil	generic	CO_2 EF for gasoil kg CO_2/litter
15	CompositionFertilizer	generic	Nutrient content of fertilizers in % (based on FAOSTAT data)
16	CropAreaCAPRI	sub-national	Crop area at NUTS2 level for EU
17	CropProperties	sub-national	only fuel use
18	DevCrop	sub-national	Percentage of manure reserved for fodder crops (grass and fodder crops)
19	EF_Fert_Prod	national	emission factors for fertilizer production
20	EF1	sub-national	emission fraction EF1 kg N-N_2O per kg N(IPCC-1997)
21	EF2	generic	emission from histosols per ha EF2 in kg N-N_2O per ha (IPCC-1997)
22	EF4	generic	factor for indirect emission due to emission of ammonia kg N-N_2O per kg NH_3 en NO_X
23	EF5	sub-national	emission fraction of indirect emissions EF5 IPCC, 1997 pr kg N-N_2O leaching and runoff
24	EffFact	generic	Over fertilization factor, based on MITERRA-Europe. Set at 1.25 for most crops, 1.1 for cereals and 1 for grass and perennial energy crops
25	EU_ani	sub-national	horses and other animal numbers at NUTS2 region for EU, only for N excr is used
26	Excr_EU	sub-national	N excretion for EU countries based on GAINS (only N excretion is used)
27	FAO_AnimalProd	national	Animal production data (mostly expressed in tonnes)
28	FAO_LandAreas	national	Land areas (1000 ha)
29	FAO_NatCA_barley	national	Area Harvested (ha) for barley
30	FAO_NatCA_maize	national	Area Harvested (ha) for maize
31	FAO_NatCA_other	national	Area Harvested (ha) for other crops
32	FAO_NatCA_soybean	national	Area Harvested (ha) for soybean
33	FAO_NatCA_wheat	national	Area Harvested (ha) for wheat
34	FAO_NatCP_barley	national	Production (tonnes) for barley
35	FAO_NatCP_maize	national	Production (tonnes) for maize
36	FAO_NatCP_other	national	Production (tonnes) for other
37	FAO_NatCP_soybean	national	Production (tonnes) for soybean
38	FAO_NatCP_wheat	national	Production (tonnes) for wheat
39	FAO_NatFertilizer	national	Fertilizer consumption (tonnes of nutrients)

[1] http://kids.fao.org/agromaps/index.html

Table 2 Selected model inputs (Continued)

40	feedset_Animals	national	Use of livestock items for animal feed per country from FAOSTAT)
41	feedset_Crops	national	Use of crop items for animal feed per country from FAOSTAT)
42	FertilizerType	national	Fertilizer consumption by fertilizer type from FAOSTAT
43	Fqatm	generic	Fraction of deposited N available for crops
44	Fqbiol	generic	Fraction of fixed N available for crops
45	FQGraz	generic	fraction of N in plant available N from grazing
46	FracR	sub-national	Fraction of crop residues that are removed (e.g. straw)
47	GrasCorrEU	sub-national	Fraction of total grassland area that is classified as rough grazing (according to FSS definition)
48	Grass_Yield	sub-national	Grassland yield (kg DM/ha)
49	GrassCorrection	sub-national	Estimated fraction of natural grassland
50	GrassYieldEstimate	sub-national	Estimated grassland yield (kg DM/ha)
51	IPCC_Nexcretion	sub-national	N excretion (kg N per animal per year)
52	LandCoverMap	sub-national	Area of land cover types
53	LD_Buffaloes	sub-national	Number of buffaloes per sub-national region
54	LD_Cattle	sub-national	Number of cattles per sub-national region
55	LD_Chickens	sub-national	Number of chickens per sub-national region
56	LD_Goats	sub-national	Number of goats per sub-national region
57	LD_Pigs	sub-national	Number of pigs per sub-national region
58	LD_Sheep	sub-national	Number of sheep per sub-national region
59	LeachingStorage	sub-national	Leaching fraction of leaching from manure storages
60	LivestockCountryTotal	national	Livestock number per country
61	ManureSU_Burned	continental	Manure management system usage (%) burned
62	ManureSU_DailySpread	continental	Manure management system usage (%) daily spread
63	ManureSU_Digester	continental	Manure management system usage (%) digester
64	ManureSU_Drylot	continental	Manure management system usage (%) drylot
65	ManureSU_Lagoon	continental	Manure management system usage (%) lagoon
66	ManureSU_Liquid	continental	Manure management system usage (%) liquid
67	ManureSU_Other	continental	Manure management system usage (%) other
68	ManureSU_Pasture	continental	Manure management system usage (%) pasture
69	ManureSU_SolidStorage	continental	Manure management system usage (%) solid storage
70	N_deposition_data	sub-national	NH_3 and NO_x deposition per region for 1860, 1993 and 2050
71	N2_animal_EF	sub-national	N_2 Emission factors from manure management based on EMEP-EEA 2009 guidebook
72	N2O_grazing	sub-national	N_2O-N emission factor for grazing
73	N2O_manure_storage	sub-national	N_2O-N emission factor (fraction) for manure management based on IPCC 2006 guidelines
74	NH3_animal_EF	sub-national	Emission factors based on EMEP-EEA 2009 guidebook
75	NH3_fert_EF	sub-national	Emission factors for total NH_3 emissions from soils due to N fertiliser volatilization
76	Nharvest	sub-national	Nutrient content of harvested crops g/kg harvested product
77	Nindex	generic	Ratio of N in harvested crop versus residues
78	NO_animal_EF	sub-national	NOx emission factors from manure management based on EMEP-EEA 2009 guidebook
79	orgNGRaz	generic	fraction organic N in grazing manure
80	OrgNLiqMan	generic	organic N (fraction) in liquid manure
81	OrgNSolMan	generic	organic N (fraction) in solid manure
82	PesticideUse	national	average level of pesticide use per ha in a country
83	Precipitation	sub-national	Precipitation data per region in mm
84	ShareFeed	national	Share of feed category per livestock type

Table 3 Selected model inputs (Continued)

85	Temperature	sub-national	Temperature data per region in mm
86	Texture_Clay	sub-national	Fraction of area with soil texture clay
87	Texture_Loam	sub-national	Fraction of area with soil texture loam
88	Texture_Sand	sub-national	Fraction of area with soil texture sand
89	TractorDensity	national	proxy parameter for the level of mechanisation of agriculture

2.3 Quantifying the model input and parameter uncertainty

The uncertainty of the model inputs are represented by probability distribution functions, extended by including the spatial-correlation (for the same MIP between different spatial locations) and cross correlation (between MIs at the same location).

To limit the complexity of calculation, the pdfs are only derived for the continuous numerical MIPs. The approach for uncertainty quantification in the MIPs is adapted Kros et al. (2012b). For each MIP, the following characteristics are identified:

- The probability distribution functions (pdfs)
- The spatial-correlations
- The cross correlations

The full table of uncertainty quantification results is provided in Annex 2.

2.3.1 The probability distribution functions

Different distributions were used depending on whether the input or parameter is measured on a continuous numerical scale on a discrete numerical scale or on a categorical scale. It matters whether the input is constant in space and time or varies in space and/or time (Kros et al., 2012b).

A sample consists of uncertain continuous numerical constant variables $((x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n))$ is characterized by its cumulative distribution function (cdf) F of a continuous numerical variable X:

$$F_X((x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)) = P(Y(x_1)\leq y_1,\ Y(x_2)\leq y_2,\ldots,Y(x_n)\leq y_n) \qquad (1)$$

The cdf is a continuous, non-decreasing function on the real numbers. The limit values are $F_X(-\infty)=0$ and $F_X(+\infty)=1$. The probability density function is the first derivative of F. The surface area below an interval describes the possibility a variable in sample Y. Figure 3 shows the plot of cdf and pdf.

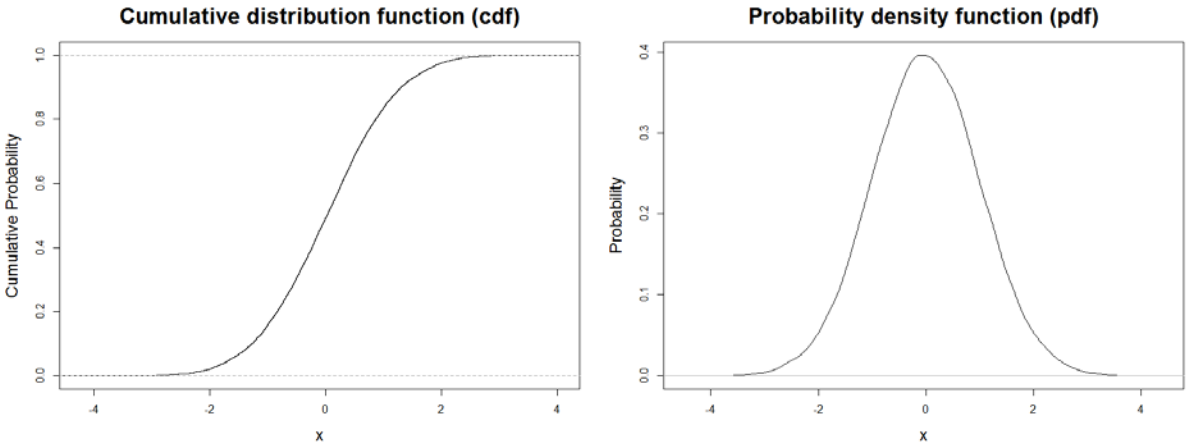


Figure 3 Plot of Cumulative distribution function and probability density function

The parameters to describe a pdf include: 1) mean (μ), 2) minimum and maximum values, 3) distribution type and 4) measure of dispersion (in terms of standard deviation (σ) or coefficient of variation (cv), i.e. σ /μ). For this research, the mean (μ) of each model input/parameter (MIP) was read

in as default value in MITERRA-Global. The minimum and maximum values of the MIPs were are set based on the characteristics of each MIP:

- For fractions, a minimum of 0 and maximum of 1 was used. Furthermore, for sets of MIs representing fractions with sum 1, a maximum sum 1 was applied.
- For other MIs the minima and maxima were not fixed, but a physical minimum and a physical maximum was used. The physical minima and maxima depend on the system boundaries of the model. The minima and maxima may also be derived by the model developers.

The Gaussian (normal) distribution is most often assumed to describe the random variation that occurs in the data from many scientific disciplines; the well-known bell-shaped curve can easily be characterized and described by two values: the arithmetic mean μ and the standard deviation σ, so that data sets are commonly described by the expression μ ± σ. However, many measurements show a more or less skewed distribution. Skewed distributions are particularly common when mean values are low, variances large, and values cannot be negative (Limpert et al., 2001). As it was described in Limpert et al. (2001) such skewed distributions closely fit the log-normal distribution. Figure 4 shows the example of a normal distribution and a lognormal distribution.

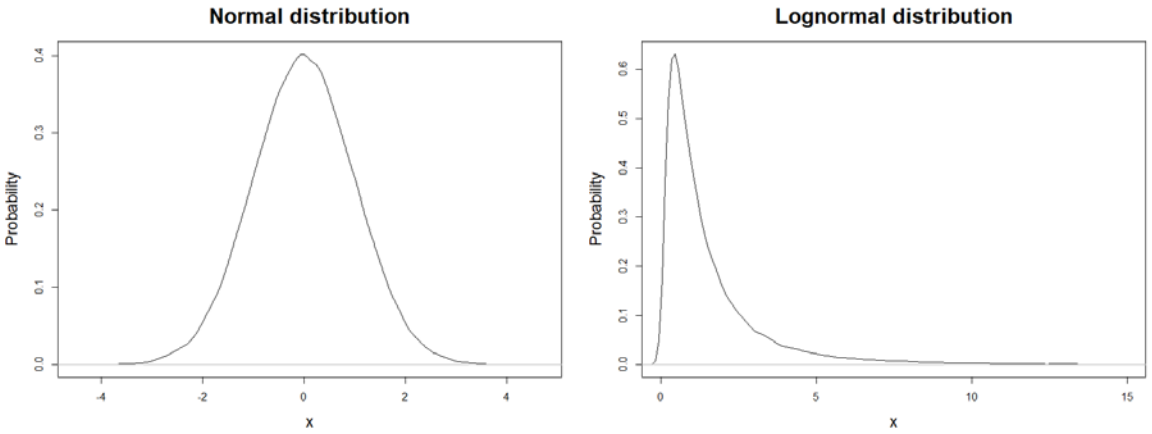


Figure 4 Normal distribution and lognormal distribution

A random variable X is log-normally distributed if log (X) has a normal distribution. Usually, natural logarithms are used, but other bases would lead to the same family of distributions, with rescaled parameters (Limpert et al., 2001).

Hence, we chose two types of distributions to describe the MIPs: normal distribution and lognormal distribution (for skewed MIPs). When the mean values are not too small and the variance is not too large, a normal distribution is assumed. When the mean values are low, the variance is large and there is a multiple effect, a lognormal distribution is assumed. For lognormal distribution natural logarithm is used.

The dispersion of normally distributed MIP is defined by the coefficient of variation (cv). A normally distributed MIP is denoted as $X \sim N(\mu, \sigma^2)$. The parameters used to describe X are the mean (μ), and the variance (σ^2). The standard deviation (σ) is the square root of the variance. Thus, the cv of a normal distribution X is:

$$cv = \sigma/\mu, X \sim N(\mu, \sigma^2) \quad (2)$$

If the MIP is lognormal distributed, it is denoted as $L \sim \Lambda(\mu^*, \sigma^{*2})$, where μ^* and σ^* denote the μ and σ of the log-transformed MIP values.

The standard R package stat[2] provides a calculation of the variance of a lognormal distribution. The probability density function (pdf) of a lognormal distribution Y, using the natural logarithm, equals to:

[2] http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Lognormal.html

$$F(L) = \frac{1}{l \cdot \sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log(l) - \mu)^2\right) \quad (3)$$

where μ and σ are the mean and standard deviation of the log transformed data.

The mean of L is:

$$\mu^* = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \quad (4)$$

The variance of L is:

$$\sigma^{*2} = \exp(2\mu + \sigma^2) \cdot (\exp(\sigma^2) - 1) \quad (5)$$

Hence, the coefficient of variation is:

$$cv^* = \sqrt{\exp(\sigma^2) - 1} \quad (6)$$

This cv* is approximately σ when that is small (e.g., σ < 0.5).

2.3.2 Deriving the pdfs for model inputs

The information regarding the pdfs of the model inputs is obtained from the MITERRA-Global group (expert knowledge) and online data base (FAOSTAT). The information for defining the spatial and cross-correlation came from expert knowledge and previous research Kros et al. (2012a).

Based on the data sources and data types, the MIPs are grouped into two categories with different uncertainty quantification methods:

1. Input and parameter uncertainty derived from census data;
2. Input and parameter uncertainty derived from literature and expert judgment

The following session will explain which inputs are included in each category, the reasoning behind and the calculated/assigned uncertainty for each parameter.

1. input and parameter pdf derived from census data

Census data from FAO do not provide the associated uncertainty. However, we used the temporal variation in the yearly data as a proxy for the uncertainty. Since for census data, the sample size is the whole population, the variation obtained from data is equivalent to the variation of the population. The variation reflects the random error in data collecting and the variance of the true value.

The database of FAOSTAT provide census data for the past. Thus, the pdfs of the MIPs which are derived from FAOSTAT will be calculated for the past 10 to 15 years (depends on the data availability), and used as the pdfs for the corresponding MIPs. Following are the parameters to describe the pdfs.

Because those MIPs are all activity data on, a physical minimum (0) and a physical maximum (infinity) are used. For the FAO inputs, the mean values are usually large (e.g. annual crop production of a country) and the variances are usually very small. According to the rule described in section 2.3.1, a normal distribution is applied for all the MIPs in this group.

The uncertainties of the inputs from the FAO database are calculated from the annual data from the past 10 to 15 years. Since there is a clear difference for each input between continents, cv`s were derived for each continent and the results are presented in Table 4 .

Note, however that this variation not only represent uncertainty, but also the “real” year-to-year variation and/or trend. For the data collected from the same country of the same category from different years, trends exist over years (see example below and Annex 1). Hence, the cv is corrected by the trend, by adding a linear least square fit line. For each input, a generic cv is calculated with the following procedures. Here the FAO crop area data (part of the original see Annex 1) is used as an example. The developed R script is provided in Annex 5.

Table 4 FAO input cv`s for each continent

Code	Eastern Europe	Western Europe	Middle East	Asia	Indian Sub-continent	North America	Latin America	Africa	Oceania
FAO_ AnimalProd	0.04	0.03	0.07	0.03	0.02	0.02	0.05	0.07	0.03
FAO_ LandAreas	0.002	0.01	0.01	0.01	0.01	0.003	0.004	0.004	0.01
FAO_ NatCA_barley	0.08	0.05	0.08	0.04	0.04	0.05	0.07	0.11	0.08
FAO_ NatCA_maize	0.08	0.05	0.08	0.04	0.04	0.05	0.07	0.11	0.08
FAO_ NatCA_other	0.08	0.05	0.08	0.04	0.04	0.05	0.07	0.11	0.08
FAO_NatCA_ soybean	0.08	0.05	0.08	0.04	0.04	0.05	0.07	0.11	0.08
FAO_NatCA_ wheat	0.08	0.05	0.08	0.04	0.04	0.05	0.07	0.11	0.08
FAO_NatCP_ barley	0.16	0.09	0.11	0.05	0.08	0.08	0.09	0.11	0.19
FAO_NatCP_ maize	0.16	0.09	0.11	0.05	0.08	0.08	0.09	0.11	0.19
FAO_NatCP_ other	0.16	0.09	0.11	0.05	0.08	0.08	0.09	0.11	0.19
FAO_NatCP_ soybean	0.16	0.09	0.11	0.05	0.08	0.08	0.09	0.11	0.19
FAO_NatCP_ wheat	0.16	0.09	0.11	0.05	0.08	0.08	0.09	0.11	0.19
FAO_ NatFertilizer	0.13	0.10	0.22	0.13	0.06	0.06	0.16	0.26	0.15
feedset_ Animals	0.14	0.15	0.10	0.14	0.03	0.09	0.15	0.17	0.18
feedset_ Crops	0.20	0.12	0.14	0.08	0.11	0.07	0.14	0.12	0.18
Fertilizer Type	0.82	0.93	0.75	0.54	0.15	0.81	0.38	0.84	0.91

Example: FAO crop area data

Annual data from 2000 to 2012 (13 years) were collected. Observations of the FAO crop area data (O) are given as a three-dimensional data set, in which every observation is denoted by a country code (k), a crop code (l) and a year variable (m). These observations were collected and stored as shown in Table 5.

Table 5 Example of a record with FAO crop area data

Row	Country	Crop	Year 1	Year 2	Year 3	Year 4	Year 5	…	Year n
i	k	l	$O_{k,l,m1}$	$O_{k,l,m2}$	$O_{k,l,m3}$	$O_{k,l,m4}$	$O_{k,l,m5}$	…	$O_{k,l,mn}$

First I analysed the data to see if trend exists. It is detected that for most of the data there is an increase in the trend. Since for the higher observation data, the uncertainty weighs more in the overall uncertainty, the highest 15 sets of data is plotted in Figure 5. From the plot, clear trend can be observed for most of the observations over years.

As a result, when calculating the variance for the model input, the effect of the trend should be removed. First, for each row (for year k, crop l and year from 2000 to 2012) the cv of the FAO trend data is estimated by using the root mean square error (RMSE), being the quadratic sum of the mean error of an observation (Oi) and the predicted (Pi) value based on a linear trend (Janssen and Heuberger, 1995):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i (P_i - O_i)^2}{N}} \qquad (7)$$

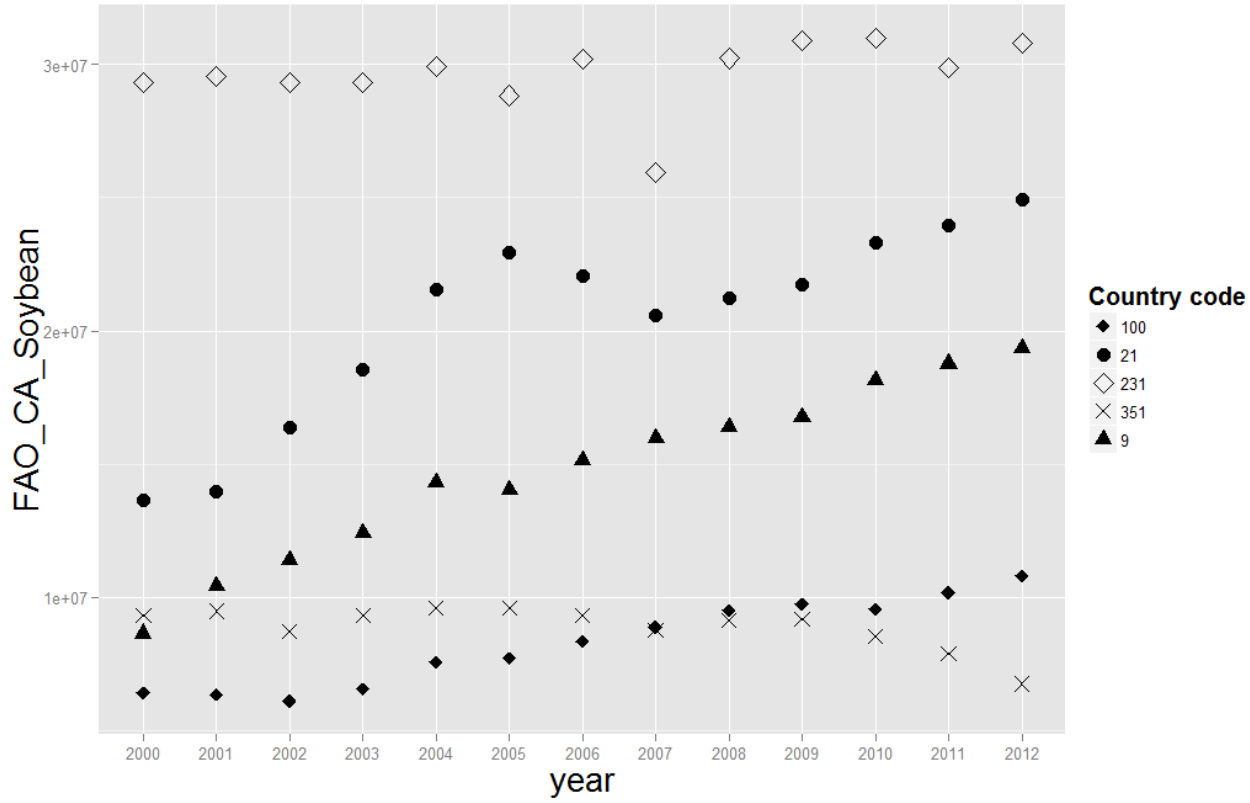


Figure 5 Plot of the highest five FAO crop area data for soybean over 13 years*

**The countries shown in Figure 5 refer to: India, Brazil, USA, China and Argentina (in the legend, from top to bottom).*

The RMSE should be approximately equal to the standard deviation of the measurement noise (Janssen and Heuberger, 1995). This means that the normalized RMSE (NRMSE), i.e. the RMSE divided by the mean of the observations equals the cv or row i:

$$cv_i \approx \mathrm{NRMSE} = \frac{\sqrt{\frac{\sum_i (P_i - O_i)^2}{N}}}{\overline{O_i}} \quad (8)$$

The cv_i is the average of $cv_{i,j}$ weight by the expected crop area $CAE_{i,j}$

Then the generic cv for crop area cv_{crop_area} is calculated as follows:

$$\mathbf{cv_{crop\_area}} = \frac{\sum \mathbf{cv_i * mean_i}}{\sum \mathbf{mean_i}} \quad (9)$$

2. Input and parameter pdf derived from literature and expert judgment

The inputs and parameters of which the uncertainties cannot be derived from historical data were estimated on the basis of literature data and expert knowledge. The minimum and maximum values, and the distribution type were estimated based on the system boundaries of MITERRA-Global model and literature. The variance is calculated or estimated differently, based on the information availability.

The estimation of the minima and maxima follows the rule described in 2.3.1. The MITERRA-Global developers delivered the estimations. As it has been discussed in 2.3.1, usually a normal distribution was assumed. However when the mean value was low, variance large, and values are not allowed to be negative, a lognormal distribution was assumed. The variance is expressed in terms of cv. There are two ways of defining the variance: a) from literature directly and b) from literature and expert judgment.

a. cv derived from literature

For some of the MIPs, uncertainty information was found from literature. This group of MIPs is summarized in Table 6 .

Table 6 Uncertainties of model inputs and parameters derived from literature

Model input code	Literature source	cv	sd	Distribution type	min	max
N2O_manure_storage	Flugsrud and Hoem (2011)	0.35		Lognormal	0	0.5
CO2_L_gasoil	Flugsrud and Hoem (2011)	0.03		Normal	0	5
CH4_EF_ManManage	Flugsrud and Hoem (2011)	0.25		Normal	0	250
CH4_EF_EntFer_cs	Flugsrud and Hoem (2011)	0.25		Normal	0	250
CH4_EF_EntFer_other	Flugsrud and Hoem (2011)	0.40		Normal	0	150
NH3_fert_EF	Monni et al. (2004)	0.30		Normal	0	0.5
EF5	Monni et al. (2004)		0.50	Lognormal	0	0.5
EF1	IPCC (2006)		0.28	Lognormal	0	0.5
EF2	IPCC (2006)		0.63	Lognormal	0	50
EF4	IPCC (2006)		0.82	Lognormal	0	0.5
N2O_grazing	IPCC (2006)		0.57	Lognormal	0	0.57

However, this uncertainty information is usually provided with lower and upper limits. In order to derive the pdfs required for this research, calculations are needed. The calculation process is explained using the example of MIPs of which the uncertainty information can be derived from the IPCC. The information is summarized in Table 7. The calculation is explained below using EF1 as an example.

Table 7 Uncertainty information derived from IPCC

	Information provided by IPCC			Log-transformed properties			
	IPCC min	IPCC mean	IPCC max	mean	min	max	sd
EF1	0.003	0.01	0.03	-4.61	-5.81	-3.51	0.28
EF2	2	8	24	2.08	0.69	3.18	0.63
EF4	0.002	0.01	0.05	-4.61	-6.21	-3.00	0.82
N2O_grazing	0.007	0.02	0.06	-3.91	-4.96	-2.81	0.57

For EF1, the IPCC (IPCC, 2006) provides a default value with minimum and maximum values. The default is assumed to be the mean of EF1. From the deviance of minimum and maximum from the default, it is clear that the distribution of EF1 is highly skewed. Based on the default method of this research, a lognormal distribution is assumed for EF1. Taking the interval determined by this minimum and maximum as the 95% confidence interval, the minimum is regarded as 2.5% percentile and the maximum as 97.5% percentile. For normal distribution $X \sim N(\mu, \sigma^2)$, there is an empirical formula for the 95% confidence interval:

$$(\mu - 1.98\,\sigma, \mu + 1.98\,\sigma) \qquad (10)$$

Thus, the difference between the log-transformed maximum and the log-transformed minimum equals to $2 \times 1.98 \times \sigma$:

$$\log(\max) - \log(\min) = 2 \times 1.98\,\sigma \qquad (11)$$

Then the sd of this log-transformed distribution can be calculated as the σ. This sd is used as the cv of EF1. A proof for this assumption can be found in section 2.3.1.

b. cv derived from expert judgment

Based on a previous study by Kros et al. (2012b), for the MIPs there is little information available on uncertainties, cv`s are derived from expert judgment. We estimate the cv`s of these MIPs fall into one of

the three categories: high uncertainty (cv=0.5), moderate uncertainty (cv=0.25) and low uncertainty (cv=0.1). The rules for assigning uncertainty category to MIPs are:

- low uncertainty is estimated for MIPs which are:
  - derived from good quality statistics data base
- moderate uncertainty is used for all MIPs, which are not indicated with high or low uncertainty
- high uncertainty is estimated for MIPs which are:
  - estimated based on expert knowledge
  - derived from other models

2.3.3 Spatial correlations

Given the limitation of data on the spatial-correlation, the spatial-correlations in the UQ is included in a pragmatic way as it is recommended by Kros et al. (2012a).

For the lowest spatial level, sub-national level, it is assumed that within each sub-national region, the spatial-correlation equals to 1. It indicate that for a model input MIP_i, it has a generic variance (σ^2) within the sub-national region. For each MIP, spatial correlation coefficients are set between plots in different:

- Sub-national regions within the same country: $\rho_{sub\text{-}national}$
- Countries within the same continent: $\rho_{national}$
- Continents within the world: $\rho_{continental}$

Dependent on the spatial dependence of the MIP, The spatial correlation coefficients are determined as five different levels, following the study by Kros et al. (2012a). The five levels are: perfect correlation (ρ =1), high correlation (ρ=0.8), moderate correlation (ρ=0.5), low correlation (ρ=0.2) and no correlation (ρ=0).

The classes of spatial-correlation were assigned to reflect the spatial dependence of the MIPs. For the same MIP, if the variances of two plots at the same spatial level (e.g. between two countries) are independent, the $\rho_{national}$ was assigned to 0. Thus, the change of one plot will not affect the other. If the variances of two plots were perfectly correlated, the spatial correlation coefficient was assigned to 1. In this case, the variances of the two plots were the same. The classes Low, Moderate and High indicate different levels of dependence. Scatter plots (Figure 6) were used to show the strength of correlation.

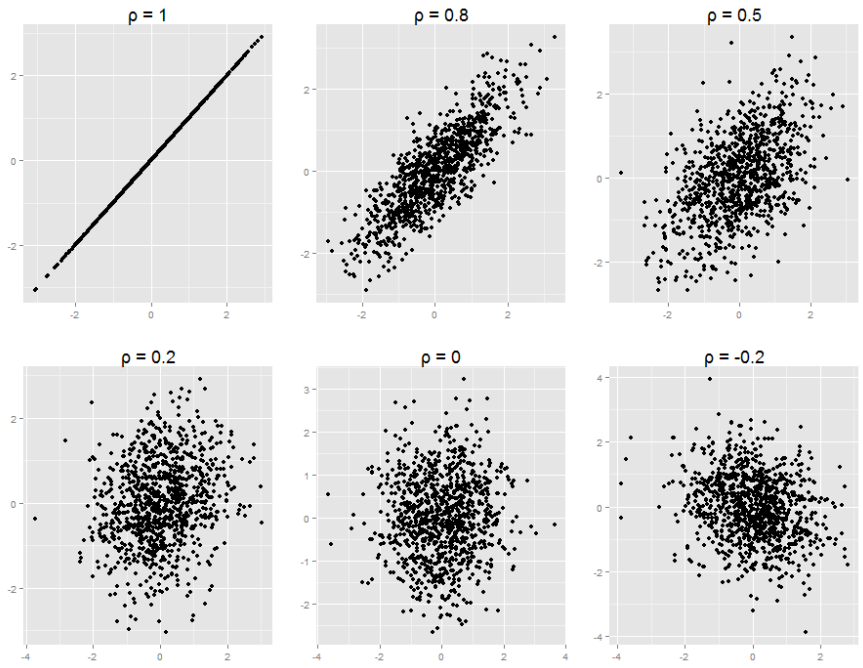


Figure 6 Scatter plots showing the strength of correlations when the correlation coefficient (ρ) takes on values from -0.2 up to 1

2.3.4 Cross correlations

Cross correlations are defined by cross correlation coefficients, between related MIPs (Kros et al., 2012a). For different MIPs, MIP_i and MIP_j, at the same location, the cross-correlations denoted as $\rho_{cc(i,j)}$, were obtained from census data or literature or expert knowledge. The cross-correlation pairs were selected based on previous research, census data and expert knowledge. The pairs of MIPs were grouped according to the method used to obtain the cross-correlation coefficient $\rho_{cc(i,j)}$ and are listed in Table 8. The strength of correlations is shown in Figure 6.

Table 8 MIPs for which cross-correlation are considered

MIP_i	MIP_j	$\rho_{cc(i,j)}$	Information source
FAO_AnimalProd	LivestockCountryTotal	0.9	Census data
FAO_NatCA_wheat	FAO_NatCP_wheat	0.81	Census data
FAO_NatCA_maize	FAO_NatCP_maize	0.81	Census data
FAO_NatCA_soybean	FAO_NatCP_soybean	0.81	Census data
FAO_NatCA_barley	FAO_NatCP_barley	0.81	Census data
FAO_NatCA_other	FAO_NatCP_other	0.81	Census data
CH4_EF_ManManage	Temperature	0.5	Previous research
FAO_LandAreas	LandCoverMap	0.5	Expert knowledge
N2O_manure_storage	N2_animal_EF	-0.2	Expert knowledge
NO_animal_EF	N2O_manure_storage	0.8	Expert knowledge
BMFac_Gras	BMFac_OtherAreable	0.5	Expert knowledge

1. Cross-correlation derived from census data

Some of the census data (including FAO data, and other annual census data) were estimated to be cross-correlated. The census data from past 10 to 15 years (depends on the data availability) were used to inspect these correlations. The calculation is illustrated using the example FAO_NatCA_wheat and FAO_NatCP_wheat.

FAO_NatCA_wheat includes annual crop area data for wheat for each country. The average crop area for wheat was calculated for each country i year j, as the average crop area $CropArea_{i,j}$. FAO_NatCP_wheat includes annual crop production data for wheat for each country. The average crop production for wheat was calculated for each country i year j, as the average crop production $CropProd_{i,j}$.

Then, for each year j, the $NatF_{i,i}$ and $CropProd_{i,j}$ at the same year are compared (see Table 9).

Table 9 Cross-correlation between FAO_NatCA_wheat and FAO_NatCP_wheat for year j

Country	FAO_NatCA_wheat	FAO_NatCP_wheat
1	CropArea1,j	CropProd1,j
2	CropArea2,j	CropProd2,j
…	…	…
i	CropAreai,i	CropProdi,j

Then the cross-correlation coefficient for year j was calculated as $\rho_{j,,}$ as well as the p value for significance. Subsequnetly, the average of the significant correlations were calculated as the cross-correlation coefficient between FAO_NatCA_wheat FAO_NatCP_wheat ρ_{cc}(FAO_NatCA_wheat, FAO_NatCP_wheat), shown as ρ_{cc} in Table 10. This ρ_{cc} is the average of the significant correlation coefficients (with p value lower than 0.05).

The strength of the correlation between FAO_NatCA_wheat and FAO_NatCP_wheat is shown in Figure 7. More examples can be found in Annex 1.

Table 10 Combing cross-correlations over years

year	Correlation coefficient	P value
2000	ρ2000	P2000
2001	ρ2001	P2001
…	…	…
2012	ρ2012	P2012
mean	ρcc	

2. Cross-correlation derived from previous research.

There is only one pair of MIPs in this group: the NO_animal_EF and N2O_manure_storage. This cross correlation was based on (Kros et al., 2012a) .

3. Cross-correlation defined based on expert knowledge

These pairs of MIs are selected based on expert knowledge. Due to the lack of data, it is not practical to assign specific cross-correlation coefficients to each pair. Thus, the approach of assigning classes to these pairs is applied. The approach has been described in the section 2.3.3 The spatial-correlation. The same classes and corresponding coefficients are used for cross-correlation. The reasoning for each pair is described below.

- N2O_manure_storage and N2_animal_EF

More N_2O (and NO) will lead to less N_2, but the correlation is not very strong. So a negative low cross-correlation was assumed.

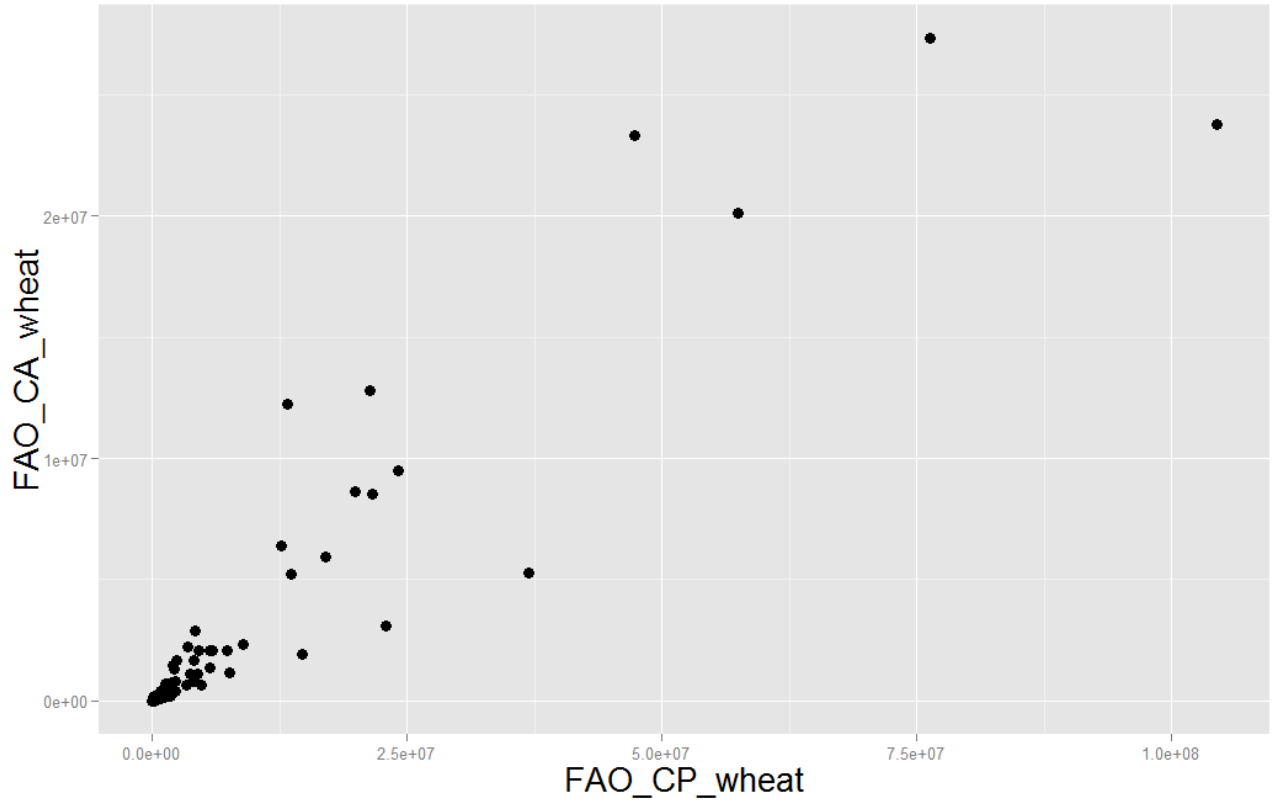


Figure 7 Scatter plot showing the cross correlation between FAO_NatCP_wheat and FAO_NatCA_wheat

- BMFac_Gras and BMFac_OtherAreable

Theis refers to the fraction of N of gross mineralization available on grass and on arable land being no grass. Since the same process is described, a medium positive correlation is assigned.

- CH4_EF_ManManage and Temperature

A medium positive correlation was assigned because CH_4 emission is temperature dependent.

- FAO_LandAreas and LandCoverMap

Satellite derived and FAOSTAT derived land use areas were, assignedto be medium positively correlated.

2.3.5 Constructing a correlation matrix

The cross-correlations and spatial-correlations were combined in a correlation matrix. For the same MIP, I and model developers assumed that A, B, C and D were different sub-national regions. A and B are from the same country, with ρ_{sub}. A and C are from different countries but in the same continents, with ρ_{nat}. A and D are from different continents, with ρ_{con}. The correlation matrix is constructed using 2 MIPs: MIP1 and MIP2 with a cross correlation coefficient ρ_{cc}. An example of constructed correlation matrix is shown in Figure 8.

		MIP1				MIP2			
		A	B	C	D	A	B	C	D
MIP 1	A	1	ρ_{sub1}	ρ_{nat1}	ρ_{con1}	ρ_{cc}	$\rho_{cc}*\frac{}{\sqrt{\rho_{sub1}*\rho_{sub}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{nat1}*\rho_{nat2}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$
	B		1	ρ_{nat1}	ρ_{con1}	$\rho_{cc}*\frac{}{\sqrt{\rho_{sub1}*\rho_{sub}}}$	ρ_{cc}	$\rho_{cc}*\frac{}{\sqrt{\rho_{nat1}*\rho_{nat2}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$
	C			1	ρ_{con1}	$\rho_{cc}*\frac{}{\sqrt{\rho_{nat1}*\rho_{nat2}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{nat1}*\rho_{nat2}}}$	ρ_{cc}	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$
	D				1	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$	$\rho_{cc}*\frac{}{\sqrt{\rho_{con1}*\rho_{con}}}$	ρ_{cc}
MIP 2	A					1	ρ_{sub2}	ρ_{nat2}	ρ_{con2}
	B						1	ρ_{nat2}	ρ_{con2}
	C							1	ρ_{con2}
	D								1

Figure 8 Building the correlation matrix

2.4 Defining the uncertain model outputs

According to the scope of this study, the uncertain model outputs were limited to CH_4, N_2O and CO_2 emissions from livestock sectors in Africa, Latin America and Europe. The GHG emissions include total emissions from the aforementioned continents, and emissions at product level and sector level. The products include cattle meat, cow milk, eggs, pig meat, poultry meat, sheep and goat meat, and sheep and goat milk. The sectors include broilers, laying hens, dairy cows, other cattle, pigs, other poultry, sheep, goats, horses, camels, turkeys and other animals. In order to show the possible cause of the uncertainties, the outputs also include the uncertainties of CH_4, N_2O and CO_2 from IPCC

categories ??shown in Table 11. All emissions were converted to CO2-eq using IPCC global warming potentials (GWP) (Lesschen et al., 2011b), which are for CH_4 and N_2O 34 and 298 times the GWP of CO_2, respectively (based on IPCC AR5, 2013 (IPCC, 2013) for a 100 years times horizon).

Table 11 IPCC categories considered in this study

Emission	code	Process
CH_4	CH4_rice	CH_4 emission from rice cultivation
CH_4	EntericFermentation	CH_4 emission from enteric fermentation
CH_4	ManureManagementCH4	CH_4 emission from manure management
CH_4 and N_2O	ManureManagement	GHG emissions from manure management
N_2O	CropArea	Crop areas for soil N_2O emission (direct + indirect) allocation
N_2O	ManureManagementN2O	N_2O emission from manure management
N_2O	N2O_ByProducts	N_2O emission from by product feeds (e.g. citrus pulp)
N_2O	N2O_grazing	N_2O emission from grazing
N_2O	N2O_soil	Direct and indirect N_2O emission
CO_2, N_2O	FertilizerProd	GHG emissions from fertilizer production (tonnes CO_2-eq)
CO_2	FuelUse	CO_2 emission from fuel use
CO_2	PesticideUse	CO_2 emission from pesticide use

2.5 Quantifying the uncertainty in the model outputs

The overall approach of obtaining the output uncertainty is illustrated in Figure 9. The methodology is explained in the following parts of this section, The results are shown in section 3.1. The R scripts are provided in Annex 6.

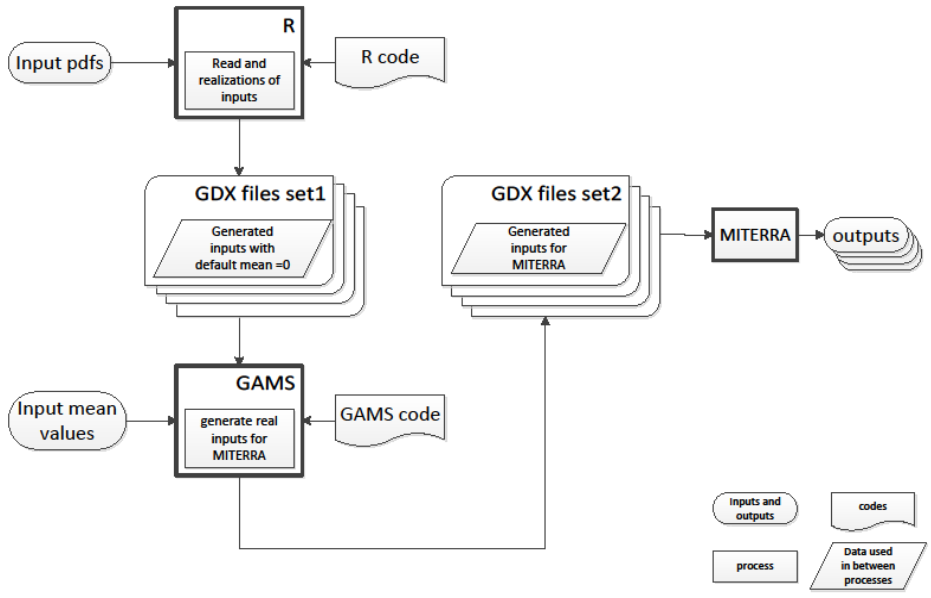


Figure 9 Process for quantifying model output uncertainty

2.5.1 Developing R-scripts for generating multiple realizations of model inputs and parameters

The Monte Carlo (MC) approach is used to generate multiple realizations of model inputs and parameters based on the pre-defined pdfs. The concept of MC is to compute repeatedly model outputs with multiple

input realizations, at model-required scales, sampled from their pdfs using stochastic spatial simulation (Truong, 2009). The realizations were generated with the statistical software environment R.

Due to the memory limits of the computer, the realizations of model inputs and parameters were divided into three different spatial levels: sub-national (containing MIPs at both sub-national level and national level), continental and generic levels, and carried out separately.

1. Sub-national level

The MIPs at sub-national and national level are all simulated at sub-national levels. For the national level MIPs, it is realized by setting the sub-national spatial-correlation coefficients to 1. After the realizations at sub-national level, the national MIPs are aggregated to national level.

The R scripts were developed based on the package "**sgsm**", which was developed for a previous research on the INTEGRATOR model (Kros et al., 2012b). After testing, it was found out that this package can only re-produce the imposed cross-correlation when the correlation matrix is not too big. Since the number of MIPs cannot be reduced, and the defined outputs are for three separated continents, this simulation at sub-national level was carried out for each continent separately. As a result, the spatial correlation at continental level is not taken into account for sub-national level and national MIPs. However, in the previous steps, these continental spatial correlations were mostly estimated to be 0, or in other cases, very small. This separating of simulation to each continent is expected to have negligible effect on the final outcomes. The process for deriving MIPs at sub-national level is shown in Figure 10.

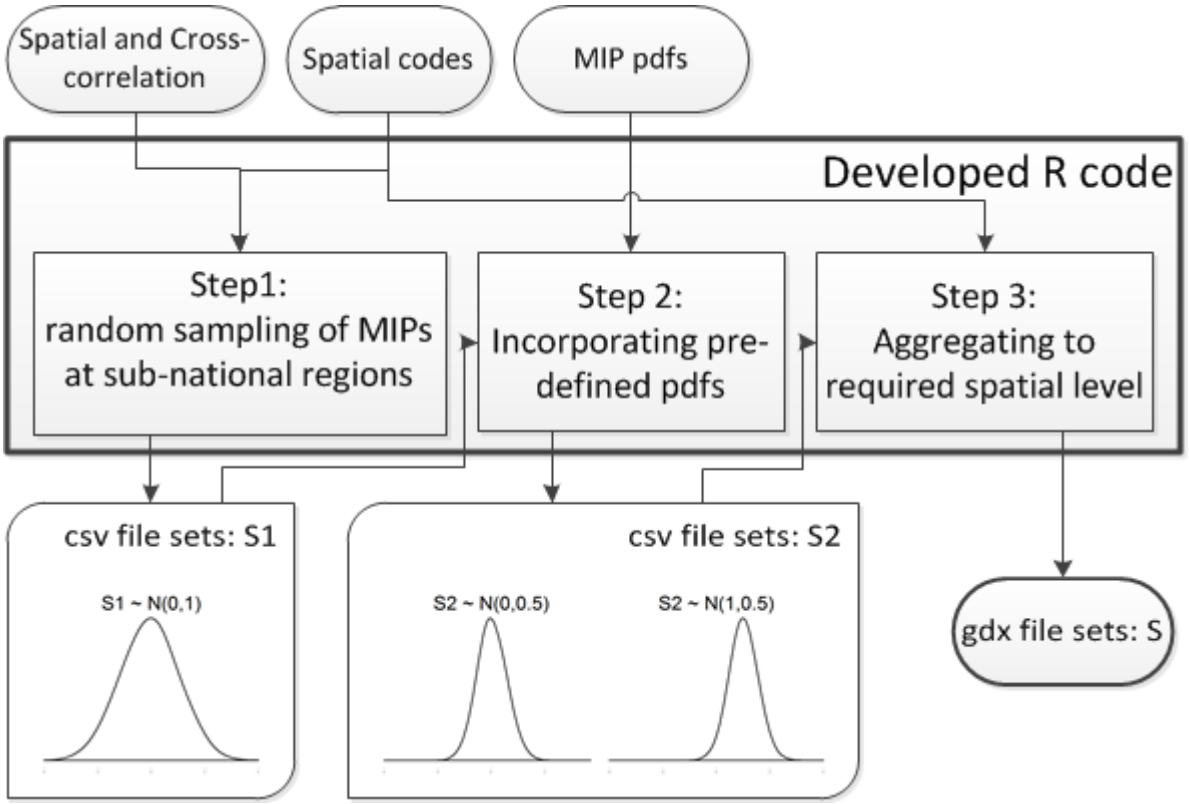


Figure 10 Flowchart showing the process of MC simulation at sub-national level

The developed R scripts are the same for each continent, including three steps:

Step 1: random sampling of MIPs at sub-national regions

In this step, the required inputs are the spatial codes, the cross correlation matrix, the spatial-correlation coefficients and the code of MIPs. A cross- and spatial-correlation matrix is constructed within the package "**sgsm**" (as described in 2.3.5). Then using Monte Carlo algorithm, each MIP was sampled from a standard normal distribution. The Monte Carlo algorithm has the feature of short-term memory. After enough MC runs, the simulation is able to produce realizations with the same pdf as designed. The

number of MC runs were determined as 1000, based on previous study (Kros et al., 2012b) and tested with cdfs. Figure 11 shows the cumulative probability function plots of the randomly generated model inputs or parameters with different number of MC runs.

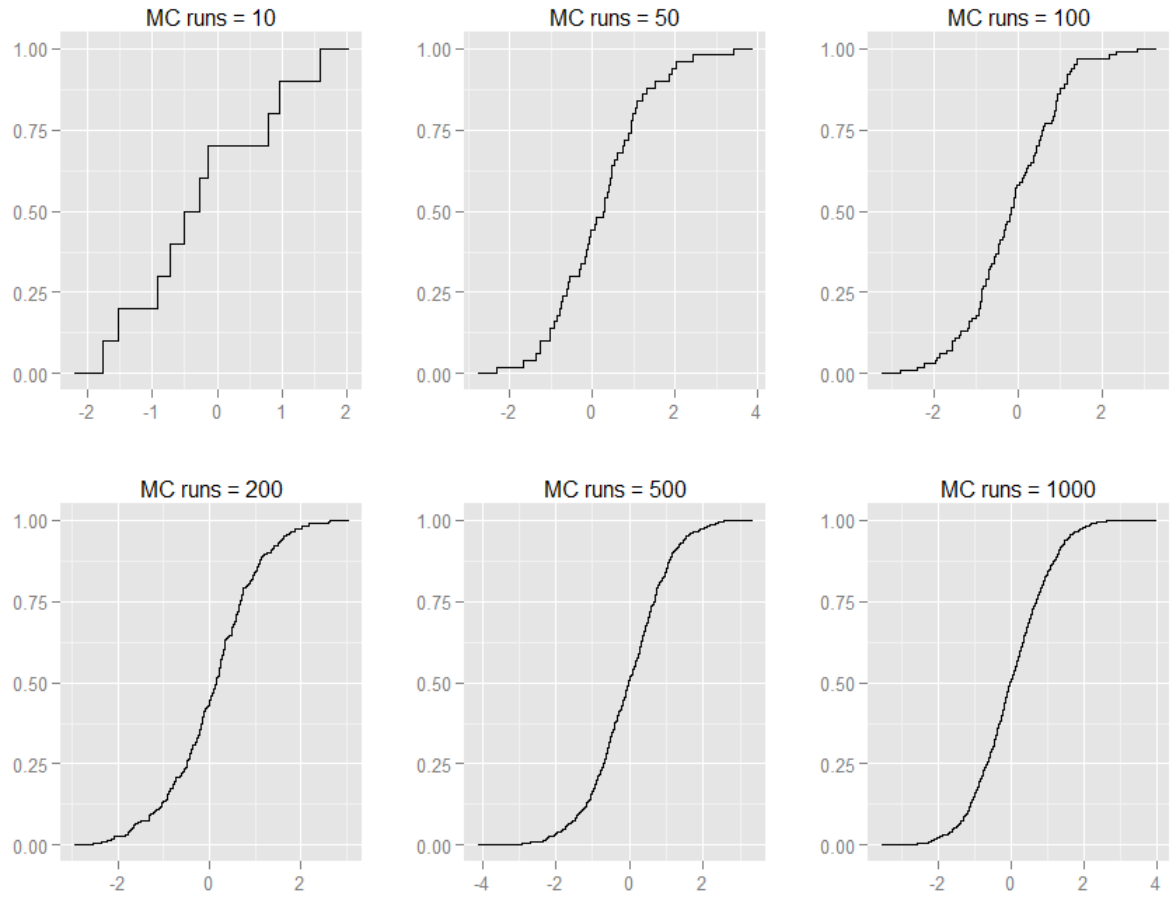


Figure 11 Plot of cdfs for different MC runs

The plots with more than 200 runs possibly produce roughly a smooth curve. With 1000 runs the simulated curve is very smooth. Thus, in this study, we choose 1000 MC runs to analyse the uncertainty in the model outputs to obtain an accurate results. Increasing from 200 runs to 1000 runs will only slightly inprove the performance, but will significantly increase the modelling time. For uncertainty contribution analysis, we only need the relative contritution instead of exact results. Thus, for uncertianty analysis, 200 MC runs were used. The representativeness is tested and the results are shown in section 3.1.

The outcomes of this step are realizations with expected value equal to 0 and unite variance (variance equal to 1), denoted as S_1.

Step 2: post processing of realizations with pre-determined pdfs

In this step, S_1 was post-processed with the pre-determined pdfs of MIPs. The inputs of this step were the distribution type (normal distributed or lognormal distributed) and the variance (in terms of cv or sd, depends on the distribution type). The outcomes of this step are denoted as S_2 and the calculation are explained for each distribution type:

For a normally distributed MIP, the cv of this distribution (cv_n) is used to represent the uncertainty of this MIP. The post process is written as:

$$S_2n = 1 + S_1 * cvn \qquad (12)$$

The result S_{2n} has a normal distribution with and expectation as 0 and cv_n as its cv.

For a lognormally distributed MIP the cv of this lognormal distribution, i.e. the sd of its log-transformed distribution (sd_n) was used to represent the uncertainty of this MIP. The post process is written as:

$$S_{2log} = S_{1*} sd_n \quad (13)$$

The result S_{2log} has a normal distribution with and expectation as 1 and sd_n as its σ.

Step 3: post processing of realizations to required spatial levels. For all the S_2`s for MIPs at national level, an aggregation of sub-national level to national level was applied. For each MIP, the average of all the $S_2$`s within one country is used as the final simulation (S) for this country. The results of this step are stored in gdx files.

2. Continental level

For the continental MIPs, a different approach was taken. This was achieved by using the "**rmultnorm**" function (Multivariate Normal Random Number Generator). The generator is based on the multivariate normal distribution, which is defined as: a multivariate random vector M with k elements, M = M (M_1,M_2,…M_k), that follows a normally distribution is denoted as M ~ N_k(μ, R), where μ = mean vector and R = variance-covariance matrix (k × k). The R matrix, with all combinations of M_i and M_j pairs, is calculated as:

$$R[i,j] = \rho_{[i,j]*} \sigma_i * \sigma_j \quad (14)$$

Where $\rho_{[i,j]}$ is the correlation coefficient matrix for M_i and M_j, and σ_i , σ_j are the sds of M_i and M_j respectively. Similar to the approach for sub-national level MIPs, the first step is to generate S_1 ~ N (0,1). The second step was to post-process this S_1 to S based on the pre-defined pdf.

For step one, the correlation matrix R had to to be constructed. Due to the memory limit, this step was also carried out separately for each continent. Thus, for each MIP, M_i and M_j represent MI at location i and j:

$$R_{[i,j]} = \rho_{[i,j]*} \sigma_{Mi*} \sigma_{Mi} \quad (15)$$

With the distribution MIP_1 ~ N (0,1), σ_{Mi} = 1, this formula can be written as

$$R_{[i,j]} = \rho_{[i,j]} \quad (16)$$

Thus, the R matrix for MIP is equal to the spatial-correlation matrix of MIP, which is represented in Figure 12

	i	...	j
i	1		ρ_{con}
...	...	1	...
j	ρ_{con}	...	1

Figure 12 Spatial correlation matrix for one continental MIP

The outcome of this step was random sampled S_1 ~ N (0,1) for each MIP at each continent. The next step is the same as the step 2 in sub-national level. The overall outcome is S with designed uncertainty information at continental level.

3. Generic level

MIPs at this level remain uniform through all the spatial levels. The same function "**rmultnorm**" is used. However, only one variance-covariance matrix R is constructed for all MIs. Since there is no spatial level needed, the R matrix equals to the cross-correlation matrix of all MIPs at generic level. Figure 13 is an example of this matrix

	MIP1	MIP2	…	MIPn
MIP1	1	ρcc12		ρcc1n
MIP2		1		ρcc2n
…			1	
MIPn				1

Figure 13 Cross correlation matrix for generic MIPs

The rest of the calculation is the same as in the continental level. The overall outcome is also S with designed uncertainty information.

2.5.2 Running the MITERRA-Global model for the generated input and parameter simulations in batch mode and storing its output simulations

The results of 2.5.1 (denoted as S) are stored in gdx format and are used as uncertain input uncertainty for MITERRA-Global.

As a first step, this S was combined with the default values of MIPs (μ_{def}) to get the "real inputs" (MCin`s) for MITERRA-Global. Depends on the distribution type of the MIP, one of the following two formulas is used to obtain the MCin:

For a normally distributed MIP, the MCin is the product of μ_{def} and S_n:

$$MCin = \mu_{def} * S_n \qquad (17)$$

The result is a normally distributed MCin ~ N (μ_{def}, σ_n^2), where $\sigma_n / \mu_{def} = cv_n$.

For a lognormally distributed MIP, the log-transformed MCin is the sum of log-transformed μ_{def} and S_{log}, so the MCin is calculated as:

$$MCin = \exp(\log(\mu_{def}) + S_{log}) \qquad (18)$$

The result is a lognormally distributed MCin ~ Λ (μ^*, σ^{*2}), with $cv^* = sd_n$.

These MCin`s are model inputs with expected value equal to the defaults and the variance determined from previous research.

The default MITERRA-Global calculation use these MCin`s as model input data. The model was run for 1000 times, same as the MC runs of the MIP simulation. The outputs from these 1000 runs were stored in gdx files.

2.6 Analysing the contribution of individual uncertainty sources to the output uncertainty

The MIPs are further grouped to nine groups (

Table 12). The MIPs are grouped in such a way what there is no correlation between MIPs from different groups. A new MC simulation was used to detect the uncertainty contribution of each group.

Two methods were used to analyse the uncertainty contribution: one group at a time method and winding stairs scheme. They will be explained separately in the following of this section. The results of this comparison are shown in section 3.2.1.

Table 12 Grouped MIPs

Code	Description	Including e.g. (group for all MIPs see Annex 2)
LAD	Livestock activity data	livestock numbers and production
CAD	Crop activity data	Crop areas and production
OAD	Other activity data	Fertilizer consumption, pesticides, etc.
BFD	Biophysical data	climate, soil data
EFC	CH4 emission factor	EF manure management and enteric fermentation
EFN	N emission factors	all N emission factors (including leaching and runoff)
CPA	Crop parameters	N content, N index, etc., grass correction
LPA	Livestock parameters	N excretion, manure system usage
OPA	Other emission factors and parameters	CO_2 emission factor, fertilizer composition

2.6.1 One group at a time

The MC run times for this step was set to 200 due to the limit of run times. Since the simulation process is time consuming, it is not practical to use 1000 runs for each group (i.e. 10000 runs in total). From the test (Figure 11) we concluded that with 200 runs the MC simulation was able to produce a satisfactory result. With the increase of the MC runs the performance only slightly improved but the time required performing the runs increased dramatically. Thus, in the end 200 MC runs were chosen for the uncertainty quantification.

For the first 200 runs, all the MIPs were treated uncertain. It is the same approach used in 2.5. For each other 200 runs, only one out of nine group was randomized. The other eight groups remain constant (using the default value stored in MITERRA-Global database). The outputs were analysed with the method analysis of variance. The strategy is illustrated in Table 13.

Table 13 One group at a time scheme used for analysing uncertainty contribution

Run numbers	Group [1)]									Variance
1~200	**LAD**	**CAD**	**OAD**	**BFD**	**EFC**	**EFN**	**CPA**	**LPA**	**OPA**	Var_{all}
201~400	**LAD**	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA	Var_{LAD}
401~600	LAD	**CAD**	OAD	BFD	EFC	EFN	CPA	LPA	OPA	Var_{CAD}
601~800	LAD	CAD	**OAD**	BFD	EFC	EFN	CPA	LPA	OPA	Var_{OAD}
801~1000	LAD	CAD	OAD	**BFD**	EFC	EFN	CPA	LPA	OPA	Var_{BFD}
1001~1200	LAD	CAD	OAD	BFD	**EFC**	EFN	CPA	LPA	OPA	Var_{EFC}
1201~1400	LAD	CAD	OAD	BFD	EFC	**EFN**	CPA	LPA	OPA	Var_{EFN}
1401~1600	LAD	CAD	OAD	BFD	EFC	EFN	**CPA**	LPA	OPA	Var_{CPA}
1601~1800	LAD	CAD	OAD	BFD	EFC	EFN	CPA	**LPA**	OPA	Var_{LPA}
1801~2000	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	**OPA**	Var_{OPA}

1) Regular font: parameters of the group are fixed to the default value (generally the mean)
Bold font: parameters of the group are treated as uncertain

The variance in Table 13 indicates the variance of the 200 outputs. The variance from the first 200 runs, Var_{all}, is used as the reference to calculate the relative contribution of each group to the overall variance in the model outputs. The relative contribution of each MIP group to model output uncertainty was expressed as percentage contribution of the variance of a group to the overall variance ($PVAR_{group}$):

$$PVAR_{group} = (Var_{group}/Var_{all})*100\% \quad (19)$$

2.6.2 Winding stairs scheme

The winding stairs scheme was adapted from (Jansen et al., 1994). In total 1000 MC runs were used in this step. For the first 100 runs, all groups were set to uncertain and randomly sampled. For the following each 100 runs, one group is copied from the first 100 runs, and the other eight groups are sampled randomly. The scheme is illustrated in Table 14.

Table 14 Uncertainty contribution winding stairs scheme

Run	Nr.	Group[1]								
1~100	W_0	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
101~200	W_1	**LAD**	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
201~300	W_2	LAD	**CAD**	OAD	BFD	EFC	EFN	CPA	LPA	OPA
301~400	W_3	LAD	CAD	**OAD**	BFD	EFC	EFN	CPA	LPA	OPA
401~500	W_4	LAD	CAD	OAD	**BFD**	EFC	EFN	CPA	LPA	OPA
501~600	W_5	LAD	CAD	OAD	BFD	**EFC**	EFN	CPA	LPA	OPA
601~700	W_6	LAD	CAD	OAD	BFD	EFC	**EFN**	CPA	LPA	OPA
701~800	W_7	LAD	CAD	OAD	BFD	EFC	EFN	**CPA**	LPA	OPA
801~900	W_8	LAD	CAD	OAD	BFD	EFC	EFN	CPA	**LPA**	OPA
901~1000	W_9	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	**OPA**

1) Regular font: parameters of the group are sampled randomly
Bold font: parameters of the group are copied from the first 100 runs

Table 14 shows that each 100 runs are grouped and numbered from W_0 to W_9. In W_0 all model inputs are randomly sampled. In W_9 group 1 to group 8 are randomly sampled, group 9 is copied from the first 100 runs (W_0).

The contribution of group G_i was calculated as:

$$\text{Perc\_G}_i = \frac{0.5 * cov(W_0, W_i)}{var(W_0)} * 100\% \qquad (20)$$

The $cov(W_0,W_i)$ is the covariance between run W_0 and run W_i. The $var(W_0)$ is the variance of run W_0. The $var(W_0)$ is the variance of run W_0. Thus, the uncertainty contribution of group G_i was quantified as half of the covariance of two runs over the total variance of the completely random run.

3. Results

3.1 Uncertainty in model inputs and parameters

In this section, I discuss a) how the test for the required sample size of simulation took place by means of cumulative frequency curves (section 3.1.1) and b) how the test for the reproduced cross correlation took place (section 3.1.2).

3.1.1 Representativeness of model input and parameter uncertainty

We have chosen model inputs and parameters from each of the four spatial levels to test the representativeness of the sampling methods. For each spatial level, the model inputs and/or parameters with the highest and lowest designed uncertainty are chosen. This is to test if the designed uncertainty will influence the representativeness. The designed uncertainty refer to the model input and/or parameter uncertainty quantified in section 2.3. Note that because for all the model inputs and parameters at continental level, the designed sd`s are all 0.25, only one model input at continental level is chosen to test the representativeness.

The seven MIPs, the designed uncertainty and the produced uncertainty are listed in Table 15:

Table 15 Selected MIPs to test the representativeness and the produced uncertainty

Code	Distribution type	Spatial level	sd*			
			designed	100 runs	200 runs	1000 runs
Fqatm	Lognormal	generic	0.25	0.24	0.25	0.25
EF4	Lognormal	generic	0.82	0.84	0.81	0.81
ManureSU_Burned	Normal	continental	0.25	0.25	0.25	0.25
bNumAniRAINS	Normal	national	0.10	0.09	0.09	0.10
ShareFeed	Normal	national	0.25	0.24	0.24	0.24
LandCoverMap	Normal	sub-national	0.10	0.10	0.10	0.10
N2O_grazing	Lognormal	sub-national	0.57	0.58	0.56	0.56

* *The uncertainty is presented in terms of sd. For MIP which is log-normally distributed, this is the sd of its normally transformed distribution.*

Table 15 shows that for most of the MIPs the sd can be reproduced, even with only 100 runs. The results from 200 runs are the same as the results from 1000 runs. This indicates that both with 200 runs and with 1000 runs the sd can be reproduced.

Figure 14 and Figure 15 show the simulated results. Figure 14 shows the simulated MIPs at generic level and continental level, which were carried out using standard R package. Figure 15 shows the results of simulated sub-national and national parameters. This simulation was carried out using “sgsm” package (details see section 2.5).

The results are shown as cumulative density functions of 100, 200 or 1000 runs. Generic MIP only has one simulated result for one run of simulation. All of the simulated within the 100, 200 or 100 runs are used for the generic parameters to produce the plots in Figure 14 and Figure 15. For national parameters, each country gets its own simulated result within each run. Figure 14 and Figure 15 show only the plots for the simulations for one country. The medians are also given in the cdf plots. Note that all of these MIPs are sampled from normal distribution. The MIPs with designed lognormal distribution will be transformed to lognormal distribution after this step. Thus, the simulations for MIPs with lognormal distribution at this step are supposed to have 50% equally distributed around 1. Simulations for MIPs with normal distribution are designed to distributed equally around 0. The representativeness can be derived from the smoothness of the cdf curve and the difference between the median from 1 or 0.

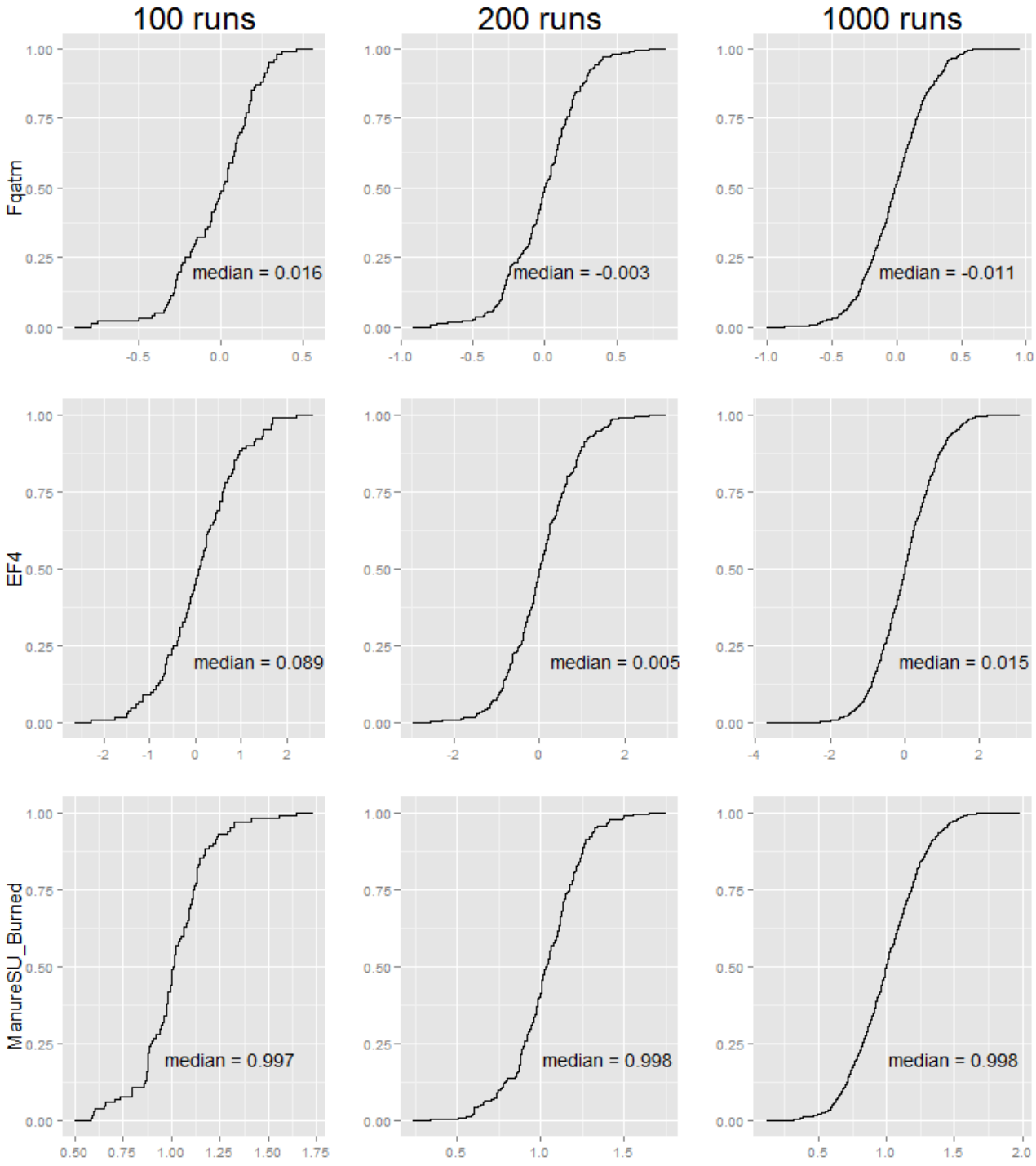


Figure 14 cdf plot for simulated MIPs using standard R package

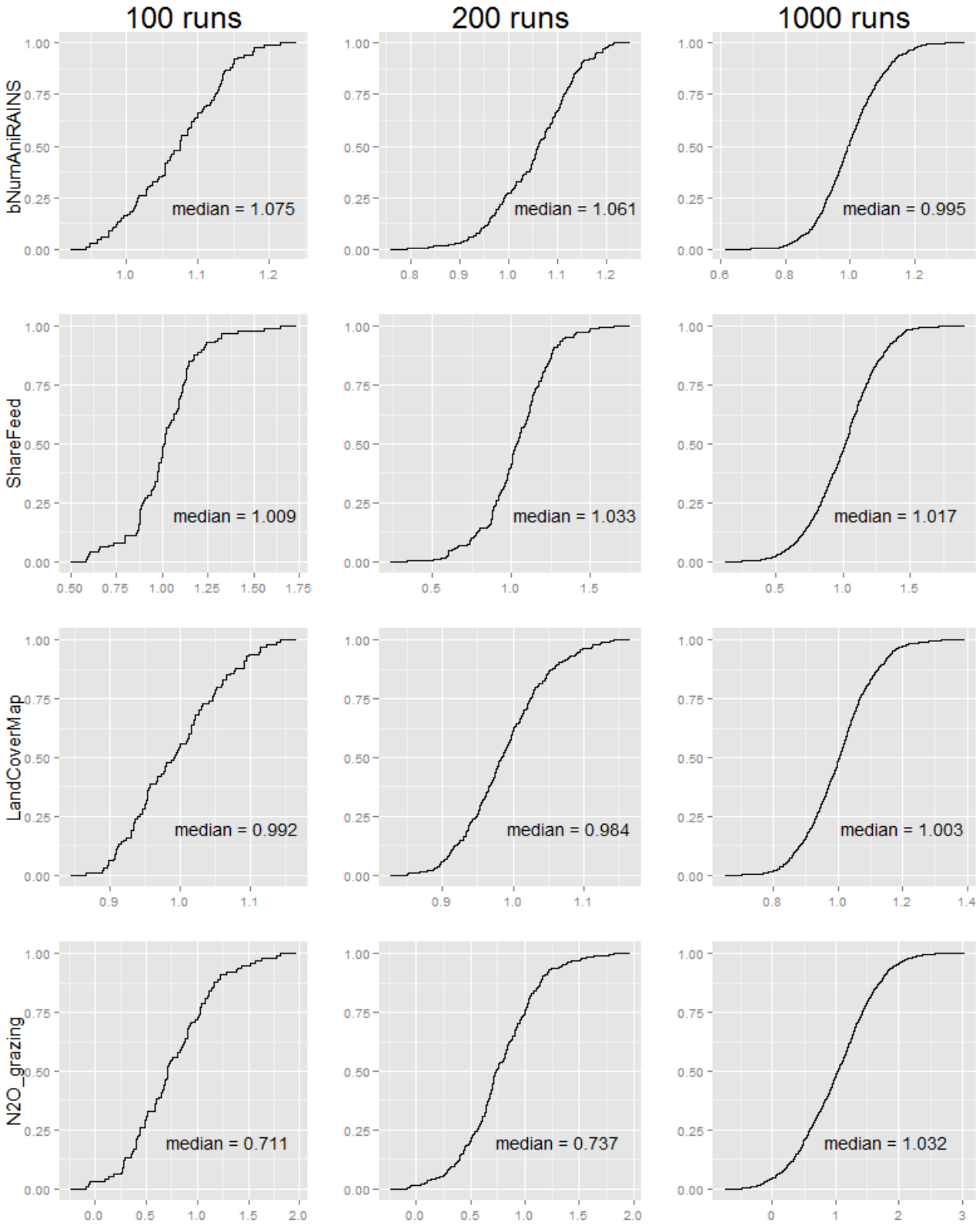


Figure 15 cdf plot for simulated MIPs using "sgsm" package

For MIPs simulated with standard R package (results presented in Figure 14), 200 runs have similar performance compared with 1000 runs. Using 200 runs is sufficient to produce the designed normal distribution. There is a clear improvement from using 100 runs to using 200 runs. But there is no clear improvement observed from using 200 runs to using 1000 runs.

For MIPs simulated with "sgsm" package, however, Figure 15 shows that the designed normal distribution can be reproduced for national parameters (medians close to 1), but might not be reproduced for sub-national parameters (in this example median can be clearly different from 1). Clear improvement can be observed for sub-national parameters when increasing the simulation MC runs. The medians become much closer to the designed 1 and the curve become very smooth. Thus, the 1000 runs used for uncertainty quantification is sufficient to reproduce the uncertainty information and the distribution type. The 200 runs used for uncertainty analysis might not be sufficient for sub-national parameters. This deficiency might lead to a bias when quantifying the uncertainty contribution. The method used to quantify uncertainty contribution was analysis of variance. This bias is indeed reflected in the results of uncertainty analysis. For more discussion please see section 3.3.1.

3.1.2 Representativeness of the cross-correlation

For the cross-correlation, the following pair is detected: FAO_CA_wheat (area for growing wheat in each FAO country) and FAO_CP_wheat (wheat production in each FAO country). The result is shown in Figure 16.

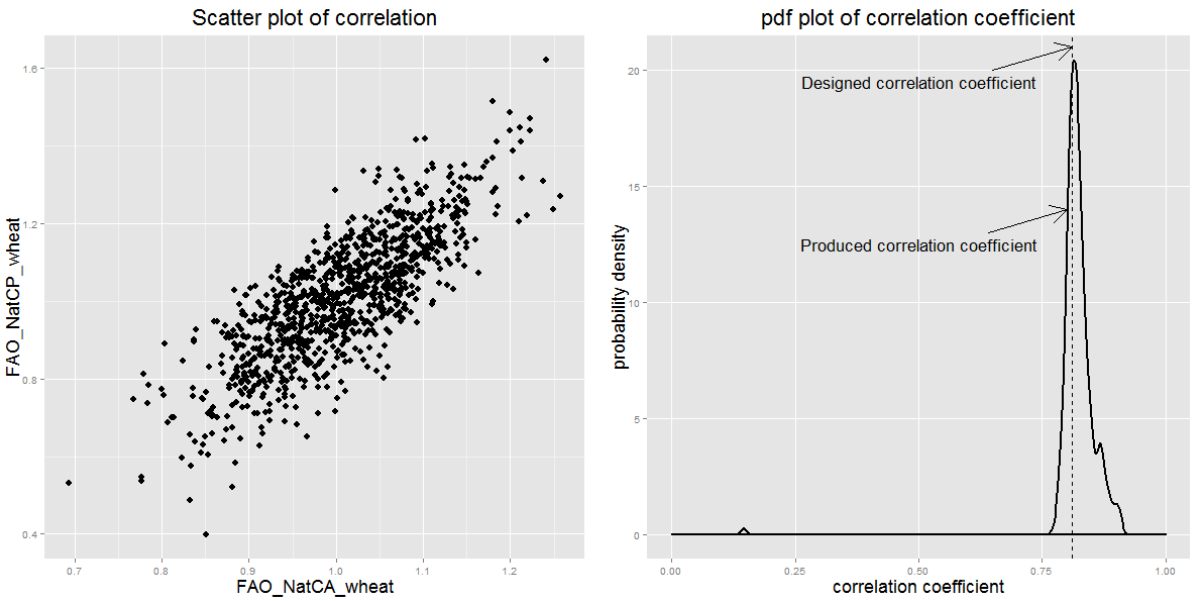


Figure 16 Scatter plot and pdf plot for produced cross-correlation

The scatter plot shows that there is a positive correlation between the model inputs FAO_CA_wheat and FAO_CP_wheat. The pdf plot shows that the probability density function of the correlation coefficients between the model inputs FAO_CA_wheat and FAO_CP_wheat. The designed correlation coefficient is 0.81 and the produced correlation coefficient has the highest probability to be 0.82, with a small chance to be around 0.15. Almost all of the produced correlation coefficients are around 0.75 to 0.92. This indicates that the methods used can re-produce the designed cross-correlation.

3.2 Uncertainties in GHG emissions from Africa, Latin America and Europe

In this section, the results of quantification of the uncertainty in the model outputs, according to methodology 2.5, are presented. The outputs including the overall greenhouse emissions from livestock sectors, the GHG emissions for different livestock sectors, the GHG emissions for different livestock products, and the GHG emissions from key categories used by IPCC. The emissions as well as their uncertainty will be discussed. The results are presented in graphs in this section, and presented with tables in Annex 3. The corresponding R are given in Annex 7.

3.2.1 Overall GHG emissions

The total GHG emission and the CH_4, N_2O, CO_2 emissions from Africa, Latin America and EU27 are presented in Figure 17. Their uncertainty (in terms of cv) is presented in Figure 18. The emission is

expressed with bar plots with error bar indicating an 95% confidence interval. The cv is visualized with bar plot.

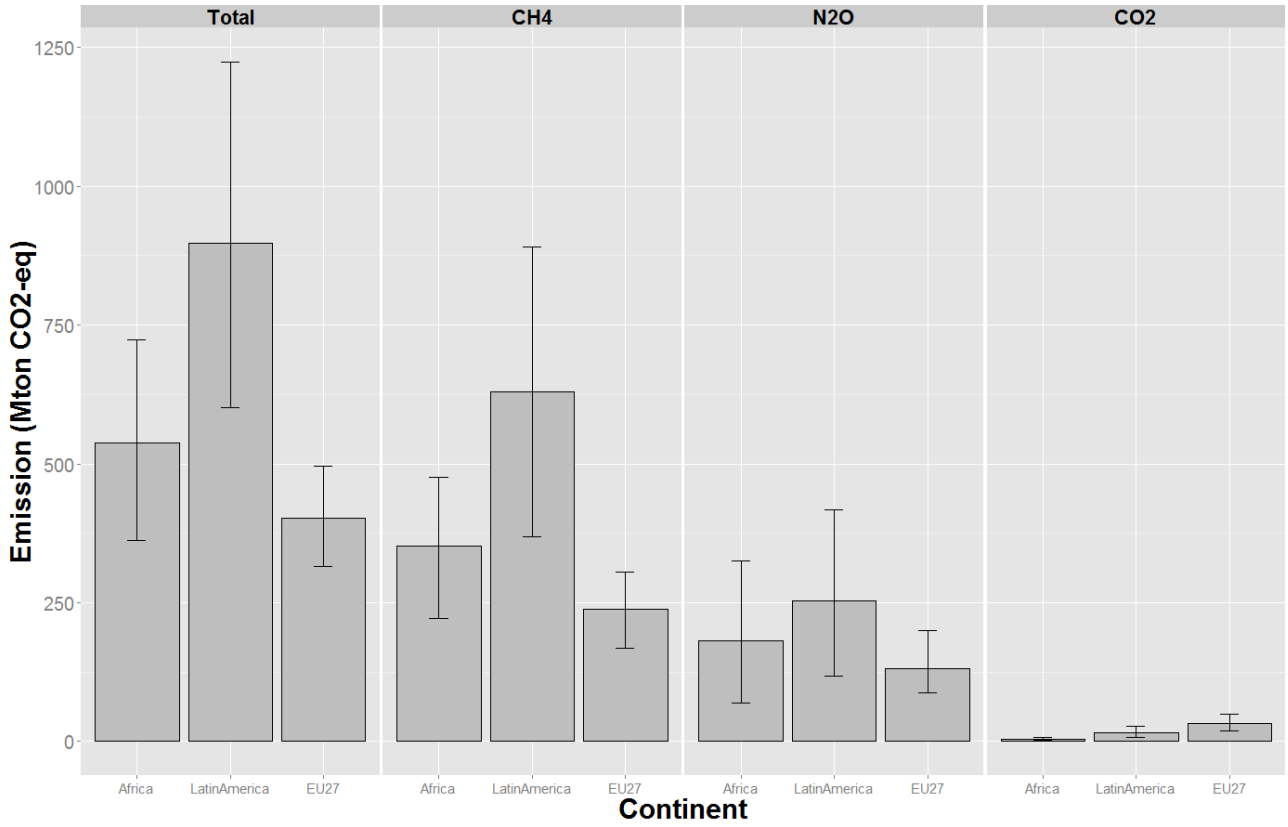


Figure 17 Overall GHG emissions from livestock production

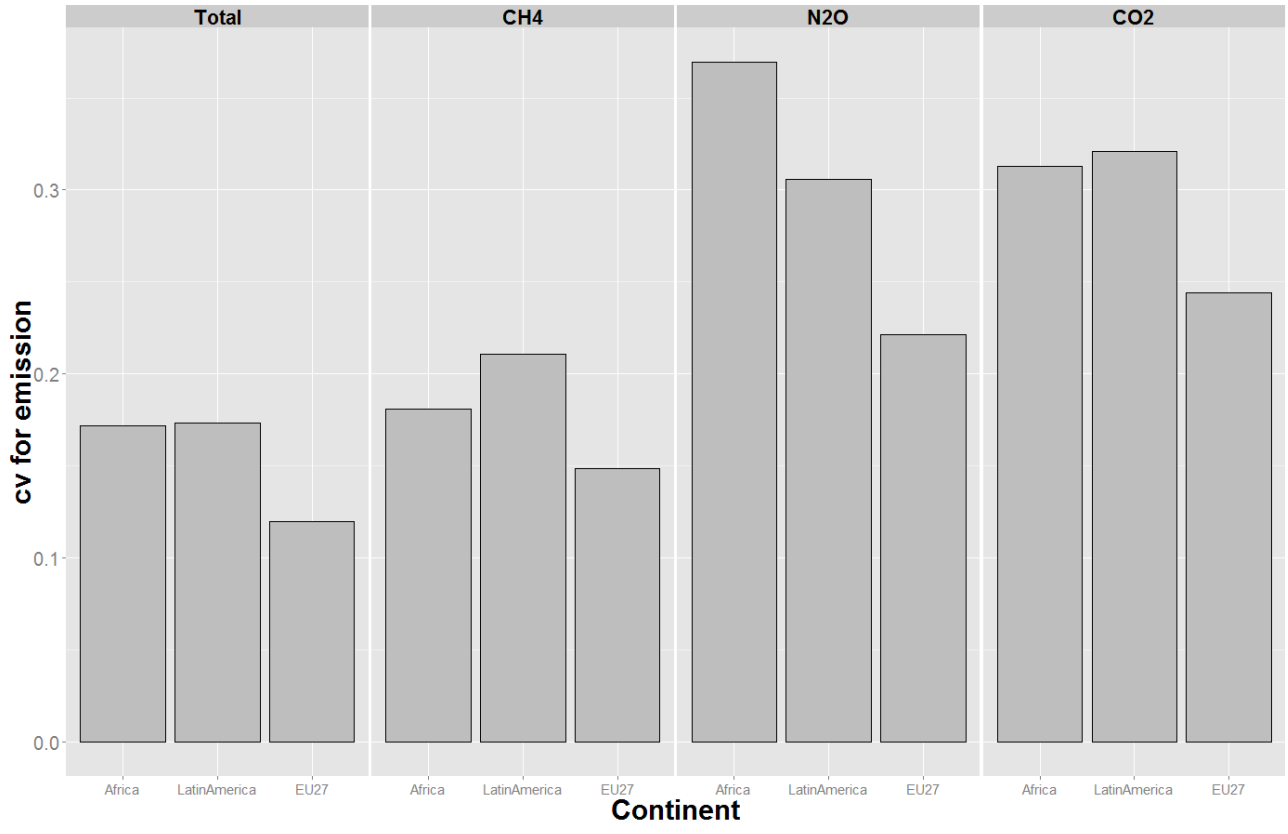


Figure 18 Uncertainty (in terms of cv) of overall GHG emissions from livestock production

Figure 17 shows that the total GHG emission is the highest in Latin America and the lowest in EU27. This trend remains the same for CH_4 emission and N_2O emission. But for CO_2 emission, it is the highest in EU 27 and lowest in Africa.

Figure 18 provide the uncertainty information of the GHG emissions in the three continents in terms of cv. It shows that the output uncertainty ranges from 0.15 to 0.37 for the three greenhouse gases, the cv increases in the following direction: Total < CH_4 < N_2O,CO_2. In terms of continents, the cv increases in the following direction: EU27 < Africa, Latin America.

3.2.2 GHG emissions from livestock sectors

The GHG emissions from different livestock sectors are presented in Figure 19. The uncertainties are of the livestock sectors are presented in Figure 20 in terms of cv. The emission is expressed with bar plots with error bar indicating an 95% confidence interval. The cv is visualized with bar plot.

Figure 20 represents the result of emission per GHG per country per livestock sector and the result of the corresponding uncertainty in terms of cv. The uncertainty ranges from 0.12 to 0.42.

For total emission, the plots show that Latin America has the highest emission for most of the sectors except for pigs. EU 27 has the lowest emissions except for pigs (Figure 19). The sector pigs has the highest emission in EU 27. The uncertainty, however, is in general higher in Africa and Latin America, lower in EU 27 (Figure 20). Africa has the highest uncertainty in turkey (0.33) and lowest in dairy caws (0.17). Latin America has the highest uncertainty in turkey (0.32) and lowest in dairy caws and sheep (0.18). EU 27 has the highest uncertainty in the sector other poultry (0.32) and lowest in dairy cattle (0.21).

For CH_4 emission, Figure 20 shows similar pattern for Africa and Latin America. Except for goats, horses and camels, the other sectors have similar uncertainties. The cv's are around 0.18 for Africa and 0.17 to 0.2 for Latin America. Goats, horses and camels have higher uncertainties. Goats have the highest uncertainty, 0.33 for both continents. Goats also have the highest uncertainty in EU 27, a cv of 0.3.

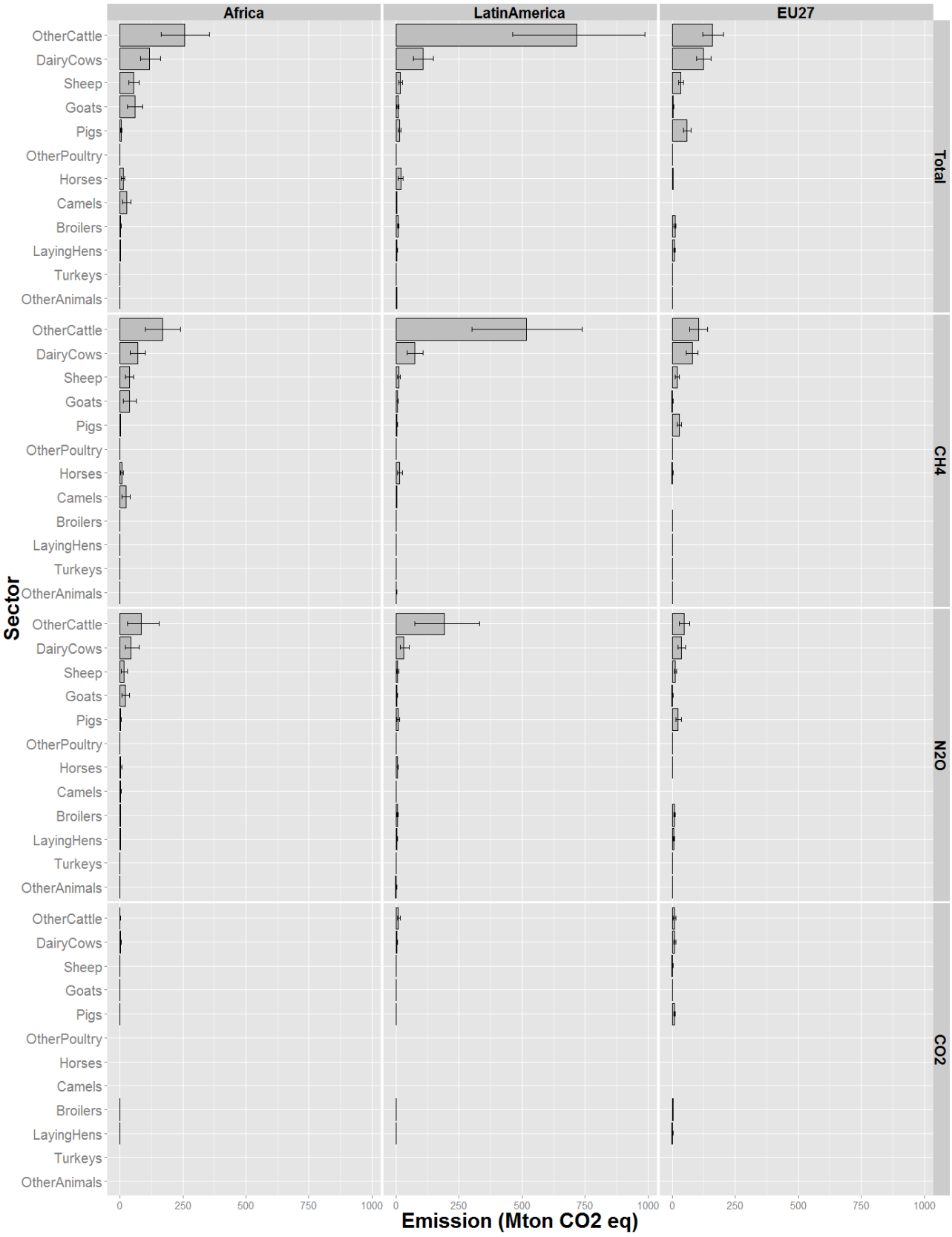


Figure 19 GHG emissions from livestock sectors

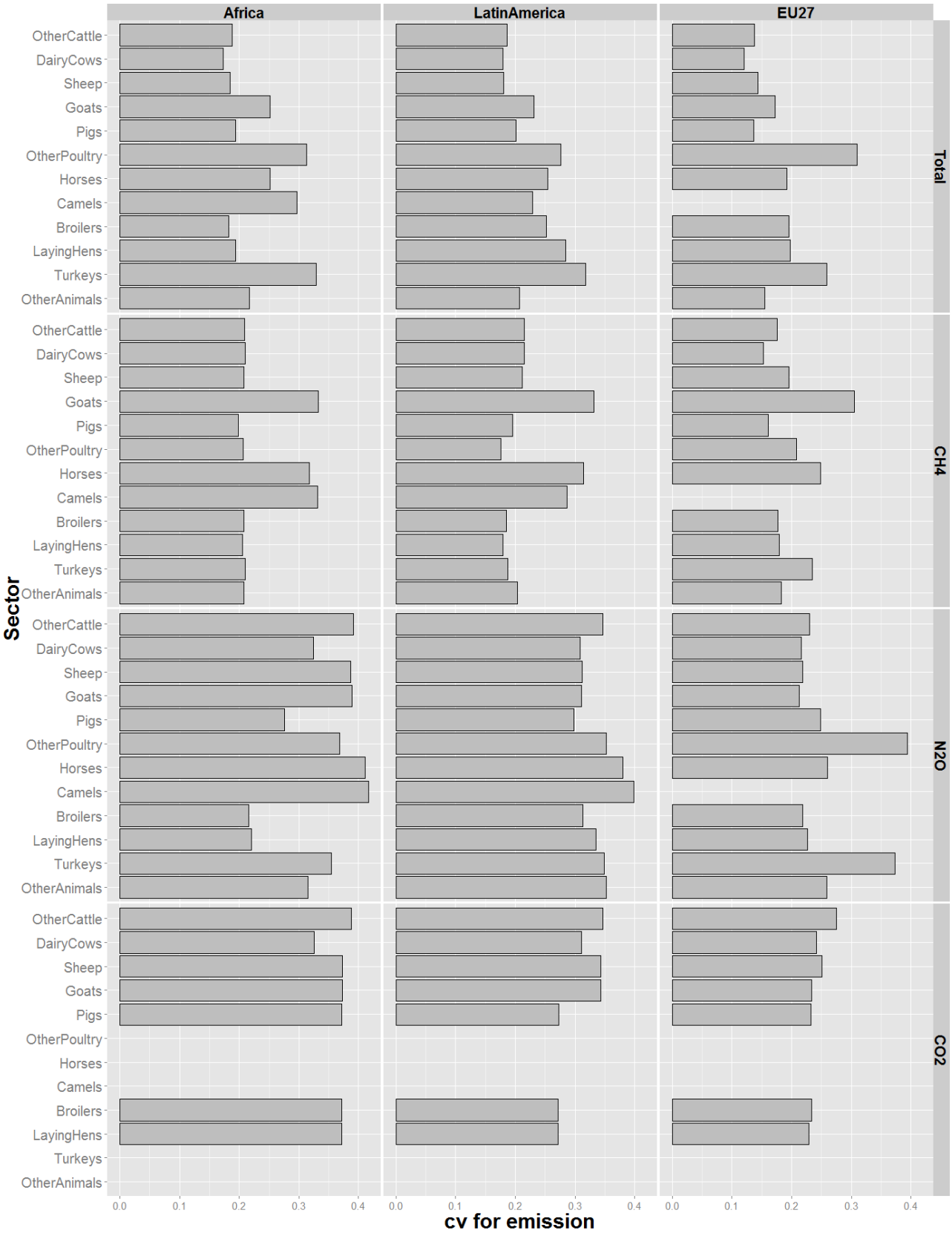


Figure 20 Uncertainty (in terms of cv) of GHG emissions from livestock sectors

3.2.3 GHG emissions from livestock products

The GHG emissions from different livestock products are presented in Figure 21. Their uncertainty is presented in Figure 22. The uncertainties are quantified in terms of cv. The emission is expressed with bar plots with error bar indicating an 95% confidence interval. The cv is visualized with bar plot.

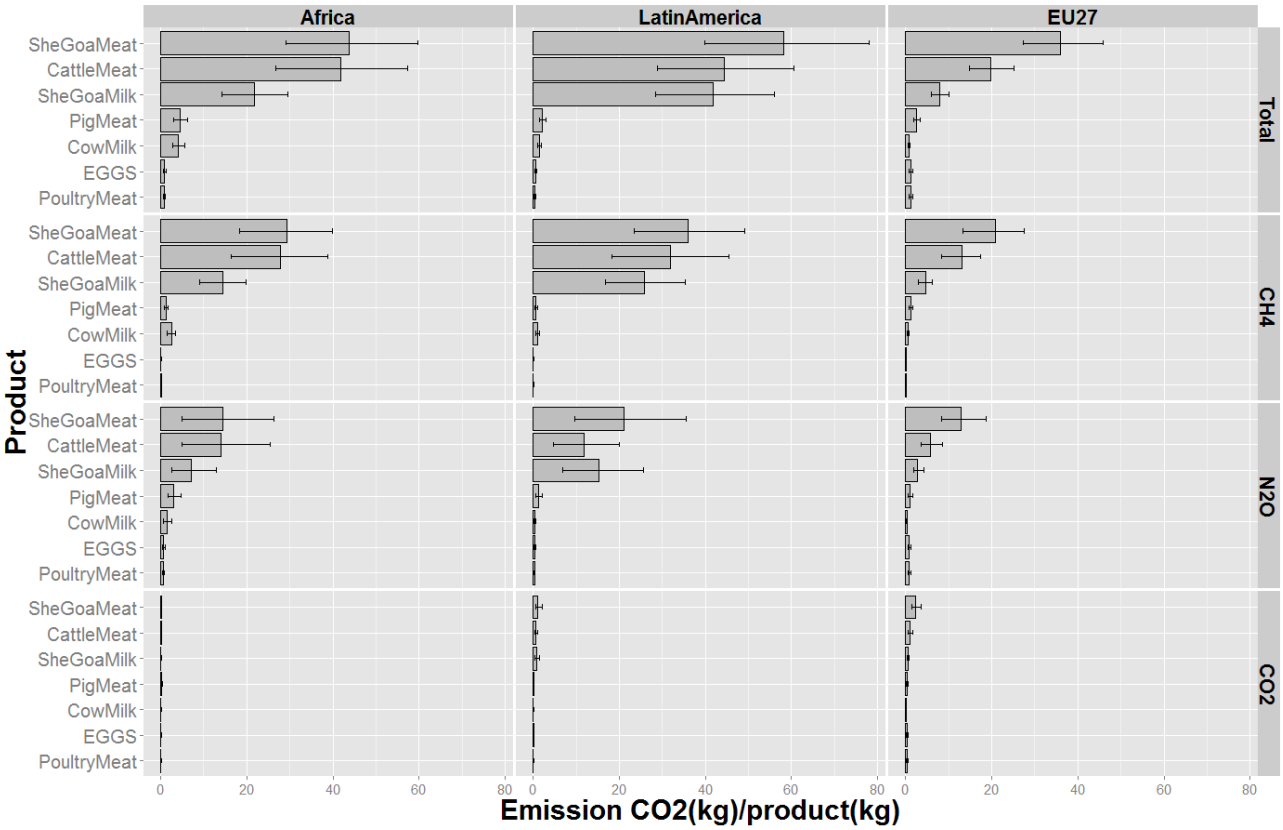


Figure 21 GHG emissions from livestock products

Figure 22 shows that for total emission, the uncertainty ranges from 0.12 to 0.28. Africa and Latin America have higher uncertainty for each livestock product. For Africa, although there is a clear difference between the GHG emissions for different livestock product, the uncertainties are similar, around 0.2. For Latin America, sheep and goat meat, cattle meat, and sheep and goat milk have the highest emission per unit product. The uncertainties for these three products are only slightly lower compared with other product. EU share the similar pattern as Latin America, but has lower uncertainties for each product.

Africa, Latin America and EU27 have similar uncertainties for CH_4 emissions for each product. The products with significant high emissions do not have lower uncertainties.

N_2O emission is in general lower than CH_4 emission for CH_4 each product for each continent. The uncertainties for N_2O are higher than that for the same product in the same continent. On contrary to the pattern found in the total emission, the products with higher emissions have higher uncertainty in Africa. In Latin America and EU27, the products with highest N_2O emission have similar uncertainties compared with other products.

In general, the uncertainty for CO_2 emission is higher for each product for each continent, and the CO_2 emissions are much lower. Africa and Latin America have higher uncertainty than EU27.

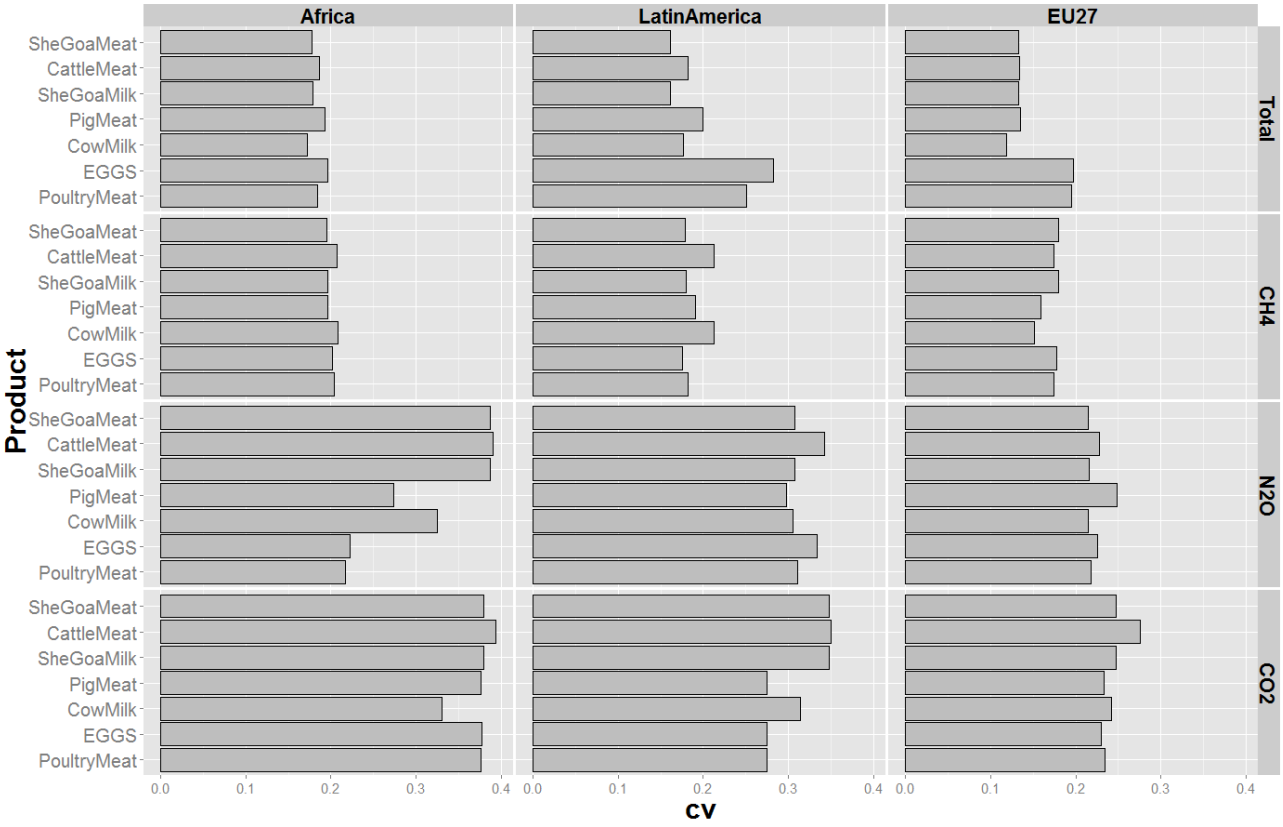


Figure 22 Uncertainty (in term of cv) of GHG emissions from livestock products

3.2.4 GHG emissions from key categories as used by the IPCC

GHG emissions from key categories as used by the IPCC (2006) in livestock production are presented in Figure 23. The uncertainty of these emissions is presented in Figure 24 The emission is expressed with bar plots with error bar indicating an 95% confidence interval. The cv is visualized with bar plot. The processes are ranked from the one with the lowest emission to the one with the highest emission. The cv is ranked in the same order.

By comparing Figure 23 and Figure 24, it is found that there is no clear correlation between the amount of emission and the relative uncertainty. The category with the highest emission, CO_2 emission from enteric fermentation, does have relative low uncertainty. However, N_2O emission from soil and grazing are the second and third highest emission, with the first two highest uncertainty.

For each category, EU27 has lower uncertainty than Africa and Latin America. Except for N_2O from by product and CO_2 from enteric fermentation, Latin America has the highest uncertainty for the other categories. For total emission, Africa and Latin America have higher uncertainty, 0.17, than EU27, 0.12.

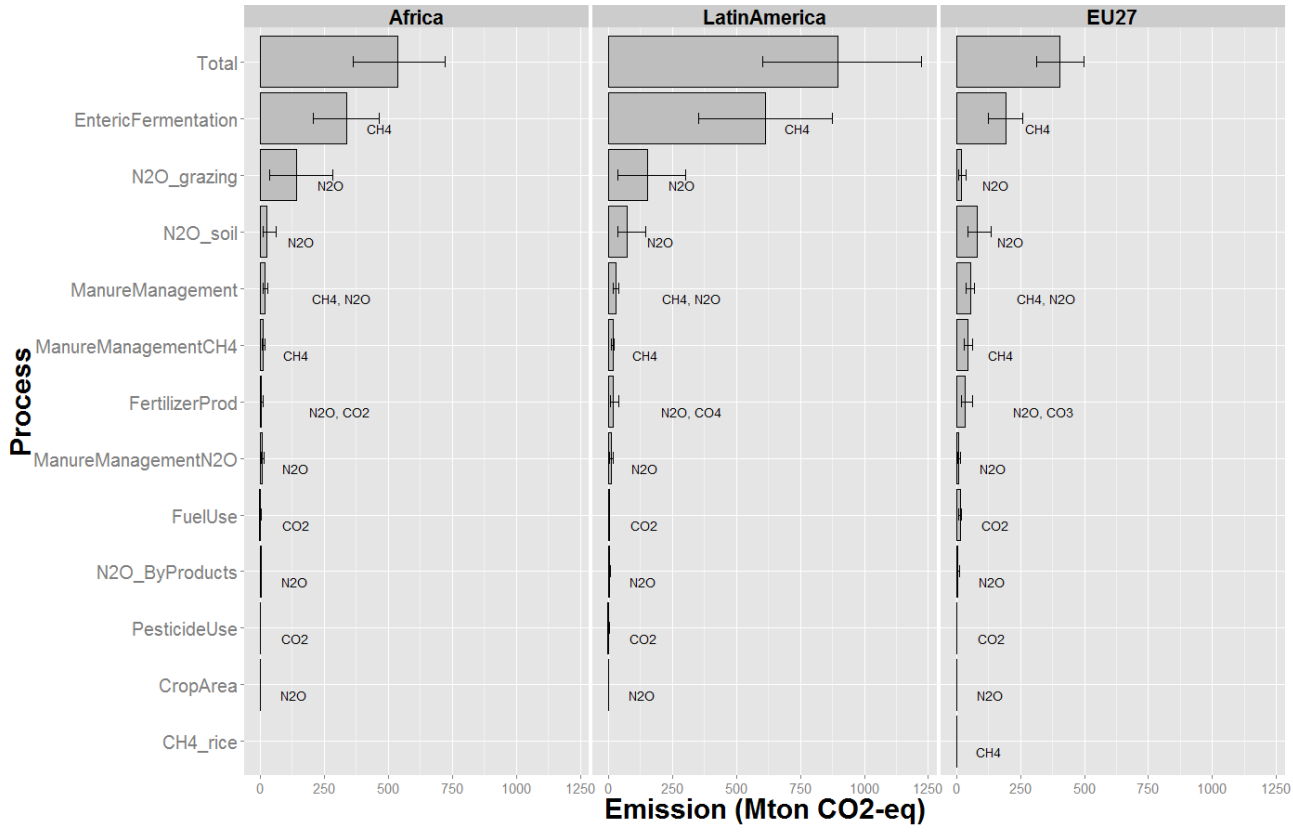


Figure 23 GHG emissions from IPCC categories

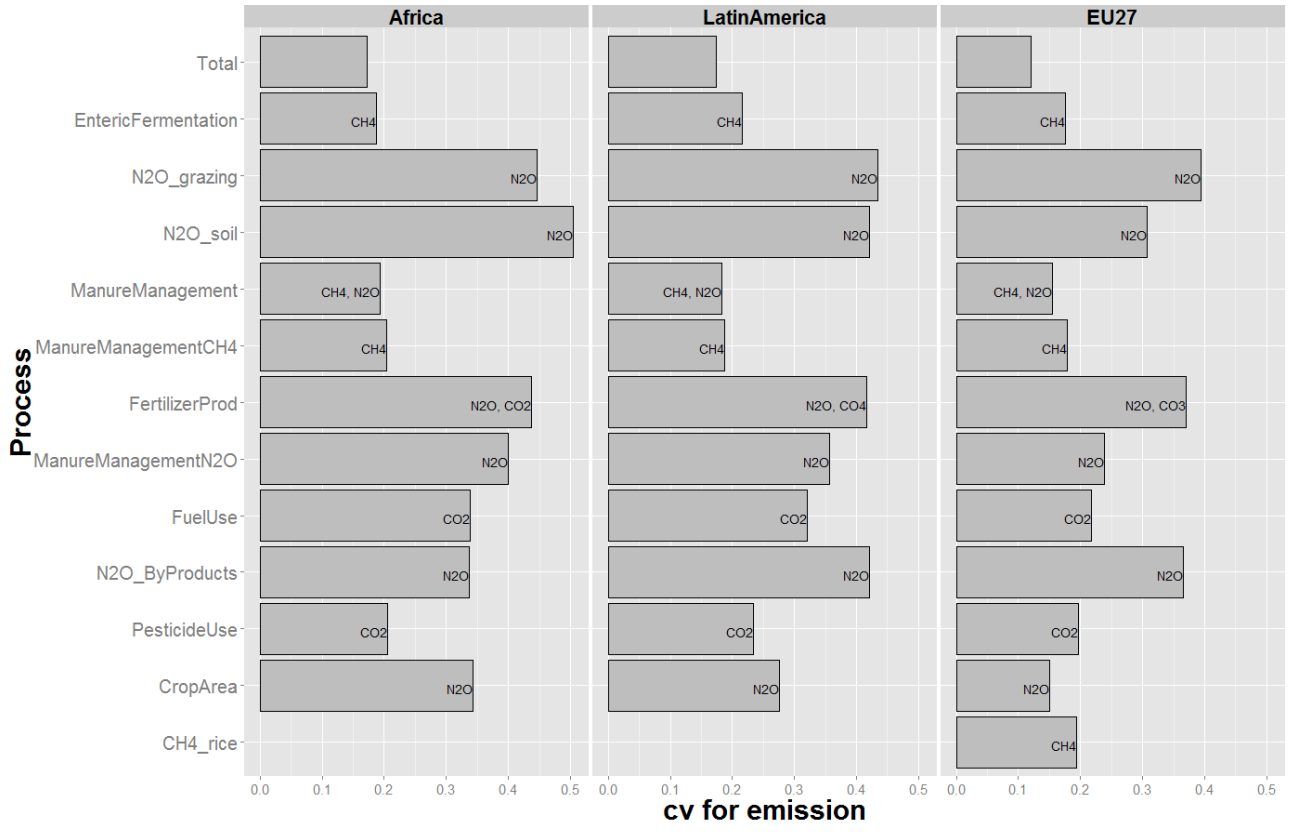


Figure 24 Uncertainty (in term of cv) of GHG emissions from IPCC categories

3.3 Uncertainty contribution of model inputs and parameters to GHG emissions from Africa, Latin America and Europe

In this section the results of the uncertainty contribution of each group to the model outputs are visualized with stacked bar plots. The results are also given with tables in Annex 4. The corresponding R scripts are given in annex 8.

In this study, two methods were applied for quantifying the uncertainty contribution of each group to the model output uncertainty, which were explained in section 2.6. For the final results I use the results from one group at a time scheme. The difference between two results are shown in section 3.3.1. The reason why I choose this method is also be explained in section 3.3.1.

3.3.1 Total GHG emissions

1. Comparing results from one group at a time scheme and winding stair scheme

Here I compared the results from two methods for the uncertainty contribution to CH_4 emission and N_2O emission from the three continents. Since in the winding stairs scheme I only used 100 runs for each group, I also compared the impact of MC run numbers on the uncertainty contribution results.

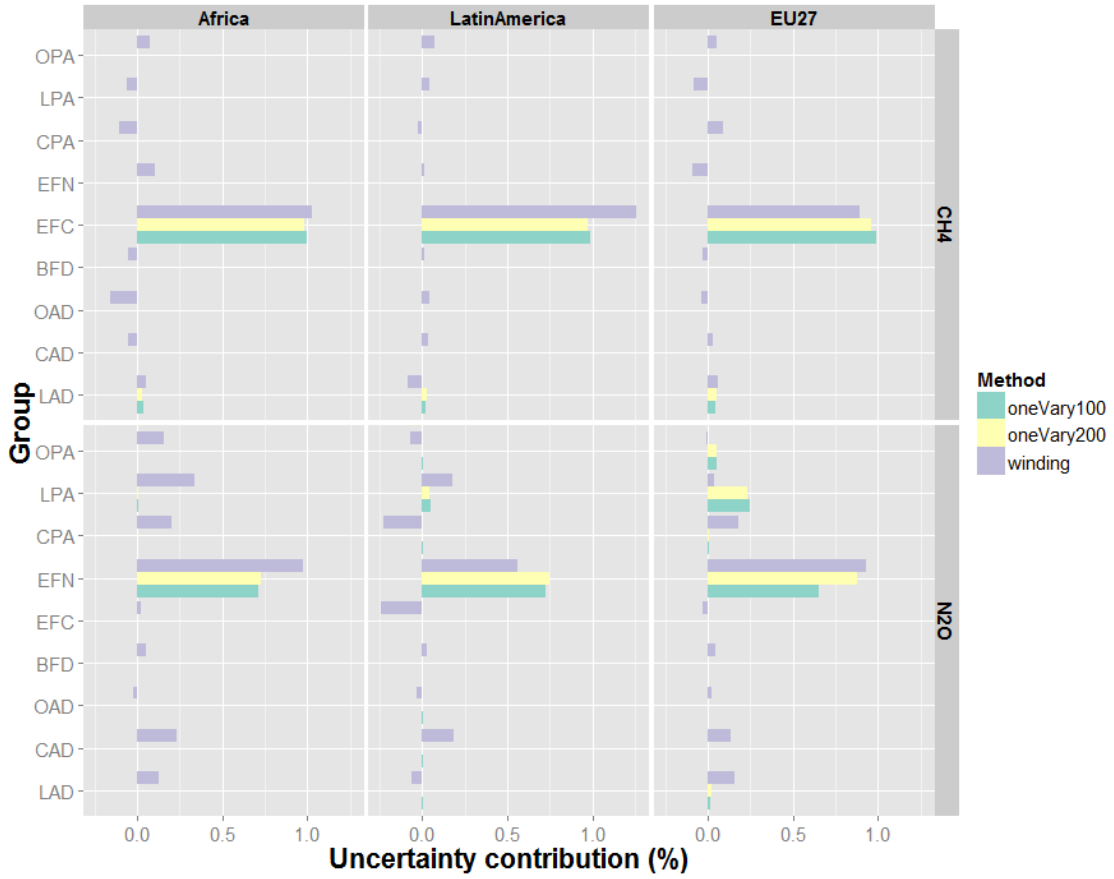


Figure 25 Comparing results from different methods

Figure 25 shows that similar results were gained from the compared two methods. Apart from N_2O emission in EU27, one group at a time scheme with 100 runs (oneVary100) and with 200 runs (oneVary200) provide almost the same results. This indicates that 100 runs with one group at a time scheme can already provide a relatively stable results.

With 100 runs, one group at a time scheme (oneVary100) and winding stairs scheme (winding) are able to indicate which group contributes the most to the model output uncertainty. However, the uncertainty contribution expressed with percentage can be different for the two methods. Moreover, the winding stairs scheme sometimes provides results that are lower than 0. In addition, the sum of uncertainty

contribution of all nine groups gained from winding stairs scheme can be much higher or lower than 100%. For example, the sum of contribution to N_2O emission from Africa, as shown in Figure 25, is much higher than 100%. This indicates that compared with one group vary at a time scheme, winding stair scheme is less suitable for this research. Thus, I decided to use the results from one group at a time scheme. More detail discussion of the methods is provided in section 4.2.

2. Uncertainty analysis results from one group at a time scheme.

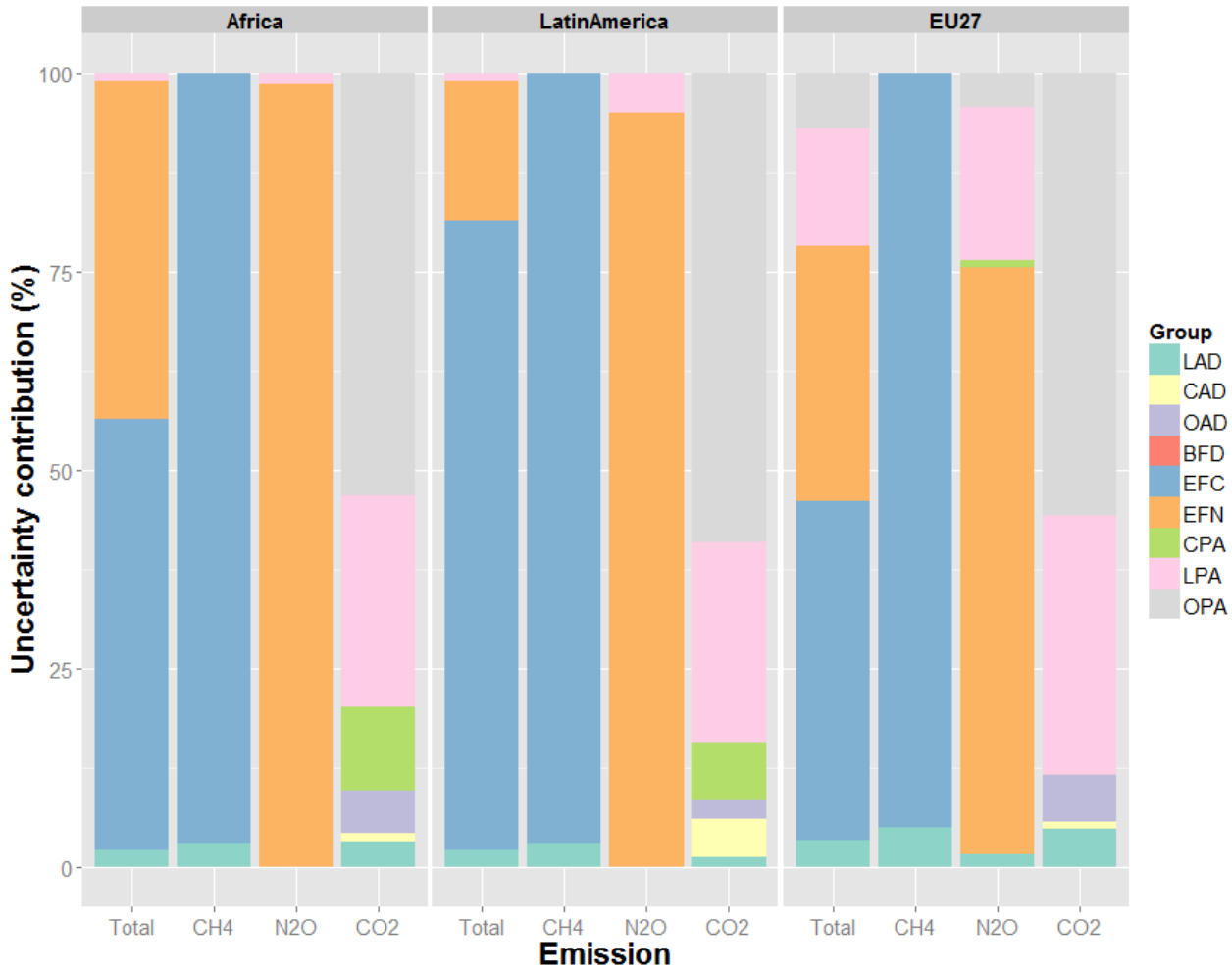


Figure 26 Uncertainty contribution of different group of MIP to total GHG emissions from livestock (see Table 12 for the explanation of the used codes)

For total emission, group N emission factors (EFN) and CH_4 emission factors (EFC) are the main factors affecting the uncertainty in Africa and Latin America. In Europe, Livestock parameters (LPA) and Other emission factors and parameters (OPA) also have a significant effect.

Uncertainties for different GHG emissions are influenced by different groups. Uncertainty for CH_4 emission is mainly caused by CH_4 emission factors, with a small fraction caused by Livestock activity data (LAD). N_2O emission uncertainty in Africa and Latin America is mainly caused by N emission factors (EFN) (>95%). But in EU27, although Livestock parameters (LPA) also contribute a substantial uncertainty (23%). Uncertainty of CO_2 emission is caused by more factors. Other emission factors and parameters (OPA) and Livestock parameters (LPA) are the main factors. Crop parameters (CPA) contribute 10% and 6% uncertainty to Africa and Latin America. However, it does not have a detectable effect on the uncertainty of CO_2 emission from EU27.

3.3.2 GHG emissions from different sectors

The results of uncertainty contribution of each group of MIPs to the GHG emissions from different livestock sectors are presented in Figure 27.

Figure 27 Uncertainty contribution of different group of MIPs to GHG emissions from different livestock sectors (see Table 12 for the explanation of the used codes)

The plot shows that the main factors contribute to the output uncertainty are in line with the ones in 3.3.1. The main MIPs that propagate uncertainty to the model outputs are N emission factors (EFN) and CH_4 emission factors (EFC). Distinguishing sectors, for total emission, Laying hens and broilers are

different from the other sectors. For those two factors, livestock parameters (LPA) and other EF and parameters (OPA) propagate more uncertainty.

For CH_4 emission, CH_4 emission factors (EFC) is still the most important MIP that cause the output uncertainty. The uncertainty contributions in the three continents are all higher than 95% (except for turkey in EU27). For N_2O emission, N emission factors (EFN) is the one contribute the most to the model output uncertainty. But for Laying hens and broilers, livestock parameters (LPA) has a higher uncertainty contribution compared with it for other sectors. For CO_2 emission, there is a difference between continents. In Africa, crop parameters (CPA) contribute to over 50% of the uncertainty in CO_2 emission from the sectors sheep, goats and other cattle. While in Latin America and EU27, for the same sectors, the contribution of group crop parameters (CPA) is much lower 1% to 8%. For the other sectors, livestock parameters (LPA) and other EF and parameters (OPA) are the main factors influent the uncertainty.

3.3.3 GHG emissions from different products

The uncertainty contribution results of groups of MIPs to GHG emissions from livestock products are presented in Figure 28.

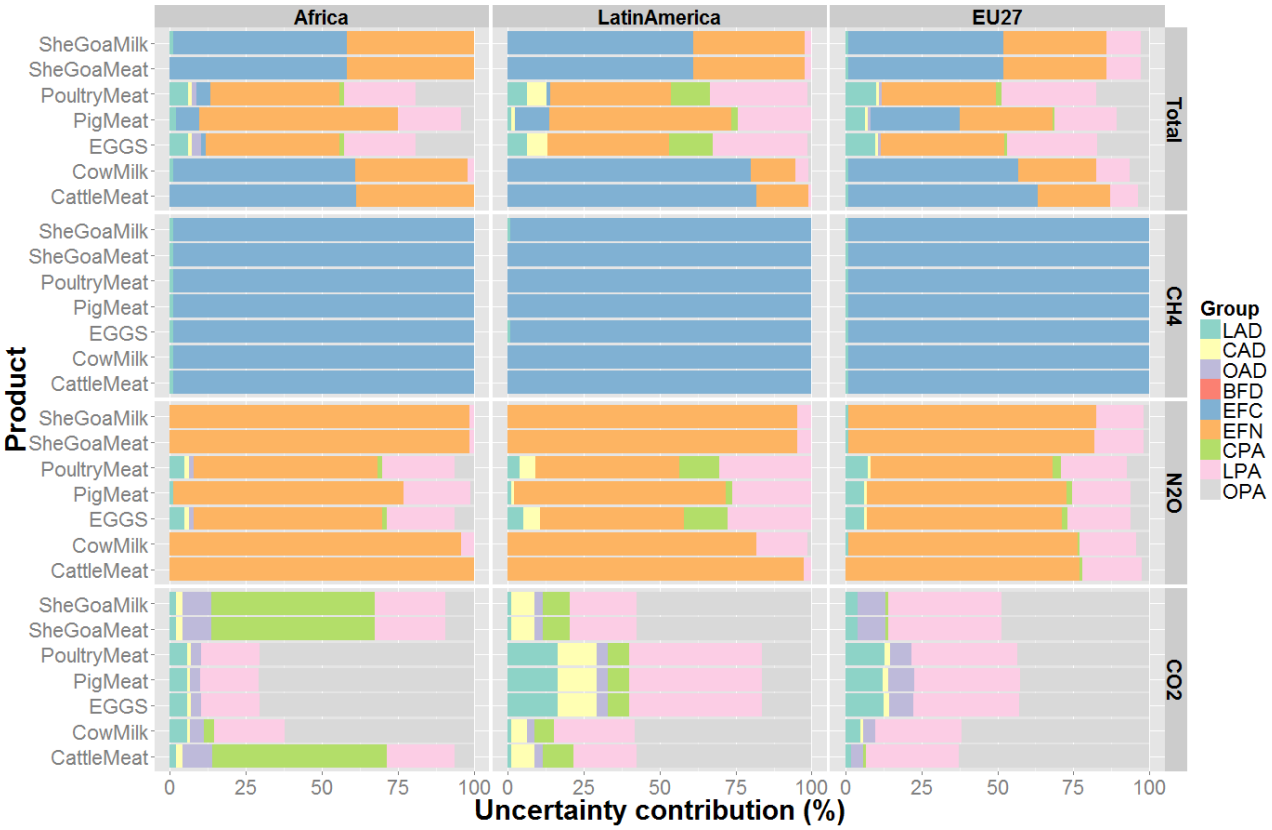


Figure 28 Uncertainty contribution of different group of MIPs to GHG emission from different livestock products (see Table 12 for the explanation of the used codes)

For total emission, the three continents show the same pattern. The main MIPs that propagate uncertainty to the model outputs are N emission factors (EFN) and CH_4 emission factors (EFC) for sheep and goat products, and cattle or cow products. For poultry meat, pig meat and eggs, the main MIPs that contribute to the model output uncertainty are N emission factors (EFN), livestock parameters (LPA) and other EF and parameters (OPA).

For CH_4 emission, CH_4 emission factor (EFC) is the dominant factor for all products in all continents. For N_2O emission, there is also a difference between two groups of products. For the products related to ruminant animals, livestock parameters (LPA) contribute more to the uncertainty than for the products from non-ruminant animals (poultry meat, pig meat and eggs). For CO_2 emission, Latin America and EU27 share similar pattern and Africa is different. In Africa, other EF and parameters (OPA) is the main

factor controlling the uncertainty for CO_2 emission from poultry meat, pig meet, eggs and cow milk. For sheep and goat meat/milk, cattle milk, crop parameters (CPA) contribute the most to the uncertainty. For the other two continents, Latin America and EU27, however, livestock parameters (LPA) and other EF and parameters (OPA) are always the major MIPs that cause the output uncertainty. For poultry meat, pig meat and eggs, livestock activity data (LAD) contribute more than they contribute to other sectors (13~14% vs. 1~5%).

3.3.1 GHG emissions from key categories as used by the IPCC

The uncertainty contribution results from MIPs to key categories as used by the IPCC are presented in Figure 29.

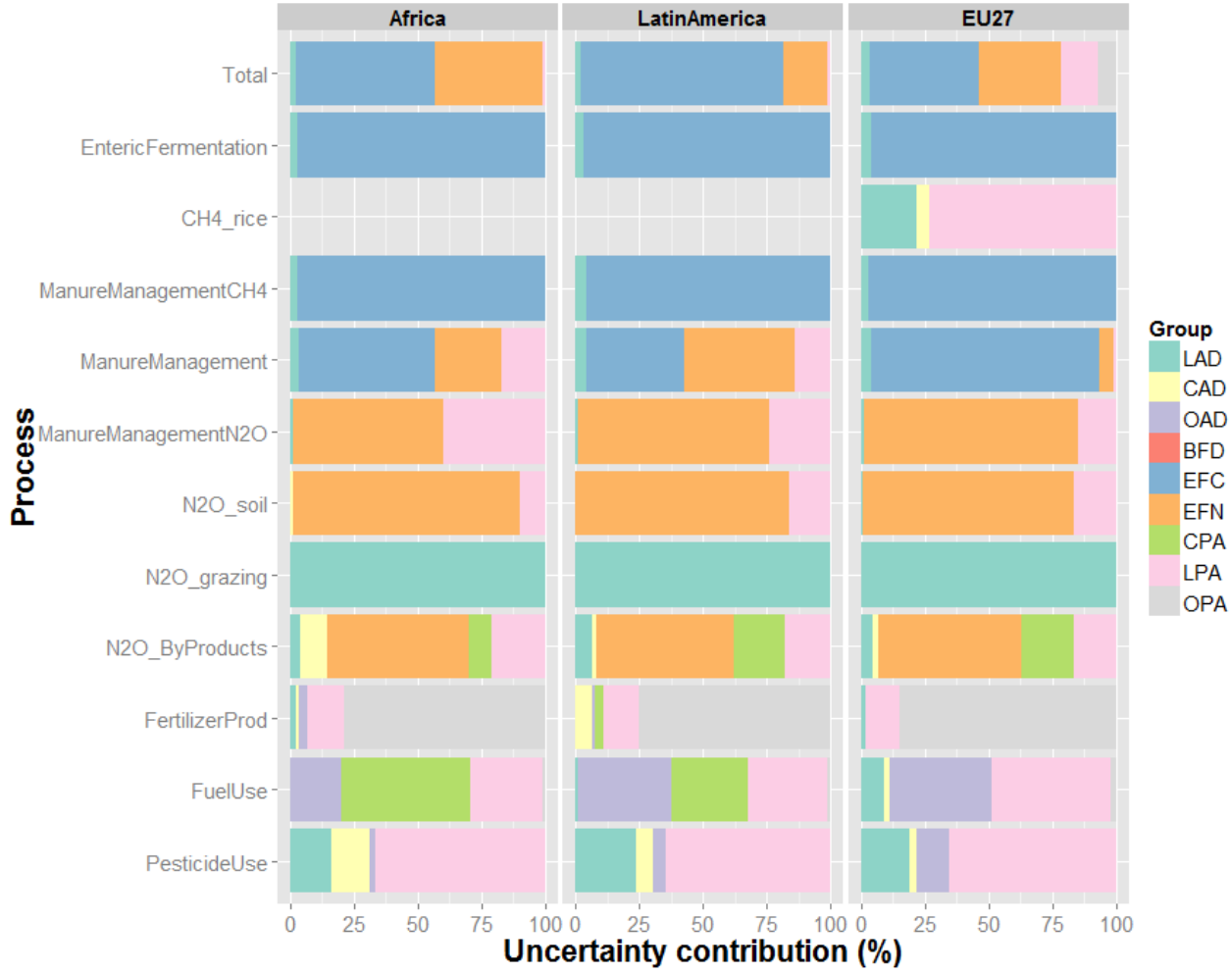


Figure 29 Uncertainty contribution of different group of MIPs to GHG emission from different IPCC categories (see Table 12 for the explanation of the used codes)

The plot shows that there is a clear difference between the different categories. For processes that only have CH_4 emissions (CH4_rice, EntericFermentation, ManureManagementCH4), the main MIPs that contribute to the uncertainty are CH_4 emission factors (EFC).

For categories that only have N_2O related emissions (N2O_ByProducts, N2O_grazing, N2O_soil, ManureManagementN2O), the causes can be different. For N2O_ByProducts, N2O_soil, and ManureManagementN2O, the uncertainty is mainly caused by N emission factors (EFN). However, for N2O emission from grazing (N2O_grazing), the uncertainty is mainly caused by livestock activity data (LAD). There is a continental difference regarding uncertainty in N_2O emission from manure management (ManureManagementN2O). The two main groups cause this uncertainty are N emission factors (EFN) and livestock parameters (LPA). While in EU27, the dominant factor is EFN (contribute 90% to the output

uncertainty), in Africa, the contribution of EFN is lower (54%) and LPA have higher uncertainty contribution (37%).

For categories that have only CO_2 emission (FuelUse and PesticideUse), the uncertainty contribution is different. For CO2 emission from fuel use (FuelUse), other activity data (OAD), crop parameters (CPA) and livestock parameters (LPA) caused the main uncertainty in Africa and Latin America. But in EU27, other activity data (OAD) does not play an important role. For CO_2 emission from pesticide use (PesticideUse), the three continents have similar patterns. The uncertainty is mainly caused by livestock parameters (LPA) and livestock activity data (LAD). The importance of crop activity data (CAD) decreases, and that of other activity data (OAD) increases in the following direction: Africa, Latin America, and EU27.

Manure management is related to both CH_4 and N_2O emission. The uncertainty is mainly caused by CH_4 emission factors (EFC) in EU27 (92%). In Latin America, CH_4 emission factors (EFC) and N_2O emission factors (EFN) are equally important (39% and 44%). In Africa, CH_4 emission factors (EFC) have a higher impact than N_2O emission factors (EFN) (50% and 24%).

4. Discussion and recommendations

In this chapter, I discuss the methods used to quantify model input and parameter uncertainty, the methods used to simulate model inputs and parameters, and the output uncertainty of MITERRA-Global model. The outputs and output uncertainty of MITERRA-Global are compared with other studies. I also provide recommendations for future research.

4.1 Statistical models used to quantify model input and parameter uncertainty

This research is limited to the uncertainty due to uncertainty in model inputs and model parameters. I do not consider the uncertainty in categorical input data, such as land use maps and soil maps nor the uncertainty due to model structure and model implementation.

Three methods were used to obtain the model input and parameter uncertainty: from census data, from literature and from expert judgement. In this section I discuss those three methods respectively.

4.1.1 Model input uncertainty calculated from census data.

As it has been introduced in Chapter 2, some model input uncertainty can be derived from census data. In this research, I applied this method to FAO data only, due to the lack of data availability on the other uncertainty sources.

The advantage of this method that it is in general objective compared with the other two methods. Based on census data, one can get insight into what the dispersion is in collected data and can also calculate the correlations. This will help modeller gain more knowledge about the data quality. One of the model inputs, FertilizerType, for example, has unexpected high uncertainty (details see Table 4). When examining the source data, I found out that for some countries, for the same type of fertilizer, the documented fertilizer consumption can have up to five orders of magnitude difference. In addition, there are many missing data and many zero values in the data set. According to the data documentation rule applied by FAO, the missing data are documented as Not Available (NA). But since it is not likely that the total fertilizer consumption by fertilizer type for a country will change dramatically from year to year, I suspected that the reason why there are so many zero values is because some missing data are also documented as zero. But as I did not find a clear explanation from the FAO website, I did not leave out the zero values when conducting this calculation. In general, this calculation confirmed the hypothesis that most of the data from FAO database have good quality and low uncertainty.

This method also has its limitations. The first limitation is that this uncertainty analysis is done for the current scenario only. It has been proved in section 2.3.1 that when using the variation of the census data, it is important to correct for the trend. Failure to correct for these trends will clearly overestimate the uncertainty. However, some other uncertainty analyses also include uncertainty in future scenario, such as Kros et al. (2012a). In this case, one should be cautious to correct the uncertainty for the trend. But what exact method should be used for the future scenarios is not within the scope of this study.

A second limitation is that accurate and comprehensive measurement data of the whole emission category are seldom available (Monni et al., 2004). For example in this research, although there are many model inputs based on census data, it was only possible to get enough data for the ones derived from FAO database. This is because in order to derive the pdf, one need more years of time serious data. The more data available, the more accurate the result is. These annual data, however, are very limited, especially for continental or global scale models like MITERRA-Global.

I would recommend future research to obtain more census data for for quantifying model input uncertainty. This research showed that the model input data quality might not be in line with the expectation. A better examination of the census data can improve the accuracy of the uncertainty assessment.

4.1.2 Model input and parameter uncertainty derived from literature

It was also possible to derive uncertainty information from literature for some model inputs and parameters. There have been many research carried out on quantifying (part of) the uncertainty related

to livestock processes. Although different models were used, those models sometimes share same parameters in certain processes. This makes it reasonable to use some of their research results.

However, there are many different ways to document the uncertainty of the model inputs/parameters and the uncertainty of the model outputs. The different ways sometimes make it difficult to interpret and compare the results. IPCC (2006) provides guideline on uncertainty documentation, suggesting that the uncertainty should be reported as the percentage lower and upper bound of the 95% confidence interval over the mean. However, different literature sources use many different documentation ways. It is not always clear how to interpret and to compare the information. Flugsrud and Hoem (2011), for example, documented the uncertainty in the model parameters as standard deviation percent with distribution type (same as the cv used in this research). Part of the results from Flugsrud and Hoem (2011) is shown in Table 16.

Table 16 Example data from Flugsrud and Hoem (2011)

IPCC Source category	Pollutant source	Gas	Standard deviation. per cent	Distribution type
4B11-12	Manure management - N2O	N_2O	Fac2	Lognormal
5A1	Forest Land remaining Forest Land, Wildfires	CH_4/N_2O	75	Lognormal

Table 16 shows the uncertainty as "Standard deviation. per cent". On might easily interpret it as standard deviation as a percentage of mean. However, under the end of the table with footnote, Flugsrud and Hoem (2011) states that Fac2 indicates that 2σ is a factor 2 below and above the mean. This means that the uncertainty is actually presented as two times the standard deviation as a percentage of mean. Moreover, it is not very clear how one should use this uncertainty information. With no further information, I interpreted the "Fac2" as the lower bound is 1/2 of the mean and the upper bound is two times the mean. It is also not clear what 75 means for the second one. In some research this information is used as the cv of the lognormal distribution, thus, the sd of the log-transformed distribution. But in some research, it is used as the range of this distribution (similar as the Fac2) in this example. Then it is actually two times the sd of the log-transformed distribution.

The different ways of documenting model output uncertainty can also cause inconvenience. Monni et al. (2004) present model output uncertainty as upper bounds of the 95% confidence interval expressed as percent relative to the mean value (roughly two times the standard deviation). Winiwarter and Rypdal (2001) also provide the model output uncertainty as two times the standard deviation over the mean. However, FAO provides the uncertainty as standard deviation as percentage of the mean in an life cycle assessment of GHG emissions from the dairy sector (Gerber et al., 2010), while also provides the uncertainty as two times deviation over the mean in a global assessment of emissions and mitigation opportunities (Gerber et al., 2013). This difference makes it difficult to compare results from different research, especially when the actually meaning of uncertainty is not stated clearly.

Both of the document methods mentioned above have its advantage and disadvantage. Coefficient of variance (standard divided by mean) is a widely accepted measure of dispersion. It can be directly used in the uncertainty calculation and calculation. However, a highly skewed distribution is less easy to be characterized by the coefficient of variance. The calculation of the coefficient of variance is more difficult (see section 2.3.2). In addition, given the coefficient of variance, the reader would still find it difficult to get a rough impression of how the distribution might look like. In this case, a 95% confidence interval might give more insight into how a highly skewed distribution looks like. However, using 95% confidence interval also has its disadvantage. The confidence interval usually has to be transformed to standard deviation in order to perform an uncertainty assessment.

In this research, using the coefficient of variance and the standard deviation to document the uncertainty is reasonable, since the model output emissions do not have highly skewed distributions. It is also clearly stated in this research how the uncertainty is presented (in terms of cv). For future research, I recommend to document the uncertainty in a way that is the most easy to conduct the research and to compare results with other studies. However, no matter in which way the uncertainty is documented, it should be clearly stated in the report and ambiguous headers should be avoided.

4.1.3 Model input and parameter uncertainty based on expert judgement

The quantification of the uncertainty in terms of statistical properties requires information and this is generally hard to get. Therefore, I must rely on expert judgement for some of the model inputs and parameters.

The estimation was mainly based discussion with experienced MITERRA-Global developers. One of the main difficulties in the discussion was how to determine the distribution type. Normal and lognormal distribution were chosen to represent the distribution type. In many research (Monni et al., 2004; Flugsrud and Hoem, 2011), lognormal distribution is applied when the uncertainty is higher than 60% (in terms of cv). However, the natural reason behind a lognormal distribution is that there should be a multiple effective instead of an additive effect which affect the value of the model input or parameter. Thus, when determining the distribution types, in addition to the magnitude of the uncertainty, we also considered the characteristics of the MIPs. Both forms of distribution cause problems. When applying normal distribution to a non-negative MIP (which is often the case in this research), the minimum value is set to zero. The negative values will be cut/off. Thus, the uncertainty might be underestimate. When applying a lognormal distribution, it is not possible to get negative values but unrealistic positive outliers can be generated. This might cause an overestimation of the uncertainty. To avoid the second situation, I applied a practical upper limit of those MIPs.

4.2 Statistical models used to derive model inputs and parameters from their pdfs for uncertainty quantification and uncertainty analysis

In this section, I discuss the adequacy of the used method (one group at a time scheme) for deriving model inputs and parameters from their pdfs. First I discuss the simulation method in general. Then I discuss the two sampling scheme: on group at a time vs. winding stairs scheme.

The representativeness of the simulated MIPs has been tested and the results are presented in section 3.1. I performed three experiments with different run numbers (100, 200 and 1000) to test if the chosen 200 MC runs were sufficient. All three groups showed a good performance on reproducing the designed uncertainty. But for the “sgsm” method, in some cases it might not be able to reproduce a completely normal distribution with 200 runs at sub-national level. The test results for cross-correlation (results presented in section 3.1.2) showed that the designed cross-correlation can be reproduced with a reasonable accuracy. It can be concluded that with 1000 MC runs the designed uncertainty can be fully reproduced. With 200 runs, the variance can still be reproduced but some of the distributions might not be reproduced. Since only the variance that is used for uncertainty analysis, this simulation method should be sufficient for this study.

For the uncertainty analysis, two sampling methods were used: on group at a time and winding stairs scheme. From the results presented in 3.3.1, it can be concluded that both sampling methods are able to allocate the major MIPs, which cause the model output uncertainty. This can be achieved even with only 100 runs. However, some strange performance showed by the winding stairs scheme: negative uncertainty contribution and overall uncertainty contribution much higher than 150%. Winding stairs scheme calculate the uncertainty contribution using the covariance:

$$\text{Perc\_}G_i = (0.5*\text{cov}(W_0,W_i))/(\text{var}(W_0)) * 100\% \text{ (details see section 2.6.2)} \quad (21)$$

In the paper where winding stairs scheme was developed (Jansen et al., 1994), it was discussed that for complex non-linear models, this scheme is not very suitable. Furthermore, the $0.5*\text{cov}(W_0,W_i)$ can only be used to represented the variance caused by group G_i when there is no correlation between groups. Although the groups were defined so that there is no correlation between groups, this is not assured because of the complexity of the model. One might suspect that the strange performance of winding stairs scheme was caused by limited run numbers used (100 MC runs). However, by comparing the results with one group at a time scheme, the clear difference shows that the run number is not the main factor that causes this performance.

Winding stairs scheme was developed in the last decade of the 20th century (Jansen et al., 1994), when the calculation capacity of computers was very limited. It is shown in this study that winding stairs scheme is able to provide reasonable results for uncertainty contribution with limited runs. However, when the number of runs is not a major limitation, winding stairs scheme might not be a best choice for this type of models. For future research, I recommend to use the on group at a time method, if there is no great need to reduce the sampling time.

4.3 Uncertainty of GHG emissions from livestock predicted by MITERRA-Global model

In this section I discuss the results on uncertainty of GHG emissions predicted by MITERRA-Global model. I reflect on what might be the main causes of uncertainties. I also compare these results with other studies. Note, however, that a detailed explanation on the uncertainties would require in-depth expertise of MITERRA-Global model from modellers.

4.3.1 GHG emissions estimated by MITERRA-Global

First I compare the results from this research with Gerber et al. (2013) on total emission from livestock production and emissions per livestock sector for the three aforementioned continents.

Table 17 Comparing results GHG emissions from livestock with Gerber et al. (2013)

	% total GHG emission			
	This study Base year average2007-2009			FAO (% total GHG emission) Base year 2005[1]
Continent	Africa	Latin America	EU27	All continents on global scale
CH_4	65	70	60	44
N_2O	33	28	33	29
CO_2	1	2	8	27

1) *Results obtained from FAO report using Global Livestock Environmental Assessment Model (GLEAM) (Gerber et al., 2013)*

Table 17 shows that CH_4 contributes the most to the GHG emissions from livestock production. However, in the study provided by FAO (Gerber et al., 2013) using GLEAM model, the contribution from CO_2 emission is much higher than the contribution in this study. The cause might be that some of the GHG emission sources are not included in MITERRA-Global model. GLEAM included more CO_2 emission sources, such as Land-use change, embedded energy related to manufacture of on-farm buildings and equipment, direct on-farm energy use for livestock (e.g. cooling, ventilation and heating), and post farm gate activities (e.g. transport of live animals and products). From the difference we can also concluded that those energy used in livestock production and in transportation contributes considerably to the GHG emissions from livestock sectors. This finding is in line with the study done by FAO (Gerber et al., 2013): CO_2 emission from land use change caused by soybean cultivation contributes 0.7% to the total GHG emissions for cattle milk and beef. CO_2 emission from land use change caused by pasture expansion contributes 14.8% to the total GHG emissions from beef. CO_2 emission cause by direct & indirect energy contributes 2.2% to GHG emissions from cattle milk production and 0.9% to GHG emissions from beef production. For the pig supply chains, CO2 emission from direct and indirect energy use contributes 3.5% to the total GHG emissions, CO2 emissions from land use change caused by soybean cultivation contributes 12.7% to the total GHG emissions.

4.3.2 Uncertainty in GHG emissions estimated by MITERRA-Global

When examining the results of uncertainty in overall GHG emission based on results from section 3.2 and section 3.3, I found that that CO_2 and N_2O have higher uncertainty than CH_4, but the uncertainty in total emission is determined mainly by CH_4 emission factors and N emission factors. This can be explained by the difference in IPCC global warming potentials (GWPs). Although the uncertainty in CH_4 emission is much lower than the uncertainty in CO_2 emission, CH_4 has 28 times global warming potentials compared with CO_2 (IPCC, 2013) and the CH_4 emission is much higher than CO_2 emission. Thus, the contribution of CH_4 to the variance in total emission is higher than CO_2.

Another observation is that the uncertainty in Europe is lower than the uncertainty in Latin America and Africa. This might be explained by the different approach used in Europe and in other continents by MITERRA-Global. This might be caused by different approach (Tier2 instead of Tier1) applied in EU27, or better data quality.

CH_4 emissions from enteric fermentation were calculated using Tier 1 emission factors (EF) derived from IPCC (2006).These Tier1 EF are animal and continent specific. For EU27 country specific emission factors were used for cattle (dairy cows and other cattle). Once the feed intake procedure is fully implemented in the model, Tier2 approach is applied for enteric fermentation for EU27. There is also difference in N_2O emissions. For N excretion factors, the IPCC Tier 1 EFs are used (IPCC_Excretion) for Latin America and Africa, while for EU27 country specific excretion factors (Excr_EU) from GAINS(Klimont and Brink, 2004) are used (Tier2 approach). This difference might have caused the much lower uncertainty in N_2O soil emission from Europe (see Figure 24).

In addition, the uncertainty derived from the census data is in general larger in Latin America and Africa than that in Europe (details see section 2.3.2). Some data are also available at a more detailed level in Europe, and this detailed spatial levels might also helped reducing the uncertainty in Europe. The crop area, for example, is at NUTS2 level(defined as sub-national level in this study) for EU and at national level for the other two continents. Figure 24 in section 3.2.4 shows a considerable decrease in the uncertainty of N_2O emission cause by the category crop area. Another example is the uncertainty in N_2O emission caused by manure management, which is largely determined by N emission factors and livestock activity data. Some of the MIPs in the livestock parameters group, such as animal numbers, livestock numbers and N excretion factor are available at more detailed spatial level in Europe. As a results, the uncertainty of N_2O emission caused by manure management is much lower in Europe, and the uncertainty contribution of the livestock parameters group is also much smaller than that in the other two continents (see Figure 29).

The uncertainty in GHG emission is largely determined by CH_4 emission factors. The results for different sectors show that horses, camels and goats have higher uncertainty in CH_4 emission than other animals. This might be because that there is a clear difference in the input uncertainty for those products. The CH_4 emission factor for enteric fermentation for cow and sheep is 0.25, while it is 0.4 for other animals. The reason why pigs do not have a higher uncertainty in CH_4 emission might be that the CH_4 emission factor is very small for pigs. Thus, the CH_4 emission from pigs concerning enteric fermentation, which does not contribute much to the overall uncertainty in CH_4 emission by pigs.

For different products, there is a clear difference between the meat/milk product and other products. Uncertainty in emissions from products by ruminant animals (beef, caw milk, sheep milk, etc) are mainly determined by CH_4 and N_2O emission factors. While for pork, it is mainly caused by N emission factors and livestock parameters (including N excretion, manure system usage). For eggs and chicken, it is determined mainly by N_2O emission factors, livestock parameters, crop parameters, other emission factors and parameters. Furthermore, the uncertainty in milk and beef products in Europe is also influenced by livestock parameters. Livestock parameters, however, does not have much effect in the other two continents. This difference might be explained by the fact that Europe has higher CH_4 emission from manure management. This is found by this research (details see Figure 23 and annex4). This result is also in line with the results from Lesschen et al. (2011b).

I also compare the uncertainty quantification results from this research with other research. A report from FAO (Gerber et al., 2013) presented that the 95 percent interval of confidence for ruminants is around ± 50 percent, while it is between ± 20 and 30 percent for monogastrics. In terms of cv, it means that around 0.25 for ruminants and 0.1 to 0.15 for monogastrics. In this study, it is 0.12 to 0.3 for both. It was also discussed in FAO report (Gerber et al., 2013). The higher uncertainty for monogastrics in this study might be associated with the high uncertainty involved in the N_2O direct and indirect soil emission and the high uncertainty in the CO_2 emission. In addition, the study provided by FAO (Gerber et al., 2013) is only performed for a few selected country. Usually, the uncertainty in country level should be lower than it in continental level (Kros et al., 2012a). However, only a few selected countries might not be representative for the uncertainty at country levels in general. This might be why the estimate from this study is not lower than the one by FAO (Gerber et al., 2013). Another study done by FAO (Gerber et al., 2010) also performed an uncertainty assessment for the GHG emissions from dairy sector. The results showed a standard deviation of 12 to 13 percent of the average value for meat and milk, in both Sweden

and Nigeria. In this study, the standard deviation for meat and milk is 0.12 to 0.13 in EU and 0.17 to 0.19 in Africa. For all of the calculation in this FAO study (Gerber et al., 2010), a Tier2 approach is used. In MITERRA-Global, Tier1 approach is used in Latin America and Tier2 approach is partly applied in EU. This indicates that the uncertainty estimated for MITERRA-Global is reasonable. The difference between the uncertainty in Nigeria (Gerber et al., 2010) and the uncertainty in Africa (this study) also confirmed that a Tier2 approach should be able to reduce the uncertainty.

5. Conclusions

The uncertainty in the model inputs and parameters (MIPs) of GHG emission profiles of livestock sectors in Africa, Latin America and Europe has been assessed. The model output uncertainty of MITERRA-Global model regarding GHG emission from livestock sectors has been assessed by a Monte Carlo analysis. The uncertainty contribution from model inputs and parameters to model outputs is also calculated and visualized.

Estimate of GHG emission from livestock sectors in Africa, Latin America and Europe from MITERRA-Global shows different levels of uncertainty for different emissions from different sectors/products/source categories. The level of uncertainty is also different for different continents. In general, the GHG emissions from livestock in Europe has a lower uncertainty than that in Africa and Latin America. In general, the CO_2 emission has the highest uncertainty and CH_4 has the lowest in relative terms. The uncertainty contribution results show that for different greenhouse gases, the uncertainty is caused by different model inputs and/or parameters. But in general, the CH_4 and N emission factors contribute the most to the uncertainty in total GHG emission.

Hereafter, I provide the conclusion of each sub-research question:

1. What are the main input data and parameters to be analysed in the uncertainty assessment?

   The main input data are livestock activity data, crop activity data, other activity data (including for example fertilizer consumption, pesticides use, etc.), and biophysical data (including climate, soil data, etc.). The main model parameters to be analysed include CH_4 and N emission factors (including leaching and runoff), crop parameters, livestock parameters and other parameters (including CO_2 EF, fertilizer composition, etc.).

2. How can a statistical model (pdf) be built that fully characterizes the uncertainty of the main selected input data for MITERRA-Global for the various continents, described as probability distributions, including cross correlation for certain pairs of model inputs and/or spatial correlation in these uncertainties where relevant.

   A statistical model (pdf) can be built based on variation in census data, information gained from literature or from expert knowledge. For uncertainty calculated from census data, it is important to take into account that the variance should be correct for the real year-to-year trend. For uncertainty derived from literature and/or expert judgement, firstly the distribution type and sd/cv should be determined. Then a maximum value and a minimum value should be set to prevent unrealistic outliers. No data were used to derive the spatial correlation. By determining if there is a dependency between plots, the spatial correlation is determined with expert knowledge. For cross correlation, it can be calculated from census data when there is enough data available. When there are no data available, the pairs of model inputs and/or parameters with inter-dependency are assigned to be correlated.

3. How can model inputs and parameters realizations be sampled efficiently from their pdfs by using stochastic simulation techniques?

   The model inputs and parameters are sampled with developed R scripts. There are two possible methods. When there is only spatial correlation or only cross correlation, the sampling can be achieved using the standard R package "**rmultnorm**" with a developed spatial or cross correlation matrix and the model input and parameter pdfs . When both cross and spatial correlation need to be taken into account, the package "sgsm" is used. For the second method, a spatial-cross-correlation matrix needs to be constructed so that all the correlations are taken into account. The model input and parameter pdfs are also needed.

4. How can the uncertainty contribution analysis be carried out in batch mode and its results be stored automatically?

   This is achieved using developed R scripts with a one group at a time scheme. In this scheme with 200 runs only one group is varying according to the predefined pdf, the other group of model inputs and parameters use the default value stored in MITERRA-Global model. The results are stored in gdx

files and read in by MITERRA-Global. In MITERRA-Global, the simulated results with the uncertainty information is combined with the default value within MITERRA-Global to obtain the real model inputs with uncertainty. The MITERRA-Global is run with these uncertain model inputs to produce the uncertain model outputs, which are also stored in gdx files.

5. What is the uncertainty in the CH_4, N_2O and CO_2 emissions of livestock sectors in the three continents and which model inputs and/or parameters contribute the most to this uncertainty?

   The uncertainty in the total GHG emissions of livestock sectors is 17% (in terms of cv) in Africa and Latin America, 12% in Europe (EU27). For CH_4 emission, the uncertainty is 18% in Africa, 21% in Latin America and 15% in Europe. The uncertainty in N_2O is 37% in Africa, 31% in Latin America and 22% in Europe. The uncertainty in CO_2 emission is 31% in Africa, 32% in Latin America and 24% in Europe. CH_4 emission factors and N emission factors contribute the most to this uncertainty.

6. How can the results of the uncertainty contribution analysis be summarized and visualized and thus efficiently communicated to end-users in the three continents?

   The results of uncertainty quantification are summarized with the following statistical parameters: mean, sd, cv, 2.5% percentile and 97.5% percentile. It is visualized with bar plot of average emission with error bars (indicating the 95% confidence interval) and with bar plot of cv`s. The bar plot of average emission with error bars provide a background information: how high is the emission and if the output is equally distributed around the mean. The cv provides the uncertainty information.

This research shows large variation in uncertainties in both model inputs/ parameters and model outputs, and large variation in uncertainty contributions of different model inputs/parameters to different model outputs. Some activity data and emission factors showed uncertainty. For calculating the uncertainty contribution to the model outputs, one group at a time approach is more feasible than winding stairs approach.

The uncertainty (% in terms of cv) of total GHG emissions of livestock sectors is higher in Africa and in Latin America (17%) than that in Europe (12%). The uncertainty of CH_4 emission (15% to 21%) is lower than that of N_2O (22% to 37%) and CO_2 (24% to 31%). CH_4 emission factors and N emission factors contribute the most to this uncertainty.

In this study, the uncertainty of total GHG emissions is less subject to the uncertainty of CO_2 emission compared with other study, because of the excluding of energy use and land use change. The Tier2 approach applied in EU substantially decreased the uncertainty of GHG emissions from Europe. Thus, further implementing the Tier2 approach may further decrease the uncertainty of the MITERRA-Global.

6. References

Boyer, E. W., R. W. Howarth, J. N. Galloway, F. J. Dentener, P. A. Green and C. J. Vörösmarty (2006). "Riverine nitrogen export from the continents to the coasts." Global Biogeochemical Cycles **20**(1).

De Boer, I., C. Cederberg, S. Eady, S. Gollnow, T. Kristensen, M. Macleod, M. Meul, T. Nemecek, L. Phong and G. Thoma (2011). "Greenhouse gas mitigation in animal production: towards an integrated life cycle sustainability assessment." Current Opinion in Environmental Sustainability **3**(5): 423-431.

Del Grosso, S., A. Mosier, W. Parton and D. Ojima (2005). "DAYCENT model analysis of past and contemporary soil N< sub> 2</sub> O and net greenhouse gas flux for major crops in the USA." Soil and Tillage Research **83**(1): 9-24.

Del Grosso, S., S. Ogle, W. Parton and F. Breidt (2010). "Estimating uncertainty in N2O emissions from US cropland soils." Global Biogeochemical Cycles **24**(1): GB1009.

Delgado, C. L. (1999). Livestock to 2020: The next food revolution, Free downloads from IFPRI.

Delgado, C. L., M. W. Rosegrant, H. Steinfeld, S. K. Ehui and C. Courbois (1999). Livestock to 2020: The next food revolution, Intl Food Policy Res Inst.

Flugsrud, K. and B. Hoem (2011). "Uncertainties in the Norwegian greenhouse gas emission inventory."

Galloway, J. N., M. Burke, G. E. Bradford, R. Naylor, W. Falcon, A. K. Chapagain, J. C. Gaskell, E. McCullough, H. A. Mooney and K. L. Oleson (2007). "International trade in meat: The tip of the pork chop." AMBIO: A Journal of the Human Environment **36**(8): 622-629.

Gerber, P., T. Vellinga, C. Opio, B. Henderson and H. Steinfeld (2010). "Greenhouse Gas Emissions from the Dairy Sector A Life Cycle Assessment." Africa, S **1**: 94.

Gerber, P. J., H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci and G. Tempio (2013). Tackling climate change through livestock – A global assessment of emissions and mitigation opportunities. Rome, Food and Agriculture Organization of the United Nations. Rome, Food and Agriculture Organization of the United Nations (FAO).

Heuvelink, G., M., M. van Oijen, W. de Vries and H. Kros (2009). Operational methodology for spatial uncertainty quantification and spatial uncertainty analysis of regional scale NEU models. Wageningen, Alterra, part of Wageningen UR.

Heuvelink, G. and E. J. Pebesma (1999). "Spatial aggregation and soil process modelling." Geoderma **89**(1): 47-65.

IPCC (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme. Kamiyamaguchi, IGES, Japan.

IPCC (2007). Synthesis Report: Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Core Writing Team, IPCC, Geneva, Switzerland.

IPCC (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA**:** 1535 pp.

Jansen, M. J., W. A. Rossing and R. A. Daamen (1994). Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. Predictability and nonlinear modelling in natural sciences and economics, Springer**:** 334-343.

Janssen, P. and P. Heuberger (1995). "Calibration of process-oriented models." Ecological Modelling **83**(1): 55-66.

Karimi-Zindashty, Y., J. D. Macdonald, R. L. Desjarins, D. E. Worth, J. J. Hutchinson and X. P. C. Vergé (2012). "Sources of uncertainty in the IPCC Tier 2 Canadian livestock model." The Journal of Agricultural Science **150**(05): 556-569.

Klimont, Z. and C. Brink (2004). "Modelling of emissions of air pollutants and greenhouse gases from agricultural sources in Europe." IIASA IR: 04-048.

Kros, J., G. Heuvelink, G. Reinds, J. Lesschen, V. Ioannidi and W. d. Vries (2012a). "Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe." Biogeosciences Discussions **9**(5): 6051-6094.

Kros, J., J. P. Lesschen, A. Bannink, P. Kuikman and O. Oenema (2012b). Review on calculation methodologies and improvement of mechanistic modelling of GHG emissions. Deliverable 3.1 of the EU project AnimalChange (http://www.animalchange.eu).

Lesschen, J., M. Van den Berg, H. Westhoek, H. Witzke and O. Oenema (2011a). "Greenhouse gas emission profiles of European livestock sectors." Animal Feed Science and Technology **166**: 16-28.

Lesschen, J. P., O. Oenema and P. Kuikman (2011b). Current GHG emissions from livestock production at regional level. Deliverable 3.2 of the EU project AnimalChange (http://www.animalchange.eu)

Lesschen, J. P., I. Staritisky and O. Oenema (2014). Improvements in MITERRA framework/tool and the extension to Africa and Latin America. Deliverable 3.5 of the EU project AnimalChange (http://www.animalchange.eu)

Limpert, E., W. A. Stahel and M. Abbt (2001). "Log-normal Distributions across the Sciences: Keys and Clues On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question." BioScience **51**(5): 341-352.

Menzi, H., O. Oenema, C. Burton, O. Shipin, P. Gerber, T. Robinson, G. Franceschini, H. Steinfeld, H. Mooney and F. Schneider (2010). "Impacts of intensive livestock production and manure management on the environment." Livestock in a Changing Landscape, Volume 1: Drivers, Consequences, and Responses **1**: 1139.

Monni, S., S. Syri and I. Savolainen (2004). "Uncertainties in the Finnish greenhouse gas emission inventory." Environmental Science & Policy **7**(2): 87-98.

Seitzinger, S., J. Harrison, E. Dumont, A. H. Beusen and A. Bouwman (2005). "Sources and delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application." Global Biogeochemical Cycles **19**(4).

Smaling, E., R. Roscoe, J. Lesschen, A. Bouwman and E. Comunello (2008). "From forest to waste: Assessment of the Brazilian soybean chain, using nitrogen as a marker." Agriculture, Ecosystems & Environment **128**(3): 185-197.

Smil, V. (2002). "Eating meat: Evolution, patterns, and consequences." Population and development review **28**(4): 599-639.

Sommer, S. G., J. E. Olesen, S. O. Petersen, M. R. Weisbjergz, L. Valli, L. Rodhe and F. Beline (2009). "Region-specific assessment of greenhouse gas mitigation with different manure management strategies in four agroecological zones." Global Change Biology **15**(12): 2825-2837.

Steinfeld, H., P. Gerber, T. Wassenaar, V. Castel, M. Rosales and C. De Haan (2006). Livestock's Long Shadow: Environmental issues and options. Food and Agriculture Organisation of the United Nations, Rome.

Steinfeld, H. and T. Wassenaar (2007). "The role of livestock production in carbon and nitrogen cycles." Annu. Rev. Environ. Resour. **32**: 271-294.

Truong, P. (2009). Uncertainty analysis of predicting terrestrial nitrousoxide emissions in Europe with INTEGRATOR. MSc, Wageningen University.

Wang, G. and S. Chen (2012). "A review on parameterization and uncertainty in modeling greenhouse gas emissions from soil." Geoderma **170**: 206-216.

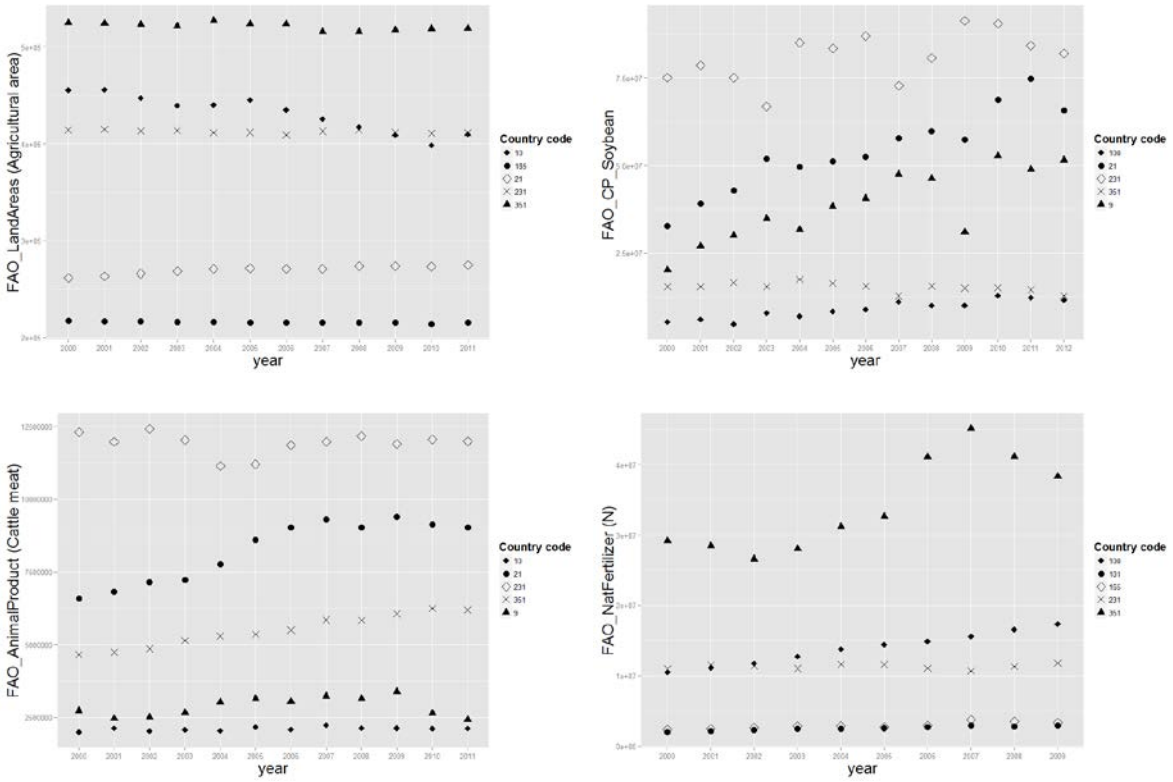
Weiss, F. and A. Leip (2012). "Greenhouse gas emissions from the EU livestock sector: A life cycle assessment carried out with the CAPRI model." Agriculture, Ecosystems & Environment **149**: 124-134.

Winiwarter, W. (2005). "The GAINS model for greenhouse gases-version 1.0: nitrous oxide (N2O)." International Institute for Applied Systems Analysis (IIASA), Laxenburg.

Winiwarter, W. and K. Rypdal (2001). "Assessing the uncertainty associated with national greenhouse gas emission inventories:: a case study for Austria." Atmospheric environment **35**(32): 5425-5440.

Annex 1 Data related to uncertainty quantification in MIPs

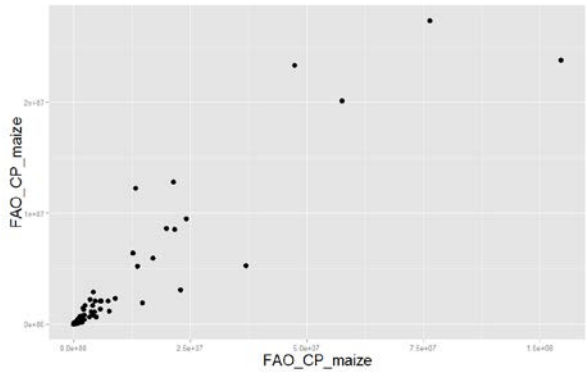
1.1 Plot of FAO data

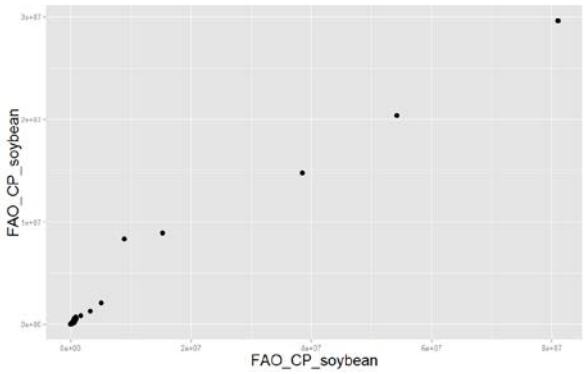


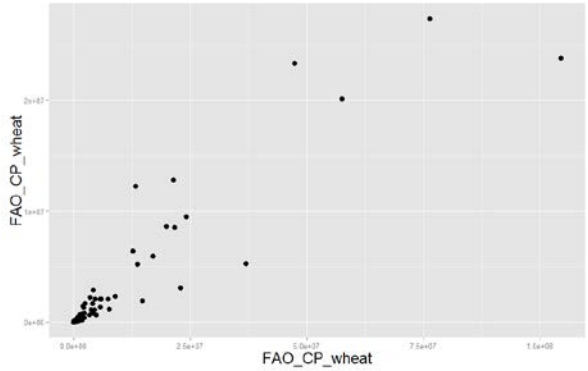
1.2 Example of FAO crop area data

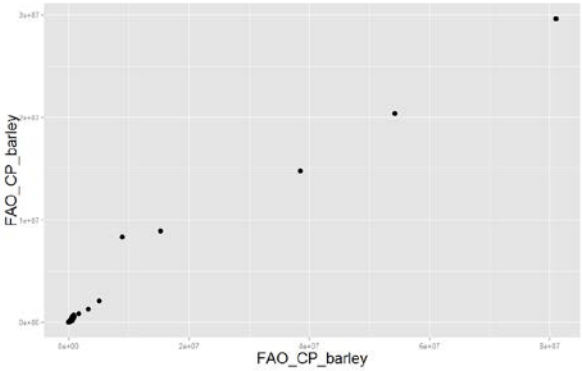
COUNTRY CODE	CROP CODE	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1	1717	156585	191331	184869	194816	197983	202111	157554	172629	170005	164014	155666	158838	170272
1	1817	156585	191331	184869	194816	197983	202111	157554	172629	170005	164014	155666	158838	170272
1	15	106440	108380	119224	126112	124479	127574	100188	98400	92810	88530	86574	77806	93476
1	1814	50145	82951	65645	68704	73504	74537	57366	74229	77195	75484	69092	81032	76796
1	44	47100	79600	61701	60392	64628	65462	49400	65180	66260	66511	60954	67873	65135
1	1801	36208	38603	34866	32901	42692	44085	45048	45680	44598	44908	46035	45553	
1	116	34202	31752	30474	32286	35690	34299	33012	31612	34298	31998	28326	28665	31243
1	1720	34202	31752	30474	32286	35690	34299	33012	31612	34298	31998	28326	28665	31243
1	1735	23427	23075	24062	27202	26273	26852	29178	31460	29663	30076	27928	30677	
1	1800	23427	23075	24062	27202	26273	26852	29178	31460	29663	30076	27928	30677	
1	560	14571	14314	12434	11301	13560	13475	14098	14097	14390	14292	14613	14478	
1	515	5959	8846	8761	8000	8000	8300	8000	8942	8298	8811	9321	8962	
1	388	5588	5351	5618	7236	6374	6291	7205	7372	6257	6231	6517	6837	
1	463	4900	5200	5300	6800	6300	6800	7500	7825	7539	7821	7397	7656	
1	521	4171	3960	2915	2700	3500	3700	3600	2905	2727	2756	2788	2729	
1	567	3386	3300	3853	4058	4030	3860	4000	5884	5446	6163	4470	5764	
1	526	3300	3243	3200	3400	5100	5300	5700	8215	8140	7808	7928	7739	
1	534	3267	3200	3220	3500	5200	5400	5700	4532	4269	4283	4317	4509	
1	358	3083	3239	3425	3290	3202	3444	3573	3531	3418	3070	3217	3453	
1	56	2552	2602	3252	3310	3127	3161	2528	3541	3963	2637	2729	3524	3075
1	826	2528	1088	890	526	480	210	80	118	170	297	417	614	
1	397	2500	2000	1949	2019	2231	2367	2360	2408	2339	2549	2237	2548	
1	1726	2024	1806	1720	1923	1984	1848	1969	1993	2199	2053	1874	1952	2005
1	176	1994	1769	1687	1777	1799	1688	1770	1884	2083	1952	1765	1864	1868
1	403	1946	1975	2116	1853	2097	2078	2394	2227	2487	2085	1869	1961	

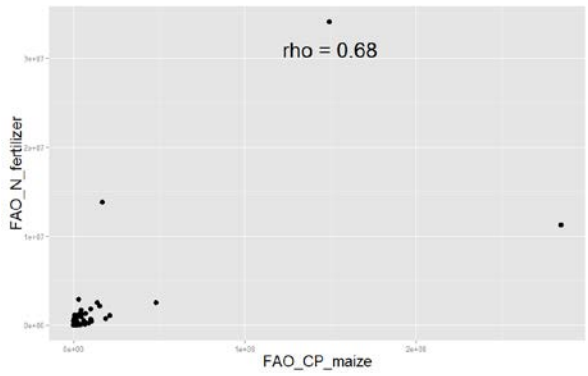
1.3 Cross correlation scatter plots of FAO data

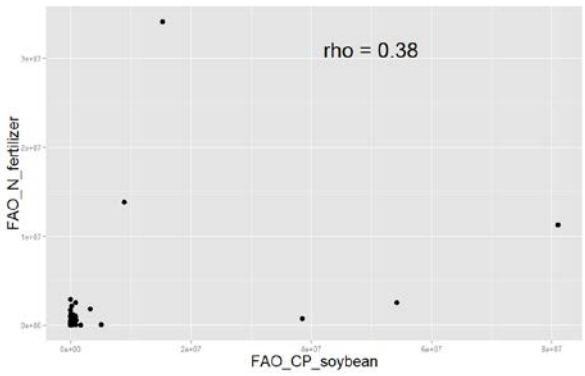


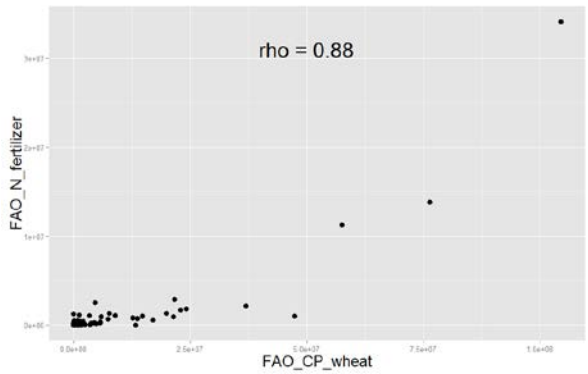


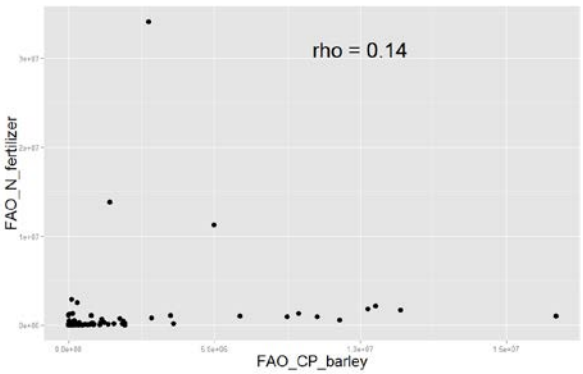


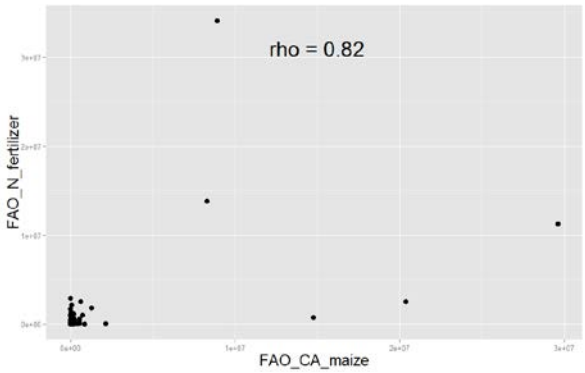


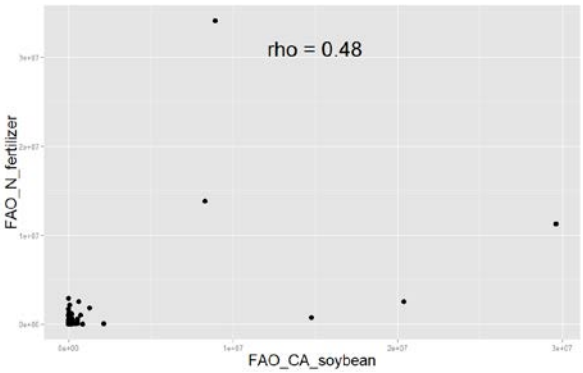


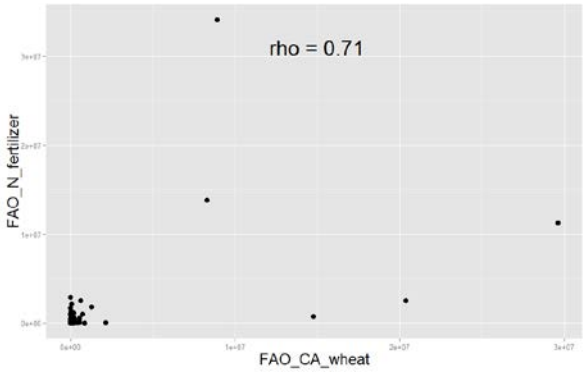


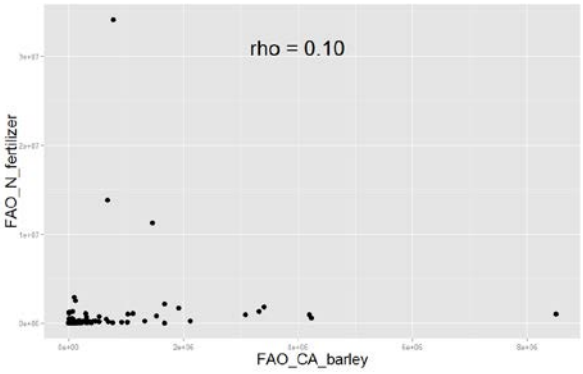





rho = 0.82
FAO_N_fertilizer
FAO_CA_maize


rho = 0.48
FAO_N_fertilizer
FAO_CA_soybean


rho = 0.71
FAO_N_fertilizer
FAO_CA_wheat


rho = 0.10
FAO_N_fertilizer
FAO_CA_barley

Annex 2 Results of uncertainty quantification in MIPs

The table below provides the complete results of uncertainty quantification for MIPs. In the code column the code for MIPs used in MITERRA-Global are listed. The Group column indicates which group this MIP belongs to. Full description of groups can be found in section 2.6. The distribution type, cv, sd, min, max are parameters describing the probability distribution function of the MIP (details see section 2.3.1). The Spatial_level indicates at what spatial level this MIP is used in the model. The rhoSUB, rhoNAT and rhoCON indicates the spatial correlation coefficient at subnational level, national level and continental level.

Nr.	Code	Group	Distribution type	cv	sd	Min	Max	Spatial_level	rhoSUB	rhoNAT	rhoCON
1	AgroMapsCA_barley	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
2	AgroMapsCA_maize	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
3	AgroMapsCA_other	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
4	AgroMapsCA_soybean	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
5	AgroMapsCA_wheat	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
6	AgroMapsCropProd	CAD	normal	0.25		0	inf	sub-national	0.8	0.5	0
7	BMFac_Gras	CPA	normal	0.25		0	1	generic	1	1	1
8	BMFac_OtherAreable	CPA	normal	0.25		0	1	generic	1	1	1
9	bNumAniRAINS	LAD	normal	0.1		0	inf	national	1	0.5	0
10	CAPRI_NumAni	LAD	normal	0.1		0	inf	sub-national	0.5	0.5	0
11	CH4_EF_EntFer_cs	EFC	normal	0.25		0	250	sub-national	0.8	0.8	0.5
12	CH4_EF_EntFer_other	EFC	normal	0.4		0	150	sub-national	0.8	0.8	0.5
13	CH4_EF_ManManage	EFC	normal	0.25		0	250	sub-national	0.8	0.8	0.5
14	CO2_L_gasoil	OPA	normal	0.03		0	5	generic	1	1	1
15	CompositionFertilizer	OPA	normal	0.25		0	100	generic	1	1	1
16	CropAreaCAPRI	CAD	normal	0.1		0	inf	sub-national	0.8	0.5	0
17	CropProperties	OAD	normal	0.25		0	500	sub-national	0.5	0.5	0.2
18	DevCrop	CPA	normal	0.25		0	100	sub-national	0.8	0.5	0.2
19	EF_Fert_Prod	OPA	normal	0.25		0	10	national	1	0.5	0.2
20	EF1	EFN	Lognormal		0.28	0	0.5	sub-national	0.5	0.5	0.2
21	EF2	EFN	Lognormal		0.63	0	50	generic	1	1	1
22	EF4	EFN	Lognormal		0.82	0	0.5	generic	1	1	1
23	EF5	EFN	Lognormal		0.5	0	0.5	sub-national	0.5	0.5	0.2
24	EffFact	CPA	normal	0.25		0	5	generic	1	1	1
25	EU_ani	LAD	normal	0.1		0	inf	sub-national	0.5	0.5	0
26	Excr_EU	LPA	normal	0.1		0	200	sub-national	0.8	0.5	0
27	FAO_AnimalProd	LAD	normal	0.021		0	inf	national	1	0.5	0
28	FAO_LandAreas	CAD	normal	0.003		0	inf	national	1	0.8	0
29	FAO_NatCA_barley	CAD	normal	0.063		0	inf	national	1	0.8	0
30	FAO_NatCA_maize	CAD	normal	0.063		0	inf	national	1	0.8	0
31	FAO_NatCA_other	CAD	normal	0.063		0	inf	national	1	0.8	0
32	FAO_NatCA_soybean	CAD	normal	0.063		0	inf	national	1	0.8	0
33	FAO_NatCA_wheat	CAD	normal	0.063		0	inf	national	1	0.8	0

34	FAO_NatCP_barley	CAD	normal	0.122		0	inf	national	1	0.5	0
35	FAO_NatCP_maize	CAD	normal	0.122		0	inf	national	1	0.5	0
36	FAO_NatCP_other	CAD	normal	0.122		0	inf	national	1	0.5	0
37	FAO_NatCP_soybean	CAD	normal	0.122		0	inf	national	1	0.5	0
38	FAO_NatCP_wheat	CAD	normal	0.122		0	inf	national	1	0.5	0
39	FAO_NatFertilizer	OAD	normal	0.082		0	inf	national	1	0.5	0
40	feedset_Animals	LAD	normal	0.068		0	inf	national	1	0.5	0
41	feedset_Crops	LAD	normal	0.075		0	inf	national	1	0.5	0.2
42	FertilizerType	OPA	normal	0.571		0	inf	national	1	0.5	0
43	Fqatm	CPA	Lognormal		0.25	0	1	generic	1	1	1
44	Fqbiol	CPA	normal	0.25		0	1	generic	1	1	1
45	FQGraz	CPA	normal	0.25		0	1	generic	1	1	1
46	FracR	CPA	normal	0.25		0	1	sub-national	0.8	0.5	0.2
47	GrasCorrEU	CPA	normal	0.1		0	1	sub-national	0.8	0.5	0
48	Grass_Yield	CAD	normal	0.25		0	20000	sub-national	0.8	0.5	0
49	GrassCorrection	CPA	normal	0.5		0	1	sub-national	0.8	0.5	0.2
50	GrassYieldEstimate	CAD	normal	0.5		0	15000	sub-national	0.8	0.5	0.2
51	IPCC_Nexcretion	LPA	normal	0.25		0	200	sub-national	0.8	0.5	0.2
52	LandCoverMap	BFD	normal	0.1		0	inf	sub-national	0.8	0.5	0
53	LD_Buffaloes	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
54	LD_Cattle	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
55	LD_Chickens	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
56	LD_Goats	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
57	LD_Pigs	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
58	LD_Sheep	LAD	normal	0.25		0	inf	sub-national	0.5	0.5	0
59	LeachingStorage	EFN	normal	0.5		0	50	sub-national	0.8	0.5	0.2
60	LivestockCountryTotal	LAD	normal	0.05		0	inf	national	1	0.5	0
61	ManureSU_Burned	LPA	normal	0.25		0	1	continental	1	1	0
62	ManureSU_DailySpread	LPA	normal	0.25		0	1	continental	1	1	0
63	ManureSU_Digester	LPA	normal	0.25		0	1	continental	1	1	0
64	ManureSU_Drylot	LPA	normal	0.25		0	1	continental	1	1	0
65	ManureSU_Lagoon	LPA	normal	0.25		0	1	continental	1	1	0
66	ManureSU_Liquid	LPA	normal	0.25		0	1	continental	1	1	0
67	ManureSU_Other	LPA	normal	0.25		0	1	continental	1	1	0
68	ManureSU_Pasture	LPA	normal	0.25		0	1	continental	1	1	0
69	ManureSU_SolidStorage	LPA	normal	0.25		0	1	continental	1	1	0
70	N_deposition_data	OAD	normal	0.25		-200	10000	sub-national	0.8	0.2	0.2
71	N2_animal_EF	EFN	Lognormal		0.5	0	1	sub-national	0.8	0.5	0.2
72	N2O_grazing	EFN	Lognormal		0.57	0	0.5	sub-national	0.8	0.5	0.2
73	N2O_manure_storage	EFN	Lognormal		0.35	0	0.5	sub-national	0.5	0.5	0.2
74	NH3_animal_EF	EFN	normal	0.25		0	1	sub-national	0.8	0.8	0.5
75	NH3_fert_EF	EFN	normal	0.3		0	0.5	sub-national	0.8	0.5	0.2
76	Nharvest	CPA	normal	0.25		0	100	sub-national	0.8	0.5	0.2

77	Nindex	CPA	normal	0.25		0	20	generic	1	1	1
78	NO_animal_EF	EFN	Lognormal		0.5	0	0.5	sub-national	0.8	0.5	0.2
79	orgNGRaz	CPA	normal	0.25		0	1	generic	1	1	1
80	OrgNLiqMan	CPA	normal	0.5		0	1	generic	1	1	1
81	OrgNSolMan	CPA	normal	0.5		0	1	generic	1	1	1
82	PesticideUse	OAD	normal	0.158		0	100	national	1	0.5	0
83	Precipitation	BFD	normal	0.1		0	10000	sub-national	0.8	0.5	0.2
84	ShareFeed	LPA	normal	0.25		0	1	national	1	0.5	0
85	Temperature	BFD	normal	0.1		-50	50	sub-national	0.8	0.8	0.2
86	Texture_Clay	BFD	normal	0.25		0	1	sub-national	0.8	0.2	0
87	Texture_Loam	BFD	normal	0.25		0	1	sub-national	0.8	0.2	0
88	Texture_Sand	BFD	normal	0.25		0	1	sub-national	0.8	0.2	0
89	TractorDensity	OAD	normal	0.1		0	10000	national	1	0.8	0

Annex 3 Results of uncertainty quantification in model outputs

The tables below provide the results of uncertainty quantification in model outputs. The results are discussed in section 3.2. The uncertainty information is presented using mean (in ton CO_2-eq) and cv of the estimated emission.

3.1 Total GHG emissions

Continent	mean (ton CO_2-eq)				cv			
	Total	CH_4	N_2O	CO_2	Total	CH_4	N_2O	CO_2
Africa	5.4E+08	3.5E+08	1.8E+08	4.7E+06	0.17	0.18	0.37	0.31
Latin America	9.0E+08	6.3E+08	2.5E+08	1.5E+07	0.17	0.21	0.31	0.32
EU27	4.0E+08	2.4E+08	1.3E+08	3.3E+07	0.12	0.15	0.22	0.24

3.2 GHG emissions for livestock sectors

Continent	Sector	mean (ton CO_2-eq)				cv			
		Total	CH_4	N_2O	CO_2	Total	CH_4	N_2O	CO_2
Africa	Broilers	3.2E+06	4.8E+05	2.4E+06	3.2E+05	0.18	0.21	0.22	0.37
	Camels	2.7E+07	2.4E+07	2.9E+06		0.30	0.33	0.42	
	DairyCows	1.2E+08	7.1E+07	4.4E+07	2.6E+06	0.17	0.21	0.32	0.33
	Goats	6.0E+07	3.9E+07	2.1E+07	1.4E+05	0.25	0.33	0.39	0.37
	Horses	1.3E+07	9.2E+06	3.9E+06		0.25	0.32	0.41	
	LayingHens	2.2E+06	2.3E+05	1.7E+06	2.3E+05	0.19	0.21	0.22	0.37
	OtherAnimals	3.7E+05	1.3E+05	2.4E+05		0.22	0.21	0.32	
	OtherCattle	2.6E+08	1.7E+08	8.5E+07	9.0E+05	0.19	0.21	0.39	0.39
	OtherPoultry	1.4E+05	2.2E+04	1.2E+05		0.31	0.21	0.37	
	Pigs	5.0E+06	1.4E+06	3.3E+06	2.9E+05	0.19	0.20	0.28	0.37
	Sheep	5.5E+07	3.8E+07	1.7E+07	1.4E+05	0.18	0.21	0.39	0.37
	Turkeys	1.1E+05	8.5E+03	1.1E+05		0.33	0.21	0.35	
EU27	Broilers	1.1E+07	3.4E+05	7.7E+06	2.8E+06	0.20	0.18	0.22	0.23
	DairyCows	1.2E+08	8.0E+07	3.6E+07	8.8E+06	0.12	0.15	0.22	0.24
	Goats	3.8E+06	1.7E+06	1.7E+06	3.5E+05	0.17	0.30	0.21	0.23
	Horses	2.6E+06	1.8E+06	7.7E+05		0.19	0.25	0.26	
	LayingHens	7.8E+06	2.8E+05	5.5E+06	2.1E+06	0.20	0.18	0.23	0.23
	OtherAnimals	1.2E+06	7.3E+05	4.8E+05		0.15	0.18	0.26	
	OtherCattle	1.6E+08	1.1E+08	4.6E+07	8.2E+06	0.14	0.18	0.23	0.28
	OtherPoultry	5.0E+02	1.1E+02	3.8E+02		0.31	0.21	0.39	
	Pigs	5.8E+07	2.7E+07	2.2E+07	8.5E+06	0.14	0.16	0.25	0.23
	Sheep	3.5E+07	2.1E+07	1.2E+07	2.1E+06	0.14	0.20	0.22	0.25
	Turkeys	2.9E+03	1.1E+03	1.8E+03		0.26	0.23	0.37	
Latin America	Broilers	7.3E+06	1.1E+06	5.5E+06	8.0E+05	0.25	0.19	0.31	0.27
	Camels	2.7E+06	1.9E+06	8.6E+05		0.23	0.29	0.40	
	DairyCows	1.1E+08	7.4E+07	3.1E+07	3.7E+06	0.18	0.22	0.31	0.31

Goats	7.4E+06	4.5E+06	2.8E+06	1.4E+05	0.23	0.33	0.31	0.34
Horses	1.9E+07	1.4E+07	4.5E+06		0.25	0.31	0.38	
LayingHens	3.5E+06	3.1E+05	2.7E+06	4.5E+05	0.28	0.18	0.33	0.27
OtherAnimals	2.8E+06	1.4E+06	1.4E+06		0.21	0.20	0.35	
OtherCattle	7.2E+08	5.2E+08	1.9E+08	9.2E+06	0.19	0.21	0.35	0.35
OtherPoultry	3.6E+04	8.2E+03	2.8E+04		0.28	0.18	0.35	
Pigs	1.3E+07	4.3E+06	7.8E+06	5.9E+05	0.20	0.19	0.30	0.27
Sheep	1.8E+07	1.1E+07	6.4E+06	3.4E+05	0.18	0.21	0.31	0.34
Turkeys	2.4E+05	2.2E+04	2.2E+05		0.32	0.19	0.35	

3.3 GHG emission for livestock products

Continent	Product	Emission (kg CO_2-eq /kg product)				cv			
		Total	CH_4	N_2O	CO_2	Total	CH_4	N_2O	CO_2
Africa	CattleMeat	41.82	27.73	13.94	0.15	0.19	0.21	0.39	0.39
	CowMilk	4.12	2.48	1.55	0.09	0.17	0.21	0.32	0.33
	EGGS	0.85	0.09	0.66	0.09	0.20	0.20	0.22	0.38
	PigMeat	4.43	1.26	2.91	0.26	0.19	0.20	0.27	0.38
	PoultryMeat	0.80	0.12	0.60	0.08	0.18	0.20	0.22	0.38
	SheGoaMeat	43.77	29.22	14.44	0.11	0.18	0.20	0.39	0.38
	SheGoaMilk	21.70	14.49	7.16	0.05	0.18	0.20	0.39	0.38
EU27	CattleMeat	19.83	13.09	5.71	1.02	0.13	0.17	0.23	0.28
	CowMilk	0.84	0.54	0.24	0.06	0.12	0.15	0.21	0.24
	EGGS	1.17	0.04	0.82	0.31	0.20	0.18	0.23	0.23
	PigMeat	2.57	1.21	0.98	0.38	0.14	0.16	0.25	0.23
	PoultryMeat	1.20	0.04	0.85	0.31	0.20	0.17	0.22	0.23
	SheGoaMeat	35.96	20.89	12.75	2.30	0.13	0.18	0.21	0.25
	SheGoaMilk	7.96	4.63	2.82	0.51	0.13	0.18	0.22	0.25
Latin America	CattleMeat	44.30	31.94	11.79	0.57	0.18	0.21	0.34	0.35
	CowMilk	1.45	0.99	0.41	0.05	0.18	0.21	0.31	0.31
	EGGS	0.53	0.05	0.41	0.07	0.28	0.18	0.33	0.27
	PigMeat	2.04	0.69	1.25	0.09	0.20	0.19	0.30	0.28
	PoultryMeat	0.39	0.06	0.29	0.04	0.25	0.18	0.31	0.27
	SheGoaMeat	58.22	35.95	21.18	1.09	0.16	0.18	0.31	0.35
	SheGoaMilk	41.79	25.81	15.20	0.78	0.16	0.18	0.31	0.35

3.4 GHG emission for IPCC categories

Category	mean (ton CO_2-eq)			cv		
	Africa	Latin America	EU27	Africa	Latin America	EU27
CH4_rice			8.5E+04			0.19
CropArea	4.6E+05	3.5E+05	9.0E+04	0.34	0.28	0.15
EntericFermentation	3.4E+08	6.1E+08	1.9E+08	0.19	0.22	0.18
FertilizerProd	5.2E+06	2.0E+07	3.4E+07	0.44	0.42	0.37
FuelUse	2.0E+06	3.3E+06	1.4E+07	0.34	0.32	0.22
ManureManagement	2.0E+07	2.9E+07	5.4E+07	0.19	0.18	0.15
ManureManagementCH4	1.3E+07	1.7E+07	4.5E+07	0.2	0.19	0.18

ManureManagementN2O	6.9E+06	1.1E+07	9.4E+06	0.4	0.36	0.24
N2O_ByProducts	2.5E+06	4.6E+06	5.7E+06	0.34	0.42	0.37
N2O_grazing	1.4E+08	1.5E+08	2.0E+07	0.45	0.44	0.39
N2O_soil	2.6E+07	7.3E+07	8.0E+07	0.51	0.42	0.31
PesticideUse	1.0E+05	1.9E+06	1.3E+06	0.2	0.23	0.2
Total	5.4E+08	9.0E+08	4.0E+08	0.17	0.17	0.12

Annex 4 Results of uncertainty contribution from MIPs to model outputs

The tables below provide the results of the uncertainty contribution from MIPs to model outputs. The results are discussed in section 3.3. The groups of MIPs are defined in section 2.6 (

Table 12).

4.1 Total GHG emission

Continent	Emission	Uncertainty contribution (%)								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
Africa	Total	2	0	0	0	50	39	0	1	0
	CH_4	3	0	0	0	98	0	0	0	0
	N_2O	0	0	0	0	0	73	0	1	0
	CO_2	3	1	5	0	0	0	10	25	50
Latin America	Total	2	0	0	0	73	16	0	1	0
	CH_4	3	0	0	0	97	0	0	0	0
	N_2O	0	0	0	0	0	75	0	4	0
	CO_2	1	4	2	0	0	0	6	21	49
EU27	Total	4	0	0	0	49	37	0	17	8
	CH_4	5	0	0	0	96	0	0	0	0
	N_2O	2	0	0	0	0	88	1	23	5
	CO_2	5	1	6	0	0	0	0	34	58

4.2 GHG emissions for livestock sectors

Continent	Sector	Uncertainty contribution (%)																	
		Total									CH_4								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
	Broilers	3	1	2	0	3	30	1	17	13	3	0	0	0	97	0	0	0	0
	Camels	1	0	0	0	97	2	0	0	0	1	0	0	0	99	0	0	0	0
	DairyCows	2	0	0	0	58	36	0	2	0	2	0	0	0	100	0	0	0	0
	Goats	1	0	0	0	69	22	0	0	0	1	0	0	0	98	0	0	0	0
	Horses	1	0	0	0	71	16	0	0	0	1	0	0	0	97	0	0	0	0
	LayingHens	3	1	2	0	1	31	1	17	14	3	0	0	0	98	0	0	0	0
	OtherAnimals	1	0	0	0	10	50	0	8	0	3	0	0	0	98	0	0	0	0
	OtherCattle	1	0	0	0	57	36	0	0	0	2	0	0	0	100	0	0	0	0
	OtherPoultry	0	0	0	0	1	55	0	4	0	3	0	0	0	98	0	0	0	0
	Pigs	2	0	0	0	7	59	0	19	4	3	0	0	0	96	0	0	0	0
	Sheep	2	0	0	0	63	30	0	0	0	2	0	0	0	100	0	0	0	0
Africa	Turkeys	0	0	0	0	0	59	0	5	0	3	0	0	0	95	0	0	0	0
	Broilers	11	1	1	0	0	41	2	34	20	3	0	0	0	98	0	0	0	0
	DairyCows	4	0	0	0	60	28	0	12	6	4	0	0	0	96	0	0	0	0
	Goats	2	0	0	0	63	30	0	10	3	1	0	0	0	100	0	0	0	0
	Horses	1	0	0	0	79	15	0	0	0	1	0	0	0	100	0	0	0	0
	LayingHens	10	1	1	0	0	46	1	34	19	3	0	0	0	99	0	0	0	0
	OtherAnimals	9	0	0	0	50	39	0	5	0	6	0	0	0	94	0	0	0	0
	OtherCattle	3	0	0	0	66	26	0	10	3	4	0	0	0	96	0	0	0	0
	OtherPoultry	2	0	0	0	2	54	0	22	0	6	0	0	0	100	0	0	0	0
	Pigs	9	1	1	0	36	38	1	25	13	4	0	0	0	100	0	0	0	0
	Sheep	3	0	0	0	59	31	0	10	3	3	0	0	0	97	0	0	0	0
EU27	Turkeys	18	0	0	0	10	58	0	18	0	22	0	0	0	71	0	0	0	0

	Broilers	6	5	0	0	1	31	10	25	1	4	0	0	0	96	0	0	0	0
	Camels	2	0	0	0	82	21	0	1	0	2	0	0	0	98	0	0	0	0
	DairyCows	2	0	0	0	76	14	0	4	1	3	0	0	0	97	0	0	0	0
	Goats	1	0	0	0	76	17	0	1	0	1	0	0	0	99	0	0	0	0
	Horses	1	0	0	0	91	8	0	0	0	1	0	0	0	99	0	0	0	0
	LayingHens	6	5	0	0	0	31	11	24	1	4	0	0	0	95	0	0	0	0
	OtherAnimals	4	0	0	0	24	48	0	23	0	4	0	0	0	100	0	0	0	0
	OtherCattle	2	0	0	0	74	16	0	1	0	3	0	0	0	97	0	0	0	0
	OtherPoultry	1	0	0	0	2	70	0	20	0	5	0	0	0	94	0	0	0	0
	Pigs	3	1	0	0	11	52	2	21	0	3	0	0	0	95	0	0	0	0
	Sheep	2	0	0	0	64	26	0	2	0	2	0	0	0	98	0	0	0	0
Latin America	Turkeys	1	0	0	0	0	70	0	22	0	4	0	0	0	96	0	0	0	0

Continent	Sector	Uncertainty contribution (%)																	
		N2O									CO2								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
	Broilers	2	1	1	0	0	40	1	15	4	4	1	3	0	0	0	0	18	65
	Camels	0	0	0	0	0	71	0	0	0	0	0	0	0	0	0	0	0	0
	DairyCows	0	0	0	0	0	72	0	3	0	3	1	4	0	0	0	4	22	59
	Goats	0	0	0	0	0	75	0	0	0	0	2	9	0	0	0	53	23	9
	Horses	0	0	0	0	0	73	0	0	0	0	0	0	0	0	0	0	0	0
	LayingHens	2	1	1	0	0	41	1	15	4	4	1	3	0	0	0	0	18	65
	OtherAnimals	0	0	0	0	0	56	0	9	0	0	0	0	0	0	0	0	0	0
	OtherCattle	0	0	0	0	0	74	0	0	0	0	2	9	0	0	0	56	22	7
	OtherPoultry	0	0	0	0	0	55	0	4	0	0	0	0	0	0	0	0	0	0
	Pigs	1	0	0	0	0	75	0	22	1	4	1	3	0	0	0	0	18	65
	Sheep	0	0	0	0	0	72	0	1	0	0	2	9	0	0	0	53	23	9
Africa	Turkeys	0	0	0	0	0	59	0	5	0	0	0	0	0	0	0	0	0	0

	Broilers	7	1	0	0	0	67	3	24	8	12	2	8	0	0	0	0	36	45
	DairyCows	1	0	0	0	0	90	1	22	5	4	1	4	0	0	0	0	30	64
	Goats	1	0	0	0	0	90	0	18	2	4	1	12	0	0	0	1	38	44
	Horses	0	0	0	0	0	81	0	2	0	0	0	0	0	0	0	0	0	0
	LayingHens	7	1	0	0	0	72	2	23	8	12	2	8	0	0	0	0	37	46
	OtherAnimals	3	0	0	0	0	86	0	12	0	0	0	0	0	0	0	0	0	0
	OtherCattle	0	0	0	0	0	92	1	23	3	1	0	4	0	0	0	1	31	65
	OtherPoultry	1	0	0	0	0	55	0	23	0	0	0	0	0	0	0	0	0	0
	Pigs	6	1	0	0	0	76	2	22	7	12	2	9	0	0	0	0	38	46
	Sheep	1	0	0	0	0	89	0	18	2	3	0	9	0	0	0	0	38	50
EU27	Turkeys	7	0	0	0	0	71	0	22	0	0	0	0	0	0	0	0	0	0
	Broilers	4	4	0	0	0	36	10	23	0	14	11	3	0	0	0	6	38	15
	Camels	0	0	0	0	0	79	0	2	0	0	0	0	0	0	0	0	0	0
	DairyCows	0	0	0	0	0	60	0	12	1	1	4	2	0	0	0	6	22	48
	Goats	0	0	0	0	0	78	0	4	0	0	6	2	0	0	0	8	18	47
	Horses	0	0	0	0	0	79	0	3	0	0	0	0	0	0	0	0	0	0
	LayingHens	4	4	0	0	0	37	11	22	0	14	11	3	0	0	0	6	38	15
	OtherAnimals	1	0	0	0	0	61	0	29	0	0	0	0	0	0	0	0	0	0
	OtherCattle	0	0	0	0	0	80	0	2	0	0	6	2	0	0	0	8	16	48
	OtherPoultry	1	0	0	0	0	72	0	20	0	0	0	0	0	0	0	0	0	0
	Pigs	2	1	0	0	0	65	2	24	0	14	11	3	0	0	0	6	38	15
	Sheep	0	0	0	0	0	79	0	4	0	0	6	2	0	0	0	8	18	47
Latin America	Turkeys	1	0	0	0	0	71	0	23	0	0	0	0	0	0	0	0	0	0

4.3 GHG emissions for livestock products

Continent	Product	Uncertainty contribution (%)																
		Total								CH_4								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
Africa	CattleMeat	0	0	0	0	54	34	0	0	1	0	0	0	97	0	0	0	0
	CowMilk	1	0	0	0	55	34	0	2	1	0	0	0	97	0	0	0	0
	EGGS	4	1	2	0	1	30	1	16	1	0	0	0	98	0	0	0	0
	PigMeat	2	0	0	0	7	60	0	19	1	0	0	0	98	0	0	0	0
	PoultryMeat	4	1	1	0	3	29	1	16	1	0	0	0	97	0	0	0	0
	SheGoaMeat	0	0	0	0	50	36	0	0	1	0	0	0	98	0	0	0	0
	SheGoaMilk	1	0	0	0	49	36	0	0	1	0	0	0	97	0	0	0	0
EU27	CattleMeat	1	0	0	0	68	26	0	10	1	0	0	0	99	0	0	0	0
	CowMilk	1	0	0	0	61	28	0	12	1	0	0	0	99	0	0	0	0
	EGGS	11	1	1	0	0	45	1	33	1	0	0	0	101	0	0	0	0
	PigMeat	8	1	1	0	36	37	1	25	1	0	0	0	101	0	0	0	0
	PoultryMeat	11	1	1	0	0	41	2	34	1	0	0	0	101	0	0	0	0
	SheGoaMeat	1	0	0	0	54	36	0	12	1	0	0	0	99	0	0	0	0
	SheGoaMilk	1	0	0	0	54	36	0	12	1	0	0	0	99	0	0	0	0
LatinAmerica	CattleMeat	0	0	0	0	76	16	0	1	0	0	0	0	99	0	0	0	0
	CowMilk	0	0	0	0	76	14	0	4	0	0	0	0	98	0	0	0	0
	EGGS	5	5	0	0	0	31	11	24	1	0	0	0	99	0	0	0	0
	PigMeat	1	1	0	0	10	52	2	21	0	0	0	0	99	0	0	0	0
	PoultryMeat	5	5	0	0	1	31	10	25	0	0	0	0	99	0	0	0	0
	SheGoaMeat	0	0	0	0	55	33	0	2	0	0	0	0	99	0	0	0	0
	SheGoaMilk	0	0	0	0	55	33	0	2	1	0	0	0	99	0	0	0	0

Continent	Product	Uncertainty contribution (%)																	
		N_2O									CO_2								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA	LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
Africa	CattleMeat	0	0	0	0	0	73	0	0	0	2	2	9	0	0	0	54	21	6
	CowMilk	0	0	0	0	0	71	0	3	0	5	1	4	0	0	0	3	21	56
	EGGS	3	1	1	0	0	39	1	14	4	5	1	3	0	0	0	0	17	62
	PigMeat	1	0	0	0	0	75	0	22	1	5	1	3	0	0	0	0	17	63
	PoultryMeat	3	1	1	0	0	38	1	15	4	5	1	3	0	0	0	0	17	62
	SheGoaMeat	0	0	0	0	0	73	0	1	0	2	2	9	0	0	0	51	22	9
	SheGoaMilk	0	0	0	0	0	72	0	1	0	2	2	9	0	0	0	51	22	9
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EU27	CattleMeat	0	0	0	0	0	91	1	23	3	2	0	4	0	0	0	1	31	64
	CowMilk	1	0	0	0	0	89	1	22	5	5	1	4	0	0	0	0	29	63
	EGGS	7	1	0	0	0	71	2	23	7	13	2	8	0	0	0	0	36	44
	PigMeat	7	1	0	0	0	75	2	22	7	13	2	9	0	0	0	0	37	45
	PoultryMeat	8	1	0	0	0	66	3	24	8	13	2	7	0	0	0	0	35	44
	SheGoaMeat	1	0	0	0	0	90	0	18	2	4	0	9	0	0	0	1	37	48
	SheGoaMilk	1	0	0	0	0	89	0	17	2	4	0	9	0	0	0	1	37	48
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LatinAmerica	CattleMeat	0	0	0	0	0	80	0	2	0	1	6	2	0	0	0	8	16	45
	CowMilk	0	0	0	0	0	59	0	12	1	1	4	2	0	0	0	5	21	46
	EGGS	4	4	0	0	0	36	11	21	0	14	11	3	0	0	0	6	37	14
	PigMeat	1	1	0	0	0	64	2	24	0	14	11	3	0	0	0	6	37	14
	PoultryMeat	3	4	0	0	0	36	10	23	0	14	11	3	0	0	0	6	37	14
	SheGoaMeat	0	0	0	0	0	79	0	4	0	1	6	2	0	0	0	7	17	45
	SheGoaMilk	0	0	0	0	0	79	0	4	0	1	6	2	0	0	0	7	17	45

4.4 GHG emissions for IPCC categories

Continent	Category	Uncertainty contribution (%)								
		LAD	CAD	OAD	BFD	EFC	EFN	CPA	LPA	OPA
Africa	CropArea	0	0	0	0	0	0	68	25	0
	EntericFermentation	3	0	0	0	99	0	0	0	0
	FertilizerProd	2	1	3	0	0	0	0	13	71
	FuelUse	0	0	17	0	0	0	43	24	1
	ManureManagement	3	0	0	0	50	24	0	16	0
	ManureManagementCH4	3	0	0	0	97	0	0	0	0
	ManureManagementN2O	1	0	0	0	0	54	0	37	0
	N2O_ByProducts	3	8	0	0	0	43	7	16	0
	N2O_grazing	76								
	N2O_soil	0	1	0	0	0	99	0	11	0
	PesticideUse	16	15	2	0	0	0	0	66	0
	Total	2	0	0	0	50	39	0	1	0
	CH4_rice									
EU27	CropArea	5	1	0	0	0	0	1	92	0
	EntericFermentation	4	0	0	0	97	0	0	0	0
	FertilizerProd	2	0	0	0	0	0	0	13	85
	FuelUse	10	2	44	0	0	0	0	51	2
	ManureManagement	4	0	0	0	92	6	0	1	0
	ManureManagementCH4	3	0	0	0	99	0	0	0	0
	ManureManagementN2O	1	0	0	0	0	90	0	16	0
	N2O_ByProducts	4	2	0	0	0	52	19	15	0
	N2O_grazing	69								
	N2O_soil	1	0	0	0	0	99	0	20	0
	PesticideUse	18	3	12	0	0	0	0	63	0
	Total	4	0	0	0	49	37	0	17	8
	CH4_rice	21	5	0	0	0	0	0	71	0
LatinAmerica	CropArea	0	0	0	0	0	0	52	36	0
	EntericFermentation	3	0	0	0	97	0	0	0	0
	FertilizerProd	0	5	1	0	0	0	3	11	61
	FuelUse	1	0	34	0	0	0	28	29	1
	ManureManagement	4	0	0	0	39	44	0	14	0
	ManureManagementCH4	4	0	0	0	97	0	0	0	0
	ManureManagementN2O	1	0	0	0	0	75	0	24	0
	N2O_ByProducts	4	1	0	0	0	33	12	11	0
	N2O_grazing	86								
	N2O_soil	0	0	0	0	0	67	0	13	0
	PesticideUse	25	7	5	0	0	0	0	68	0
	Total	2	0	0	0	73	16	0	1	0
	CH4_rice									

Annex 5 R script for uncertainty quantification with census data

```
############################## Uncertainty quantification ####################
# script used for calculating the cv (RMSE method)  for data from FAO data base and to analysis the trend of the data
# SETTING ENVIRONMENT  ##
#clean-up memory:
rm(list = ls())
#load packages:
library(reshape)
library(matrixStats)
#set work dictionary:
setwd("M:/Thesis/1.UQ_in/FAOdata")
# set options (print warnings as they occur)
options(warn = 2)
############################## LOADING POINT
#1. [crpar]FAO_NatCropArea
# read input data
fao_crpar <- read.csv("FAO_NatCropArea.csv",header = TRUE,sep=",")
# make a year factor to document the years from which the data is used
year_crpar<-as.numeric(2000:2012)
#create a new data frame for storing the statistics
fao_crpar_stat<-fao_crpar
# calculate the mean and the sd
fao_crpar_stat$mean<-rowMeans(fao_crpar_stat[,3:15], na.rm = TRUE)
fao_crpar_stat$sd<-rowSds(as.matrix(fao_crpar_stat[,3:15]), na.rm = TRUE)
fao_crpar_stat$cv.original<-fao_crpar_stat$sd/fao_crpar_stat$mean
cv_old_crpar<-sum(fao_crpar_stat$sd,na.rm = TRUE)/
          sum(fao_crpar_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_crpar_stat$RMSE<-fao_crpar_stat$mean
fao_crpar_stat$cv.NRMSE<-fao_crpar_stat$mean
fao_crpar_stat$cv<-fao_crpar_stat$mean
fao_crpar_stat$cv.weight<-fao_crpar_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_crpar<-nrow(fao_crpar_stat)
for (i in 1:nr_crpar){
  if (!is.na(fao_crpar_stat[i,"sd"])){
    observation.i<-as.numeric(fao_crpar_stat[i,3:15],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_crpar,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_crpar","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_crpar
    #define the length of the year factor
    ylength<-length(year_crpar)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_crpar[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_crpar
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
```

```
    }}
   #define the length of RMSE.i withouth NAs
   N<-length(RMSE.i[!is.na(RMSE.i)])
   #calculate the RMSE
   fao_crpar_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
   #calculate the cv.NRMSE for each row
   fao_crpar_stat[i,"cv.NRMSE"]<-fao_crpar_stat[i,"RMSE"]/
                          fao_crpar_stat[i,"mean"]
   #choose the smaller one from cv.NRMSE and the cv.original as the cv
   #because the cv from being removed trend should be smaller
   fao_crpar_stat[i,"cv"]<-min(fao_crpar_stat[i,"cv.original"],
                         fao_crpar_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 13075:nr_crpar){
 if (is.na(fao_crpar_stat[i,"cv.original"])){
  fao_crpar_stat[i,"cv"]<-NA
 }
 #weigh the cv by the mean of this row
 fao_crpar_stat[i,"cv.weight"]<-fao_crpar_stat[i,"cv"]*
                         fao_crpar_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_crpar<-sum(fao_crpar_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_crpar<-sum(fao_crpar_stat$cv.weight,na.rm=TRUE)/sum_mean_crpar
#write to csv
write.csv(x = fao_crpar_stat,file = "stat_FAO_NatCropArea.csv",row.names=FALSE)
# 2. [crpyi]FAO_NatCropProd
#read input data
fao_crpyi<-read.csv("FAO_NatCropProd.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_crpyi<-as.numeric(2000:2012)
#create a new data frame for storing the statistics
fao_crpyi_stat<-fao_crpyi
#calculate the mean and the sd
fao_crpyi_stat$mean<-rowMeans(fao_crpyi_stat[,3:15], na.rm = TRUE)
fao_crpyi_stat$sd<-rowSds(as.matrix(fao_crpyi_stat[,3:15]), na.rm = TRUE)
fao_crpyi_stat$cv.original<-fao_crpyi_stat$sd/fao_crpyi_stat$mean
cv_old_crpyi<-sum(fao_crpyi_stat$sd,na.rm = TRUE)/
          sum(fao_crpyi_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_crpyi_stat$RMSE<-fao_crpyi_stat$mean
fao_crpyi_stat$cv.NRMSE<-fao_crpyi_stat$mean
fao_crpyi_stat$cv<-fao_crpyi_stat$mean
fao_crpyi_stat$cv.weight<-fao_crpyi_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_crpyi<-nrow(fao_crpyi_stat)
for (i in 1:nr_crpyi){
 if (!is.na(fao_crpyi_stat[i,"sd"])){
   observation.i<-as.numeric(fao_crpyi_stat[i,3:15],na.rm=TRUE)
   #get the linear least square fit line,skip the NA values
   fitline.i<-lm(observation.i~year_crpyi,na.action = na.omit)
   #get the formula of the fit line, expressed as y=ax+b
   summary<-summary(fitline.i)
   coefficients<-summary$coefficients
   a<-coefficients["year_crpyi","Estimate"]
   b<-coefficients["(Intercept)","Estimate"]
   #generic a list for storing the predictions
   prediction.i<-year_crpyi
   #define the length of the year factor
   ylength<-length(year_crpyi)
   #fill the list prediction.i with predicted values
   for (j in 1:ylength){
     prediction.i[[j]]<-a*year_crpyi[[j]]+b
   }
```

```
    #generic a list for storing the RMSE
    RMSE.i<-year_crpyi
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
      }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_crpyi_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_crpyi_stat[i,"cv.NRMSE"]<-fao_crpyi_stat[i,"RMSE"]/
                              fao_crpyi_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_crpyi_stat[i,"cv"]<-min(fao_crpyi_stat[i,"cv.original"],
                            fao_crpyi_stat[i,"cv.NRMSE"])
  }}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:nr_crpyi){
  if (is.na(fao_crpyi_stat[i,"cv.original"])){
    fao_crpyi_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_crpyi_stat[i,"cv.weight"]<-fao_crpyi_stat[i,"cv"]*
                            fao_crpyi_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_crpyi<-sum(fao_crpyi_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_crpyi<-sum(fao_crpyi_stat$cv.weight,na.rm=TRUE)/sum_mean_crpyi
#write to csv
write.csv(x = fao_crpyi_stat,file = "stat_FAO_NatCropProd.csv",row.names=FALSE)
# 3. [aninu]FAO_NatAnimals
#read input data
fao_aninu<-read.csv("FAO_NatAnimals.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_aninu<-as.numeric(2000:2011)
#create a new data frame for storing the statistics
fao_aninu_stat<-fao_aninu
#calculate the mean and the sd
fao_aninu_stat$mean<-rowMeans(fao_aninu_stat[,4:15], na.rm = TRUE)
fao_aninu_stat$sd<-rowSds(as.matrix(fao_aninu_stat[,4:15]), na.rm = TRUE)
fao_aninu_stat$cv.original<-fao_aninu_stat$sd/fao_aninu_stat$mean
cv_old_aninu<-sum(fao_aninu_stat$sd,na.rm = TRUE)/
          sum(fao_aninu_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_aninu_stat$RMSE<-fao_aninu_stat$mean
fao_aninu_stat$cv.NRMSE<-fao_aninu_stat$mean
fao_aninu_stat$cv<-fao_aninu_stat$mean
fao_aninu_stat$cv.weight<-fao_aninu_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_aninu<-nrow(fao_aninu_stat)
for (i in 1:nr_aninu){
  if (!is.na(fao_aninu_stat[i,"sd"])){
    observation.i<-as.numeric(fao_aninu_stat[i,4:15],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_aninu,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_aninu","Estimate"]
```

```
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_aninu
    #define the length of the year factor
    ylength<-length(year_aninu)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_aninu[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_aninu
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_aninu_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_aninu_stat[i,"cv.NRMSE"]<-fao_aninu_stat[i,"RMSE"]/
                            fao_aninu_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_aninu_stat[i,"cv"]<-min(fao_aninu_stat[i,"cv.original"],
                           fao_aninu_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_aninu){
  if (is.na(fao_aninu_stat[i,"cv.original"])){
    fao_aninu_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_aninu_stat[i,"cv.weight"]<-fao_aninu_stat[i,"cv"]*
    fao_aninu_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_aninu<-sum(fao_aninu_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_aninu<-sum(fao_aninu_stat$cv.weight,na.rm=TRUE)/sum_mean_aninu
#write to csv
write.csv(x = fao_aninu_stat,file = "stat_FAO_NatAnimals.csv",row.names=FALSE)
# 4. [lndar]FAO_LandAreas
#read input data
fao_lndar<-read.csv("FAO_LandAreas.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_lndar<-as.numeric(2000:2011)
#create a new data frame for storing the statistics
fao_lndar_stat<-fao_lndar
#calculate the mean and the sd
fao_lndar_stat$mean<-rowMeans(fao_lndar_stat[,3:14], na.rm = TRUE)
fao_lndar_stat$sd<-rowSds(as.matrix(fao_lndar_stat[,3:14]), na.rm = TRUE)
fao_lndar_stat$cv.original<-fao_lndar_stat$sd/fao_lndar_stat$mean
cv_old_lndar<-sum(fao_lndar_stat$sd,na.rm = TRUE)/
          sum(fao_lndar_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_lndar_stat$RMSE<-fao_lndar_stat$mean
fao_lndar_stat$cv.NRMSE<-fao_lndar_stat$mean
fao_lndar_stat$cv<-fao_lndar_stat$mean
fao_lndar_stat$cv.weight<-fao_lndar_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_lndar<-nrow(fao_lndar_stat)
```

```
for (i in 1:nr_lndar){
  if (!is.na(fao_lndar_stat[i,"sd"])){
    observation.i<-as.numeric(fao_lndar_stat[i,3:14],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_lndar,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_lndar","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_lndar
    #define the length of the year factor
    ylength<-length(year_lndar)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_lndar[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_lndar
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
      }
    }
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_lndar_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_lndar_stat[i,"cv.NRMSE"]<-fao_lndar_stat[i,"RMSE"]/
      fao_lndar_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_lndar_stat[i,"cv"]<-min(fao_lndar_stat[i,"cv.original"],
                                fao_lndar_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_lndar){
  if (is.na(fao_lndar_stat[i,"cv.original"])){
    fao_lndar_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_lndar_stat[i,"cv.weight"]<-fao_lndar_stat[i,"cv"]*
                                 fao_lndar_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_lndar<-sum(fao_lndar_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_lndar<-sum(fao_lndar_stat$cv.weight,na.rm=TRUE)/sum_mean_lndar
#write to csv
write.csv(x = fao_lndar_stat,file = "stat_FAO_LandAreas.csv",row.names=FALSE)
# 5. [anipr]FAO_AnimalProd
#read input data
fao_anipr<-read.csv("FAO_AnimalProd.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_anipr<-as.numeric(2000:2011)
#create a new data frame for storing the statistics
fao_anipr_stat<-fao_anipr
#calculate the mean and the sd
fao_anipr_stat$mean<-rowMeans(fao_anipr_stat[,3:14], na.rm = TRUE)
fao_anipr_stat$sd<-rowSds(as.matrix(fao_anipr_stat[,3:14]), na.rm = TRUE)
fao_anipr_stat$cv.original<-fao_anipr_stat$sd/fao_anipr_stat$mean
```

```
cv_old_anipr<-sum(fao_anipr_stat$sd,na.rm = TRUE)/
          sum(fao_anipr_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_anipr_stat$RMSE<-fao_anipr_stat$mean
fao_anipr_stat$cv.NRMSE<-fao_anipr_stat$mean
fao_anipr_stat$cv<-fao_anipr_stat$mean
fao_anipr_stat$cv.weight<-fao_anipr_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_anipr<-nrow(fao_anipr_stat)
for (i in 1:nr_anipr){
  if (!is.na(fao_anipr_stat[i,"sd"])){
    observation.i<-as.numeric(fao_anipr_stat[i,3:14],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_anipr,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_anipr","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_anipr
    #define the length of the year factor
    ylength<-length(year_anipr)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_anipr[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_anipr
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_anipr_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_anipr_stat[i,"cv.NRMSE"]<-fao_anipr_stat[i,"RMSE"]/
                              fao_anipr_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_anipr_stat[i,"cv"]<-min(fao_anipr_stat[i,"cv.original"],
                            fao_anipr_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_anipr){
  if (is.na(fao_anipr_stat[i,"cv.original"])){
    fao_anipr_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_anipr_stat[i,"cv.weight"]<-fao_anipr_stat[i,"cv"]*
                              fao_anipr_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_anipr<-sum(fao_anipr_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_anipr<-sum(fao_anipr_stat$cv.weight,na.rm=TRUE)/sum_mean_anipr
#write to csv
write.csv(x = fao_anipr_stat,file = "stat_FAO_AnimalProd.csv",row.names=FALSE)
# 6. [prdani]FAO_ProducingAnimals
#read input data
```

```
fao_prdani<-read.csv("FAO_ProducingAnimals.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_prdani<-as.numeric(2000:2011)
#create a new data frame for storing the statistics
fao_prdani_stat<-fao_prdani
#calculate the mean and the sd
fao_prdani_stat$mean<-rowMeans(fao_prdani_stat[,3:14], na.rm = TRUE)
fao_prdani_stat$sd<-rowSds(as.matrix(fao_prdani_stat[,3:14]), na.rm = TRUE)
fao_prdani_stat$cv.original<-fao_prdani_stat$sd/fao_prdani_stat$mean
cv_old_prdani<-sum(fao_prdani_stat$sd,na.rm = TRUE)/
           sum(fao_prdani_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_prdani_stat$RMSE<-fao_prdani_stat$mean
fao_prdani_stat$cv.NRMSE<-fao_prdani_stat$mean
fao_prdani_stat$cv<-fao_prdani_stat$mean
fao_prdani_stat$cv.weight<-fao_prdani_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_prdani<-nrow(fao_prdani_stat)
for (i in 1:nr_prdani){
  if (!is.na(fao_prdani_stat[i,"sd"])){
    observation.i<-as.numeric(fao_prdani_stat[i,3:14],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_prdani,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_prdani","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_prdani
    #define the length of the year factor
    ylength<-length(year_prdani)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_prdani[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_prdani
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_prdani_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_prdani_stat[i,"cv.NRMSE"]<-fao_prdani_stat[i,"RMSE"]/
                              fao_prdani_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_prdani_stat[i,"cv"]<-min(fao_prdani_stat[i,"cv.original"],
                             fao_prdani_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_prdani){
  if (is.na(fao_prdani_stat[i,"cv.original"])){
    fao_prdani_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_prdani_stat[i,"cv.weight"]<-fao_prdani_stat[i,"cv"]*
                              fao_prdani_stat[i,"mean"]
```

```
}
#get the sum of the means of each row from the input
sum_mean_prdani<-sum(fao_prdani_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_prdani<-sum(fao_prdani_stat$cv.weight,na.rm=TRUE)/sum_mean_prdani
#write to csv
write.csv(x = fao_prdani_stat,file = "stat_FAO_ProducingAnimals.csv",row.names=FALSE)
# 7. [natfrt]FAO_NatFertilizer
#read input data
fao_natfrt<-read.csv("FAO_NatFertilizer.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_natfrt<-as.numeric(2002:2011)
#create a new data frame for storing the statistics
fao_natfrt_stat<-fao_natfrt
#calculate the mean and the sd
fao_natfrt_stat$mean<-rowMeans(fao_natfrt_stat[,3:12], na.rm = TRUE)
fao_natfrt_stat$sd<-rowSds(as.matrix(fao_natfrt_stat[,3:12]), na.rm = TRUE)
fao_natfrt_stat$cv.original<-fao_natfrt_stat$sd/fao_natfrt_stat$mean
cv_old_natfrt<-sum(fao_natfrt_stat$sd,na.rm = TRUE)/
          sum(fao_natfrt_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_natfrt_stat$RMSE<-fao_natfrt_stat$mean
fao_natfrt_stat$cv.NRMSE<-fao_natfrt_stat$mean
fao_natfrt_stat$cv<-fao_natfrt_stat$mean
fao_natfrt_stat$cv.weight<-fao_natfrt_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_natfrt<-nrow(fao_natfrt_stat)
for (i in 1:nr_natfrt){
  if (!is.na(fao_natfrt_stat[i,"sd"])){
    observation.i<-as.numeric(fao_natfrt_stat[i,3:12],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_natfrt,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_natfrt","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_natfrt
    #define the length of the year factor
    ylength<-length(year_natfrt)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_natfrt[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_natfrt
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_natfrt_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_natfrt_stat[i,"cv.NRMSE"]<-fao_natfrt_stat[i,"RMSE"]/
                            fao_natfrt_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_natfrt_stat[i,"cv"]<-min(fao_natfrt_stat[i,"cv.original"],
                           fao_natfrt_stat[i,"cv.NRMSE"])
```

```
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_natfrt){
  if (is.na(fao_natfrt_stat[i,"cv.original"])){
    fao_natfrt_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_natfrt_stat[i,"cv.weight"]<-fao_natfrt_stat[i,"cv"]*
                                  fao_natfrt_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_natfrt<-sum(fao_natfrt_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_natfrt<-sum(fao_natfrt_stat$cv.weight,na.rm=TRUE)/sum_mean_natfrt
#write to csv
write.csv(x = fao_natfrt_stat,file = "stat_FAO_NatFertilizer.csv",row.names=FALSE)
# 8. [frttp]FertilizerType
#read input data
fao_frttp<-read.csv("FAO_FertilizerType.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_frttp<-as.numeric(2002:2011)
#create a new data frame for storing the statistics
fao_frttp_stat<-fao_frttp
#calculate the mean and the sd
fao_frttp_stat$mean<-rowMeans(fao_frttp_stat[,3:12], na.rm = TRUE)
fao_frttp_stat$sd<-rowSds(as.matrix(fao_frttp_stat[,3:12]), na.rm = TRUE)
fao_frttp_stat$cv.original<-fao_frttp_stat$sd/fao_frttp_stat$mean
cv_old_frttp<-sum(fao_frttp_stat$sd,na.rm = TRUE)/
           sum(fao_frttp_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_frttp_stat$RMSE<-fao_frttp_stat$mean
fao_frttp_stat$cv.NRMSE<-fao_frttp_stat$mean
fao_frttp_stat$cv<-fao_frttp_stat$mean
fao_frttp_stat$cv.weight<-fao_frttp_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_frttp<-nrow(fao_frttp_stat)
for (i in 1:nr_frttp){
  if (!is.na(fao_frttp_stat[i,"sd"])){
    observation.i<-as.numeric(fao_frttp_stat[i,3:12],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_frttp,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_frttp","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_frttp
    #define the length of the year factor
    ylength<-length(year_frttp)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_frttp[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_frttp
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
      }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
```

```
    #calculate the RMSE
    fao_frttp_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_frttp_stat[i,"cv.NRMSE"]<-fao_frttp_stat[i,"RMSE"]/
                            fao_frttp_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_frttp_stat[i,"cv"]<-min(fao_frttp_stat[i,"cv.original"],
                           fao_frttp_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_frttp){
  if (is.na(fao_frttp_stat[i,"cv.original"])){
    fao_frttp_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_frttp_stat[i,"cv.weight"]<-fao_frttp_stat[i,"cv"]*
                            fao_frttp_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_frttp<-sum(fao_frttp_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_frttp<-sum(fao_frttp_stat$cv.weight,na.rm=TRUE)/sum_mean_frttp
#write to csv
write.csv(x = fao_frttp_stat,file = "stat_FAO_FertilizerType.csv",row.names=FALSE)
# 9. [feedcr]feedset_Crops
#read input data
fao_feedcr<-read.csv("FAO_feedset_Crops.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_feedcr<-as.numeric(2000:2009)
#create a new data frame for storing the statistics
fao_feedcr_stat<-fao_feedcr
#calculate the mean and the sd
fao_feedcr_stat$mean<-rowMeans(fao_feedcr_stat[,3:12], na.rm = TRUE)
fao_feedcr_stat$sd<-rowSds(as.matrix(fao_feedcr_stat[,3:12]), na.rm = TRUE)
fao_feedcr_stat$cv.original<-fao_feedcr_stat$sd/fao_feedcr_stat$mean
cv_old_feedcr<-sum(fao_feedcr_stat$sd,na.rm = TRUE)/
  sum(fao_feedcr_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_feedcr_stat$RMSE<-fao_feedcr_stat$mean
fao_feedcr_stat$cv.NRMSE<-fao_feedcr_stat$mean
fao_feedcr_stat$cv<-fao_feedcr_stat$mean
fao_feedcr_stat$cv.weight<-fao_feedcr_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_feedcr<-nrow(fao_feedcr_stat)
for (i in 1:nr_feedcr){
  if (!is.na(fao_feedcr_stat[i,"sd"])){
    observation.i<-as.numeric(fao_feedcr_stat[i,3:12],na.rm=TRUE)
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_feedcr,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_feedcr","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_feedcr
    #define the length of the year factor
    ylength<-length(year_feedcr)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_feedcr[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_feedcr
    for (j in 1:ylength){
```

```
    if (!is.na(observation.i[[j]])){
      #if observation(i,j) is not NA,
      #calculate the (Pi-Oi)2 and store the result in list RMSE.i
      RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
    }else{
      RMSE.i[[j]]<-NA
  }}
  #define the length of RMSE.i withouth NAs
  N<-length(RMSE.i[!is.na(RMSE.i)])
  #calculate the RMSE
  fao_feedcr_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
  #calculate the cv.NRMSE for each row
  fao_feedcr_stat[i,"cv.NRMSE"]<-fao_feedcr_stat[i,"RMSE"]/
                          fao_feedcr_stat[i,"mean"]
  #choose the smaller one from cv.NRMSE and the cv.original as the cv
  #because the cv from being removed trend should be smaller
  fao_feedcr_stat[i,"cv"]<-min(fao_feedcr_stat[i,"cv.original"],
                         fao_feedcr_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_feedcr){
 if (is.na(fao_feedcr_stat[i,"cv.original"])){
  fao_feedcr_stat[i,"cv"]<-NA
 }
 #weigh the cv by the mean of this row
 fao_feedcr_stat[i,"cv.weight"]<-fao_feedcr_stat[i,"cv"]*
                         fao_feedcr_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_feedcr<-sum(fao_feedcr_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_feedcr<-sum(fao_feedcr_stat$cv.weight,na.rm=TRUE)/sum_mean_feedcr
#write to csv
write.csv(x = fao_feedcr_stat,file = "stat_FAO_feedset_Crops.csv",row.names=FALSE)
# 10. [feedani]feedset_Animals
#read input data
fao_feedani<-read.csv("FAO_feedset_Animals.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_feedani<-as.numeric(2000:2009)
#create a new data frame for storing the statistics
fao_feedani_stat<-fao_feedani
#calculate the mean and the sd
fao_feedani_stat$mean<-rowMeans(fao_feedani_stat[,3:12], na.rm = TRUE)
fao_feedani_stat$sd<-rowSds(as.matrix(fao_feedani_stat[,3:12]), na.rm = TRUE)
fao_feedani_stat$cv.original<-fao_feedani_stat$sd/fao_feedani_stat$mean
cv_old_feedani<-sum(fao_feedani_stat$sd,na.rm = TRUE)/
            sum(fao_feedani_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_feedani_stat$RMSE<-fao_feedani_stat$mean
fao_feedani_stat$cv.NRMSE<-fao_feedani_stat$mean
fao_feedani_stat$cv<-fao_feedani_stat$mean
fao_feedani_stat$cv.weight<-fao_feedani_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_feedani<-nrow(fao_feedani_stat)
for (i in 1:nr_feedani){
 if (!is.na(fao_feedani_stat[i,"sd"])){
  observation.i<-as.numeric(fao_feedani_stat[i,3:12],na.rm=TRUE)
  #get the linear least square fit line,skip the NA values
  fitline.i<-lm(observation.i~year_feedani,na.action = na.omit)
  #get the formula of the fit line, expressed as y=ax+b
  summary<-summary(fitline.i)
  coefficients<-summary$coefficients
  a<-coefficients["year_feedani","Estimate"]
  b<-coefficients["(Intercept)","Estimate"]
  #generic a list for storing the predictions
  prediction.i<-year_feedani
```

```
    #define the length of the year factor
    ylength<-length(year_feedani)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_feedani[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_feedani
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_feedani_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_feedani_stat[i,"cv.NRMSE"]<-fao_feedani_stat[i,"RMSE"]/
                                fao_feedani_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_feedani_stat[i,"cv"]<-min(fao_feedani_stat[i,"cv.original"],
                               fao_feedani_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_feedani){
  if (is.na(fao_feedani_stat[i,"cv.original"])){
    fao_feedani_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_feedani_stat[i,"cv.weight"]<-fao_feedani_stat[i,"cv"]*
                               fao_feedani_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_feedani<-sum(fao_feedani_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_feedani<-sum(fao_feedani_stat$cv.weight,na.rm=TRUE)/sum_mean_feedani
#write to csv
write.csv(x = fao_feedani_stat,file = "stat_FAO_feedset_Animals.csv",row.names=FALSE)
# 11.[pstuse]PesticideUse
#read input data
fao_pstuse<-read.csv("FAO_PesticideUse.csv",header = TRUE,sep=",")
#make a year factor to document the years from which the data is used
year_pstuse<-as.numeric(2000:2010)
#create a new data frame for storing the statistics
fao_pstuse_stat<-fao_pstuse
#calculate the mean and the sd
fao_pstuse_stat$mean<-rowMeans(fao_pstuse_stat[,2:12], na.rm = TRUE)
fao_pstuse_stat$sd<-rowSds(as.matrix(fao_pstuse_stat[,2:12]), na.rm = TRUE)
fao_pstuse_stat$cv.original<-fao_pstuse_stat$sd/fao_pstuse_stat$mean
cv_old_pstuse<-sum(fao_pstuse_stat$sd,na.rm = TRUE)/
            sum(fao_pstuse_stat$mean,na.rm = TRUE)
#define other columns for storing the results
fao_pstuse_stat$RMSE<-fao_pstuse_stat$mean
fao_pstuse_stat$cv.NRMSE<-fao_pstuse_stat$mean
fao_pstuse_stat$cv<-fao_pstuse_stat$mean
fao_pstuse_stat$cv.weight<-fao_pstuse_stat$mean
#write a loop to get individual cv for each row in the input data
#define the length of rows
nr_pstuse<-nrow(fao_pstuse_stat)
for (i in 1:nr_pstuse){
  if (!is.na(fao_pstuse_stat[i,"sd"])){
    observation.i<-as.numeric(fao_pstuse_stat[i,2:12],na.rm=TRUE)
```

```
    #get the linear least square fit line,skip the NA values
    fitline.i<-lm(observation.i~year_pstuse,na.action = na.omit)
    #get the formula of the fit line, expressed as y=ax+b
    summary<-summary(fitline.i)
    coefficients<-summary$coefficients
    a<-coefficients["year_pstuse","Estimate"]
    b<-coefficients["(Intercept)","Estimate"]
    #generic a list for storing the predictions
    prediction.i<-year_pstuse
    #define the length of the year factor
    ylength<-length(year_pstuse)
    #fill the list prediction.i with predicted values
    for (j in 1:ylength){
      prediction.i[[j]]<-a*year_pstuse[[j]]+b
    }
    #generic a list for storing the RMSE
    RMSE.i<-year_pstuse
    for (j in 1:ylength){
      if (!is.na(observation.i[[j]])){
        #if observation(i,j) is not NA,
        #calculate the (Pi-Oi)2 and store the result in list RMSE.i
        RMSE.i[[j]]<-(prediction.i[[j]]-observation.i[[j]])^2
      }else{
        RMSE.i[[j]]<-NA
    }}
    #define the length of RMSE.i withouth NAs
    N<-length(RMSE.i[!is.na(RMSE.i)])
    #calculate the RMSE
    fao_pstuse_stat[i,"RMSE"]<-sqrt(sum(RMSE.i,na.rm=TRUE)/N)
    #calculate the cv.NRMSE for each row
    fao_pstuse_stat[i,"cv.NRMSE"]<-fao_pstuse_stat[i,"RMSE"]/
                                fao_pstuse_stat[i,"mean"]
    #choose the smaller one from cv.NRMSE and the cv.original as the cv
    #because the cv from being removed trend should be smaller
    fao_pstuse_stat[i,"cv"]<-min(fao_pstuse_stat[i,"cv.original"],
                               fao_pstuse_stat[i,"cv.NRMSE"])
}}
#but if the cv.original is NA, the cv should remail NA
for (i in 1:1:nr_pstuse){
  if (is.na(fao_pstuse_stat[i,"cv.original"])){
    fao_pstuse_stat[i,"cv"]<-NA
  }
  #weigh the cv by the mean of this row
  fao_pstuse_stat[i,"cv.weight"]<-fao_pstuse_stat[i,"cv"]*
                               fao_pstuse_stat[i,"mean"]
}
#get the sum of the means of each row from the input
sum_mean_pstuse<-sum(fao_pstuse_stat[,"mean"],na.rm=TRUE)
#get average cv for fao crop area data
ave_weightcv_pstuse<-sum(fao_pstuse_stat$cv.weight,na.rm=TRUE)/sum_mean_pstuse
#write to csv
write.csv(x = fao_pstuse_stat,file = "stat_FAO_PesticideUse.csv",row.names=FALSE)
# write overview to text
sink('overview_RMSE.txt')
cat('######################fao data overview######################')
cat('\n')
cat('\n')
cat('[FAO_NatCropArea]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_crpar)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_crpar)
cat('\n')
cat('\n')
cat('[FAO_NatCropProd]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_crpyi)
```

```
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_crpyi)
cat('\n')
cat('\n')
cat('[FAO_NatAnimals]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_aninu)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_aninu)
cat('\n')
cat('\n')
cat('[FAO_LandAreas]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_lndar)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_lndar)
cat('\n')
cat('\n')
cat('[FAO_AnimalProd]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_anipr)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_anipr)
cat('\n')
cat('\n')
cat('[FAO_ProducingAnimals]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_prdani)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_prdani)
cat('\n')
cat('\n')
cat('[FAO_NatFertilizer]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_natfrt)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_natfrt)
cat('\n')
cat('\n')
cat('[FertilizerType]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_frttp)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_frttp)
cat('\n')
cat('\n')
cat('[feedset_Crops]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_feedcr)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_feedcr)
cat('\n')
cat('\n')
cat('[feedset_Animals]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_feedani)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_feedani)
cat('\n')
cat('\n')
cat('[PesticideUse]')
cat('\n')
cat('average weight cv without corrected by trend',sep='\t',cv_old_pstuse)
cat('\n')
cat('average weight cv corrected by trend with RMSE method',sep='\t',ave_weightcv_pstuse)
sink()
```

```
############################## Cross correlation calculation ####################
# script used for calculating the cross correlation for potential correlated model input pairs
#  SETTING ENVIRONMENT
#clean-up memory:
rm(list = ls())
#load packages:
library(psych)
library(matrixStats)
#set work dictionary:
setwd("D:/Thesis_BQ/1.UQ_in/cross correlation")
# set options (print warnings as they occur)
options(warn = 2)
############################## LOADING POINT
# 1. [FAO_NatCropArea](crpar) and [FAO_NatCropProd](crppr)
#read input data
fao_crpar<-read.csv("FAO_NatCropArea.csv",header = TRUE,sep=",")
fao_crppr<-read.csv("FAO_NatCropProd.csv",header = TRUE,sep=",")
#merge the two data set by country code and crop code
crpar_crppr<-merge(fao_crpar,fao_crppr,by=c('COUNTRYCODE','CROPCODE'))
#write a loop to get the cross correlation coefficient between crop area and crop production
#compare all the area and production for each year
#generate a data frame to store the results
corr_crpar_crppr<-data.frame("year"=2000:2012,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:13){
  crpar<-as.numeric(crpar_crppr[,i+2])
  crppr<-as.numeric(crpar_crppr[,i+15])
  crpar_crppr_inloop<-cbind(crpar,crppr)
  crpar_crppr_inloop<-na.omit(crpar_crppr_inloop)
  crpar<-crpar_crppr_inloop[,1]
  crppr<-crpar_crppr_inloop[,2]
  corr<- cor.test(crpar,crppr)
  corr_crpar_crppr[i,"cross_correlation"]<-corr$estimate
  corr_crpar_crppr[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_crpar_crppr<-na.omit(corr_crpar_crppr)
corr_crpar_crppr[i+1,"year"]<-"mean"
corr_crpar_crppr[i+1,"cross_correlation"]<-mean(omit.corr_crpar_crppr[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_crpar_crppr<-subset(corr_crpar_crppr,p_value<=0.05)
sig_corr_crpar_crppr<-na.omit(sig_corr_crpar_crppr)
corr_crpar_crppr[i+2,"year"]<-"mean_significant"
corr_crpar_crppr[i+2,"cross_correlation"]<-mean(sig_corr_crpar_crppr[,"cross_correlation"])
# 2. [FAO_NatFertilizer](natfrt) and [FAO_NatCropProd](crppr
#read input data
fao_natfrt<-read.csv("FAO_NatFertilizer.csv",header = TRUE,sep=",")
fao_crppr<-read.csv("FAO_NatCropProd.csv",header = TRUE,sep=",")
#aggregate each input by the country code
agg_natfrt<-aggregate(fao_natfrt,by=list(fao_natfrt$FaoTerritory),FUN=mean, na.rm=TRUE)
#agg_natfrt<-fao_natfrt[fao_natfrt$FaoItemCodes_IA==3102,]
#agg_natfrt<-fao_natfrt[fao_natfrt$FaoItemCodes_IA==3103,]
#agg_natfrt<-fao_natfrt[fao_natfrt$FaoItemCodes_IA==3104,]
agg_crppr<-aggregate(fao_crppr,by=list(fao_crppr$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#rename the natfrt first two columns
colnames(agg_natfrt)[1]<-"COUNTRYCODE"
#merge the two data set by country code and crop code
natfrt_crppr<-merge(agg_natfrt,agg_crppr,by='COUNTRYCODE')
natfrt_crppr<-na.omit(natfrt_crppr)
#generate a data frame to store the results
corr_natfrt_crppr<-data.frame("year"=2002:2011,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:10){
  natfrt<-as.numeric(natfrt_crppr[,i+3])
  crppr<-as.numeric(natfrt_crppr[,i+17])
  natfrt_crppr_inloop<-cbind(natfrt,crppr)
  natfrt_crppr_inloop<-na.omit(natfrt_crppr_inloop)
  natfrt<-natfrt_crppr_inloop[,1]
  crppr<-natfrt_crppr_inloop[,2]
```

```
  corr<- cor.test(natfrt,crppr)
  corr_natfrt_crppr[i,"cross_correlation"]<-corr$estimate
  corr_natfrt_crppr[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_natfrt_crppr<-na.omit(corr_natfrt_crppr)
corr_natfrt_crppr[i+1,"year"]<-"mean"
corr_natfrt_crppr[i+1,"cross_correlation"]<-mean(omit.corr_natfrt_crppr[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_natfrt_crppr<-subset(corr_natfrt_crppr,p_value<=0.05)
sig_corr_natfrt_crppr<-na.omit(sig_corr_natfrt_crppr)
corr_natfrt_crppr[i+2,"year"]<-"mean_significant"
corr_natfrt_crppr[i+2,"cross_correlation"]<-mean(sig_corr_natfrt_crppr[,"cross_correlation"])
# 3. [FAO_AnimalProd](anipr) and [FAO_ProducingAnimals](prani
#read input data
fao_anipr<-read.csv("FAO_AnimalProd.csv",header = TRUE,sep=",")
fao_prani<-read.csv("FAO_ProducingAnimals.csv",header = TRUE,sep=",")
#aggregate each input by the country code
agg_anipr<-aggregate(fao_anipr,by=list(fao_anipr$FaoTerritory),FUN=mean, na.rm=TRUE)
agg_prani<-aggregate(fao_prani,by=list(fao_prani$FaoTerritory),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
anipr_prani<-merge(agg_anipr,agg_prani,by='FaoTerritory')
anipr_prani<-na.omit(anipr_prani)
#generate a data frame to store the results
corr_anipr_prani<-data.frame("year"=2000:2011,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:12){
  anipr<-as.numeric(anipr_prani[,i+3])
  prani<-as.numeric(anipr_prani[,i+17])
  anipr_prani_inloop<-cbind(anipr,prani)
  anipr_prani_inloop<-na.omit(anipr_prani_inloop)
  anipr<-anipr_prani_inloop[,1]
  prani<-anipr_prani_inloop[,2]
  corr<- cor.test(anipr,prani)
  corr_anipr_prani[i,"cross_correlation"]<-corr$estimate
  corr_anipr_prani[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_anipr_prani<-na.omit(corr_anipr_prani)
corr_anipr_prani[i+1,"year"]<-"mean"
corr_anipr_prani[i+1,"cross_correlation"]<-mean(omit.corr_anipr_prani[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_anipr_prani<-subset(corr_anipr_prani,p_value<=0.05)
sig_corr_anipr_prani<-na.omit(sig_corr_anipr_prani)
corr_anipr_prani[i+2,"year"]<-"mean_significant"
corr_anipr_prani[i+2,"cross_correlation"]<-mean(sig_corr_anipr_prani[,"cross_correlation"])
# 4. [FAO_AnimalProd](anipr) and [FAO_NatAnimals](natani)
#read input data
fao_anipr<-read.csv("FAO_AnimalProd.csv",header = TRUE,sep=",")
fao_natani<-read.csv("FAO_NatAnimals.csv",header = TRUE,sep=",")
#rename the natfrt first two columns
colnames(fao_anipr)[1]<-"COUNTRYCODE"
colnames(fao_natani)[1]<-"COUNTRYCODE"
#aggregate each input by the country code
agg_anipr<-aggregate(fao_anipr,by=list(fao_anipr$COUNTRYCODE),FUN=mean, na.rm=TRUE)
agg_natani<-aggregate(fao_natani,by=list(fao_natani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
anipr_natani<-merge(agg_anipr,agg_natani,by='COUNTRYCODE')
anipr_natani<-na.omit(anipr_natani)
#generate a data frame to store the results
corr_anipr_natani<-data.frame("year"=2000:2011,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:12){
  anipr<-as.numeric(anipr_natani[,i+3])
  natani<-as.numeric(anipr_natani[,i+18])
  anipr_natani_inloop<-cbind(anipr,natani)
  anipr_natani_inloop<-na.omit(anipr_natani_inloop)
  anipr<-anipr_natani_inloop[,1]
  natani<-anipr_natani_inloop[,2]
```

```
  corr<- cor.test(anipr,natani)
  corr_anipr_natani[i,"cross_correlation"]<-corr$estimate
  corr_anipr_natani[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_anipr_natani<-na.omit(corr_anipr_natani)
corr_anipr_natani[i+1,"year"]<-"mean"
corr_anipr_natani[i+1,"cross_correlation"]<-mean(omit.corr_anipr_natani[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_anipr_natani<-subset(corr_anipr_natani,p_value<=0.05)
sig_corr_anipr_natani<-na.omit(sig_corr_anipr_natani)
corr_anipr_natani[i+2,"year"]<-"mean_significant"
corr_anipr_natani[i+2,"cross_correlation"]<-mean(sig_corr_anipr_natani[,"cross_correlation"])
# 5. [FAO_ProducingAnimals](prani) and [FAO_NatAnimals](natani)
#read input data
fao_prani<-read.csv("FAO_ProducingAnimals.csv",header = TRUE,sep=",")
fao_natani<-read.csv("FAO_NatAnimals.csv",header = TRUE,sep=",")
#rename the natfrt first two columns
colnames(fao_prani)[1]<-"COUNTRYCODE"
colnames(fao_natani)[1]<-"COUNTRYCODE"
#aggregate each input by the country code
agg_prani<-aggregate(fao_prani,by=list(fao_prani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
agg_natani<-aggregate(fao_natani,by=list(fao_natani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
prani_natani<-merge(agg_prani,agg_natani,by='COUNTRYCODE')
prani_natani<-na.omit(prani_natani)

#generate a data frame to store the results
corr_prani_natani<-data.frame("year"=2000:2011,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:12){
  prani<-as.numeric(prani_natani[,i+3])
  natani<-as.numeric(prani_natani[,i+18])
  prani_natani_inloop<-cbind(prani,natani)
  prani_natani_inloop<-na.omit(prani_natani_inloop)
  prani<-prani_natani_inloop[,1]
  natani<-prani_natani_inloop[,2]
  corr<- cor.test(prani,natani)
  corr_prani_natani[i,"cross_correlation"]<-corr$estimate
  corr_prani_natani[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_prani_natani<-na.omit(corr_prani_natani)
corr_prani_natani[i+1,"year"]<-"mean"
corr_prani_natani[i+1,"cross_correlation"]<-mean(omit.corr_prani_natani[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_prani_natani<-subset(corr_prani_natani,p_value<=0.05)
sig_corr_prani_natani<-na.omit(sig_corr_prani_natani)
corr_prani_natani[i+2,"year"]<-"mean_significant"
corr_prani_natani[i+2,"cross_correlation"]<-mean(sig_corr_prani_natani[,"cross_correlation"])
# 6. [FAO_feedset_Crops](feedcr) and [FAO_NatAnimals](natani)
#read input data
fao_feedcr<-read.csv("FAO_feedset_Crops.csv",header = TRUE,sep=",")
fao_natani<-read.csv("FAO_NatAnimals.csv",header = TRUE,sep=",")
#rename the natfrt first two columns
colnames(fao_feedcr)[1]<-"COUNTRYCODE"
colnames(fao_natani)[1]<-"COUNTRYCODE"
#aggregate each input by the country code
agg_feedcr<-aggregate(fao_feedcr,by=list(fao_feedcr$COUNTRYCODE),FUN=mean, na.rm=TRUE)
agg_natani<-aggregate(fao_natani,by=list(fao_natani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
feedcr_natani<-merge(agg_feedcr,agg_natani,by='COUNTRYCODE')
feedcr_natani<-na.omit(feedcr_natani)
#generate a data frame to store the results
corr_feedcr_natani<-data.frame("year"=2000:2009,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:10){
  feedcr<-as.numeric(feedcr_natani[,i+3])
  natani<-as.numeric(feedcr_natani[,i+16])
```

```
  feedcr_natani_inloop<-cbind(feedcr,natani)
  feedcr_natani_inloop<-na.omit(feedcr_natani_inloop)
  feedcr<-feedcr_natani_inloop[,1]
  natani<-feedcr_natani_inloop[,2]
  corr<- cor.test(feedcr,natani)
  corr_feedcr_natani[i,"cross_correlation"]<-corr$estimate
  corr_feedcr_natani[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_feedcr_natani<-na.omit(corr_feedcr_natani)
corr_feedcr_natani[i+1,"year"]<-"mean"
corr_feedcr_natani[i+1,"cross_correlation"]<-mean(omit.corr_feedcr_natani[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_feedcr_natani<-subset(corr_feedcr_natani,p_value<=0.05)
sig_corr_feedcr_natani<-na.omit(sig_corr_feedcr_natani)
corr_feedcr_natani[i+2,"year"]<-"mean_significant"
corr_feedcr_natani[i+2,"cross_correlation"]<-mean(sig_corr_feedcr_natani[,"cross_correlation"])
# 7. [FAO_feedset_Animals](feedani) and [FAO_NatAnimals](natani)
#read input data
fao_feedani<-read.csv("FAO_feedset_Animals.csv",header = TRUE,sep=",")
fao_natani<-read.csv("FAO_NatAnimals.csv",header = TRUE,sep=",")
#rename the natfrt first two columns
colnames(fao_feedani)[1]<-"COUNTRYCODE"
colnames(fao_natani)[1]<-"COUNTRYCODE"
#aggregate each input by the country code
agg_feedani<-aggregate(fao_feedani,by=list(fao_feedani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
agg_natani<-aggregate(fao_natani,by=list(fao_natani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
feedani_natani<-merge(agg_feedani,agg_natani,by='COUNTRYCODE')
feedani_natani<-na.omit(feedani_natani)
#generate a data frame to store the results
corr_feedani_natani<-data.frame("year"=2000:2009,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:10){
  feedani<-as.numeric(feedani_natani[,i+3])
  natani<-as.numeric(feedani_natani[,i+16])
  feedani_natani_inloop<-cbind(feedani,natani)
  feedani_natani_inloop<-na.omit(feedani_natani_inloop)
  feedani<-feedani_natani_inloop[,1]
  natani<-feedani_natani_inloop[,2]
  corr<- cor.test(feedani,natani)
  corr_feedani_natani[i,"cross_correlation"]<-corr$estimate
  corr_feedani_natani[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_feedani_natani<-na.omit(corr_feedani_natani)
corr_feedani_natani[i+1,"year"]<-"mean"
corr_feedani_natani[i+1,"cross_correlation"]<-mean(omit.corr_feedani_natani[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_feedani_natani<-subset(corr_feedani_natani,p_value<=0.05)
sig_corr_feedani_natani<-na.omit(sig_corr_feedani_natani)
corr_feedani_natani[i+2,"year"]<-"mean_significant"
corr_feedani_natani[i+2,"cross_correlation"]<-mean(sig_corr_feedani_natani[,"cross_correlation"])
# 8. [LivestockDensity](lstden) and [FAO_natani_lsdmals](natani_lsd)
#read input data
livestockDensity<-read.csv("LivestockDensity.csv",header = TRUE,sep=",")
fao_natani_lsd<-read.csv("FAO_NatAnimals_lsd.csv",header = TRUE,sep=",")
spatial_code<-read.csv("spatial_code.csv",header = TRUE,sep=",")
#re-order the livestockDensity input
livestockDensity<-merge(livestockDensity,spatial_code,by="region")
livestockDensity[[1]]<-livestockDensity$FAOcountry
colnames(livestockDensity)[1]<-"COUNTRY"
livestockDensity<-livestockDensity[,-8]
#aggregate the livestockdensity data into country level
agglstden<-aggregate(livestockDensity[,2:6],by=list(livestockDensity$COUNTRY),"sum")
lstden_chickens<-aggregate(as.numeric(livestockDensity[,7]),by=list(livestockDensity$COUNTRY),"sum")
agglstden[,"Chicken"]<-lstden_chickens[,2]
colnames(agglstden)[1]<-"COUNTRY"
```

```
#merge the two data set by country code and crop code
lstden_natani_lsd<-merge(agglstden,fao_natani_lsd,by='COUNTRY')
lstden_natani_lsd<-na.omit(lstden_natani_lsd)
#generate a data frame to store the results
corr_lstden_natani_lsd<-data.frame("AnimalType"=c("Buffaloes","Cattle","Goats","Pigs","Sheep",
                                    "Chicken","mean","mean_significant"),
                                "cross_correlation"=NA,"p_value"=NA)
#rownames(corr_lstden_natani_lsd)<-1:8
for (i in 1:6){
  lstden<-as.numeric(lstden_natani_lsd[,i+1])
  natani_lsd<-as.numeric(lstden_natani_lsd[,i+7])
  lstden_natani_lsd_inloop<-cbind(lstden,natani_lsd)
  lstden_natani_lsd_inloop<-na.omit(lstden_natani_lsd_inloop)
  lstden<-lstden_natani_lsd_inloop[,1]
  natani_lsd<-lstden_natani_lsd_inloop[,2]
  corr<- cor.test(lstden,natani_lsd)
  corr_lstden_natani_lsd[i,"cross_correlation"]<-corr$estimate
  corr_lstden_natani_lsd[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
corr_lstden_natani_lsd[7,"cross_correlation"]<-mean(corr_lstden_natani_lsd[1:6,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_lstden_natani_lsd<-subset(corr_lstden_natani_lsd,p_value<=0.05)
corr_lstden_natani_lsd[8,"cross_correlation"]<-mean(sig_corr_lstden_natani_lsd[,"cross_correlation"])
# 8.1 [LivestockDensity](mlstden) and [FAO_natani_lsdmals](natani_lsd) ###  mean not sum
#read input data
livestockDensity<-read.csv("LivestockDensity.csv",header = TRUE,sep=",")
fao_natani_lsd<-read.csv("FAO_NatAnimals_lsd.csv",header = TRUE,sep=",")
spatial_code<-read.csv("spatial_code.csv",header = TRUE,sep=",")
#re-order the livestockDensity input
livestockDensity<-merge(livestockDensity,spatial_code,by="region")
livestockDensity[[1]]<-livestockDensity$FAOcountry
colnames(livestockDensity)[1]<-"COUNTRY"
livestockDensity<-livestockDensity[,-8]
#aggregate the livestockdensity data into country level, take the mean(density!)
aggmlstden<-aggregate(livestockDensity[,2:6],by=list(livestockDensity$COUNTRY),"mean")
mlstden_chickens<-
aggregate(as.numeric(livestockDensity[,7]),by=list(livestockDensity$COUNTRY),"sum")
aggmlstden[,"Chicken"]<-mlstden_chickens[,2]
colnames(aggmlstden)[1]<-"COUNTRY"
#merge the two data set by country code and crop code
mlstden_natani_lsd<-merge(aggmlstden,fao_natani_lsd,by='COUNTRY')
mlstden_natani_lsd<-na.omit(mlstden_natani_lsd)
#generate a data frame to store the results
corr_mlstden_natani_lsd<-data.frame("AnimalType"=c("Buffaloes","Cattle","Goats","Pigs","Sheep",
                                    "Chicken","mean","mean_significant"),
                                "cross_correlation"=NA,"p_value"=NA)
for (i in 1:6){
  mlstden<-as.numeric(mlstden_natani_lsd[,i+1])
  natani_lsd<-as.numeric(mlstden_natani_lsd[,i+7])
  mlstden_natani_lsd_inloop<-cbind(mlstden,natani_lsd)
  mlstden_natani_lsd_inloop<-na.omit(mlstden_natani_lsd_inloop)
  mlstden<-mlstden_natani_lsd_inloop[,1]
  natani_lsd<-mlstden_natani_lsd_inloop[,2]
  corr<- cor.test(mlstden,natani_lsd)

  corr_mlstden_natani_lsd[i,"cross_correlation"]<-corr$estimate
  corr_mlstden_natani_lsd[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
corr_mlstden_natani_lsd[7,"cross_correlation"]<-
mean(corr_mlstden_natani_lsd[1:6,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_mlstden_natani_lsd<-subset(corr_mlstden_natani_lsd,p_value<=0.05)
corr_mlstden_natani_lsd[8,"cross_correlation"]<-
mean(sig_corr_mlstden_natani_lsd[,"cross_correlation"])
# 9. [feedset_Crops](feedcr) and [feedset_Animals](feedani)
```

```
#read input data
fao_feedcr<-read.csv("FAO_feedset_Crops.csv",header = TRUE,sep=",")
fao_feedani<-read.csv("FAO_feedset_Animals.csv",header = TRUE,sep=",")
#rename the natfrt first two columns
colnames(fao_feedani)[1]<-"COUNTRYCODE"
colnames(fao_feedcr)[1]<-"COUNTRYCODE"
#aggregate each input by the country code
agg_feedcr<-aggregate(fao_feedcr,by=list(fao_feedcr$COUNTRYCODE),FUN=mean, na.rm=TRUE)
agg_feedani<-aggregate(fao_feedani,by=list(fao_feedani$COUNTRYCODE),FUN=mean, na.rm=TRUE)
#merge the two data set by country code and crop code
feedset<-merge(agg_feedcr,agg_feedani,by='COUNTRYCODE')
feedset<-na.omit(feedset)
#generate a data frame to store the results
corr_feedset<-data.frame("year"=2000:2009,"cross_correlation"=NA,"p_value"=NA)
for (i in 1:10){
  feedcr<-as.numeric(feedset[,i+3])
  feedani<-as.numeric(feedset[,i+15])
  feedset_inloop<-cbind(feedcr,feedani)
  feedset_inloop<-na.omit(feedset_inloop)
  feedcr<-feedset_inloop[,1]
  feedani<-feedset_inloop[,2]
  corr<- cor.test(feedcr,feedani)
  corr_feedset[i,"cross_correlation"]<-corr$estimate
  corr_feedset[i,"p_value"]<-corr$p.value
}
#get the mean of the cross correlation
omit.corr_feedset<-na.omit(corr_feedset)
corr_feedset[i+1,"year"]<-"mean"
corr_feedset[i+1,"cross_correlation"]<-mean(omit.corr_feedset[,"cross_correlation"])
#get the mean of the significant cross correlations
sig_corr_feedset<-subset(corr_feedset,p_value<=0.05)
sig_corr_feedset<-na.omit(sig_corr_feedset)
corr_feedset[i+2,"year"]<-"mean_significant"
corr_feedset[i+2,"cross_correlation"]<-mean(sig_corr_feedset[,"cross_correlation"])
# write the result to overview
sink('overview_fao_cross_correlation.txt')
cat('################## fao cross correlation overview ##################')
cat('\n')
cat('\n')
cat('1. [FAO_NatCropArea](crpar) and [FAO_NatCropProd](crppr)')
cat('\n')
print(corr_crpar_crppr)
cat('\n')
cat('\n')
cat('2. [FAO_NatFertilizer](natfrt) and [FAO_NatCropProd](crppr)')
cat('\n')
print(corr_natfrt_crppr)
cat('\n')
cat('\n')
cat('3. [FAO_AnimalProd](anipr) and [FAO_ProducingAnimals](prani)')
cat('\n')
print(corr_anipr_prani)
cat('\n')
cat('\n')
cat('4. [FAO_AnimalProd](anipr) and [FAO_NatAnimals](natani)')
cat('\n')
print(corr_anipr_natani)
cat('\n')
cat('\n')
cat('5. [FAO_ProducingAnimals](prani) and  [FAO_NatAnimals](natani)')
cat('\n')
print(corr_prani_natani)
cat('\n')
cat('\n')
cat('6. [feedset_Crops](feedcr) and  [FAO_NatAnimals](natani)')
cat('\n')
print(corr_feedcr_natani)
```

```
cat('\n')
cat('\n')
cat('7. [feedset_Animals](feedani) and  [FAO_NatAnimals](natani)')
cat('\n')
print(corr_feedani_natani)
cat('\n')
cat('\n')
cat('8. [LivestockDensity](lstden) and [FAO_natani_lsdmals](natani_lsd)')
cat('\n')
print(corr_lstden_natani_lsd)
cat('9. [feedset_Crops](feedcr) and [feedset_Animals](feedani)')
cat('\n')
print(corr_feedset)
sink()
```

Annex 6 R script for simulating input data

```
#####################  sub-national and national parameters  ##################
# SETTING ENVIRONMENT  ##
#clean-up memory:
rm(list = ls())
# set work directory
setwd("D:/Thesis_BQ/2.In_sim/simulations/3rd_run/")
# set options (print warnings as they occur)
options(warn = 2)
############################## LOADING POINT
# create a continent factor and define the length
continents<-
c("Eastern_Europe","Western_Europe","Africa","Latin_America","Asia","Indian_Subcontinent",
          "Middle_east","North_America","Oceania")
lcon<-length(continents)
#define the MC runs
K <- 1
# time the process
system.time({
  for (con in 1:lcon){
    continent<-continents[[con]]
    ## call needed files
    source("new_sub_1.sim.R")
    source("new_sub_2_sample_post_aggregrate.R")
}})
# combining continents together
source("new_sub_3_combine_continents.R")
# read inputs
# -input 1: relation between ncu, nuts, country
# -input 2: ncu     = correlation between sub-nationals in the country
#         nuts    = correlation between sub-nationals from different countries but in the same
#                   continent
#         country = correlation between sub-nationals from different countries (and continents)
# -input 3: crosscorrelation between parameters
input1 <- read.csv(file = "sim_input/Spatial_code.csv")
spatcor <- read.csv(file = "sim_input/Parameter_ref.csv", header = TRUE, row.names = 1)
crosscor <- read.csv(file = "sim_input/Crosscor_ref_sub.csv", header = TRUE)
#choose the sub-national and national parameters
spatcor<-spatcor[spatcor$Level=="NCU"|spatcor$Level == "NUTS",]
#choose the needed continent
input1<-input1[input1$IPCCcontinent== paste(continent),]
# contruct cross-correlation matrices for cross-corelated and non-cross-corelated parameters
rn <- rownames(spatcor)
nr <- length(rn)
crossmat <- matrix(data = 0, nrow = nr, ncol = nr, dimnames = list(rn,rn))
colid <- match( crosscor$Par2, rn)
rowid <- match( crosscor$Par1, rn)
crosssel <- matrix(data = 0, nrow = length(crosscor$Par2), ncol = length(crosscor$Par1), dimnames =
list(crosscor$Par2,crosscor$Par1))
selparnm <- unique(append(levels(crosscor$Par1), levels(crosscor$Par2)))
selparid <- sort(match(selparnm, rn))
selspatcor <- spatcor[selparid,]
nselspatcor <- spatcor[-selparid,]

crossmat[cbind(colid,rowid)] <- crosscor$Crossc
crossmat[cbind(rowid,colid)] <- crosscor$Crossc
diag(crossmat) <- 1
selcrossmat <- crossmat[selparid,selparid]
nselcrossmat <- crossmat[-selparid,-selparid]
# write cross-correlations to file
write.csv(x=crossmat,file=paste("sim_in_process/",continent, "_crosscor_matrix.csv", sep=""))
write.csv(x=selcrossmat,file=paste("sim_in_process/",continent,"_selcrosscor_matrix.csv",sep=""))
write.csv(x=nselcrossmat,file=paste("sim_in_process/",continent,"_nselcrosscor_matrix.csv",sep=""))
input3 <- read.csv(file = paste("sim_in_process/",continent,"_selcrosscor_matrix.csv",sep=""),
              row.names = 1)
input3n<-read.csv(file = paste("sim_in_process/",continent,"_nselcrosscor_matrix.csv",sep=""),
```

```
              row.names = 1)
input2 <- subset(spatcor, select = rhoNCU:rhoCOUNTRY)
input2 <- subset(selspatcor, select = rhoNCU:rhoCOUNTRY)
input2n <- subset(nselspatcor, select = rhoNCU:rhoCOUNTRY)
# rename column names (to make things consistent with memo Gerard)
names(input1) <- c("ncu", "nuts", "country")
names(input2) <- names(input1)
names(input2n) <- names(input1)
# adjust spatial correlations
input2n[, 1] <- pmin(0.9999,       input2n[, 1])
input2n[, 2] <- pmin(input2n[, 1], input2n[, 2])
input2n[, 3] <- pmin(input2n[, 2], input2n[, 3])
input2[, 1] <- pmin(0.9999,       input2[, 1])
input2[, 2] <- pmin(input2[, 1], input2[, 2])
input2[, 3] <- pmin(input2[, 2], input2[, 3])
## convert inputs to method arguments (matrices)
input3 <- as(input3, "matrix")
input3n <- as(input3n, "matrix")
save(input1,file = paste("sim_in_process/",continent,"_input1.bin",sep=""))
#time the simulation process
#system.time({
  # set seed (reproducibility)
  set.seed(19700124)
  # construct spatial structure
  ss <- spatialStructure(topology = input1, R_auto = input2, R_cross = input3)
  # construct Gaussian Random Field
  gaussianRandomVector <- krige(ss, maxNeighbors = 12, verbose = FALSE)

  save( gaussianRandomVector, input2, input3, file = paste("sim_in_process/",continent,
                                          "_realisation_ccpar.bin",sep=""))
  gRS_lst <- list()
  for (parnm in rownames(input3n)) {
    inp2 <- input2n[parnm,,drop = FALSE]
    inp3 <- input3n[parnm, parnm, drop = FALSE]
    selss <- spatialStructure(topology = input1, R_auto = inp2, R_cross = inp3)
    selgaussianRandomVector <- krige(selss, maxNeighbors = 12, verbose = TRUE)
    #save(            selgaussianRandomVector,            inp2,            inp3,            file            =
paste("sim_in_process/",continent,"_",parnm,".bin",sep = ""))
    #load(file = paste(parnm,".bin"))
    gRS_lst <- c(list(selgaussianRandomVector), gRS_lst)
  }
  save( gRS_lst, file = paste("sim_in_process/",continent, "_gRS_list.bin",sep=""))
#})
############ preparation for sampling ############

# load the prepared files from step 1
load(file = paste("sim_in_process/",continent, "_gRS_list.bin",sep=""))
load(file = paste("sim_in_process/",continent,"_realisation_ccpar.bin",sep=""))
load(file = paste("sim_in_process/",continent,"_input1.bin",sep=""))

############ preparation for post processing ############
# read the input reference data
par_ref_post<- read.csv(file = "sim_input/Parameter_ref.csv", header = TRUE,row.names=1)
# select the parameters and sub and nat level
par_ref_post<-par_ref_post[par_ref_post$Level=="NCU"|par_ref_post$Level == "NUTS",]
#select the cv for this continent
npar<-nrow(par_ref_post)
for (r in 1:npar){
  if(!is.na(par_ref_post[r,continent]))
    par_ref_post[r,"CV"]<-par_ref_post[r,continent]
}
############ preparation for aggregrating ############
# read the input reference data
spatial_code <- read.csv("sim_input/Spatial_code.csv",header = TRUE)
par_ref_agg <- read.csv(file = "sim_input/Parameter_ref.csv", header = TRUE,row.names=1)
# select the parameters and sub and nat level
sub_par <- rownames(par_ref_agg[par_ref_agg$scode == "sub", ])
```

```
nat_par <- rownames(par_ref_agg[par_ref_agg$scode == "nat", ])
con_par <- rownames(par_ref_agg[par_ref_agg$scode == "con" & par_ref_agg$Level == "NCU", ])
#start the MC run
for (i in 1:K) {
  cat(continent,i, "\n")
   # step2: sample Gaussian Random Field
  realisations <- input1[,1,drop=FALSE]
  for (p in 1:length(gRS_lst)) {
    realisation <- sample(gRS_lst[[p]])
    realisations <- cbind(realisations, realisation[,2, drop=FALSE])
  }
  realisation <- sample(gaussianRandomVector)
  realisations <- cbind(realisations, realisation[,-1, drop=FALSE])
  #change the sim of FAO_NatFertilizer to the mean of FAO_NatProd
  realisations$FAO_NatFertilizer<-
rowMeans(subset(realisations,select=c("FAO_NatCP_wheat","FAO_NatCP_maize",
                                        "FAO_NatCP_soybean","FAO_NatCP_barley",
                                        "FAO_NatCP_other")),na.rm = TRUE)
  #rename the spatial code of the realisations to sub national level
  colnames(realisations)[1] <- "sub"
  # result: realisations (zero mean, unit variance)
  # step3: post process results to designed variance
  # use the simulated results from step 2 (zero mean, unit variance): realisations
  #define the column/row names and the column/row number of the realisations
  lsub_reg<-nrow(realisations)
  df<-realisations[,1,drop=FALSE]
  #convert the results based on it`s uncertainty character (either cv or sd)
  for(parname in colnames(realisations)[-1]){
    #define a generic data frame for each column of the realisations
    dfn<-realisations[,paste(parname),drop=FALSE]
    if(!is.na(par_ref_post[paste(parname),"CV"]))          # normal distribution
      dfn[,paste(parname)] <- dfn[,paste(parname)] * par_ref_post[paste(parname),"CV"]+1
    else                                  # lognormal distribution
      dfn[,paste(parname)] <- dfn[,paste(parname)] * par_ref_post[paste(parname),"SD"]

    df<-cbind(df,dfn)
  }
  # result:df (designed variance)
  #write to csv file
  write.csv(x = df,
          file = paste("sim_in_process/sub/",continent,"/result2_real_variance/real_post" ,
                    formatC(x = i, width = nchar(K), flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  # step4: aggregrate results to designed spatial level
  nat<-cbind(realisations[,"sub",drop=F],realisations[,nat_par,drop=F])
  # aggregrate to the designed levels
  nat<-merge(nat,spatial_code,by="sub")
  nat$sub<-NULL
  nat$IPCCcontinent<-NULL
  nat<-aggregate(nat,by=list(nat$FAOcountry),FUN=mean)
  nat$Group.1<-nat$FAOcountry
  nat$FAOcountry<-NULL
  colnames(nat)[1]<-"FAOcountry"
  #write to csv
  write.csv(x = sub,
          file = paste("sim_output/", continent, "/sim_sub",formatC(x = i, width = nchar(K),
                                                  flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  write.csv(x = nat,
          file = paste("sim_output/", continent,"/sim_nat",formatC(x = i, width = nchar(K),
                                                flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  #write to gdx
  #symDim <- 2
  #attr(sub, "symName") <-"sim"
  #sub <- wgdx.reshape(sub, symDim)
```

```
  #wgdx.lst(paste("sim_output/gdx/", continent ,"/sim_sub", formatC(x = i, width = nchar(K), flag =
"0"),
  #".gdx", sep = ""),sub)
  #attr(nat, "symName") <-"sim"
  #nat <- wgdx.reshape(nat, symDim)
  #wgdx.lst(paste("sim_output/gdx/",continent, "/sim_nat", formatC(x = i, width = nchar(K), flag = "0"),
  #".gdx", sep = ""),nat)
  gc()
}
for (i in 1: K){
  cat("combine",i, "\n")
  # creat an empty data frame to store the results
  sub_sim<-data.frame()
  nat_sim<-data.frame()
  for (con in 1:lcon){
    continent<-continents[[con]]
    # read simulated data
    sub_con<-read.csv(file = paste("sim_output/", continent, "/sim_sub",
                          formatC(x = i, width = nchar(K),flag = "0"), ".csv", sep = ""),
                header = TRUE)
    sub_sim<-rbind(sub_sim,sub_con)

    nat_con<-read.csv(file = paste("sim_output/", continent, "/sim_nat",
                          formatC(x = i, width = nchar(K),flag = "0"), ".csv", sep = ""),
                header = TRUE)
    nat_sim<-rbind(nat_sim,nat_con)
  }
  # write the final result to csv
  write.csv(x = sub_sim,
          file = paste("sim_output_final/sim_sub",formatC(x = i, width = nchar(K),
                                        flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  write.csv(x = nat_sim,
          file = paste("sim_output_final/sim_nat",formatC(x = i, width = nchar(K),
                                        flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  #write to gdx
  symDim <- 2
  attr(sub_sim, "symName") <-"sim"
  sub_sim <- wgdx.reshape(sub_sim, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_sub", formatC(x = i, width = nchar(K), flag = "0"),
             ".gdx", sep = ""),sub_sim)
  attr(nat_sim, "symName") <-"sim"
  nat_sim <- wgdx.reshape(nat_sim, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_nat", formatC(x = i, width = nchar(K), flag = "0"),
             ".gdx", sep = ""),nat_sim)
}
#  national parameters
for (i in 1: K){
  cat("nat",i, "\n")
  # creat an empty data frame to store the results
  nat_sim<-data.frame()
  for (con in 1:lcon){
    continent<-continents[[con]]
    # read simulated data
    nat_con<-read.csv(file = paste("sim_output/", continent, "/sim_nat",
                          formatC(x = i, width = nchar(K),flag = "0"), ".csv", sep = ""),
                header = TRUE)
    nat_sim<-rbind(nat_sim,nat_con)
  }
  # write the final result to csv
  write.csv(x = nat_sim,
          file = paste("sim_output_final/sim_nat",formatC(x = i, width = nchar(K),
                                        flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  #write to gdx
  symDim <- 2
```

```
  attr(nat_sim, "symName") <-"sim"
  nat_sim <- wgdx.reshape(nat_sim, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_nat", formatC(x = i, width = nchar(K), flag = "0"),
             ".gdx", sep = ""),nat_sim)
}
# sub national parameters
for (i in 1: K){
  cat("sub",i, "\n")
  # creat an empty data frame to store the results
  sub_sim<-data.frame()
  for (con in 1:lcon){
    continent<-continents[[con]]
    # read simulated data
    sub_con<-read.csv(file = paste("sim_output/", continent, "/sim_sub",
                                   formatC(x = i, width = nchar(K),flag = "0"), ".csv", sep = ""),
                      header = TRUE)
    sub_sim<-rbind(sub_sim,sub_con)
  }
  # write the final result to csv
  write.csv(x = sub_sim,
            file = paste("sim_output_final/sim_sub",formatC(x = i, width = nchar(K),
                                                            flag = "0"), ".csv", sep = ""),
            row.names = FALSE)
  #write to gdx
  symDim <- 2
  attr(sub_sim, "symName") <-"sim"
  sub_sim <- wgdx.reshape(sub_sim, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_sub", formatC(x = i, width = nchar(K), flag = "0"),
             ".gdx", sep = ""),sub_sim)
}

############################ continental parameters ##########################
# SETTING ENVIRONMENT
#clean-up memory:
rm(list = ls())
#load packages:
library(MSBVAR) #"mtvnorm" package
library(reshape)
library(gdxrrw) #load the library for writing to gdx
# set work directory
setwd("D:/Thesis_BQ/2.In_sim/simulations/3rd_run/")
#set the work place for gdx
igdx("I:/Metabase_NEW/GAMS64/GAMS24.1")
# set options (print warnings as they occur)
options(warn = 2)
############################## LOADING POINT
# define the number of the MC runs
K=1000
# read input data
spcode<-read.csv(file = "sim_input/Spatial_code.csv", header = TRUE)
parlist <- read.csv(file = "sim_input/Parameter_ref.csv", header = TRUE)
#select the needed spacial code (continental level) and parameters
spcode = data.frame(unique(spcode$IPCCcontinent))
colnames(spcode)<-"continent"
parlist<-parlist[parlist$Level == "NAT",]
nsp<-nrow(spcode)
npar<-nrow(parlist)

# make a loop for each MC run
for (m in 1:K) {
  # create output data frame with only one column of continent
  con = spcode
  cat(m, "\n")
  #make a loop to get each parameter
  for (i in 1:npar){
    par<-parlist[i,"Code"]
    #make a spatial correlation matrix for this parameter
```

```
   RHO = matrix(data = parlist$rhoCOUNTRY[i],nrow= nsp, ncol = nsp)
   # replace diagonal with ones:
   for (j in 1:nsp)  {
    RHO[j,j] = 1
   }
   # simulate from the multivariate normal distribution:
   sim = rmultnorm(n = 1, rep(0, times=nsp), vmat = RHO, tol = 1e-10)
   # post-process the simulations of ksi depending on parameters characterized by CV or SD
   if(!is.na(parlist$CV[i])) {
    sim = sim * parlist$CV[i] + 1
   } else {
    sim = sim * parlist$SD[i]
   }
   # transpose sim to a column
   sim<-t(sim)
   #rename sim
   colnames(sim)<-par
   # store simulated values in a dataframe
   con = data.frame(con, sim)
  }
  #write result to csv
  write.csv(x = con,
          file = paste("sim_output_final/sim_con", formatC(x = m, width = nchar(K),
                                                flag = "0"), ".csv", sep = ""),
          row.names = FALSE)
  #write to gdx
  symDim <- 2

  attr(con, "symName") <-"sim"
  con <- wgdx.reshape(con, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_con", formatC(x = m, width = nchar(K), flag = "0"),
              ".gdx", sep = ""), con)
}

############################  generic parameters ############################
# SETTING ENVIRONMENT
#clean-up memory:
rm(list = ls())
#load packages:
library(MSBVAR) #"mtvnorm" package
library(reshape)
library(gdxrrw) #load the library for writing to gdx
# set work directory
setwd("D:/Thesis_BQ/2.In_sim/simulations/3rd_run/")
#set the work place for gdx
igdx("I:/Metabase_NEW/GAMS64/GAMS24.1")
# set options (print warnings as they occur)
options(warn = 2)
############################## LOADING POINT
# define the number of the MC runs
K=1000
# read input data
parlist <- read.csv(file = "sim_input/Parameter_ref.csv", header = TRUE)
crosscor <- read.csv(file = "sim_input/Crosscor_ref_gen.csv", header = TRUE)
# select the needed parameters (generic ones)
parlist<-parlist[parlist$Level == "GEN",]
npar<-nrow(parlist)
cc<-nrow(crosscor)
# make a cross correlation matrix for these parameters
RHO = matrix(data = 0,nrow= npar, ncol = npar)
dimnames(RHO)<-list(parlist[,"Code"],parlist[,"Code"])
# replace diagonal with ones:
for (n in 1:npar) {
 RHO[n,n] = 1
}
# replace the cross correlations indicated by crosscor
for (c in 1:cc){
```

```
  Par1<-as.character(crosscor[c,"Par1"])
  Par2<-as.character(crosscor[c,"Par2"])
  Crossc<-crosscor[c,"Crossc"]
  RHO[Par1,Par2]<-Crossc
  RHO[Par2,Par1]<-Crossc
}

# simulate from the multivariate normal distribution:
sim = rmultnorm(n = K, rep(0, times=npar), vmat = RHO, tol = 1e-10)
# post-process the simulations depending on parameters characterized by CV or SD
for (i in 1: 1:npar)
if(!is.na(parlist$CV[i])) {
  sim[,i] = sim[,i] * parlist$CV[i] + 1
} else {
  sim[,i] = sim[,i] * parlist$SD[i]
}
# add a row to indicate the runs
run<-data.frame(1:K)
colnames(run)<-"run"
gen<-cbind(run,sim)

#write result to csv
write.csv(x = gen,"sim_output_final/sim_gen.csv",row.names = FALSE)
#write to gdx
#symDim <- 2

#attr(gen, "symName") <-"sim"
#gen <- wgdx.reshape(gen, symDim)
#wgdx.lst("new_uncertainty_result/sim_gen.gdx", gen)

# seperate generic parameter for each run
gen <- read.csv("sim_output_final/sim_gen.csv", header = TRUE)
#define the number of rows
for (r in 1:K){
  cat("gen",r, "\n")
  gen_row <- gen[r,,drop=FALSE]
  gen_row[,1]<-"sim_mean"
  # write to csv
  write.csv(x = gen_row,
          file = paste("sim_output_final/sim_generic",formatC(x = r, width = nchar(K), flag = "0"),
                  ".csv", sep = ""),row.names = FALSE)
  # write to gdx
  symDim <- 2
  attr(gen_row, "symName") <-"sim"
  gen_row <- wgdx.reshape(gen_row, symDim)
  wgdx.lst(paste("sim_output_final/gdx/sim_generic",formatC(x = r, width = nchar(K), flag = "0"),
              ".gdx", sep = ""),gen_row)

}
```

Annex 7 R script for analysing and visualising the uncertainty quantification results

```
# 1. total emissions for each continent
data_file = "9_pAggr_LS_GHG_total_AC_JP.gdx"
LS_GHG_sum<-rgdx.param (data_file,"Livestock_GHG_sum_AC",names =
c("Run","Continent","Emission"),compress=TRUE)
total_emission <-LS_GHG_sum
#processing the data
# with ddply
ddply_total<-ddply(total_emission, c("Continent", "Emission"), summarise,
              mean = mean(value, na.rm=TRUE)*10^-6, # convert result to Mton
              sd = sd(value, na.rm=TRUE)*10^-6,
              cv = sd/mean,
              n = sum(!is.na(value)),
              q_025=quantile(value, p=0.025)*10^-6,
              q_975=quantile(value, p=0.975)*10^-6)
write.csv(ddply_total,"ddply_stat_total.csv",row.names=FALSE)
ddply_total<-read.csv("ddply_stat_total.csv",header=T)
ddply_total$Continent<-factor(ddply_total$Continent, c("Africa","LatinAmerica","EU27"))
ddply_total$Emission<-factor(ddply_total$Emission, c("Total","CH4","N2O","CO2"))
# plot mean
png("one_total_mean.png",width=1400,height=900)
ggplot(na.omit(ddply_total),aes(x=Continent,y=mean))+
  ylab("Emission (Mton CO2-eq)")+
  geom_bar(stat="identity",fill="grey", colour="black")+
  geom_errorbar(aes(ymin=q_025, ymax=q_975), width=.2)+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_wrap(~Emission,nrow=1)
dev.off()
#plot the cv
png("one_total_cv.png",width=1400,height=900)
ggplot(na.omit(ddply_total),aes(x=Continent,y=cv))+
  ylab("cv for emission")+
  geom_bar(stat="identity",fill="grey", colour="black")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_wrap(~Emission,nrow=1)
dev.off()
# 2. GHG emissions for different sectors
data_file = "10_pAggr_LS_GHG_sector_AC.gdx"
LGHG_AC<-
rgdx.param(data_file,"Livestock_GHG_sum_AC",names=c("Run","Continent","Sector","Emission"),
compress=TRUE)
write.csv(LGHG_AC,"data_sector.csv",row.names=F)
LGHG_AC<-read.csv("data_sector.csv",header=T)
LGHG_AC$Continent<-factor(LGHG_AC$Continent,c("Africa","LatinAmerica","EU27"))
LGHG_AC$Sector<-factor(LGHG_AC$Sector,
                       c("OtherAnimals","Turkeys","LayingHens","Broilers",
                         "Camels","Horses","OtherPoultry","Pigs",
                         "Goats","Sheep","DairyCows","OtherCattle"))
LGHG_AC$Emission<-factor(LGHG_AC$Emission, c("Total","CH4","N2O","CO2"))
# processing the data
# stat with ddply
stat_sector<-ddply(LGHG_AC, c("Continent", "Emission","Sector"), summarise,
              mean = mean(value, na.rm=TRUE)*10^-6,
```

```
                sd = sd(value, na.rm=TRUE)*10^-6,
                cv = sd/mean,
                n = sum(!is.na(value)),
                q_025=quantile(value, p=0.025)*10^-6,
                q_975=quantile(value, p=0.975)*10^-6)
write.csv(stat_sector,"ddply_stat_sector.csv",row.names=FALSE)
stat_sector<-read.csv("ddply_stat_sector.csv",header=TRUE)
stat_sector$Continent<-factor(stat_sector$Continent,c("Africa","LatinAmerica","EU27"))
stat_sector$Sector<-factor(stat_sector$Sector,
                      c("OtherAnimals","Turkeys","LayingHens","Broilers",
                        "Camels","Horses","OtherPoultry","Pigs",
                        "Goats","Sheep","DairyCows","OtherCattle"))
stat_sector$Emission<-factor(stat_sector$Emission,
                        c("Total","CH4","N2O","CO2"))
# plot for mean
png("one_sector_mean.png",width=1400,height=1800)
ggplot(stat_sector, aes(x=Sector, y=mean))+
  geom_bar(stat="identity",fill="grey",color="black")+
  geom_errorbar(aes(ymin=q_025, ymax=q_975), width=.2)+
  ylab("Emission (Mton CO2 eq)")+
  xlab("Sector")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(Emission~ Continent,scales="free")+
  coord_flip()
dev.off()
# plot for cv
png("one_sector_cv.png",width=1400,height=1800)
ggplot(stat_sector, aes(x=Sector, y=cv))+
  geom_bar(stat="identity",fill="grey",color="black")+
  ylab("cv for emission")+
  xlab("Sector")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(Emission~ Continent,scales="free")+
  coord_flip()
dev.off()
# 3. GHG emissions for different products
data_file = "8_pAggr_LS_GHG_product_AC.gdx"
LS_GHG_p_AC <- rgdx.param (data_file,"Aggr_LS_GHG_product_AC",names =
c("Run","Continent","Product","Emission"), compress=TRUE)
write.csv(LS_GHG_p_AC,"data_product.csv",row.names=FALSE)
LS_GHG_p_AC<-read.csv("data_product.csv",header=T)
LS_GHG_p_AC$Continent<-factor(LS_GHG_p_AC$Continent,c("Africa","LatinAmerica","EU27"))
LS_GHG_p_AC$Emission<-factor(LS_GHG_p_AC$Emission,c("Total","CH4","N2O","CO2"))
LS_GHG_p_AC$Product<-reorder(LS_GHG_p_AC$Product,LS_GHG_p_AC$value,FUN=mean)
#processing the data
# with ddply
ddply_product<-ddply(LS_GHG_p_AC, c("Continent", "Product","Emission"), summarise,
                mean = mean(value, na.rm=TRUE),
                sd = sd(value, na.rm=TRUE),
                cv = sd/mean,
                n = sum(!is.na(value)),
                q_025=quantile(value, p=0.025),
                q_975=quantile(value, p=0.975))
write.csv(ddply_product,"ddply_stat_product.csv",row.names=FALSE)
ddply_product<-read.csv("ddply_stat_product.csv",header=T)
ddply_product$Continent<-factor(ddply_product$Continent,c("Africa","LatinAmerica","EU27"))
ddply_product$Emission<-factor(ddply_product$Emission, c("Total","CH4","N2O","CO2"))
ddply_product$Product<-reorder(ddply_product$Product, ddply_product$mean,FUN=mean)
# plot th mean
```

```
png("one_product_mean.png",width=1400,height=900)
ggplot(ddply_product, aes(x=Product, y=mean))+
  geom_bar(stat="identity",fill="grey",color="black")+
  geom_errorbar(aes(ymin=q_025, ymax=q_975), width=.2)+
  ylab("Emission CO2(kg)/product(kg)") + xlab("Product")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(Emission~ Continent)+
  coord_flip()
dev.off()
# plot cv
png("one_product_cv.png",width=1400,height=900)
ggplot(ddply_product, aes(x=Product, y=cv))+
  geom_bar(stat="identity",fill="grey",color="black")+
  ylab("cv")+xlab("Product")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(Emission~ Continent)+
  coord_flip()
dev.off()
# 4. GHG emissions for different processes
data_file = "9_pAggr_LS_GHG_total_AC_JP.gdx"
LS_GHG_total <- rgdx.param(data_file,"Livestock_GHG_total_AC",names =
c("Run","Continent","Process"), compress=TRUE)
write.csv(LS_GHG_total,"data_process.csv",row.names=F)
LS_GHG_total<-read.csv("data_process.csv",header=T)
LS_GHG_total$Continent<-
factor(LS_GHG_total$Continent, c("Africa","LatinAmerica","EU27"))
#processing the data
# with ddply
ddply_process<-ddply(LS_GHG_total, c("Continent", "Process"), summarise,
              mean = mean(value, na.rm=TRUE)*10^-6,
              sd = sd(value, na.rm=TRUE)*10^-6,
              cv = sd/mean,
              n = sum(!is.na(value)),
              q_025=quantile(value, p=0.025)*10^-6,
              q_975=quantile(value, p=0.975)*10^-6)
write.csv(ddply_process,"ddply_stat_process.csv",row.names=FALSE)
ddply_process<-read.csv("ddply_stat_process.csv",header=T)
ddply_process$Continent<-factor(ddply_process$Continent,
                          c("Africa","LatinAmerica","EU27"))
ddply_process$Process<-reorder(ddply_process$Process,
                       ddply_process$mean,FUN=mean)
png("one_process_mean.png",width=1400,height=900)
ggplot(na.omit(ddply_process), aes(x=Process, y=mean))+
  geom_bar(stat="identity",fill="grey",color="black")+
  geom_errorbar(aes(ymin=q_025, ymax=q_975), width=.2)+
  ylab("Emission (Mton CO2-eq)")+ xlab("Process")+
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(.~ Continent)+
  coord_flip()
dev.off()
png("one_process_cv.png",width=1400,height=900)
ggplot(na.omit(ddply_process), aes(x=Process, y=cv))+
  geom_bar(stat="identity",fill="grey",color="black")+
  ylab("cv for emission")+
  xlab("Process")+
```

```
  theme(axis.text.y=element_text(size=20),
        axis.text.x=element_text(size=15),
        axis.title.y=element_text(size=30,face="bold"),
        axis.title.x=element_text(size=30,face="bold"),
        strip.text = element_text(face="bold", size=22))+
  facet_grid(.~ Continent)+
  coord_flip()
dev.off()
```

Annex 8 R script for analysing and visualising the uncertainty analysis results

```
# 1. total emissions for each continent
# read the data
stat_total_emission<-read.csv("stat_total_emission.csv",header=T)
stat_total_emission<-melt(stat_total_emission,id=1:2)
colnames(stat_total_emission)[3]<-"Group"
stat_total_emission$Continent<-factor(stat_total_emission$Continent,
                                      c("Africa","LatinAmerica","EU27"))
stat_total_emission$Emission<-factor(stat_total_emission$Emission,
                                     c("Total","CH4","N2O","CO2"))
stat_total_emission<- ddply(stat_total_emission, .(Continent,Emission), transform,
                            percent_contribution = value / sum(na.omit(value)) * 100)
png("UA_total_emission.png",height=700,width=900)
ggplot(na.omit(stat_total_emission), aes(x=Emission, y=percent_contribution, fill=Group)) +
  geom_bar(stat="identity")+
  facet_grid(.~Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=22,face="bold"),
        axis.title.y=element_text(size=22,face="bold"),
        axis.text.x=element_text(size=15),
        axis.text.y=element_text(size=15),
        legend.title=element_text(size=15),
        legend.text=element_text(size=15),
        strip.text = element_text(face="bold", size=15))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))
dev.off()
# plot for two
tt_p2<-stat_total_emission[stat_total_emission$Emission=="CH4"|
                           stat_total_emission$Emission=="N2O",,drop=F]
png("UA_total_200.png",height=700,width=900)
ggplot(tt_p2, aes(x=Group, y=value)) +
  geom_bar(stat="identity",position="dodge",fill="grey60")+
  facet_grid(Emission ~Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=22,face="bold"),
        axis.title.y=element_text(size=22,face="bold"),
        axis.text.x=element_text(size=15),
        axis.text.y=element_text(size=15),
        legend.title=element_text(size=15),
        legend.text=element_text(size=15),
        strip.text = element_text(face="bold", size=15))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))
dev.off()
# plot the result of all (comparing winding and one group at a time)
stat_total_all<-read.csv("stat_total_all.csv",header=T)
stat_total_all<-melt(stat_total_all,id=1:3)
colnames(stat_total_all)[4]<-"Group"
stat_total_all$Continent<-factor(stat_total_all$Continent,
                                 c("Africa","LatinAmerica","EU27"))
stat_total_all$Emission<-factor(stat_total_all$Emission,
                                c("Total","CH4","N2O","CO2"))

stat_total_all<- ddply(stat_total_all, .(Continent,Emission), transform,
                       percent_contribution = value / sum(na.omit(value)) * 100)
stat_total_all<-stat_total_all[stat_total_all$Emission=="CH4"|
                               stat_total_all$Emission=="N2O",,drop=F]
png("UA_total_all.png",height=700,width=900)
ggplot(stat_total_all, aes(x=Group, y=value,fill=Method)) +
  geom_bar(stat="identity",position="dodge")+
  facet_grid(Emission ~ Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=22,face="bold"),
        axis.title.y=element_text(size=22,face="bold"),
        axis.text.x=element_text(size=15),
```

```
      axis.text.y=element_text(size=15),
      legend.title=element_text(size=15),
      legend.text=element_text(size=15),
      strip.text = element_text(face="bold", size=15))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))+
  coord_flip()
dev.off()
# 2. GHG emissions for different sectors
# read the data
stat_sector<-read.csv("stat_sector.csv",header=T)
stat_sector<-melt(stat_sector,id=1:3)
colnames(stat_sector)[4]<-"Group"
stat_sector$Continent<-factor(stat_sector$Continent,
                      c("Africa","LatinAmerica","EU27"))
stat_sector$Emission<-factor(stat_sector$Emission,
                      c("Total","CH4","N2O","CO2"))
stat_sector<- ddply(stat_sector, .(Continent,Emission,Sector), transform,
               percent_contribution = value / sum(na.omit(value)) * 100)
png("UA_sector.png",height=1800,width=1400)
ggplot(na.omit(stat_sector), aes(x=Sector, y=percent_contribution, fill=Group)) +
  geom_bar(stat="identity")+
  facet_grid(Emission~Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=30,face="bold"),
      axis.title.y=element_text(size=30,face="bold"),
      axis.text.x=element_text(size=22),
      axis.text.y=element_text(size=22),
      legend.title=element_text(size=22),
      legend.text=element_text(size=22),
      strip.text = element_text(face="bold", size=22))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))+
  coord_flip()
dev.off()
# 3. GHG emissions for different products
# read the data
stat_product<-read.csv("stat_product.csv",header=T)
stat_product<-melt(stat_product,id=1:3)
colnames(stat_product)[4]<-"Group"
stat_product$Continent<-factor(stat_product$Continent,
                       c("Africa","LatinAmerica","EU27"))
stat_product$Emission<-factor(stat_product$Emission,
                      c("Total","CH4","N2O","CO2"))
stat_product<- ddply(stat_product, .(Continent,Emission,Product), transform,
               percent_contribution = value / sum(na.omit(value)) * 100)
png("UA_product.png",height=1800,width=1400)
ggplot(na.omit(stat_product), aes(x=Product, y=percent_contribution, fill=Group)) +
  geom_bar(stat="identity")+
  facet_grid(Emission~Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=30,face="bold"),
      axis.title.y=element_text(size=30,face="bold"),
      axis.text.x=element_text(size=22),
      axis.text.y=element_text(size=22),
      legend.title=element_text(size=22),
      legend.text=element_text(size=22),
      strip.text = element_text(face="bold", size=22))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))+
  coord_flip()
dev.off()
# 4. GHG emissions for different processes
stat_process<-read.csv("stat_process.csv",header=T)
stat_process<-melt(stat_process,id=1:2)
colnames(stat_process)[3]<-"Group"
stat_process$Continent<-factor(stat_process$Continent,
                       c("Africa","LatinAmerica","EU27"))
stat_process<- ddply(stat_process, .(Continent,Process), transform,
               percent_contribution = value / sum(na.omit(value)) * 100)
```

```
png("UA_process.png",height=700,width=900)
ggplot(na.omit(stat_process), aes(x=Process, y=percent_contribution, fill=Group)) +
  geom_bar(stat="identity")+
  facet_grid(.~Continent)+
  ylab("Uncertainty contribution (%)")+
  theme(axis.title.x=element_text(size=22,face="bold"),
        axis.title.y=element_text(size=22,face="bold"),
        axis.text.x=element_text(size=15),
        axis.text.y=element_text(size=15),
        legend.title=element_text(size=15),
        legend.text=element_text(size=15),
        strip.text = element_text(face="bold", size=15))+
  guides(fill = guide_legend(keywidth = 1.5, keyheight = 1.5))+
  coord_flip()
dev.off()
```