



**4th International Workshop
on Uncertainty in Atmospheric Emissions**
7-9 October 2015, Krakow, Poland

PROCEEDINGS



**4th International Workshop
on Uncertainty
in Atmospheric Emissions**

7–9 October 2015, Kraków, Poland

PROCEEDINGS

Warszawa 2015

4th International Workshop on Uncertainty in Atmospheric Emissions
7- 9 October 2015, Cracow, Poland

Printed from the material submitted by the authors.

ISBN 83-894-7557-X
EAN 9788389475572

© Systems Research Institute, Polish Academy of Sciences, Warszawa, Poland 2015

About the Workshop

The assessment of greenhouse gases and air pollutants (indirect GHGs) emitted to and removed from the atmosphere is high on the political and scientific agendas. Building on the UN climate process, the international community strives to address the long-term challenge of climate change collectively and comprehensively, and to take concrete and timely action that proves sustainable and robust in the future. Under the umbrella of the UN Framework Convention on Climate Change, mainly developed country parties to the Convention have, since the mid-1990s, published annual or periodic inventories of emissions and removals, and continued to do so after the Kyoto Protocol to the Convention ceased in 2012. Policymakers use these inventories to develop strategies and policies for emission reductions and to track the progress of those strategies and policies. Where formal commitments to limit emissions exist, regulatory agencies and corporations rely on emission inventories to establish compliance records.

However, as increasing international concern and cooperation aim at policy-oriented solutions to the climate change problem, a number of issues circulating around uncertainty have come to the fore, which were undervalued or left unmentioned at the time of the Kyoto Protocol but require adequate recognition under a workable and legislated successor agreement. Accounting and verification of emissions in space and time, compliance with emission reduction commitments, risk of exceeding future temperature targets, evaluating effects of mitigation versus adaptation versus intensity of induced impacts at home and elsewhere, and accounting of traded emission permits are to name but a few.

The *4th International Workshop on Uncertainty in Atmospheric Emissions* is jointly organized by the *Systems Research Institute of the Polish Academy of Sciences*, the Austrian-based *International Institute for Applied Systems Analysis*, and the *Lviv Polytechnic National University*. The 4th Uncertainty Workshop follows up and expands on the scope of the earlier Uncertainty Workshops – the *1st Workshop* in 2004 in Warsaw, Poland; the *2nd Workshop* in 2007 in Laxenburg, Austria; and the *3rd Workshop* in 2010 in Lviv, Ukraine.

Steering Committee

Rostyslav BUN (Lviv Polytechnic National University, UA)
Matthias JONAS (International Institute for Applied Systems Analysis, AT)
Zbigniew NAHORSKI (Polish Academy of Sciences, PL) – Chair

Scientific Committee

Evgueni GORDOV (Siberian Center for Environmental Research & Training, RU)
Piotr HOLNICKI-SZULC (Polish Academy of Sciences, PL)
Joanna HORABIK-PYZEL (Polish Academy of Sciences, PL)
Olgiard HRYNIEWICZ (Polish Academy of Sciences, PL)
Katarzyna JUDA-REZLER (Warsaw University of Technology, PL)
Petro LAKYDA (National University of Life and Environmental Sciences of Ukraine, UA)
Myroslava LESIV (Lviv Polytechnic National University, UA)
Gregg MARLAND (Appalachian State University, USA)
Sten NILSSON (Forest Sector Insights AB, SE)
Tom ODA (Univ. Space Research Association, NASA Goddard Space Flight Center, USA)
Stefan PICKL (Universität der Bundeswehr München, Germany)
Elena ROVENSKAYA (International Institute for Applied Systems Analysis, AT)
Kazimierz RÓŻAŃSKI (AGH University of Science and Technology in Cracow, PL)
Dmitry SCHEPASCHENKO (International Institute for Applied Systems Analysis, AT)
Anatoly SHVIDENKO (International Institute for Applied Systems Analysis, AT)
Jacek SKOŚKIEWICZ (National Centre for Emissions Management, PL)
Philippe THUNIS (EC Joint Research Centre Ispra, EU)
Marialuisa VOLTA (University of Brescia, IT)

Local Organizing Committee

Joanna HORABIK-PYZEL
Jolanta JARNICKA - Chair
Weronika RADZISZEWSKA
Jörg VERSTRAETE

Table of Contents

FORESTRY

- Full verified carbon account of forest ecosystems as a fuzzy system: An attempt to assess uncertainty**..... 1
Anatoly Shvidenko, Dmitry Schepaschenko, Florian Kraxner, Steffen Fritz
- Forest map and its uncertainty as an important input for carbon sink estimation for Poland and Ukraine**..... 9
Myroslava Lesiv, Anatoly Shvidenko, Dmitry Schepaschenko, Linda See, Steffen Fritz

AGRICULTURE

- Spatial GHG inventory in the Agriculture sector and uncertainty analysis: A case study for Poland**..... 16
Nadiia Charkovska, Rostyslav Bun, Olha Danylo, Joanna Horabik-Pyzel, Matthias Jonas
- Conditionally autoregressive model for spatial disaggregation of activity data in GHG inventory: Application for agriculture sector in Poland**..... 25
Joanna Horabik-Pyzel, Nadiia Charkovska, Olha Danylo, Zbigniew Nahorski, Rostyslav Bun
- Modeling uncertainty in ammonia emissions from agriculture: Regional upscaling by Monte Carlo analysis**..... 32
Bettina Schäppi, Jürg Heldstab, Thomas Kupper

ENERGY AND INDUSTRY

- High resolution spatial inventory of GHG emissions from stationary and mobile sources in Poland: summarized results and uncertainty analysis**..... 41
Rostyslav Bun, Zbigniew Nahorski, Joanna Horabik-Pyzel, Olha Danylo, Nadiia Charkovska, Petro Topylko, Mariia Halushchak, Myroslava Lesiv, Oleksandr Striamets
- Spatial Greenhouse Gas (GHG) inventory and uncertainty analysis: A case study for electricity generation in Poland and Ukraine**..... 49
Petro Topylko, Mariia Halushchak, Rostyslav Bun, Tomohiro Oda, Myroslava Lesiv, Olha Danylo
- Uncertainty analysis of GHG spatial inventory from the industrial activity: A case study for Poland**..... 57
Nadiia Charkovska, Mariia Halushchak, Rostyslav Bun, Matthias Jonas

**Spatial inventory of GHG emissions from fossil fuels extraction and processing:
An uncertainty analysis..... 64**
Mariia Halushchak, Rostyslav Bun, Matthias Jonas, Petro Topylko

Carbon emission inventory calculation and analysis based on coal lifecycle..... 71
Xiangyang Xu, Junlian Gao

UNCERTAINTY, LEARNING, AND POTENTIAL POLICY IMPACTS

**A metric for the prognostic outreach of scenarios: Learning from the past to establish a
standard in applied systems analysis..... 78**
Matthias Jonas, Elena Rovenskaya, Piotr Żebrowski

**Assessing the improvement of greenhouse gases inventories: can we capture diagnostic
learning?..... 90**
Piotr Żebrowski, Matthias Jonas, Elena Rovenskaya

**A method for estimating time evolution of precision and accuracy of greenhouse gases
inventories from revised reports..... 97**
Jolanta Jarnicka, Zbigniew Nahorski

**How uncertainty of air emission inventories impacts policy decisions at sub-national
level. A Shift-Share Analysis undertaken in Piedmont (Italy).....103**
Alessandra La Notte, Stefania Tonin, Silvio Nocera

GLOBAL IMPACTS OF GHG EMISSIONS

Performance of global black carbon emission inventories in the Arctic..... 110
Ville-Veikko Paunu, Kaarle Kupiainen

Global anthropogenic particle number emissions and their size distributions.....111
*Pauli Paasonen, Kaarle Kupiainen, Zbigniew Klimont, Antoon Visschedijk, Hugo Denier
van der Gon, Markus Amann*

GRIDDED EMISSIONS ESTIMATES

Uncertainty in gridded CO₂ emissions estimates..... 117
*Susannah Hogue, Eric Marland, Robert J. Andres, Gregg Marland, Charles Robison,
Dawn Woodard*

**Uncertainty associated with fossil fuel carbon dioxide (CO₂) gridded emission
datasets..... 124**
*Tomohiro Oda, Lesley Ott, Petro Topylko, Mariia Halushchak, Rostyslav Bun, Myroslava
Lesiv, Olha Danylo, Joanna Horabik-Pyzel*

Remapping gridded data using Artificial Intelligence: real world challenges..... 130
Jörg Verstraete

Integration of multi-source information in disaggregation of spatial emission data... 137
Joanna Horabik-Pyzel, Zbigniew Nahorski

PRICING UNCERTAINTY AND TRADING

Sensitivity of marginal abatement cost curves to variation of G4M parameters..... 152
Mykola Gusti, Nikolay Khabarov, Nicklas Forsell

Multi-agent auction simulation of the GHG international emission permit trading... 159
Piotr Pałka, Jarosław Stańczak, Weronika Radziszewska, Zbigniew Nahorski

URBAN EMISSIONS

Urban metabolism: atmospheric loads and fluxes of major greenhouse gases (CO₂, CH₄) in Krakow, southern Poland..... 177
Mirosław Zimnoch, Jarosław Necki, Jarosław Chmura, Alina Jasek, Michał Galkowski, Tadeusz Kuc, Zbigniew Gorczyca, Jakub Bartyzel, Kazimierz Rozanski

Inventories of domestic heating sources and their emissions in urban areas – methods, results and uncertainty..... 187
Marek Rosicki, Magdalena Załupka

Accounting uncertainty for spatial modeling of greenhouse gas emissions in the residential sector: fuel combustion and heat production..... 193
Olha Danylo, Rostyslav Bun, Linda See, Petro Topylko, Xiangyang Xu, Nadiia Charkovska, Przemysław Tymków

On the uncertainty in modeling urban air quality under imprecise emission data.....201
Piotr Holnicki, Zbigniew Nahorski

On the possibility of selenium air emission inventory from small domestic sources... 208
Damian Zasina, Jarosław Zawadzki

Full verified carbon account of forest ecosystems as a fuzzy system: An attempt to assess uncertainty

Anatoly Shvidenko^{1,2}, Dmitry Schepaschenko^{1,3}, Florian Kraxner¹, Steffen Fritz¹

¹ International Institute for Applied Systems Analysis,
Laxenburg, Austria

²V.N. Sukachev Institute of Forest, SB RAS, 660036 Krasnoyarsk,

³Moscow State Forest University, Mytishi, 141005 Moscow, Russia
shvidenk@iiasa.ac.at, schepd@iiasa.ac.at, kraxner@iiasa.ac.at, fritz@iiasa.ac.at

Abstract

Carbon cycling of terrestrial ecosystems is a fuzzy (underspecified) system that imposes substantial constraints on possibility to get unbiased estimates of basic intermediate components (e.g., Net Primary Production, Heterotrophic Respiration) and final results (e.g., Net Ecosystem Carbon Budget) of the account within strictly defined confidential intervals based on any individually used carbon cycling method or model. We present a methodology attempting at minimizing possible biases and restricting the multivariate uncertainty's space. The methodology follows the principles of applied systems analysis and is based on integration of major independent methods of carbon cycling study (landscape-ecosystem approach, process-based models, eddy covariance and inverse modelling) with following harmonizing and mutual constraints of the results. Based on a case study for Russia's forests, we discuss strengths and limitations of the outlined methodology.

Keywords: Carbon cycle, uncertainty, fuzzy systems, Northern Eurasian forests

1. Introduction

Assessment of carbon budget of terrestrial ecosystems (FCA) requires obtaining two equally important outputs: 1) an unbiased proxy value, e.g. Net Ecosystem Carbon Budget (NECB) in a spatial and temporal explicit way and 2) uncertainties of NECB and its major components. A possible bias of the results depends upon the method used and completeness of the FCA. The latter is usually estimated based on expert estimates and professional judgements. Consideration of numerous interacting processes, which control NECB, in many models are often limited by a few such as climate change, impact of elevated CO₂, sometimes disturbances, nitrogen limitation and deposition [1]. Based on previous assessments of uncertainties' range of major components of the FCA, we consider the carbon account as full if the accounting schemes include $\geq 98\%$ of all recognized processes. A *verified* account of NECB supposes reliable and complete assessment of uncertainties, i.e. judgments about "uncertainty of uncertainties" would be possible [2]. However, the full carbon account of terrestrial ecosystems, particularly at large spatial scales is a typical fuzzy (underspecified) system, of which membership function is inherently stochastic, with some typical features of full complexity problems [3] and to some extent - wicked problems [e.g., 4]. This predetermines a principle impossibility of formally strict assessment of structural uncertainties within any method individually used. Thus "within method" uncertainty inevitably presents only part of "full" uncertainties. Posterior independent empirical validation of NECB is difficult to be realized in practice due to large resources required. This necessitates development of a methodology, which would be able to assess the "full uncertainties" of a studied system.

We attempt to outline such a methodology based on major principles of applied systems analysis [2,5], considering combination of major methods of carbon cycling understanding: landscape-ecosystem approach (LEA), process-based models, inverse modeling, and eddy covariance. Use of remote sensing methods in the FCA is crucial and two-faced because those deliver important input data (such as land cover at its biophysical parameters like above-ground live biomass) for different methods, but also some components of FCA directly (e.g., NPP). The principle of integration is applied at all stages and for all modules of the account - from development of the information base to uncertainty assessment of final results. Some ideas of the considered approach have been presented in previous publications [5,6,7] but the descriptions of methods used were lacked a common system basis. The approach was applied to the FCA of Russian forests as the most complicated by structure and processes terrestrial ecosystem that allows to highlight the methodology's strengths, weaknesses and potential. We also discuss system requirements to different methods of FCA, relevant scales and required details, information and research needs, as well as obtained and potential levels of uncertainties.

2. Methods

Basic methods of studying the carbon cycling of terrestrial ecosystems differ by specifics of cognition of biogeochemical processes, amount of information required, spatial and temporal details of consideration, and possibility of uncertainties' assessments. In an ideal case, each method should satisfy a minimum of system requirements that would allow to reliably assess "within method" uncertainties including monosemantic (and potentially consistent) definitions and classification schemes; explicit structuring of the account including strict spatial, temporal and process boundaries; explicit algorithmic description of the FCA for all steps and modules including that of assumptions, expert estimates and other "soft knowledge"; matching the temporal dimensions of the FCA with characteristic times of processes considered. Effectiveness of potential integration of results obtained by different methods depends on compatibility and amount of information comprising by each method. Structure of the FCA is outlined in Figure.

2.1 Landscape-ecosystem approach as empirical background of FCA

Landscape-ecosystem approach (LEA) plays specific role in the FCA as its empirical basis. In essence, it combines two basic backgrounds of any carbon cycling study - pool-based and flux-based approaches in a possibly complimentary way. The LEA serves for strict designing the studied system, defining the inter- and intra- boundaries, and contains spatially distributed accumulated information about ecosystems and landscapes (data of measurements *in situ*, diverse empirical and semi-empirical aggregations, data of forest inventory and different surveys, empirical aggregations and models etc.). LEA's information background is presented in form of an Integrated Land Information System as multi-layer and multi-scale GIS by polygons of a hybrid land cover (HLC). The HLC uses a hierarchical classification of land cover with details, needed for carbon cycling assessment. Land cover is developed using diversity of relevant remote sensing products, geographically-weighted regression and validation by Geo-Wiki tool. For instance, the last version of forest mask for Russia (resolution 230 m) was based on 12 remote sensing products, 5300 control points for the algorithm training and 730 for validation points; this allowed to minimize the possible biases in assessment of the forest area and its distribution providing accuracy of the forest

mask >95% [8]. By-pixel parametrization of forest cover is provided based on multi-sensor remote sensing data, data of forest inventory, soil and landscape characteristics and other diverse relevant sources using a special optimization algorithm [9].

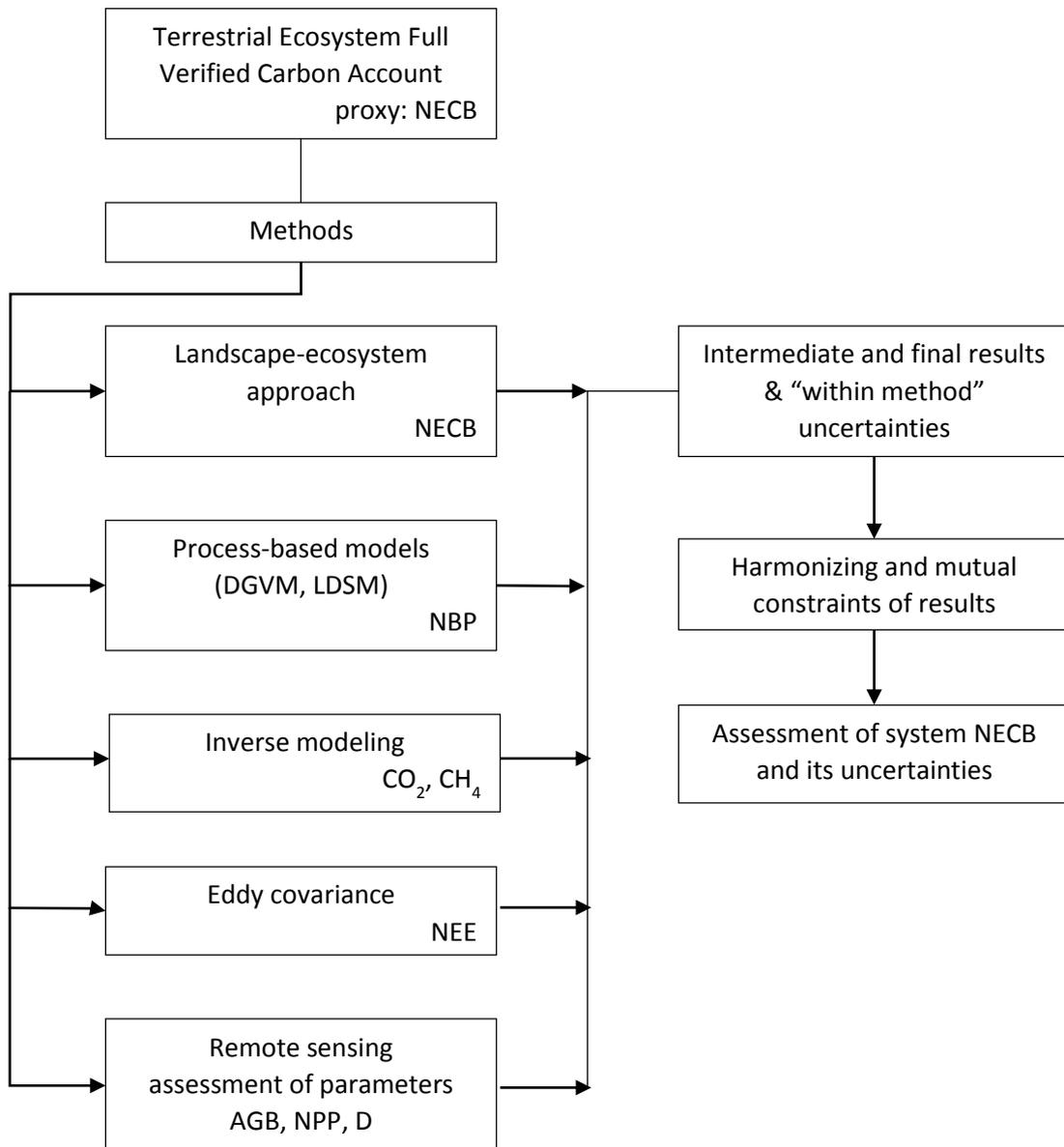


Figure 1. Structural scheme of full verified carbon account of forest ecosystems.

An important requirement is providing a system consistency between resolution (spatial scale of land cover and its parametrization) and certainty of attributive data. It could be shown that accuracy of major part of input data and empirically based models are logically consistent with resolution of 200-500 m at the country's scale. This provides a minimal level of uncertainty which presumably would be available for policy makers [10] but requires, e.g. for forests, by-pixel knowledge of dominant tree species, age, average height and diameter, site index, relative stocking, growing stock volume, and stock of dead wood. At the level of forest enterprises (of the total amount of ~1700 for Russia) the algorithm provides consistency of aggregated ILIS data with the most accurate available information sources (e.g., data of recent forest inventory). The assigned by-pixel parameters are presented by the most likely values based on indexes

of suitability which are calculated based on ILIS data aggregating the system characteristics of site and growth conditions (such as elevation and exposure in mountains, soils, hydrological regimes etc.).

Pools of organic carbon include live biomass, dead wood, and soil carbon. Live biomass is calculated based on regionally distributed multi-dimensional regressions of Biomass Extension Factors which include region, aggregated forest type, dominant species, age, site index and relative stocking [10]. These regressions are based on ~7000 sample plots and allow to assess live biomass by 7 components (stem wood, branches, foliage, coarse roots, fine roots, understory (undergrowth + shrubs), and green forest floor). Coarse woody debris that includes logs, snags, stumps, and dry branches of living trees is assessed based on field measurements on sample plots and relevant data of forest inventory. Soil carbon is assessed for on-ground organic layer and 1m top layer of mineral soil based on soil map at scale at 1:2.5 M and corresponding database of typical soil profiles [11].

Major carbon fluxes that directly describe production process include Net Primary Production (NPP), Soil Heterotrophic Respiration (SHR), decomposition of coarse woody debris (DEC), fluxes due to disturbances (D), and lateral fluxes. By definition, NECB also includes other carbon contained substances like methane (CH₄), carbon oxide (CO), Volatile Organic Compounds (VOC) and particulates. NPP is assessed by a tentatively unbiased semi-empirical method which is based on modelling of full production of live biomass by components presented in models of bioproductivity [11]. A special empirically based modelling system was used for assessing SHR [11]. Decomposition of dead wood is described by kinetic models of the 1st order. Fluxes due to disturbances include fire, outbreaks of insects and deceases and impacts of unfavorable weather and environmental conditions [2,5]. Harvest and later fluxes of wood products (import, export) were assessed following Ciais et al. [13]. Fluxes to the hydrosphere are estimated based on measurements of DOC in water reservoirs including estimation of outgassing [14,15]. Emissions of methane and VOC were estimated based on dataset of field measurements and simplified models of dependences of emissions on different classes of forest cover.

A disputable and not finally solved question is relevance of the account of impacts of elevated concentration of CO₂ and deposition of nitrogen on vast and to a substantial impact unmanaged forests of Russia. The data on this topic for Russian forests are scarce and not consistent. At this stage, we used an aggregated approach which combined recognized but not accounted impacts on forest health and productivity. Observation on permanent sample plots [e.g. 16] and analysis of data of forest inventory [17,18] indicated that during the last 4 decades the increase of productivity (expressed in terms of growing stock volume) was 0.2-0.4% yr⁻¹. Such corrections were implemented when updating forest inventory data for input them in the ILIS was provided.

All fluxes which depend on climatic or environmental conditions and are calculated based on databases of measurement *in situ* are corrected for seasonal weather and environment conditions.

2.2 Assessment of uncertainty

Uncertainties within LEA were calculated in the following way: 1) analysis and numerical attribution of accuracy of input data; 2) calculation of precision of intermediate and final results; 3) use of error propagation theory (assuming the Gaussian distribution) and/or numerical differentiation for assessing the precision of intermediate

and final results; 4) expert estimation of completeness of the account and „transformation” of precision in uncertainty using the sensitivity analysis. Note that in practice basically „summarized errors” of input data, i.e. a mixture of random and systematic errors are available. Two end points of the assessment were considered: assessment of the unknown „fixed true value” and unknown true distribution.

The situation with assessment of uncertainties of parameters obtained by other methods is more diverse. Such results are usually derived from different studies which are not coordinated each other in any way. While *process-based models* (e.g., DGVMs) remain practically a sole method for explanation of processes and prediction, they have a number of specific features which should be taken into account: 1) as a proxy, DGVMs present only part of NECB (either Net Biome Production or Net Ecosystem Production); 2) they use a very simplified land cover classification with a limited number of plant functional types; part of these classification do not consider such important land classes as wetland or agricultural land; 3) substantial part of DGVMs is based on modelling „potential” vegetation and consider in very simplified way (or not consider) disturbances; 4) as global models, they are not able to properly describe some important regional features, e.g., specifics of impacts of processes on permafrost on forests of high latitudes [19]. *Eddy covariance method* presents a direct „bottom-up” estimate the Net Ecosystem Exchange (NEE) is widely used for parametrization of different models but at this stage cannot be used for upscaling for forests of the entire country due to very small amount of measurements (totally only in 17 different sites of which 13 were in forests). *Inverse modelling* is an only methods of a „top-down” control of NEE. Uncertainty of measurements of some components of the FCA by remote sensing (e.g., NPP) substantially depends on completeness of regional validation and reliability of the models used at the regional level. Very often, the proper assessment of this type of uncertainties requires additional regional validation.

Harmonizing and mutual constraints of the results obtained by different methods have some specifics. First, the methods estimate different final indicators of carbon cycling: LEA – NECB, DGVMs – NBP, eddy covariance and inverse modeling - NEE. Second, the estimated uncertainties for DGVMs and inverse modelling differ from those of LEA and eddy covariance because they are usually calculated as standard deviation between different models of the ensembles used. This impacts the essence of the final (system) results constrained by the Bayesian approach, particularly in the judgment about confidential intervals.

3. Results and discussion

Application of the LEA to Russian forests for 2007-2009 gave the following major results. NECB was estimated as the net sink of $546 \pm 120 \text{ Tg C yr}^{-1}$ with substantial spatial variability: significant areas on permafrost and in disturbed forests serve as a carbon source. Uncertainties of major carbon pools were estimated (CI is equal 0.9, here and below) : live biomass $\pm 5.0\%$ and dead wood $\pm 9.7\%$. Soil carbon pool could be estimated only very approximately (at level of 7-10%) that – taken into account a high size of this pool - limits the potential use of pool-based methods in the FCA. Uncertainties of major fluxes were estimated: NPP $\pm 6\%$, HSR $\pm 8\%$, DEC $\pm 12\%$, fire $\pm 23\%$, biotic factors $\pm 25\%$, forest harvest and use of forest products $\pm 25\%$, flux to the hydrosphere and hydrosphere $\pm 33\%$. These data were obtained assuming that the estimates do not have significant systematic errors.

Other published results of carbon budget of Russian forests are diverse. Using the pool-based method and the FAO definition of forest (the LEA used the Russian national definition) Pan et al. [20] defined the sink of Russian forests at $463 \pm 116 \text{ Tg C yr}^{-1}$ during 1990-2007. Transition to the Russian definition of forests gives the forest sink at $\sim 530 \text{ Tg C yr}^{-1}$, i.e. very close to the above flux-based estimates by the LEA. However, this publication calculated change of soil carbon by using models based on one-shot measurements of estimated indicators that allows to assume that uncertainty of this result is underestimated. There are a number of other „inventory” based estimates of the carbon sink for different years. These estimates reported NBP in the range from 200-800 TgC yr^{-1} . However these studies do not report any uncertainties and often contain simplified approaches.

Based on inverse modeling, carbon sink estimates for Russia (all land classes) are rather consistent. Within the Global Carbon Project Dolman et al. [7] used 12 different inversion schemes for different periods between 1992 and 2008 and reported the average sink at $-690 \text{ Tg C yr}^{-1}$ although the inter-model variation is high – the standard deviation was $\pm 246 \text{ Tg C yr}^{-1}$. Sink for 2000-2004 that was received for vegetative land of Russia by four different inversion models on average reported $-0.65 \pm 0.12 \text{ Pg C yr}^{-1}$ (P.Ciais, personal communication). These results are in line with a majority of previous studies for large Russian regions like Boreal Asia or Central Siberia [21,22,6].

Results presented by DGVMs are less consistent. While NPP estimates by ensembles of DGVMs is very close to major part of “semi-empirical” assessments (e.g., about 7% of the LEA results), the NBP differs for about 50% [5,6,7,23]. The reason of this may be found in a balance between NPP and HR that to a significant extent is prescribed by DGVM approaches. However, this is not a case for high latitudes with their low intensive rates of decomposition of dead organic where fire is an important regulator. In addition, some substantial components of the FCA are omitted in current generations of DGVMs [1].

Upscaling the direct measurements of NEE by eddy covariance is very uncertain. One of a very few attempts realized in [7] gave the estimate in range from -760 to $-1097 \text{ Tg C yr}^{-1}$. However, the certainty of this conclusion is basically in field of expert judgement.

Application of the Bayesian approach to results received by the LEA, pool-based methods from [20] and inverse modelling from different publications resulted in $560 \pm 117 \text{ Tg C yr}^{-1}$. Note that confidence interval of such an estimate, like and possible bias, could be estimated only in a very approximate way.

Taking into account the estimates of uncertainties obtained in this study, the following overall conclusions could be done: 1) with a high probability Russian forests served as a net carbon sink with NECB at $550\text{-}650 \text{ Tg C yr}^{-1}$ during the last decade; uncertainty of this average is in limits of 15-20%; forests provide at 90-95% of net sink of the total land flux; 2) temporal and spatial variability of the carbon sink is high, particularly for individual region of the country; this variability is basically explained by interannual variability of seasonal weather and connected to this natural disturbances like fire and insect outbreaks; 3) in spite of the high average sink, there are vast areas (mostly in disturbed forests and in forest on permafrost) which serve as a carbon source or are close to the neutral state; 4) the last decade demonstrate a weak trend of decreasing the NECB.

In spite of substantial decrease of uncertainties of the FCA and increase of formal strictness of the results in this study for Russian forests, a number of expert estimates and unrecognized biases remain. Evidently, this is inevitably at this stage of cognition of impacts of terrestrial ecosystems on global biogeochemical cycles. However, the

approach used allows to exclude the clear outliers from intermediate results or to stress a need to pay a special attention to questionable results of other studies. At the same time, this study highlighted a number of system requirements to major methods of studying the carbon cycle. The initial important consideration is a relevance of development of an integrated information base which could be used by all the major methods developed for understanding emissions to, and removals out, greenhouse gases by the terrestrial biosphere. An experience of development of the Integrated Land Information System seems very promising for that. Using such a system might substantially improve information capacity of process-based models and generate a solid basis for upscaling of „point” measurements, e.g. in eddy covariance applications. Another lesson is a clear evidence and need of a system improvements of practically all methods of study of the biospheric role of terrestrial vegetation if an integrated analysis would be used. Finally, an important and unresolved question is a search of relevant tools for harmonizing and mutual constraints of independently obtained results. In current applications, the Bayesian methods is limited by the normal theory but experiences show that empirical distributions, which are usual in the considered system, might be very far from any normal regularities.

References

- [1] Ciais, P. et al. (2015). Observed regional carbon budgets imply reduced soil heterotrophic respiration. *Science* (submitted).
- [2] Nilsson S. et al. (2007). Uncertainties of a regional terrestrial biota full carbon account: A systems analysis. *Water, Air, and Soil Pollution*, 7, 425-441.
- [3] Schellnhuber, H.J. (2015). Integration assessment of adaptation and mitigation. *World Climate Change Conference, Moscow*, pp. 94-95.
- [4] Rittel, H.W. and M.M. Weber (1973). Wicked problems, *Management Science* 4, 155-169.
- [5] Shvidenko et al. (2010). Can the uncertainty of full carbon accounting of forest ecosystems be made acceptable to policy makers? *Climatic Change*, 103 (1-2), 137-157.
- [6] Quegan, S. et al. (2011). Estimating the carbon balance of central Siberia using a landscape-ecosystem approach, atmospheric inversion and Dynamic Global Vegetation Models. *Global Change Biology*, 17 (1), pp. 351-365.
- [7] Dolman, A.J. et al. (2012). An estimate of the terrestrial carbon budget of Russia using inventory-based, eddy covariance and inversion methods. *Biogeosciences*, 9, 5323-5340.
- [8] Schepaschenko, D.G. et al. (2015). Area of Russian forests and its dynamics estimated at basis of synthesis of remote sensing products. *Forest Science*, 3, 163-171 [in Russian].
- [9] Schepaschenko, D.G. et al. (2011). A new hybrid land cover dataset for Russia: a methodology for integrating statistics, remote sensing and in situ information. *Journal of Land Use Science*, iFirst, doi: 10.1080/1747423X.2010.511681, 1-15.
- [10] Shvidenko, A. et al. (2008). Tables and models of growth and biological productivity of forests of major forest forming species of Northern Eurasia (standard and reference

- data), 2nd edition, supplemented. Federal Forest Service of Russia and International Institute for Applied Systems Analysis, Moscow, 886 pp. [in Russian].
- [11] Mukhortova L. et al. 2015. Soil contribution to carbon budget of Russian forests. *Agricultural and Forest Meteorology*, 200, 97-108.
- [12] Shvidenko, A. et al. 2011. Impacts of wildfire in Russia between 1998-2010 on ecosystems and the global carbon budget. *Proceedings of the Russian Academy of Sciences (Doklady Earth Sciences)*, Vol. 441, part 2, pp. 1678-1682).
- [13] Ciais P. et al. 2008. The impact of lateral carbon fluxes on the European carbon balance. *Biogeosciences*, 5, 1259-1271.
- [14] Raymond, P.A. et al. (2013). Global carbon dioxide emissions from inland water. *Nature*, 503, 355-359.
- [15] Lauerwald, R. et al. (2015). Spatial patterns in CO₂ evasion from the global river network. *Global Biogeochem. Cycles*, 29, 534-554.
- [16] Sennov, S.N. (1999). Results of 60-year observations of natural dynamics of forest. Saint-Petersburg Forest Research Institute, 97 pp. [in Russian].
- [17] Alexseyev, V.A., Markov, M.V. (2003). Statistical data about forest fund and change of productivity of Russian forests in the second half of the XX century. Saint-Petersburg Forest Research Institute, 271 pp. [in Russian].
- [18] Shvidenko, A. et al. (2007). Materials for understanding of productivity of Russian forests. In: *Basic Problems of Transition of Sustainable Forest Management in Russia*, Proceedings of the Int. Workshop, 3-35 [in Russian].
- [19] Shvidenko, A.Z., Schepaschenko, D.G. (2014). Carbon balance of Russian forests. *Siberian Forest Journal*, 1, 69-92.
- [20] Pan, Y. et al. (2011). 2011. A Large and Persistent Carbon Sink in the World's Forests. *Science* 19 August 2011: 988-993. Published online 14 July 2011 [DOI:10.1126/science.1201609]
- [21] Maksyutov, S. et al. (2003). Effect of recent observation on Asian CO₂ flux estimate by transport model inversions, *Tellus B*, 55, 522-529.
- [22] Gurney, K.R. et al. (2003). TransCom3 CO₂ inversion intercomparison; 1, Annual mean control results and sensitivity to transport and prior flux information, *Tellus B*, 55, 555-579.
- [23] Cramer, W. et al. (1999), Comparing global models of terrestrial net production: overview and key results. *Global Change Biology*, 5, 1-15.

Forest map and its uncertainty as an important input for carbon sink estimation for Poland and Ukraine

Myroslava Lesiv¹, Anatoly Shvidenko¹, Dmitry Schepaschenko^{1,2}, Linda See¹, and Steffen Fritz¹

¹International Institute for Applied Systems Analysis,
Laxenburg, A-2361, Austria

lesiv@iiasa.ac.at, shvidenk@iiasa.ac.at; schepd@iiasa.ac.at; see@iiasa.ac.at; fritz@iiasa.ac.at

²Moscow State Forest University,
Mytischi, 141005 Moscow, Russia

Abstract

Improving knowledge on the land cover and forest ecosystems is of a high importance for carrying out spatial inventories of emissions and removals in forestry as the best way to achieve reliable results of forest carbon account. The region of the study is the territory of Poland and Ukraine, covering a substantial part of European diversity of natural landscapes. In addition, Ukraine and Poland have a high potential to sequester carbon through afforestation. The accuracy of available forest maps varies considerably over space. We have applied the method of geographically weighted regression to generate a hybrid forest map for Poland and Ukraine. This method predicts land cover types based on crowdsourced data obtained from the Geo-Wiki project, and land cover/forest cover products derived from remote sensing. The hybrid forest cover was found to be more accurate than the individual forest maps extracted from global remote sensing land cover products.

Keywords: forest cover, carbon sink, remote sensing.

1. Introduction

Improving knowledge on the land cover and forest ecosystems is of a high importance for carrying out spatial inventories of emissions and removals in forestry as the best way to achieve reliable results of forest carbon account. Not every country provide a full information on forest area and forest spatial distribution, including distribution of tree species and their age [1]. The reasons for this are different, e.g. absence of forest inventory in the territories that do not belong to forest enterprises; unavailability of data about private forests; obsolete data of forest inventories; existence of territories with rapid changes of forest cover, e.g. encroachment of forests in abandoned agricultural land. Providing an accurate data on forest spatial distribution is one of the steps towards an appropriate estimation of full carbon account [2]. One of the ways to complement the forest data is involving remote sensing data in the estimation of forest area and forest parameters. In this study, we developed a new forest map at a resolution of 60 m by fusing available data derived from remote sensing.

The region of the study is the territory of Poland and Ukraine, covering a substantial part of East-European diversity of natural landscapes. Forest in Poland covers more than 30% of the total area of the country while Ukraine is a forest-poor country with less than 16%. This provides a contrasting set of countries for analysis. In addition, Ukraine and Poland have a high potential to sequester carbon through afforestation [3].

As input data we used a number of global land cover products as well as global forest maps that have become recently available. The accuracy of these maps varies considerably over space [4]. We have applied data fusion methods to combine available sources of forest allocation in order to produce a hybrid product of higher accuracy than

any of the individual input maps. Particularly, we have applied the method of geographically weighted regression (GWR) to generate a hybrid forest (raster) map for Poland and Ukraine. This method predicts land cover types based on (1) crowdsourced data obtained from the Geo-Wiki project (<http://geo-wiki.org/>), which are assumed to be true, and (2) land cover/forest products derived from remote sensing (e.g., LANDSAT-based Hansen's forest change, Globeland 30m, JAXA forest presence/absence). The year of reference of the input data is 2010.

The paper includes methodology description and analysis of the results.

2. Methodology

2.1 Input layers

Recently a number of remote sensing products has emerged. The overall trend has been towards higher spatial resolution such as the 30-meter resolution maps of percentage forest cover, forest cover gain and loss by Hansen [5], and the 30m Globeland product [6]. These maps were developed from Landsat high resolution satellite imagery, which has recently become freely available [7]. Another example is a new JAXA forest/non forest map at a resolution of 25m [8]. A resolution of other available remote sensing datasets is much higher, e.g. Globcover 2009 with a resolution 300m [9], MODIS vegetation continuous fields 250m [10], etc. Disaggregation of the medium resolution products increases uncertainty of forest distribution in space. Therefore Hansen's tree cover, Globeland 30m and JAXA forest/non-forest products have been chosen to develop a hybrid forest map at a resolution of 60m for the year 2010. The short description of the input products is below.

Landsat-based tree cover 2000 by Hansen is a global forest cover change product for the years 2000–2012 with a spatial resolution of 30 m [5]. The product is based on Landsat imagery and has three components: forest cover 2000, forest gain 2000–2012 and annual forest loss. We created a forest map for 2010 by combining the data from three levels: a basis – forest map 2000 – plus forest gain and minus forest loss for the time period of 2000-2010.

The 30m Globeland product 2000/2010 is provided by National Geomatics Center of China [11]. It is based on Landsat imagery with the combination of land resource information and HJ-1 satellite image. The product is freely available and comprises ten land cover classes including forest. We extracted the forest mask from Globeland 30m 2010 for Poland and Ukraine.

Japan Aerospace Exploration Agency (JAXA) has produced the 25 m forest/non-forest map based on imagery from the Phased Array type L-band Synthetic Aperture Radar (PALSAR) aboard the Advanced Land Observing Satellite "DAICHI" (ALOS) [8]. The product is available also at a resolution of 10 m.

The three forest maps were aggregated to the resolution of 60m in order to minimize the spatial errors while comparing different grids. We then calculated the average percentage of forest cover in a 60m pixel for every product.

2.2 Reference data from Geo-wiki

Reference data on forest cover were collected through the Geo-Wiki project [12], which aims at validating, correcting and enhancing land cover products. Five forestry and remote sensing experts collected the data by visually estimating land cover visible in cells of a grid overlaid onto high resolution Google Earth imagery. Figure 1 illustrates the example of collecting forest data through a customised Geo-wiki application. The

60m grid was used as the basis for the output map. Our samples of training data and validating data were randomly generated in forest and non-forest areas.

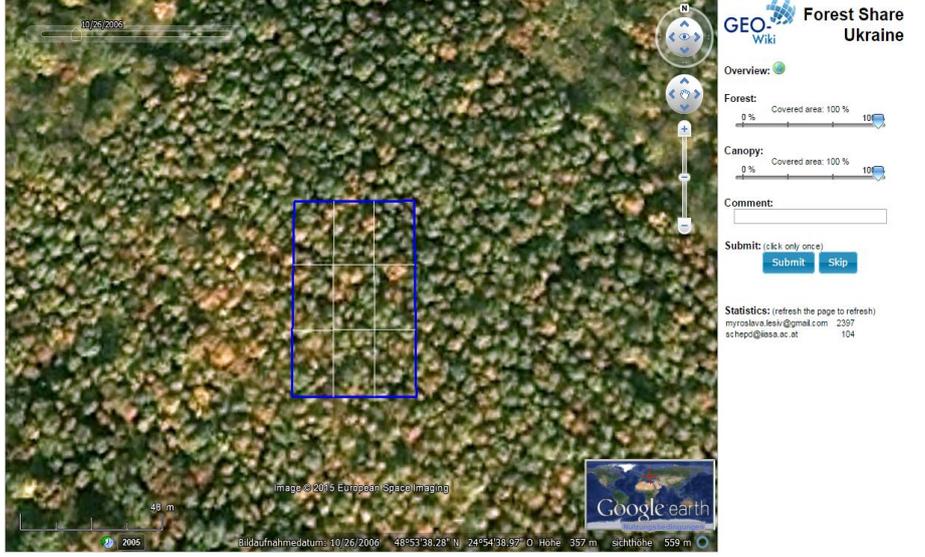


Figure 1. A customised geo-wiki application for collecting forest cover data

The final training data set contains approximately 14 K and 6 K pixels of land cover information (presence/absence of forest) for Ukraine and Poland, respectively. The validation datasets include approximately 4 K and 2 K pixels for Ukraine and Poland, respectively.

2.3 Geographically weighted regression

To combine the three above land cover products and Geo-wiki training data on forest presence/absence, geographically weighted regression (GWR) is employed for development of forest cover map [13]. GWR estimates model parameters at each geographical location by using a kernel. In addition, the observations are weighted by distance, so those closer to the studied location have more influence on the parameter estimates.

The probability of forest presence was then estimated using logistic GWR where the probabilities of correspondence between the Geo-Wiki training data and the input layers were calculated as follows:

$$\logit(P(y_i = 1)) = b_{0(u_i, v_i)} + b_{1(u_i, v_i)}x_{1(i, j)} + b_{2(u_i, v_i)}x_{1(i, j)} + \dots + b_{n(u_i, v_i)}x_{n(i, j)}$$

where $P(y_i = 1)$ is the probability of forest at each location i ; *logit* is a logistic regression; (u_i, v_i) is the two-dimensional vector of location i ; $b_{0(u_i, v_i)}$ is the interception term; $b_j, j = 1, n$ are coefficients of logistic regression model; $x_j, j = 1, n$ indicate the presence of forest cover by global land cover product j ; n is a number of input datasets. and n is the number of input datasets.

Maps of forest probabilities were converted to forest presence/absence maps by applying a threshold of 50%, following the example of the usage of logistic regression models in [14]. The hybrid forest map was developed in the R environment, which is a free statistical software with various geographical libraries.

3. Results

We developed a hybrid forest map for the year 2010 for Poland and Ukraine. It is a first forest map for those countries at a resolution of 60m. Figure 2 presents the forest distribution of Poland and Figure 3 corresponds to the forest distribution of Ukraine.

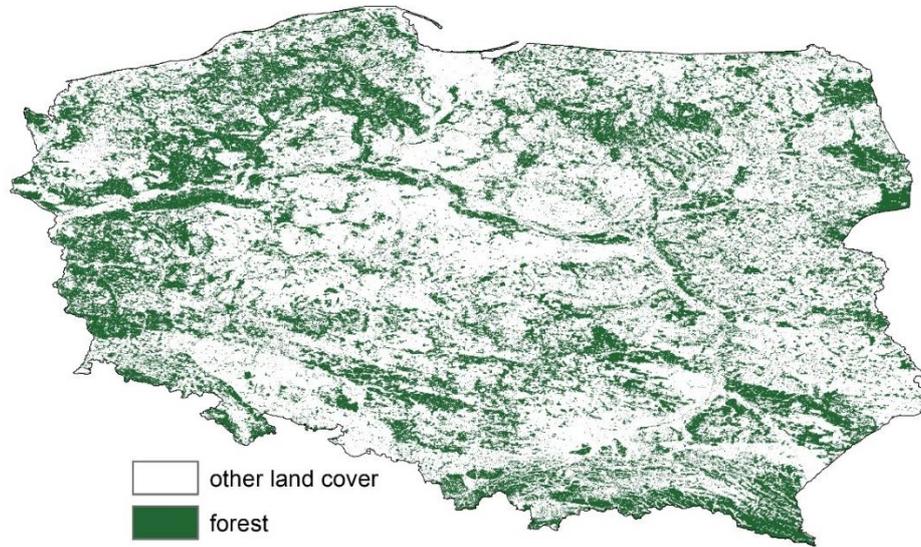


Figure 2. Forest cover map of Poland, 2010

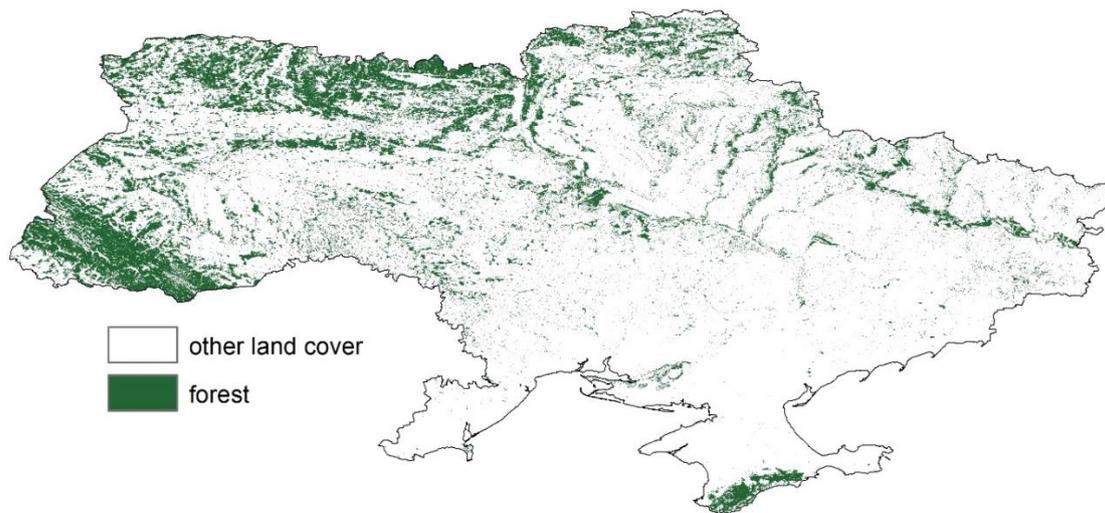


Figure 3. Forest cover map of Ukraine, 2010

The accuracy of the resultant map was assessed by using an independent validation dataset for Poland and Ukraine. Table 1 includes the estimated uncertainty, sensitivity and specificity of the hybrid forest map, and aggregated Hansen's tree cover, Globeland 30m and JAXA forest/non-forest maps. Sensitivity is calculated as the proportion of true positives, and specificity is calculated as the proportion of true negatives.

The hybrid forest maps are more accurate than the input layers. Globeland 30m has the lowest accuracy and, therefore, cannot be used for identification of forest changes during the period 2000-2010. Hansen's tree cover and JAXA forest/non-forest also need

to be improved by the producers as they cannot be used for detection of forest changes in space neither.

Table 1. Accuracy, sensitivity and specificity of the hybrid forest map and the input

Forest maps	Ukraine			Poland		
	Accuracy %	Sensitivity %	Specificity %	Accuracy %	Sensitivity %	Specificity %
A hybrid map	90,1	91,0	87,0	91,2	95,6	86,1
Hansen's map	86,7	95,0	77,6	89,0	87,0	91,3
Globeland 30m	88,2	92,3	83,1	82,6	85,5	80,5
JAXA map	84,7	84,0	85,2	83,2	91,6	72,2

The total forest area from the hybrid maps has been found to be approximately 9.56 mln ha and 9.7 mln ha for Ukraine and Poland, respectively. In official statistics, Ukraine reports to have 9.57 mln ha of forest land [15]. Such a high consistency seems surprising taking into account that reliable inventory data is available only for 8.5 mln ha of forest. In addition, official Ukrainian reports do not account forest land on abandoned agricultural land and contain obsolete data about protective forests and shelterbelts on agricultural land, particularly in steppe and forest steppe zones of the country. These processes are revealed on regional level providing increase the forest area in the northern part and decrease – in the southern one. The simplified calculation allows us to conclude that the hybrid map estimates the total forest area in Ukraine with uncertainty in limits of 2-3%, while regional estimates are more uncertain and likely less biased than forest inventory data.

According to the official forest reports of Poland, the country has 9.2 mln ha of forest land [16]. Taking into account that the hybrid map also covers the settlement areas covered by trees (e.g., parks and garden), this also could be a reason of some, relatively small discrepancy in our estimates and official data of forest areas in Poland and Ukraine.

4. Conclusions

The hybrid forest cover for Poland and Ukraine was found to be more accurate than the individual forest maps extracted from global remote sensing products. Overall, these estimates are rather close to the countries' official statistics taken into account some inconsistency in the forest definitions used by official statistics and by this study. The two major current processes of rapid changes of forest area of Ukraine are: 1) restoration of forest vegetation on abandoned agricultural land in the forest zone and 2) impoverishment of protective forests in the southern part of the country. These processes provide different impacts on the change of forest area are not satisfactory reflected by the official forest inventory.

For countries that do not currently have an accurate enough land cover data, the presented methodology provides an opportunity to develop forest maps that can be further used in different national, regional and global applications, including accounting and verification of emissions of greenhouse gases in space and time. This study shows that uncertainties of such maps do not exceed uncertainties of other components of carbon budget of forest ecosystems.

Acknowledgement

The work was supported by Marie Curie grant FP7-MC-IIF: SIFCAS Project no. 627481.

References

- [1] Schepaschenko, D., L. See, M. Lesiv, I. McCallum, S. Fritz, C. Salk, E. Moltchanova, C. Perger, A. Shvidenko, F. Albrecht, F. Kraxner, A. Bun, M. Duerauer, S. Maksyutov, A. Sokolov, M. Obersteiner, V. Karminov, and P. Ontikov (2015). Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics. *Remote Sens. Environ.*, 162:208-220.
- [2] Schepaschenko, D., L. See, I. McCallum, C. Schill, C. Perger, A. Baccini, H. Gallaun, G. Kindermann, F. Kraxner, S. Saatchi, M. Obersteiner, M. Santoro, C. Schmullius, A. Shvidenko, and M. Schepaschenko (2012). Observing Forest Biomass Globally. *Earthzine*, <http://earthzine.org/2012/06/09/observing-forest-biomass-globally/>.
- [3] Galos, B., A. Hansler, G. Kinderman, D. Rechid, K. Sieck, and D. Jacob (2012). The role of forests in mitigating climate change - A case study for Europe. *Acta Silv. Lignaria Hung.*, vol. 8, pp. 87–102.
- [4] Fritz S., P. Havlik, U. A. Schneider, E. Schmid, R. Skalský, and M. Obersteiner (2009). Uncertainties in global land cover data and its implications for climate change mitigation policies assessment. Presented at the 33rd International Symposium on Remote Sensing of Environment (ISRSE-33), Stresa, Italy.
- [5] Hansen, M., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, vol. 342, no. 6160, pp. 850–853.
- [6] Yu L., J. Wang, and P. Gong (2013). Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: a segmentation-based approach. *Int. J. Remote Sens.*, vol. 34, no. 16, pp. 5851–5867.
- [7] Wulder, M.A., J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.*, vol. 122, pp. 2–10.
- [8] Shimada, M., T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas (2014). New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sens. Environ.*, vol. 155, pp. 13–31.
- [9] Bontemps, S., P. Defourny, E. van Bogaert, O. Arino, V. Kalogirou, and J. R. Perez (2011). GLOBCOVER 2009: Products Description and Validation Report. European Space Agency.
- [10] DiMiceli, M., M. L. Carroll, R. A. Sohlberg, C. Huang, M. C. Hansen, and J. R. G. Townshend (2011). Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover. University of Maryland, College Park, MD, USA.
- [11] Jun, C., Y. Ban, and S. Li (2014). China: Open access to Earth land-cover map. *Nature*, vol. 514, no. 7523, pp. 434–434.

- [12] Fritz, S., I. McCallum, C. Schill, C. Perger, L. See, D. Schepaschenko, M. van der Velde, F. Kraxner, and M. Obersteiner (2012). Geo-Wiki: An online platform for improving global land cover. *Environ. Model. Softw.*, vol. 31, pp. 110–123.
- [13] Schepaschenko D., L. See, M. Lesiv, S. Fritz, I. McCallum, C. Perger, A. Shvidenko, and F. Kraxner (2013). Global hybrid forest mask: synergy of remote sensing, crowd sourcing and statistics. Presented at the AGU Fall Meeting, AGU Fall Meeting. San Francisco, CA.
- [14] Pampel, F. C, *Logistic Regression: A Primer*. SAGE, 2000.
- [15] Book on Forest Fund of Ukraine (state at 01.01.2011) (2012). Irepen', Ukrstateforestproject, 130 pp. [in Ukrainian].
- [16] Raport o stanie lasów w Polsce 2010 (2011). Centrum Informacyjne Lasów Państwowych, Warszawa, ISSN 1641-3229, 84 pp. [in Polish]

Spatial GHG inventory in the Agriculture sector and uncertainty analysis: A case study for Poland

Nadiia Charkovska¹, Rostyslav Bun^{1,2}, Olha Danylo^{1,3},
Joanna Horabik-Pyzel⁴, Matthias Jonas³

¹Lviv Polytechnic National University, St.Bandery, 12, Lviv, 79013, Ukraine,
mail: charkovska.n@gmail.com;

²Academy of Business in Dąbrowa Górnicza, Poland

³International Institute for Applied Systems Analysis, Laxenburg, Austria;

⁴Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland

Abstract

Estimation of uncertainties is an important part of complete inventory of greenhouse gas (GHG) emissions. Information on uncertainty is intended not only to question the reliability of inventory estimates, but to assist in the identifying priority measures to improve the quality of future inventories. This article discusses bottom-up inventory from the agricultural sector in Poland. Accordingly to the developed geoinformation approach area-type sources of emission (arable lands, rural localities) were investigated. In implemented mathematical models for the estimation of GHG emissions from agricultural activity the statistical data on animal and crop production, as well as specific emission factors were used. Methods for the spatial inventory of GHG emissions from agricultural sources, taking into account the specifics of animal nutrition, are described. Monte-Carlo method was applied for a detailed estimation of uncertainty "from category to category," because uncertainties of input parameters (CH₄ and N₂O emission factors) are large and non-normally distributed (95% confidence interval). The land use map is used to calculate the territorial distribution of GHG emissions. The structure of total GHG emissions on different categories of animal sector and agricultural soils sector by type of GHG is presented and visualised as digital maps. Analysis of uncertainty of GHG inventory results were carried out for voivodeships. Results are presented as sets of numerical values of the bounds of confidence intervals for the main GHGs and at different levels of spatial disaggregation. The improving of knowledge on territories, where emissions took places, enables us to better inventory process and reduce the overall uncertainty.

Keywords: GHG emission, spatial GHG inventory, agriculture sector, uncertainty analysis, Monte-Carlo method.

1. Introduction

During the last century the environment has experienced a lot of irreversible changes. Equally serious impact of global climate change felt the economies of many world countries and humanity in general. Most of scientists in the field of climate changes research affirm that climate change is largely, except natural factors, influenced by results of anthropogenic action. According to the latest assessment report of the IPCC the human activity from 95-100% degree of confidence is the main reason of climate changes after 1950. First of all anthropogenic factors include increasing the concentration of greenhouse gases (GHG) in the Earth's atmosphere and its pollution with the tiniest solid particles. For example, in Ukraine and Poland we are watching more frequent droughts and floods, which are the main reason of agriculture productivity reduction. Apart from the energy sector, a significant share in terms of GHG emissions belongs to agricultural activity.

The IPCC has developed a universal traditional methodology of GHG inventory in different categories of anthropogenic activity [9]. Using of these methods makes it possible to form national reports about GHG emissions and provides emissions assessment at the level of the whole country. General methods are ineffective for evaluation of emissions at the regional level, because they don't take into account the specifics of emission processes and irregularity of territorial distribution of the emission sources. At the same time, it's more useful to implement the essentially new spatial inventory of GHG emissions with the possibility of assessment on small areas of territory and building spatial emission inventories in order to plan the strategic development of individual regions. It's also important that GHG inventory loses its significance, without the uncertainty analysis of input and output data (statistical information about the results of anthropogenic activity, the emission factors, the emission estimates) [4].

Below an approach is presented for spatial inventory of GHG emissions in agriculture sector in Poland. For all categories of this sector covered by IPCC Guidelines [9], we analyzed the sources of emissions in terms of their spatial representation. Such emission sources can be analyzed as area-type (diffused) objects. We built the digital maps of the sources using Corine Land Cover vector map [7], and analyzed them as polygons without using any regular grid, as it is often made. Such elementary objects are split by administrative boundaries regions/voivodeships, districts/powiats, and municipalities/gminas. It gives us a possibility to keep administrative assignment of each elementary object. Then we created the algorithms for calculating GHG emissions from these objects using activity data and emission coefficient. For the activity data assessment, we have developed the algorithms for disaggregation of available statistical data (at the lowest level as possible) to the level of elementary objects.

Using created digital maps and mathematical models we carried out spatial inventory of emissions for each elementary object and got sets of geospatial data on GHG emissions caused by enteric fermentation, manure management, agricultural soils etc. (according to the agriculture sector structure in the IPCC Guidelines [9]). Maximum resolution is determined by the resolution of used digital maps of land use and does not exceed 100m. Below, this approach is illustrated on the example of animal sector only.

2. The specificity of greenhouse gases emissions processes

Animal sector, as one of the subsectors of agriculture, plays a very important ecological, economic and social role in various parts of the world. The emissions of GHG from animal sector occur as a result of the animals enteric fermentation (dairy and non-dairy cattle, sheep, goats, horses and pigs), and also the decomposition, collection, storage and use of animal manure in various storage systems (manure reservoir in solid and liquid forms separately). However, the scientific literature has not evaluated the long-term trend of GHG emissions from animal sector separately for developed and developing countries [4].

Except animals, the cultivated lands (arable lands), where agricultural crops grow that are manured by various kinds of fertilizers, and thanks to them the processes of leaching and runoff of nitrogen take place, and it can be considered polygonal (area-type) sources of emissions. The changes in agricultural production and, consequently, changes in GHG emissions since the mid-1990s were mainly caused by adaptation to the demand in the domestic market, priorities of international trade, the prices of

agricultural means of production, such as machinery and agricultural prices. Since 2004, when Poland joined the European Union agricultural subsidies started to influence on the development of promising tendencies of agriculture.

An analysis of statistical information of livestock numbers in Poland in 2010 showed that in one municipality/gmina the number of pigs was over 800 thousands [1, 2]. Despite strong criticism of environmentalists in this gmina in 2004 it was opened two large pig farms. This case and many others show that emission territorial distribution in animal subsector is essentially non-uniform. Therefore we need tools for spatial analysis of GHG emissions, which will give an opportunity for experts and authorities to take effective measures to reduce emissions in areas where they are high [6].

3. Mathematical models for spatial inventory

During modeling the emission processes in animal subsector in Poland (in categories "Enteric Fermentation" and "Decomposition, collection, storage and use of animal manure") the several presumptions were used. Especially, because there is no possibility to monitor emissions from individual animals, so we estimate total emissions from all animals of one species within each rural locality in general. In proposed mathematical models was taken into account the fact that the Polish statistical data on livestock and poultry served separately for agricultural enterprises and households (population) in gminas/municipalities. It's assumed that the number of animals in the households are distributed geographically between rural settlements in proportion to gmina rural population.

The ratio of the population in the analyzed elementary object to the population in gmina can be calculated as:

$$V(\delta_n) = \frac{p(\delta_n) \cdot \text{area}(R_{3,n_3} \cap \delta_n)}{P(R_{3,n_3})}, \quad n = \overline{1, N}, \quad (1)$$

where $V(\delta_n)$ is the desired share of the population in the n -th elementary object δ_n ; N is the total number of such objects in Poland; $p(\delta_n)$ is the population density in the n -th elementary object; $P(R_{3,n_3})$ is the number of people in gmina; R_{3,n_3} is the third level of administrative unit, which includes the n -th elementary object, that is $\delta_n \subset R_{3,n_3}$ (geographical object δ_n is within the geographic object R_{3,n_3}), besides that $n_3 \in \overline{[1, N_3]}$; N_3 is the number of gminas in Poland; $\text{area}(x)$ is the area of object x , \cap is the operation of intersection of the common area of two geographic objects. Further, this parameter $V(\delta_n)$ is used as an indicator for disaggregation of known statistical data on the number of animal livestock within gmina to the level of elementary objects.

Geographically the farms are located on agricultural lands, that's why statistical data on livestock and poultry within these farms are disaggregated to the level of elementary objects in proportion to the area of agricultural land (arable land, grassland, etc.) using the formula:

$$S(\delta_n) = \frac{\sum_{f_i \in F} \text{area}(f_i \cap \delta_n)}{\sum_{f_j \in F} \text{area}(f_j \cap R_{3,n_3})}, \quad \forall f_i \cap \delta_n \neq 0, \quad f_j \cap R_{3,n_3} \neq 0, \quad n = \overline{1, N}, \quad (2)$$

where $S(\delta_n)$ is the ratio of the sum of areas of agricultural lands $f_i \in F$, that are located within elementary area δ_n , to the sum of such areas of lands in the gmina R_{3,n_3} , which contains this elementary object, that is $\delta_n \subset R_{3,n_3}$, F is the set of elements of digital map of land use of the whole country that are agricultural lands.

The methane emissions from enteric fermentation of animals, which are owned by population and by agricultural enterprises, can be calculated using mathematical model:

$$E_{EntFerm}^{CH_4}(\delta_n) = \sum_{t=1}^T [A_t^{ind}(R_{3,n_3}) \times V_t(\delta_n) + A_t^{agr}(R_{3,n_3}) \times S_t(\delta_n)] \times K_t^{CH_4}(\delta_n), \quad n = \overline{1, N}, \quad (3)$$

where $E_{EntFerm}^{CH_4}(\delta_n)$ is the total annual emissions of methane in the n -th elementary object δ_n ; $A_t^{ind}(R_{3,n_3})$ and $A_t^{agr}(R_{3,n_3})$ are the statistical data on the number of the t -th animal species (dairy cattle, non-dairy cattle, sheep, goats, horses, pigs, poultry) in individual households (rural population) (*ind*) and agricultural enterprises (*agr*) for the appropriate year in gmina R_{3,n_3} , which contains this elementary object δ_n ; $V_t(\delta_n)$ and $S_t(\delta_n)$ are the coefficients calculated using formulas (1) and (2) for disaggregation of statistical data on livestock of the t -th animal species, accordingly, in households and agricultural farms, from R_{3,n_3} gmina level to the level of elementary object δ_n ; $K_t^{CH_4}(\delta_n)$ is the coefficient of methane emission from enteric fermentation for the t -th animal species in the n -th elementary object (in fact, this coefficient depends on the climate zone, in which this object is located); *EntFerm* is the index that means emissions from enteric fermentation.

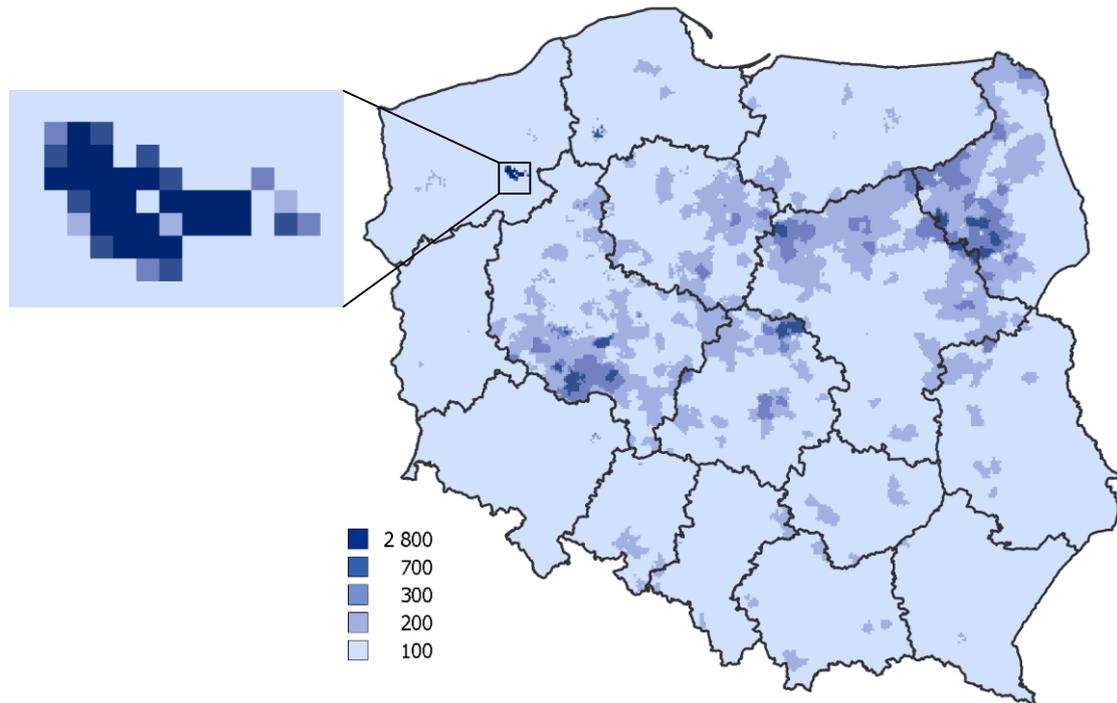


Figure 1. The specific total GHG emissions in animal sector in Poland (elementary areas 2 x 2 km; Mg/km², CO₂-equivalent, 2010)

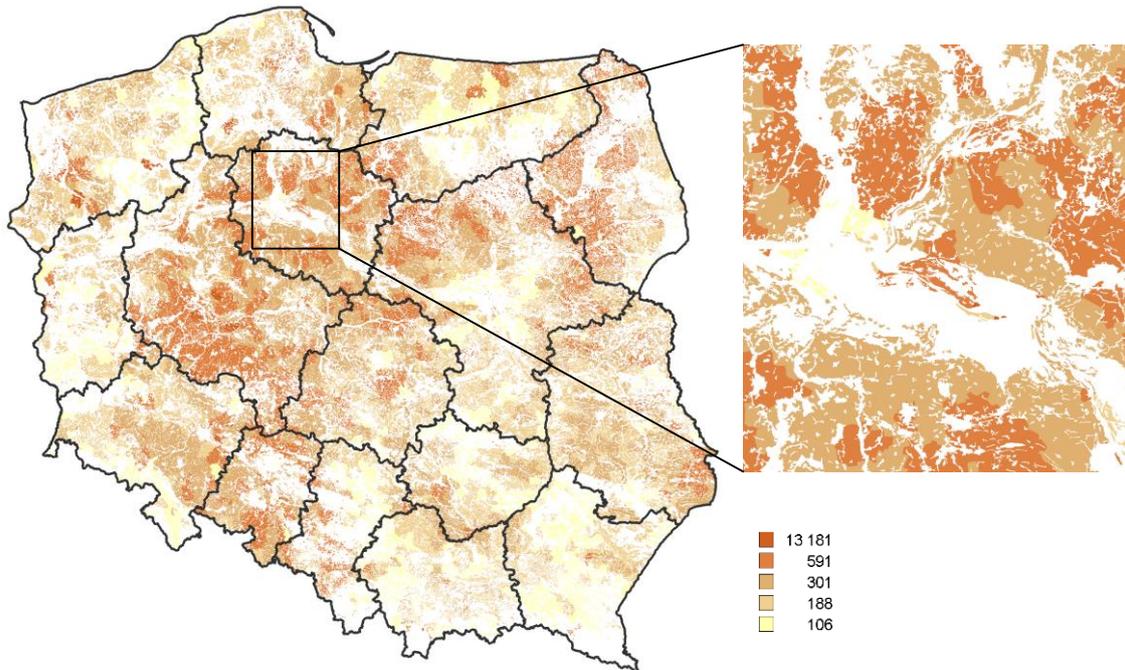


Figure 2. Specific N₂O emissions from fertilization of arable lands in Poland (kg/km², 2010)

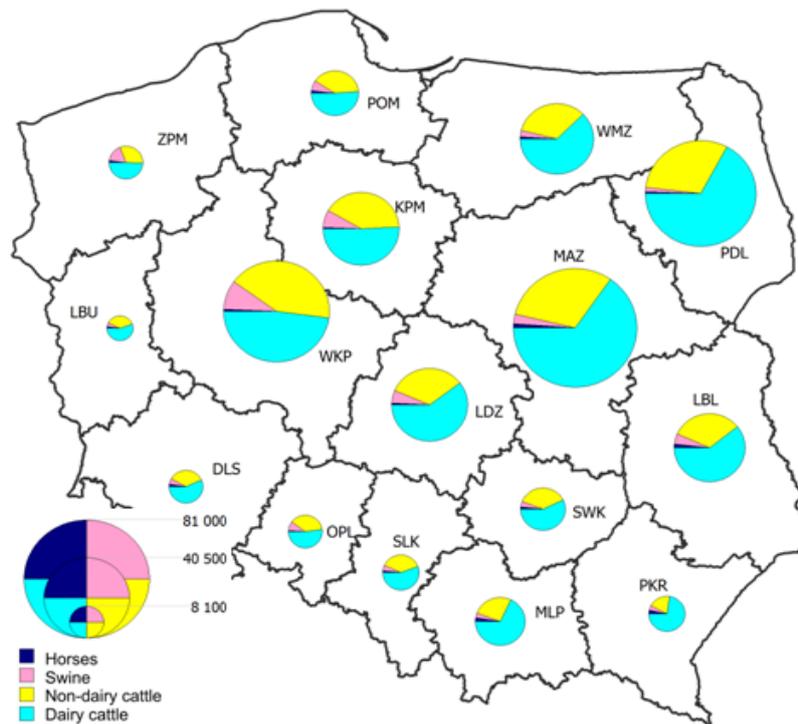


Figure 3. Annual emissions of methane from enteric fermentation of agricultural animals in the voivodeships in Poland (tons, 2010)

4. The results of spatial inventory

Developed mathematical models gave the opportunity to obtain spatial estimates of GHG emissions for each source category in the agricultural sector. The results of computational experiments showed that the largest methane emissions in the agricultural

sector occurred as a result of enteric fermentation of farm animals, such as dairy and non-dairy cattle. In such a way, the results of spatial inventory were obtained at the level of elementary areas (see an example in Figure 1), at the level of arable lands or rural settlements (see an example in Figure 2). The spatial inventory results can be aggregated to the larger area-type objects like the voivodeship in Poland (see Figure 3). The total GHG emissions in the agriculture sector are presented in Figure 4.

As we can see in Figure 3, the biggest emissions of methane in the animal subsector are in the Mazovian voivodeship (80,694 tons), Greater Poland (60,956 tons), and Podlaskie (66,266 tons), but the least is in the Lubusz voivodeship (5,190 tons). The total emissions of methane from enteric fermentation of all species in 2010 amounted to 434.7 ths. tons, that is 75% of total emissions of this gases in animal sector and the rest of 25% is caused by decomposition of manure.

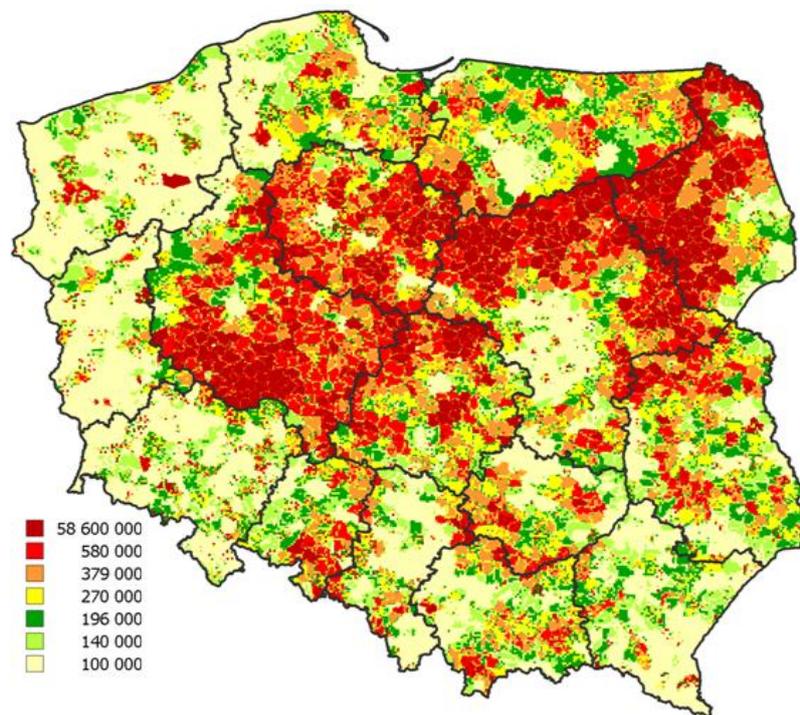


Figure 4. The specific total GHG emissions in the agriculture sector in Poland (elementary areas 2 x 2 km, kg, CO₂-equivalent, 2010)

5. Uncertainty analysis

Input data for developed mathematical models of spatial inventory are not known exactly, and they can be simulated as random variables. For example, the statistical data on livestock population and the specific animal species' GHG emission factors can be attributed to random variables. Currently, one of the main methods of modelling GHG emissions taking into account uncertainty, is Monte Carlo method. Its advantage is the ability of using the information based on uncertainty of input parameters of mathematical models to estimate the level of uncertainty in GHG emissions for different areas, regions and the country as a whole.

The resulting emissions uncertainties in the agricultural sector were analyzed at the level of voivodeships/regions, particularly from enteric fermentation of farm animals (cows, non-dairy cattle, sheep, goats, horses and pigs). As for the uncertainty of statistical data on these animal livestock, it should be noted that the accuracy of the data depends greatly on

the completeness and reliability of the national census methods. In addition, in the census there are different rules for accounting of agricultural animals that don't live during the year, such as pigs, so this should be considered during analysis of emissions uncertainty. Another source of emissions uncertainty from livestock is the use of various data in the formulas to calculate methane emission factor [5].

In the implemented mathematical models of GHG emissions evaluation the agriculture statistical data are used, which uncertainty range for animals is 5% (symmetrical distribution). For modeling GHG emissions in the category "Enteric Fermentation" by Monte-Carlo method the methane emissions factor for agricultural animals (IPCC Guidelines [9]) and appropriate uncertainty ranges (50%, symmetrical [8]) were used. On the basis of implemented geospatial database and developed approach to analysis of uncertainties of GHG emissions were realized computational experiments with the using Monte Carlo method on the investigation of GHG emissions uncertainty from enteric fermentation of agricultural livestock. The results were obtained at the level of voivodeships in Poland (according to statistical data of 2010). Results are presented in Table 1.

Table 1. Input data for the uncertainty analysis of methane emissions from enteric fermentation in region of Poland (2010)

Voivodeship	CH ₄ emissions, tons					
	The limits of uncertainty range, %					
	Dairy cattle	Non-dairy cattle	Pigs	Horses	Sheep	Goats
Lower Silesian	4674,4 ±50,3	3186,1 ±50,1	419,7 ±50,2	203,1 ±50,2	102,6 ±50,2	32,3 ±50,3
Kuyavian-Pomeranian	17143,3 ±50,3	14177,9 ±50,2	2684,2 ±50,2	172,1 ±50,3	111,4 ±50,2	15,0 ±50,2
Lublin	18223,2 ±50,4	14156,3 ±50,2	1510,1 ±50,3	546,6 ±50,3	133,5 ±50,3	62,5 ±50,3
Lubusz	2879,5 ±50,3	2114,8 ±50,4	300,6 ±50,2	107,2 ±50,3	33,6 ±50,2	9,6 ±50,2
Łódź	21064,7 ±50,3	11696,9 ±50,4	1959,4 ±50,1	271,5 ±50,3	120,7 ±50,2	25,6 ±50,2
Lesser Poland	10986,5 ±50,3	4371,4 ±50,3	541,2 ±50,2	385,1 ±50,2	575,4 ±50,3	89,5 ±50,2
Masovian	52734,1 ±50,4	25303,7 ±50,2	2115,5 ±50,1	856,4 ±50,2	72,9 ±50,3	31,6 ±50,2
Opole	4698,3 ±50,3	3674,8 ±50,2	901,3 ±50,2	72,9 ±50,3	23,6 ±50,2	14,1 ±50,3
Subcarpathian	7266,6 ±50,3	2081,6 ±50,3	448,7 ±50,2	318,1 ±50,3	152,8 ±50,3	76,2 ±50,3
Podlaskie	44430,2 ±50,3	20639,0 ±50,3	827,5 ±50,3	363,2 ±50,2	173,0 ±50,2	15,8 ±50,2
Pomeranian	7428,6 ±50,3	5941,1 ±50,2	1262,6 ±50,1	257,4 ±50,3	133,6 ±50,3	14,8 ±50,2
Silesian	5230,6 ±50,2	3670,7 ±50,1	524,8 ±50,2	155,4 ±50,3	110,9 ±50,2	42,6 ±50,2
Świętokrzyskie	7761,7 ±50,4	5056,4 ±50,2	603,4 ±50,2	213,6 ±50,3	33,1 ±50,2	26,3 ±50,3
Warmian-Masurian	20538,9 ±50,4	11384,5 ±50,3	1025,1 ±50,1	300,3 ±50,2	84,5 ±50,2	19,6 ±50,2
Greater Poland	29543,7 ±50,3	26487,1 ±50,2	5879,3 ±50,2	376,8 ±50,2	196,0 ±50,2	92,0 ±50,2
West Pomeranian	4225,2 ±50,4	3042,0 ±50,1	1815,9 ±50,1	159,5 ±50,2	103,8 ±50,2	15,8 ±50,2

The Monte Carlo method was also used for estimation of emissions uncertainty from applying mineral ammonia fertilizers to soils in Poland (on data of 2010). Based on the results of modelling in main categories of animal sector and agricultural soils sector, the uncertainty ranges of emissions amounted to $\pm 12,7\%$ for CH_4 emissions from enteric fermentation and $-51,2\% : + 64,1\%$ for N_2O emissions from synthetic fertilizers applied to soils (symmetric normal distribution and asymmetric log-normal distribution are used). The verification of the correctness of realized mathematical and software tools was carried out using Polish national annual reports [10] on GHG emission at the country level as a whole. The obtained results show a high uncertainty of inventory results in the agricultural sector in 2010.

This should positively affect the total uncertainty of regional or national emissions for all categories of anthropogenic activities and give the authorities the opportunity to take into account this factor in the verification of the fulfilment of international arrangements on reduction of GHG emissions.

Thus the problem of determining of the categories of economic activities, which are important in terms of sensitive analysis, is very interesting. It means that overall uncertainty of inventory results is the most sensitive to the changes in uncertainty of input parameters [2]. Figure 5 illustrates graphically a sensitivity of uncertainty of CO_2 -equivalent emissions from the agricultural activity. The results show that the relative uncertainty for methane emissions is the more dependent on the uncertainty of statistical data on livestock numbers than on the uncertainty of CH_4 emission factor. The uncertainty of total CH_4 emissions in animal sector depends on improving the knowledge about census results. For example, the reduction of uncertainty ranges of animal population into 40% causes the decreasing of CH_4 emissions uncertainty in a half.

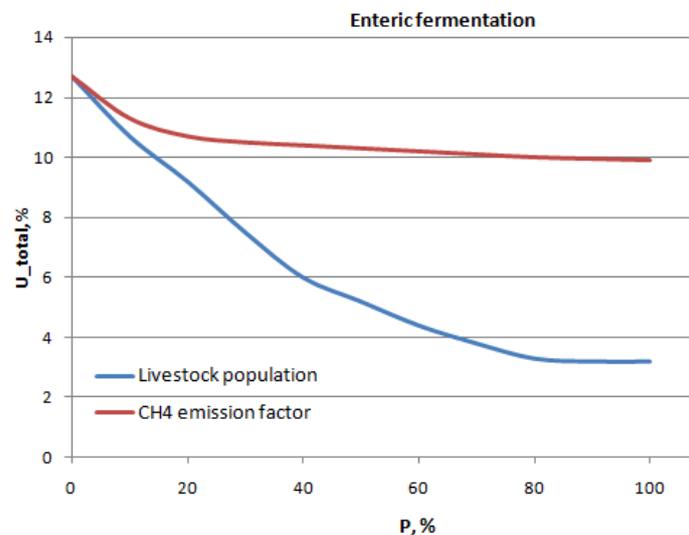


Figure 5. Dependence of uncertainty of CH_4 emissions in enteric fermentation of livestock during decreasing uncertainty of input data into P percent (total as for Table 1; Monte Carlo method)

6. Conclusions

The main GHG emission sources in the animal sector in Poland, in particular enteric fermentation, are analyzed in this paper. Mathematical models of emission processes from these sources at the level of elementary objects of fixed size are useful for spatial

inventory of GHG emissions. Using geoinformation system tools, the geospatial database of statistical information on the number of livestock in Polish regions is formed. As a result of numerical experiments, the estimates of methane emissions by type of animals at the level of elementary areas 2 x 2 km and at the level of voivodeships are obtained.

The obtained results of the spatial analysis of GHG emissions have been showed not so high uncertainties for emissions from enteric fermentation by respective animal species. It has a positive impact on the uncertainty of total regional or national emissions from all categories of anthropogenic activity.

References

- [1] Agricultural census 2010 by holdings headquarter; Livestock (cattle, pigs, horses, poultry), Available online at: <http://www.stat.gov.pl>
- [2] Bank Danych Lokalnych (Local Data Bank), GUS, Warsaw, Poland, Available at: <http://stat.gov.pl/bdl>
- [3] Boychuk Kh., Bun R. (2014) Regional spatial cadastres of GHG emissions in Energy sector: Accounting for uncertainty, *Climatic Change*, 124(3), 561-574.
- [4] Bun R., Hamal Kh., Gusti M., Bun A. (2010) A spatial GHG inventory on regional level: Accounting for uncertainty, *Climatic Change*. Springer Netherlands, 103(1), 227-244.
- [5] Caro D., Davis S.J., Bastianoni S., Caldeira K. (2014) Global and regional trends in greenhouse gas emissions from livestock, *Climatic Change*, 126, 203-216.
- [6] Charkovska N.V., Bun R.A. (2014) Mathematical modeling and spatial analysis of GHG emissions processes from Agriculture sector of Poland, *Proceedings of the International Conference on Environmental Observations, Modeling and Information Systems (ENVIROMIS-2014)*, Tomsk, Russia, SCERT, 76-77.
- [7] Corine Land Cover 2006 raster data, Available at: <http://www.eea.europa.eu/>
- [8] Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, Penman Jim, Dina Kruger, Ian Galbally et al.; IPCC, 2001.
- [9] IPCC (2006) IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K (eds)
- [10] Poland's National Inventory Report 2012, KOBIZE, Warsaw, 2012, 358 p., Available online at: http://unfccc.int/national_reports

Conditionally autoregressive model for spatial disaggregation of activity data in GHG inventory: Application for agriculture sector in Poland

Joanna Horabik-Pyzel¹, Nadiia Charkovska², Olha Danylo^{2,3}, Zbigniew Nahorski¹,
and Rostyslav Bun^{2,4}

¹ Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
Joanna.Horabik@ibspan.waw.pl

² Lviv Polytechnic National University, Lviv, Ukraine

³ International Institute for Applied Systems Analysis, Laxenburg, Austria

⁴ Academy of Business in Dąbrowa Górnicza, Poland

Abstract

This report presents a novel approach for allocation of spatially correlated data, such as emission inventories, into finer spatial scales conditional on covariate information observable in a fine grid. Spatial dependence is modelled with the conditional autoregressive structure introduced into a linear model as a random effect. The maximum likelihood approach to inference is employed, and the optimal predictors are developed to assess missing values in a fine grid. The usefulness of the proposed technique is shown for agricultural sector of GHG inventory in Poland. An example of allocation of livestock data (a number of horses) from district to municipality level is analysed. The results indicate that the proposed method outperforms a naive and commonly used approach of proportional distribution.

Keywords: GHG inventory, agricultural sector, spatial correlation, disaggregation, conditional autoregressive model

1. Introduction

Spatially resolved inventories of greenhouse gases (GHG) contribute valuable information for an assessment of carbon sources and sinks. Various authors point out that a regional or local formulation improves accuracy of the assessment. Quality of these inventories is subject to various conditions; particularly, it depends on availability of high resolution activity data.

In case of national GHG inventories, relevant information about low resolution activity data needs to be acquired from national/regional totals. A procedure of allocation into smaller spatial units (like districts, municipalities, and finally 2x2km grid cells) differs among various emission sectors. Basically, all the emission sources are categorised as line, area or large point emission sources; further steps differ significantly for each group. Area sources comprise e.g. agricultural fields, urban areas as well as highly dense urban transportation network. In this case, a procedure of spatial allocation depends on methods and technologies of fossil fuel combustion in a considered sector [1]. A common approach though, is a spatial allocation made in a proportion to some related indicators that are available in a finer grid.

In this study, the statistical scaling method is developed in order to support the procedure of compiling high resolution activity data. We propose the method for allocating GHG activity data to finer spatial scales, conditional on covariate

information, such as land use, observable in a fine grid. The proposition is suitable for spatially correlated, area emission sources.

Regarding an assumption on residual covariance, we apply the structure suitable for area data, i.e. the conditional autoregressive (CAR) model. Although the CAR specification is typically used in epidemiology [2], it was also successfully applied for modelling air pollution over space [3]. We demonstrate usefulness of the proposed technique for the agricultural sector of GHG national inventory in Poland. The example considers an allocation of livestock data (a number of horses) from district to municipality level.

A part of the methodology described in section 3.1 was already presented in [4]. This contribution extends the basic model for the case of various regression models in each region (here voivodeship); see section 3.2. Performance of the method for livestock data in agricultural sector of GHG inventory is presented in section 4.

2. Inventory livestock dataset

Considered is a livestock dataset (cattle, pigs, horses, poultry, etc.) for the territory of Poland, based on the agricultural census 2010, and available from the Central Statistical Office of Poland - Local Data Bank [5]. The goal is to allocate relevant livestock amounts from districts (*powiaty*) to municipalities (*gminy*).

In particular, for horses the data are available also in municipalities, and this fact enables us to verify the proposed disaggregation method. Therefore, in what follows we consider the task of disaggregation of number of horses reported for 314 districts into 2171 municipalities, taking advantage of covariate information observable for municipalities. Only rural municipalities are considered in the study.

As explanatory variables we use population density (denoted x_1) and land use information. For the latter, the CORINE Land Cover map, available from the European Environment Agency [6], was employed. For each rural municipality we calculate the area of agricultural classes, which may be related to livestock farming. Three CORINE classes were considered (the CORINE class numbers are given in brackets):

- Arable land (2.1); denoted x_2
- Pastures (2.3); denoted x_3
- Heterogeneous agricultural areas (2.4); denoted x_4 .

The results of the disaggregation with the proposed procedure are further compared with the results of allocation proportional to population of municipalities. This naive approach, however, gave rise for a modification of the basic version of the method. Namely, we account for the fact that a relationship of farmed livestock with available covariates is diversified across the country - we allow for various regression models for regions. In this case study, we treat 16 voivodeships (*województwa*) as regions.

3. The disaggregation framework

3.1 The basic model

First, the model is specified on a level of *fine* grid. Let Y_i denote a random variable associated with an unknown value of interest y_i defined at each cell i for $i=1, \dots, n$ of a fine grid (n denotes the overall number of cells in a fine grid). The random variables Y_i are assumed to follow the Gaussian distribution with the mean μ_i and variance σ_Y^2

$$Y_i | \mu_i \sim \text{Gau}(\mu_i, \sigma_Y^2)$$

Given the values μ_i and σ_Y^2 , the random variables Y_i are assumed independent. The mean $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^n$ represents the true process underlying emissions, and the (unknown)

observations are related to this process through a measurement error with the variance σ_Y^2 . The approach to modeling μ_i expresses an assumption that available covariates explain part of the spatial pattern, and the remaining part is captured through a spatial dependence. The CAR scheme follows an assumption of similar random effects in adjacent cells, and it is given through the specification of full conditional distribution functions of μ_i for $i = 1, \dots, n$

$$\mu_i | \boldsymbol{\mu}_{-i} \sim \text{Gau} \left(\mathbf{x}_i^T \boldsymbol{\beta} + \rho \sum_{\substack{j=1 \\ j \neq i}}^n \frac{w_{ij}}{w_{i+}} (\mu_j - \mathbf{x}_j^T \boldsymbol{\beta}), \frac{\tau^2}{w_{i+}} \right)$$

where $\boldsymbol{\mu}_{-i}$ denotes all elements in $\boldsymbol{\mu}$ but μ_i , w_{ij} are the adjacency weights ($w_{ij} = 1$ if j is a neighbour of i and 0 otherwise, also $w_{ii} = 0$); $w_{i+} = \sum_j w_{ij}$ is the number of neighbours of an area i ; $\mathbf{x}_i^T \boldsymbol{\beta}$ is a regression component with proxy information available for area i and a respective vector of regression coefficients; τ^2 is a variance parameter. Thus, the mean of the conditional distribution $\mu_i | \boldsymbol{\mu}_{-i}$ consists of the regression part and the second summand, which is proportional to the average values of remainders $\mu_j - \mathbf{x}_j^T \boldsymbol{\beta}$ for neighbouring sites (i.e. when $w_{ij} = 1$). The proportion is calibrated with the parameter ρ , reflecting strength of a spatial association. Furthermore, the variance of the conditional distribution $\mu_i | \boldsymbol{\mu}_{-i}$ is inversely proportional to a number of neighbours w_{i+} .

The joint distribution of the process $\boldsymbol{\mu}$ is the following (for the derivation see [2])

$$\boldsymbol{\mu} \sim \text{Gau}_n(\mathbf{X}\boldsymbol{\beta}, \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1}) \quad (1)$$

where \mathbf{D} is an $n \times n$ diagonal matrix with w_{i+} on the diagonal; and \mathbf{W} is an $n \times n$ matrix with adjacency weights w_{ij} . Equivalently, we can write (1) as

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \text{Gau}_n(\mathbf{0}, \boldsymbol{\Omega}) \quad (2)$$

with $\boldsymbol{\Omega} = \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1}$.

The model for a *coarse* grid of (aggregated) observed data is obtained by multiplication of (2) with the $N \times n$ *aggregation matrix* \mathbf{C} , where N is a number of observations in a coarse grid

$$\mathbf{C}\boldsymbol{\mu} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\varepsilon}, \quad \mathbf{C}\boldsymbol{\varepsilon} \sim \text{Gau}_N(\mathbf{0}, \mathbf{C}\boldsymbol{\Omega}\mathbf{C}^T) \quad (3)$$

The aggregation matrix \mathbf{C} consists of 0's and 1's, indicating which cells have to be aligned together. The random variable $\boldsymbol{\lambda} = \mathbf{C}\boldsymbol{\mu}$ is treated as the mean process for variables $\mathbf{Z} = \{Z_i\}_{i=1}^N$ associated with observations $\mathbf{z} = \{z_i\}_{i=1}^N$ of the aggregated model (in a coarse grid)

$$\mathbf{Z} | \boldsymbol{\lambda} \sim \text{Gau}_N(\boldsymbol{\lambda}, \sigma_Z^2 \mathbf{I}_N)$$

Also at this level, the underlying process $\boldsymbol{\lambda}$ is related to \mathbf{Z} through a measurement error with variance σ_Z^2 .

Model parameters $\boldsymbol{\beta}$, σ_Z^2 , τ^2 and ρ are estimated with the maximum likelihood method based on the joint unconditional distribution of observed random variables \mathbf{Z}

$$\mathbf{Z} \sim \text{Gau}_N(\mathbf{C}\mathbf{X}\boldsymbol{\beta}, \sigma_Z^2 \mathbf{I}_N + \mathbf{C}\boldsymbol{\Omega}\mathbf{C}^T) \quad (4)$$

The log likelihood function associated with (4) is formulated, and the analytical derivation is limited to the regression coefficients $\boldsymbol{\beta}$; further maximization of the profile log likelihood is performed numerically.

As to the prediction of missing values in a fine grid, the underlying mean process $\boldsymbol{\mu}$ is of our primary interest. The predictors optimal in terms of the mean squared error are given by the conditional expected value $E(\boldsymbol{\mu} | \mathbf{z})$. The joint distribution of $(\boldsymbol{\mu}, \mathbf{Z})$ is

$$\begin{bmatrix} \boldsymbol{\mu} \\ \mathbf{Z} \end{bmatrix} \sim \text{Gau}_{n+N} \left(\begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{C}\mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega} & \boldsymbol{\Omega}\mathbf{C}^T \\ \mathbf{C}\boldsymbol{\Omega} & \sigma_Z^2 \mathbf{I}_N + \mathbf{C}\boldsymbol{\Omega}\mathbf{C}^T \end{bmatrix} \right) \quad (5)$$

The distribution (5) yields both the predictor $E(\widehat{\boldsymbol{\mu}} | \mathbf{z})$ and its error $\text{Var}(\widehat{\boldsymbol{\mu}} | \mathbf{z})$

maximum values of d_i as well as the sample correlation coefficient r between the predicted and observed values. From here, it is obvious that the spatial CAR structure considerably improves the results obtained with the model of independent errors LM. For comparison, we also include the results obtained with the allocation proportional to population in municipalities; this setting is called NAIVE. It is a straightforward and commonly used approach in this area of application. Here we note that the NAIVE approach provides reasonable results, but the CAR model outperforms it in terms of all the reported criteria. The decrease of the mean squared error is from 3374.4 for NAIVE to 3069.4 for CAR, which gives 9% improvement. From the maps of predicted values for the models CAR and NAIVE (Figure 1), it is difficult to spot a meaningful difference.

Table 2. Analysis of residuals ($d_i = y_i - y_i^*$).

	<i>mse</i>	$\min(d_i)$	$\max(d_i)$	<i>r</i>
CAR	3069.4	-275	469	0.784
LM	5641.2	-357	522	0.555
NAIVE	3374.4	-475	403	0.766
CAR*	3437.0	-258	512	0.763
LM*	4876.1	-374	546	0.651
CAR**	3124.9	-256	446	0.783
LM**	4427.6	-352	472	0.674

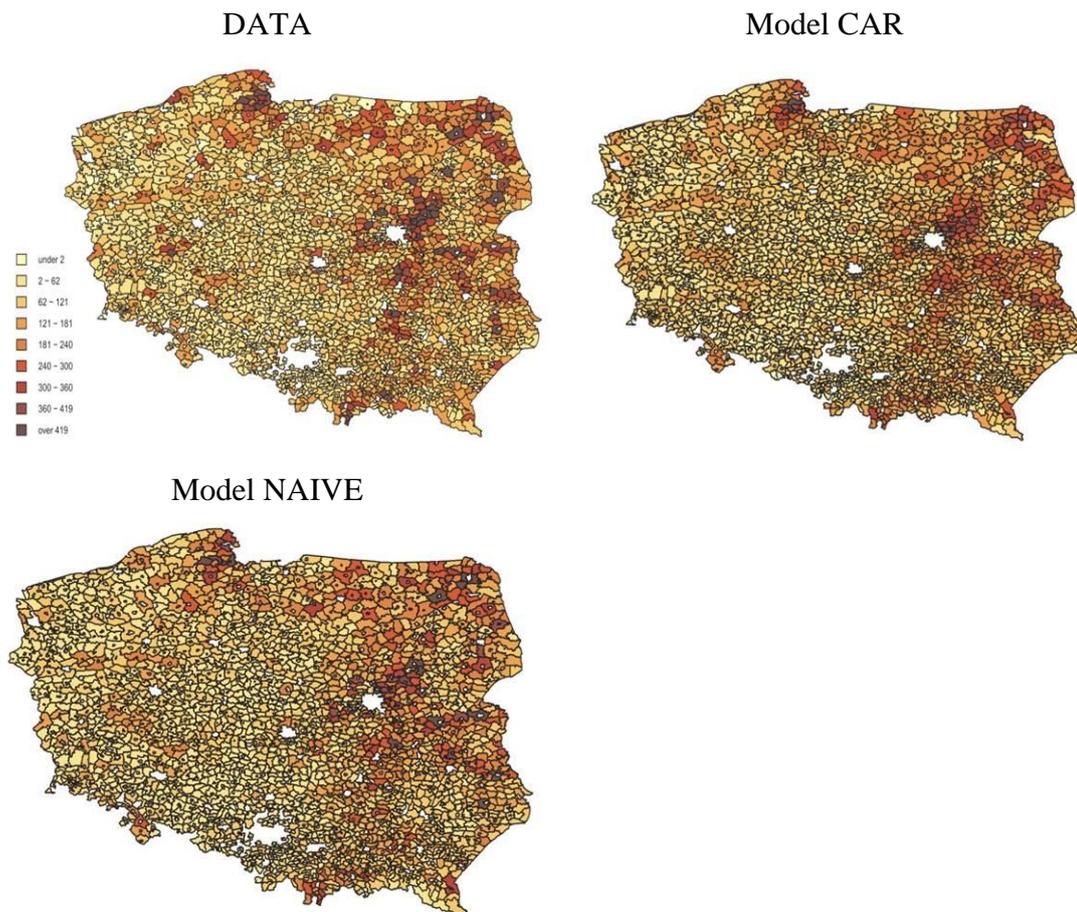


Figure 1. Original data in municipalities as well as predicted values for the models NAIVE and CAR.

Next, we considered the models with various regression coefficients in regions (voivodeships) but having the same set of covariates ($\beta_0, \beta_1, \beta_3, \beta_4$); the models are denoted CAR* and LM*, respectively, for the spatial and non-spatial approaches. Note that the model CAR* gives much worse results than the models CAR and NAIVE.

Further, considered were the models (CAR** and LM**) where, both, the coefficients as well as sets of covariates vary across the regions. Only the statistically significant covariates were chosen. Due to a lack of space, we do not provide here the table with the regression coefficients and their standard errors for all the considered regions. We only report that the values of estimated parameters for CAR** and LM** showed considerable differences across the voivodeships, not only in terms of the estimated values, but also in terms of their significance. From Table 2 we note that this setting (CAR**) provides the results comparable to that of CAR.

5. Concluding remarks

The study presents the first attempt to apply the spatial scaling model for the GHG inventory in Poland. The task was to allocate spatially correlated data to finer spatial scales, conditional on covariate information observable in a fine grid. The results of the disaggregation with the proposed procedure were compared with the allocation proportional to population; an improvement of 9% in terms of the mean squared error was reported. The model was extended to allow for various regression covariates in regions (here voivodeships). Numerous features of the method require further investigation.

The proposed method provided good results for livestock activity data of agricultural sector. Apart from the study reported above, the approach was also applied to a residential sector for disaggregation of natural gas consumption in households. In that case, with disaggregation featured from voivodeships to municipalities, the results turned out to be quite modest. This was partly due to a limited spatial correlation of the analysed process, and too large extent of disaggregation. The method is feasible for disaggregation from districts to municipalities, but not from voivodeships to municipalities.

Acknowledgements

Joanna Horabik-Pyzel was supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing; project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

References

- [1] Boychuk, K. and R. Bun (2014) Regional spatial inventories (cadastres) of GHG emissions in the Energy sector: Accounting for uncertainty, *Climatic Change* 124:561-574.
- [2] Banerjee, S., B.P. Carlin and A.E. Gelfand (2004) Hierarchical modeling and analysis for spatial data. Chapman & Hall/CRC.
- [3] Kaiser, M.S., M.J. Daniels, K. Furakawa and P. Dixon (2002) Analysis of particulate matter air pollution using Markov random field models of spatial dependence, *Environmetrics* 13: 615-628.

- [4] Horabik, J. and Z. Nahorski (2014) Improving resolution of a spatial air pollution inventory with a statistical inference approach, *Climatic Change* 124:575-589.
- [5] Główny Urząd Statystyczny (2012) Bank Danych Lokalnych.
http://www.stat.gov.pl/bdlen/app/strona.html?p_name=indeks
- [6] European Environment Agency (2000) Corine Land Cover 2000.
<http://www.eea.europa.eu/data-and-maps/data>
- [7] Horabik, J. and Z. Nahorski (2014) The Cramér-Rao lower bound for the estimated parameters in a spatial disaggregation model for areal data. In: D. Filev et al.: *Intelligent Systems 2014*, Springer International Publishing, pp. 661-668.

Modeling uncertainty in ammonia emissions from agriculture: Regional upscaling by Monte Carlo analysis

Bettina Schächli¹, Jürg Heldstab¹, Thomas Kupper²

¹ INFRAS, Research and Consulting, 8045 Zürich, Switzerland

² Bern University of Applied Sciences, School of Agricultural, Forest and Food Sciences, 3052
Zollikofen, Switzerland
bettina.schaeppli@infrass.ch

Abstract

Assessment of NH₃ emissions and related uncertainties is required for both the inventory of air pollutants as well as the inventory of greenhouse gases, since N deposition leads to formation of indirect N₂O emissions. In Switzerland, the nitrogen mass-flow model Agrammon provides data on farm-specific NH₃ emissions and derives the national total by upscaling based on total livestock numbers. So far, related uncertainties relied solely on expert judgement.

We show an approach for assessing model uncertainty by a combination of Monte Carlo simulations and Gaussian error propagation. This approach allows accounting for large, asymmetric uncertainties and correlations across regional scales and therefore permits a robust assessment of aggregated uncertainties. A particular focus lies on aggregation of uncertainties in process-specific model parameters to the categories that are reported to UNECE.

The new approach permits a more detailed analysis of model uncertainties and thus a more accurate reporting of NH₃ emissions and indirect N₂O emissions.

Keywords: Monte Carlo, Ammonia emissions, Nitrogen mass-flow model, Inventory uncertainty

1. Introduction

Atmospheric nitrogen deposition currently exceeds critical loads in a large part of natural ecosystems in Switzerland [1]. Additionally, ammonia (NH₃) emissions increase the formation of secondary aerosols. Atmospheric nitrogen depositions induce substantial indirect gaseous nitrogen losses due to microbial processes in the soil, thereby leading to an increase of indirect nitrous oxide (N₂O) emissions ([2], [3]). Emissions of NH₃ and N₂O have to be reported annually by Switzerland to Convention on Long-range Transboundary Air Pollution (CLRTAP/UNECE) and United Nations Framework Convention on Climate Change (UNFCCC) in the respective inventories.

Thus, climate change mitigation and air pollution control require measures to reduce NH₃ emissions. Knowledge of related uncertainties is an important prerequisite for designing effective abatement measures.

In Switzerland, NH₃ emissions from agriculture amounted to 57.3 kt in 2013. With a share of 93.1%, agriculture is by far the largest source of Switzerland's total NH₃-emissions. Within this source, the category 3B Manure management contributes with 46% to the agricultural emissions in the year 2013 and the remaining 54% occur in category 3D Crop production and agricultural soils [4]. Thus, accurate assessment of agricultural NH₃ emissions and related uncertainties is of particular importance.

So far, considerable effort was invested in modelling agricultural ammonia emissions in Switzerland by means of the nitrogen mass flow model Agrammon, which simulates Switzerland's national NH₃ emission that are reported to UNECE and that

provide a basis for calculating N₂O emissions reported to UNFCCC. Since NH₃ volatilization is highly dependent on manure management techniques as well as environmental parameters (e.g. [5], [6], [7], [8], [9]), it is crucial to take into account individual farm characteristics as much as possible. Therefore, the Agrammon model applies a detailed bottom-up approach that accounts for technical aspects in the manure handling, housing and yard characteristics as well as composition of animal feed ([10], [11], [12]). Due to the large number of model parameters, assessment of underlying model uncertainties and their aggregation to the national level is not straightforward. So far, the uncertainties reported to UNECE relied solely on expert judgement.

Commissioned by the Swiss Federal Office for the Environment, we developed a model that assesses uncertainties of Switzerland's nitrogen mass-flow model Agrammon and aggregates uncertainties in the emissions at the farm level to the national scale as required in the annual reporting to UNECE. It addresses the issue of correlated model parameters and large uncertainties by Monte Carlo simulations. Subsequently, it performs a stepwise aggregation of process specific uncertainties at the farm level to the national scale by a combination of Monte Carlo simulations and Gaussian error propagation.

2. Data and Methods

Agricultural NH₃ emissions from livestock production in Switzerland are estimated from the nitrogen mass flow model Agrammon (www.agrammon.ch, [10], [12]). In this study, we developed a model that assesses related uncertainties as part of the post-processing of the Agrammon model output. Even though only the model output is used, a brief overview of the Agrammon model is provided in section 2.2 for illustrative purposes. Subsection 2.3 describes the methodological approach implemented in the uncertainty simulation model and subsection 2.4 shows how the process specific emission factors are aggregated to the CLRTAP categories.

2.1 Data

The Agrammon model simulations are based on data from a regularly conducted survey, which covers around 3000 farms, representing around 5% of all farms in Switzerland. The survey provides detailed data on farm-specific technical parameters that are influencing emission factors, such as timing and method of manure application, type of manure storage as well as composition of animal feed. The survey is stratified according to three geographical regions (East, Central, West/South), three altitude zones (valley, hills, mountains) and five farm types. Detailed information on the survey conducted in 2010 is provided in [12]. The present study applies data from this survey. The Swiss Federal Statistical Office conducts an annual census on livestock numbers, which provides the necessary activity data for the present study. Related uncertainties are estimated to be in the order of 6% [13].

A previous study performed sensitivity analyses with respect to the technical parameters for selected farm classes and livestock categories [14]. In the present study, we used the resulting sensitivities of the simulated emissions as input to the uncertainty simulation model.

2.2 NH₃ emission modeling

The NH₃ emission model applied in Switzerland simulates emissions from livestock farming by partitioning total excretion into different processes that are relevant for

simulating emissions. The model distinguishes different 24 livestock categories and 32 farm classes, which were derived from 3 geographic regions and 3 altitude zones and 5 farm types. In addition, it accounts for emissions related to use of fertilizers in crop production. It calculates farm-specific NH_3 emission from nitrogen fluxes along the manure management chain (housing, storage, grazing, manure application) based on data gathered from stratified surveys on farm and manure management (see subsection 2.1). For each stage in the manure management chain, specific emission factors are defined as a share of total soluble nitrogen (total ammoniacal N – TAN) present at a given stage. The model also allows for adjustment of these standard emission factors by a set of correction factors that take into account farm-specific manure management practices. These parameters account for differences in composition of animal feed (e.g. protein contents), manure storage systems (size, type, mixing frequency and coverage), manure application (timing, application rate and technique) as well as technical aspects of housing and yard that are influencing NH_3 emissions.

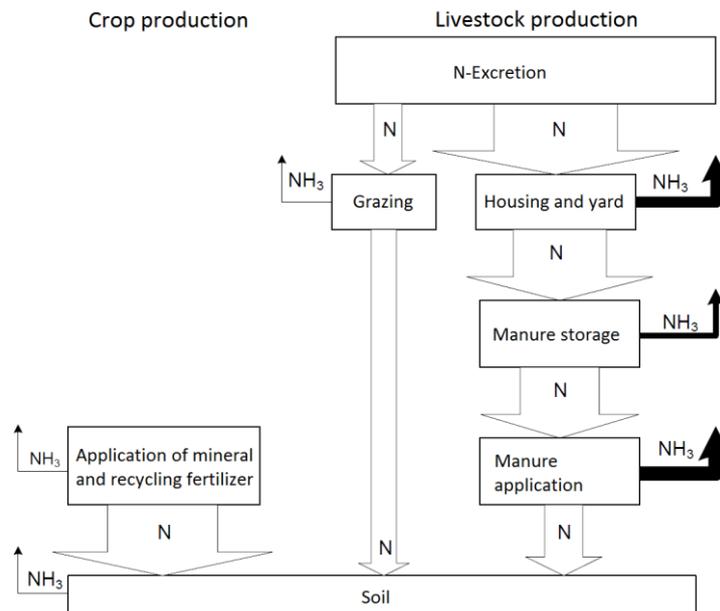


Figure 1. Model illustration: simulations account for NH_3 emissions from different stages such as housing and yard, storage of solid and liquid manure, application of manure and grazing.

Based on the survey data, the Agrammon model simulates farm-specific emissions for each livestock category (l) and each emission stage (s) in the manure management chain. Within a given farm class, a mean emission factor ($EF_{s,l}$) is derived by a linear regression of farm-specific emissions ($Em_{f,s,l}$) and corresponding activity data ($AD_{f,l}$), which consist of the livestock numbers of the surveyed farms.

$$Em_{f,s,l} = EF_{s,l} AD_{f,l}, \quad (1)$$

Regression analysis is performed separately for each manure management stage (s) and each livestock category (l), resulting in stage specific emission factors for each farm class ($EF_{s,l}$). Total emissions from a specific livestock category and manure management stage in a given farm class result from multiplying these mean emission factors with the total livestock numbers (AD_{tot}) of this class. The regional total of NH_3

emissions (Em_{tot}) consists of the sum of the simulated emissions over all livestock categories and manure management stages.

$$Em_{tot} = \sum_s \sum_l EF_{s,l} AD_{tot} , \quad (2)$$

While deriving the national total emissions is straightforward, assessment of the propagation of related model uncertainties is more challenging. The following section presents the simulation model that estimates related model uncertainties for each livestock category by means of Monte Carlo simulations.

2.3 Uncertainty assessment by Monte Carlo simulation

As described in the previous subsection, the nitrogen mass flow model Agrammon provides farm-specific (f) emission data ($Em_{f,s,l}$) for each livestock category (l) and manure management stage (s). Uncertainties in the mean emission factors estimated by linear regression of farm-specific emissions and corresponding livestock numbers are the result of uncertainties in the simulated emissions at the level of individual farms. Previous research shows that uncertainties in the emissions at the farm level are dependent on farm types and thus the data set exhibits a non-constant variance [14].

Linear regression models require that the errors in the data set fulfill certain assumptions, which are the statistical independence of the errors, constant variance in the errors and normality of the error distribution. Farm-specific emissions simulated by Agrammon violate in particular the second assumption since variance in the error terms is larger for farms with high emissions. In addition, there are correlations in the error terms since some of the underlying technical parameters are identical for all farm classes. Thus, confidence intervals estimated by conventional linear regression analysis are biased.

Therefore, we implemented an approach based on Monte Carlo simulations that provides a robust estimate of the standard errors and estimates confidence intervals. The uncertainty simulation model is implemented in the statistics program R. The model allows accounting for correlated error terms and non-constant variance.

When the assumption of constant variance is violated, conventional estimation of standard errors can be biased. Therefore, we adopted an approach for estimating robust standard errors in the estimated coefficients that takes into account the heteroscedasticity in the data. Instead of using the root mean square error, the standard error is estimated based on the squared residual (e_i) of each observation [15].

$$SE = \sqrt{\sum_i w_i^2 e_i^2} , \quad (3)$$

Where w_i indicates the observation weights. Besides accounting for heteroscedasticity in the data set, the uncertainty simulation model addresses the issue of correlated error terms. In each simulation run, emission data (Em_{sim}) are generated by adding a farm-specific error term (ϵ), which is uncorrelated, and a correlated error term (ϵ_{corr}), which is identical for all data points in a given model run.

$$Em_{sim} = Em_{s,l} (1 + \epsilon + \epsilon_{corr}) , \quad (4)$$

From the distribution of the simulated emission data, we estimate the 95% confidence interval, from which we derive relative uncertainties for each emission factor provided by the Agrammon model.

In order to estimate the total uncertainty in NH₃ emissions in Switzerland for the national inventory, uncertainties are required at the national scale as a total for each livestock category. Thus in a next step, uncertainties in the specific emission factors have to be aggregated to the national scale and to the categories required for the reporting under the CLRTAP.

2.4 Aggregation of uncertainties

In analogy to the aggregation of emissions from specific farm classes and emission stages to the national level and to the categories required for the informative inventory report (see equation (2)), related uncertainties need to be aggregated as well. Uncertainties in emission factors of different manure management stages and across regions are correlated and therefore Gaussian error propagation is not applicable. Instead, the model aggregates uncertainties in a stepwise procedure by means of an additional Monte Carlo simulation that allows accounting for correlation in the uncertainties at the regional scale. This step results in uncertainties of stage and livestock specific emission factors at the national scale.

Since the Agrammon model distinguishes more livestock categories and manure management stages than required for the reporting under CLRTAP, further aggregation of manure management stages and livestock categories is required. Thus, in a next step the model aggregates uncertainties over all emission stages (s) and livestock categories (l) of a given CLRTAP Category (Cat). For example, category 3 B 1 b Cattle non-dairy subsumes all manure management stages (except manure application and grazing) and several livestock categories such as calves, heifers and beef cattle. At this level, uncertainties are assumed to be independent and they are aggregated by means of Gaussian error propagation.

$$U_{EF,Cat} = \sqrt{\sum_{l \in Cat} \sum_{s \in Cat} U_{EF,l,s}^2} \quad (5)$$

In the final step, the uncertainties in the resulting emissions ($U_{Em,Cat}$) are estimated for each CLRTAP category again by means of Gaussian error propagation from the uncertainties in the corresponding emission factors ($U_{l,Cat}$) and in the livestock numbers ($U_{l,Cat}$), which is estimated to be in the order of 6% [13].

$$U_{Em,Cat} = \sqrt{U_{l,Cat}^2 + U_{l,Cat}^2} \quad (6)$$

3. Results and Discussion

The simulations provide model uncertainties according to livestock categories as defined in the CLRTAP (see Figure 2). Relative uncertainties range between 20% and 80%. The results show largest uncertainties for poultry, goats as well as mules and asses. Emissions of cattle, swine and horses have considerably lower uncertainties.

Generally, high uncertainties are observed for those categories that are modelled by Agrammon with a low degree of regional differentiation. For example in the category turkeys, the Agrammon model does not apply any regional stratification at all, since the

number of observations would not allow for a regionalization. This explains the rather large uncertainties observed for this livestock category.

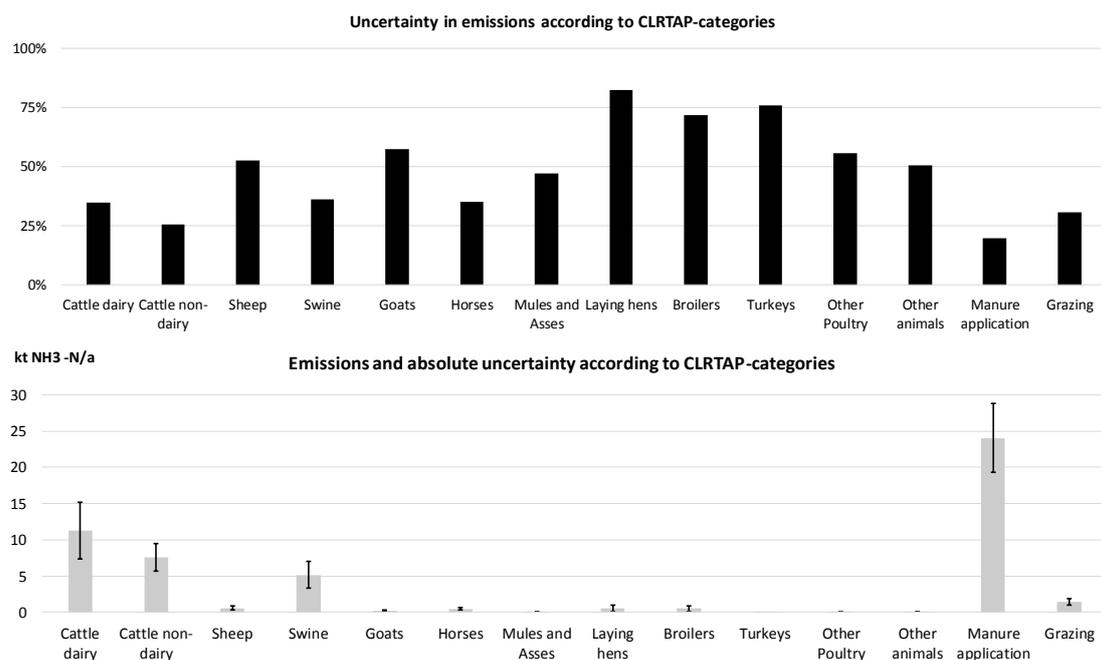


Figure 2. Upper part: Relative uncertainty in NH₃ emissions according to CLRTAP categories. Lower part: Annual emissions in kilotonnes (kt NH₃-N/a) and absolute uncertainty (as error bars).

Table 1. NH₃ emissions (Em) in kilotonnes reported in 2011 and related uncertainties in activity data (U_{AD}), emission factors (U_{EF}) and Emissions (U_{Em}) according to UNECE/CLRTAP categories.

Category UNECE/CLRTAP	Em in kt	U _{AD}	U _{EF}	U _{Em}
3B1a Manure management - Dairy cattle	11.3	6%	34%	35%
3B1b Manure management - Non-dairy cattle	7.6	6%	25%	26%
3B2 Manure management - Sheep	0.6	6%	52%	53%
3B3 Manure management - Swine	5.2	6%	36%	36%
3B4d Manure management - Goats	0.2	6%	57%	57%
3B4e Manure management - Horses	0.5	6%	34%	35%
3B4f Manure management - Mules and asses	0.1	6%	47%	47%
3B4gi Manure management - Laying hens	0.6	6%	82%	82%
3B4gii Manure management - Broilers	0.5	6%	72%	72%
3B4giii Manure management - Turkeys	0.0	6%	76%	76%
3B4giv Manure management - Other poultry	0.1	6%	55%	56%
3B4h Manure management - Other animals	0.0	6%	50%	50%
3Da1 Inorganic N-fertilizers (includes also urea application)	2.0	25%	50%	56%
3Da2a Animal manure applied to soils	24.1	6%	19%	20%
3Da2c Other organic fertilisers applied to soils	0.4	6%	50%	50%
3Da3 Urine and dung deposited by grazing animals	1.4	6%	30%	31%
3Db Indirect emissions from managed soils	2.8	6%	50%	50%
Total Uncertainty in agricultural NH₃-Emissions	57.3			12%

Due to the large share of dairy and non-dairy cattle and swine in the total livestock production of Switzerland, the contribution to the overall uncertainty is dominated by these categories [4]. The total uncertainty in Switzerland's ammonia emissions from livestock production amounts to about 13%. Uncertainties from dairy cattle account for 43% of the total variance, other cattle 21%, swine 20%, laying hens and broilers 9% and all other animal categories account for the remaining 7%.

In addition, the emissions from use of synthetic fertilizers and farm-level agricultural operations are estimated based on the statistics provided by the Swiss farmer's association and related uncertainties are based on expert judgement [16] (see Table 1). This results in a total uncertainty in agricultural NH₃-Emissions of around 12%.

4. Conclusions and Outlook

The new uncertainty simulation model permits a robust and standardized assessment analysis of model uncertainties, as it is able to account for large, asymmetric uncertainties and correlations among the technical model parameters and across regional scales. Thus, it allows a more accurate monitoring and reporting of NH₃ emissions, which indirectly improves also the assessment of related indirect N₂O emissions. By identifying the most uncertain sources and their contribution to the total uncertainty in NH₃ emissions, the new uncertainty simulation model can serve as a basis for further improvements in Switzerland's air pollutant and greenhouse gas inventories.

Previously reported uncertainties of Switzerland's NH₃ emissions from livestock production based on expert judgement were estimated to be in the order of about 50% in each category [16]. The results of the present study indicate that uncertainties are considerably lower for cattle and swine. Since these livestock categories contribute substantially to the total uncertainty of NH₃-emissions from livestock production, the results indicate that total uncertainty has been overestimated in previous inventories. Future research aiming at reducing existing uncertainties should therefore primarily address those livestock categories.

The updated uncertainties of the ammonia emissions were integrated into the uncertainty level and trend analyses of Switzerland's air pollutant inventory in 2013 [17] in line with the reporting obligations to UNECE under the CLRTAP [18]. For submission in 2015 [4], we modified the uncertainty simulation model such that it aggregates uncertainties to the new categories in line with the new EMEP/EEA Guidelines of 2013 [19]. Future work will focus on a refined assessment of uncertainties at the level of process specific parameters and on assessing correlations among technical parameters.

5. Acknowledgements

We thank the Air Quality Management Section of the Federal Office for the Environment (Switzerland) for financial support and Bonjour Engineering SA for technical support and for providing data.

References

- [1] Dentener, F., Drevet, J., Lamarque, J.F., Bey, I., Eickhout, B., Fiore, A.M., Hauglustaine, D., Horowitz, L.W., Krol, M., Kulshrestha, U.C., Lawrence, M., Galy-

- Lacaux, C., Rast, S., Shindell, D., Stevenson, D., Van Noije, T., Atherton, C., Bell, N., Bergman, D., Butler, T., Cofala, J., Collins, B., Doherty, R., Ellingsen, K., Galloway, J., Gauss, M., Montanaro, V., Müller, J.F., Pitari, G., Rodriguez, J., Sanderson, M., Solomon, F., Strahan, S., Schultz, M., Sudo, K., Szopa, S., Wild, O., 2006. Nitrogen and sulfur deposition on regional and global scales: a multimodel evaluation. *Global Biogeochem. Cycles* 20. 2006
- [2] Leip, A., The European Nitrogen Assessment - Integrating nitrogen fluxes at the European scale, chapter 16, ed. Mark A. Sutton, Clare M. Howard, Jan Willem Erisman, Gilles Billen, Albert Bleeker, Perine Grennfelt, Hans van Grinsven and Bruna Grizzetti. Published by Cambridge University Press. 2011
- [3] Bühlmann T., Hiltbrunner E., Körner C., Rihm B., Achermann B., Induction of indirect N₂O and NO emissions by atmospheric nitrogen deposition in (semi-)natural ecosystems in Switzerland, *Atmospheric Environment* 103 (2015) 94-101. 2015
- [4] Switzerland's Informative Inventory Report 2015 (IIR). Submission under the UNECE Convention on Long-range Transboundary Air Pollution. Submission of March 2015 to the United Nations ECE Secretariat. Federal Office for the Environment, Bern. 2015 [http://www.ceip.at/ms/ceip_home1/ceip_home/status_reporting/2015_submissions/\[14.09.2015\]](http://www.ceip.at/ms/ceip_home1/ceip_home/status_reporting/2015_submissions/[14.09.2015])
- [5] Huijsmans J.F.M., Hol J.M.G., Hendricks M.M.W.B, Effect of application technique, manure characteristics, weather and field conditions on ammonia volatilization from manure applied to grassland, *Netherlands Journal of Agricultural Science* 49 (2001) 323-342. 2001
- [6] Huijsmans J.F.M., Hol J.M.G., Vermeulen G.D., Effect of application method, manure characteristics, weather and field conditions on ammonia volatilization from manure applied to arable land, *Atmospheric Environment* 37 (2003) 3669–3680. 2003
- [7] Misselbrook T.H., Van Der Weerden T.J., Pain B.F., Jarvis S.C., Chambers B.J., Smith K.A., Phillips V.R., Demmers T.G.M., Ammonia emission factors for UK agriculture, *Atmospheric Environment* 34 (2000) 871-880. 2000
- [8] Misselbrook, T.H., Nicholson, F.A., Chambers B.J., Johnson R.A., Measuring ammonia emissions from land applied manure: an intercomparison of commonly used samplers and techniques, *Environmental Pollution* 135 (2005) 389–397. 2005
- [9] Sommer S.G., Jensen L. S., Clausen S. B., Sogaard H. T., Ammonia volatilization from surface-applied livestock slurry as affected by slurry composition and slurry infiltration depth, *Journal of Agricultural Science*, Page 1 of 7. Cambridge University Press, 2006
- [10] Kupper, T., Bonjour, C., Zaucker, F., Achermann, B., Menzi, H. 2010. Agrammon: An internet based model for the estimation of ammonia emissions. In: Cordovil, C., Ferreira, L., (eds.). 14th RAMIRAN International Conference; Lisboa Portugal. p 334-337, 2010
- [11] Berner Fachhochschule, Hochschule für Agrar-, Forst- und Lebensmittelwissenschaften, Documentation of technical parameters, Agrammon-Modell, Version 20.03.2013 <http://www.agrammon.ch/assets/Downloads/Dokumentation-Technische-Parameter-20130320.pdf>

- [12] Kupper, T., Bonjour, C., Menzi, H. 2015. Evolution of farm and manure management and their influence on ammonia emissions from agriculture in Switzerland between 1990 and 2010. *Atmos. Environ.* 103(0): 215-221.
- [13] Bretscher, D. and Leifeld, J. Uncertainty of agricultural CH₄ and N₂O emissions in Switzerland. Internal documentation by, Agroscope Reckenholz-Tänikon Research Station, Zürich. 2008
<http://www.bafu.admin.ch/klima/13879/13880/14577/15536/index.html?lang=en>
(see ART 2008a) [14.09.2015]
- [14] Bonjour C., Sensitivity analysis for simulated ammonia emissions in 2010 for selected farm classes. *Bonjour Engineering*. 2013
- [15] Barreto H., Howland F., *Introductory Econometrics, Using Monte Carlo Simulation with Microsoft Excel*, Chapter 19 - Heteroskedasticity pp. 508-557, Online ISBN: 9780511809231, DOI: <http://dx.doi.org/10.1017/CBO9780511809231>, Cambridge University Press.
- [16] Expert judge on uncertainty estimates of emission factors of the AGRAMMON model, oral communication to S. Liechti and B. Achermann (FOEN), 13 March 2012.
- [17] Switzerland's Informative Inventory Report 2013 (IIR). Submission under the UNECE Convention on Long-range Transboundary Air Pollution. Submission of March 2013 to the United Nations ECE Secretariat. Federal Office for the Environment, Bern. 2013
http://www.ceip.at/ms/ceip_home1/ceip_home/status_reporting/2013_submissions/
[14.09.2015]
- [18] EMEP/EEA air pollutant emission inventory guidebook. 2009
<http://www.eea.europa.eu/publications/emep-eea-emission-inventory-guidebook-2009> [14.09.2015]
- [19] EMEP/EEA air pollutant emission inventory guidebook, 2013
<http://www.eea.europa.eu/publications/emep-eea-guidebook-2013>

High resolution spatial inventory of GHG emissions from stationary and mobile sources in Poland: summarized results and uncertainty analysis

Rostyslav Bun^{1,2}, Zbigniew Nahorski³, Joanna Horabik-Pyzel³, Olha Danylo^{1,4},
Nadiia Charkovska¹, Petro Topylko¹, Mariia Halushchak¹, Myroslava Lesiv⁴,
Oleksandr Striamets¹

¹Lviv Polytechnic National University, Lviv, Ukraine, mail: rbun@org.lviv.net;

²Academy of Business in Dąbrowa Górnicza, Poland;

³Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland;

⁴International Institute for Applied Systems Analysis, Laxenburg, Austria

Abstract

Greenhouse gases (GHG) inventories at national or regional levels include the total emissions and emissions for many categories of economic activity. The aim of our research is to analyze the high resolution spatial distributions of emissions for all categories of economic activity in Poland. GHG emission sources are classified into point-, line- and area-type sources. We created maps of such sources for all categories of economic activities covered by IPCC Guidelines, using official information of companies, administrative maps, Corine Land Cover maps, and other available data. The worst resolution is for area-type sources (100 m). We used statistical data at the lowest level as possible (regions, districts, and municipalities). We created the algorithms for these data disaggregation to the level of elementary objects for GHG spatial inventory. These algorithms depend on category of economic activity and cover all categories under investigation. We analyzed emissions of CO₂, CH₄, N₂O, SO₂, NMVOC, and others, and we calculated the total emissions in CO₂-equivalent. We used a grid to calculate the summarizing emissions from the all categories. The grid size depends on the aim of spatial inventory, but it can't be less than 100 m. For uncertainty analysis we used uncertainty of statistical data, uncertainty of calorific values, and uncertainty of emission factors, with symmetric and asymmetric (lognormal) distributions. On this basis and using Monte-Carlo method the 95% confidence intervals of results' uncertainties were estimated for big point-type emission source, the regions, and the subsectors.

Keywords: GHG emissions, high resolution spatial inventory, uncertainty, Monte Carlo method

1. Introduction

Preventing climate changes requires of the humanity to reduce greenhouse gas (GHG) emissions. To control the performance of obligations to reduce or limit GHG emissions, we need to make inventories of emissions and absorptions of these gases. GHG inventories at national or regional levels include the total emissions and emissions for many categories of economic activity. But for deeper study of emission processes as well as their structure, it is more reasonably to make a spatial inventory of GHG. Such an inventory reflects the emissions linked to the territory where they appear. Scientists are constantly trying to reduce the spatial resolution of the inventory results to better reflect the specifics of territorial emission processes [1, 8, 9, 11]. The aim of our research is to analyze the high resolution spatial distributions of emissions for all categories of economic activity in Poland.

2. Spatial inventory approach

For all sectors and categories of anthropogenic activity covered by IPCC Guidelines [7], we analyzed the sources of emissions or sinks in terms of their spatial representation for inventory procedures. GHG emission sources are classified into point-, line- and

area-type sources [3] (see below). Then we built digital maps of sources / sinks for each category. For some categories they are digital maps of point objects, for other categories they are digital maps of linear objects or area-type objects (polygons). We further analyzed the emission / absorption from these diverse elementary objects (points, lines, and areas) without using any regular grid, as it is often made. The line- and area-type (diffused) elementary objects are split by administrative boundaries. It gives us a possibility to keep administrative assignment of each elementary object to the regions (voivodeship in Poland), districts (powiat), and municipalities (gminas).

The next step was to create algorithms for calculating GHG emissions from these elementary objects. Basically, these algorithms reflect the main principles of IPCC Guidelines [7], according to which the emission is a product of activity data and emission factors. However, a common problem is to obtain data about the activities at the level of elementary objects. For this purpose we have developed algorithms for disaggregation of available statistical data for regions (or even for municipalities in some categories) to the level of elementary objects. These algorithms are different for each category. They take into account the available statistics relevant to corresponding administrative level, and use other parameters, that can be considered as indicators for disaggregation of statistical data.

The specific feature of the approach is an ability to use different emission factors for separate elementary objects (or even part of objects), if such data are available, as opposed to using averaged default values. The results of calculating emissions in each category of anthropogenic activity for elementary objects can be visualized in the form of digital maps using different approaches, depending on the source type. Since this approach 'saves' information about administrative assignment of each elementary object (emission source), it is possible to aggregate emissions for whole administrative units (even for small units like municipalities) without loss of accuracy.

3. High resolution maps of emission sources

Examples of the point-type emission sources are electricity or combined electricity and heat production plants, cement plants, production of glass, ammonia, iron and steel, aluminium, pulp and paper, petroleum refining, mining etc. Using the official information on these companies we determined their addresses, and then, using Google Earth (TM) we searched for their production facilities treated as point emission sources (stacks of power plants, for example), and we fixed their geographical coordinates (latitude and longitude). Thus the digital maps of emission sources were built for the categories of human activity, in which the emission sources can be presented as point-type objects. The spatial resolution of these maps is of the order of several meters, that can be considered very accurate for spatial inventory of GHG emissions.

The roads and railways are examples of the line-type emission sources. To construct the maps of these sources, we used the OpenStreetMap. This digital map is created using GPS navigators. Therefore its spatial resolution is also high for GHG spatial inventory. Information on road category was used as one of the indicators for disaggregation of data on fossil fuel combustion by various categories of vehicles in the transport sector.

Area-type (or diffused) sources or sinks are croplands, settlements, industrial areas, forests and others. We created digital maps of such sources / sinks for all categories of human activities under investigation. Elementary objects for GHG spatial analysis are represented as polygons. To build these maps we used Corine Land Cover vector maps [4]. These maps are obtained on the basis of processing raster maps with a resolution

of 100 m. This resolution determines the resolution of all created digital maps of area-type (diffused) sources. It further defines the resolution of the summarized results of GHG spatial inventory.

4. Input data and disaggregation algorithms

As activity data, we used statistical data at the lowest level as possible (levels of voivodeships/regions, powiats/districts, and gminas/municipalities, as some statistical data are available even at the level of gminas) [2]. The amount of fossil fuel used, data on production, the number of animals in agriculture, and any more are examples of such activity data in different emission categories. We created the algorithms for these data disaggregation to the level of elementary objects for GHG spatial inventory (level of point-, line-, and area-type emission sources). These algorithms depend on category of anthropogenic activity and cover all categories under investigation. As input data we also used the values of certain indicators, which are needed for activity data disaggregation to the level of elementary objects. The population density, data on access to energy sources, power of electricity generation plants, gross value production, and many others are examples of such indicators.

In cases where it was possible, the emission coefficients and parameters that reflect the territorial specifics of emission and absorption processes were applied. For example, when calculating accumulated carbon in forests, we used the information of Local Data Bank (BDL) [2] on species composition, age structure, etc. at the level of districts/powiats and communes/gminas.

5. Results of spatial inventory

By using created digital maps of GHG emission sources / sinks in Poland and the algorithms for activity data disaggregation, we formed the geospatial database needed for the spatial inventory. Then we calculated the GHG emissions using appropriate mathematical models. We analyzed emissions of CO₂, CH₄, N₂O, SO₂, NMVOC, and others, and we calculated the total emissions in CO₂-equivalent, using Global Warming Coefficients. These results we obtained at the level of elementary objects (point-, line- and area-type sources of emissions).

We used a grid to calculate the sum of emissions from all categories under investigation. But the cells of this grid are split by administrative map. The grid size depends on the aim of spatial inventory, but it cannot be less than 100 m. This is due to the fact that the worst spatial resolution of the digital maps is for the area-type sources of emissions and it is 100 m. For visualization of results on the maps we calculated also a specific emission, i.e. emissions divided by area, because areas of these objects are not equal.

As an example, the map of total specific GHG emissions in Poland and the Silesian voivodeship, which is the most industrialized region, is presented in Figure 1.

As a rule, the emission from point sources are significant, therefore it causes problems when we visualize the spatial inventory results. Figure 2 presents the results using the square root scale for visualization of spatial inventory results for the Silesian region in Poland. For the purpose of this figure the results of spatial inventory at the level of point-, line-, and area-type elementary sources are aggregated to the regular grid 2 km in size.

Based on the results of GHG spatial inventory at the level of elementary objects or regular grid we can calculate the total emissions in the administrative units like gmina/municipality, powiat/district or voivodeship/region. These results reflect

a structure of emissions by sectors or certain categories of activity, by types of fossil fuel used, and the structure of certain emissions of greenhouse gases. In Figure 3 we can see the structure of total GHG emissions from all sectors in Poland at the regional level, and Figure 4 presents the structure of emissions in the Energy sector, which is the most influential one.

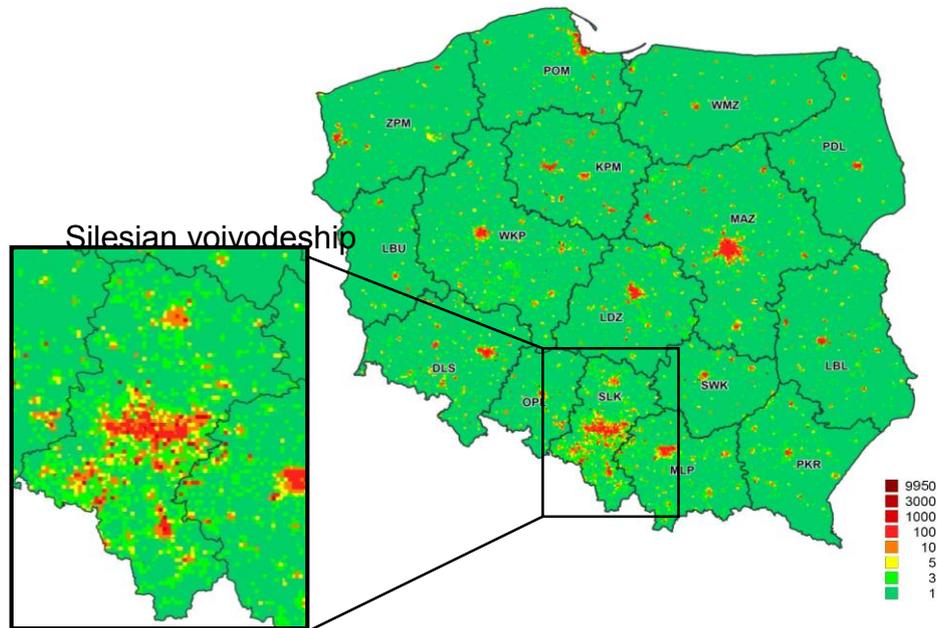


Figure 1. The total specific GHG emissions in Poland and the Silesian region (the all categories without LULUCF, 2010, Gg/cell, CO₂-equivalent, 2 km grid size)

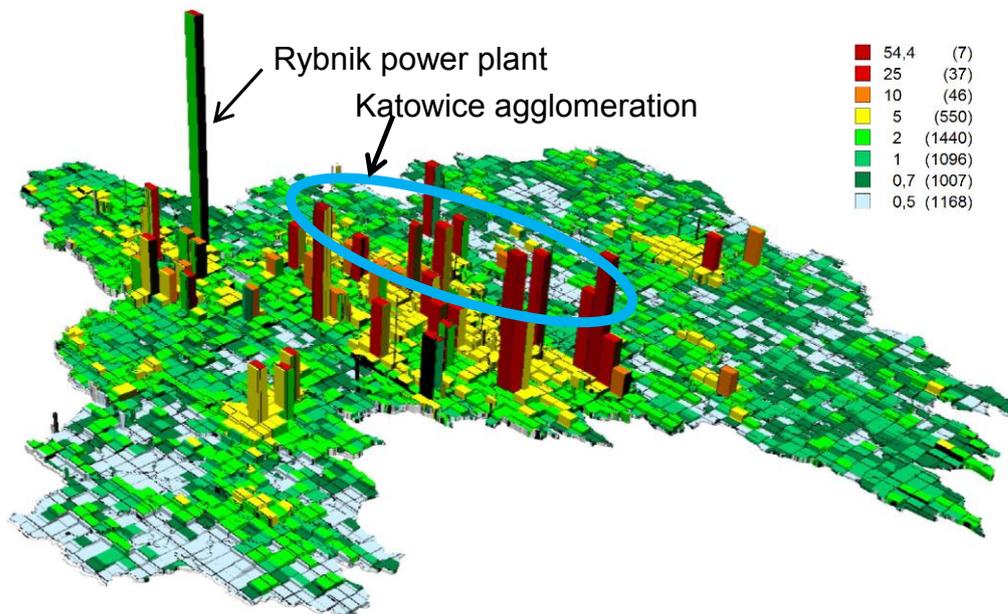


Figure 2. Prism-map of specific GHG emissions from all anthropogenic sectors without LULUCF in the Silesia region at the level of elementary objects (CO₂-equivalent, Gg/km², square root scale, 2 x 2 km, 2010)

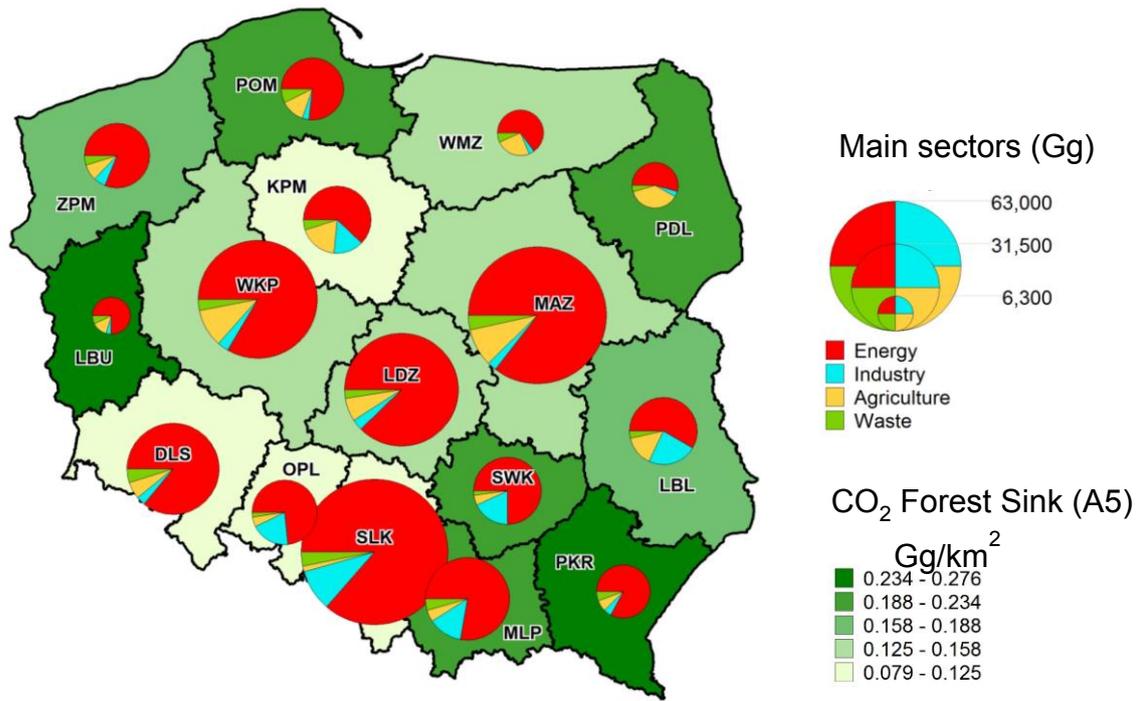


Figure 3. The total GHG emissions structure by sector (Poland, CO₂-equivalent, 2010)

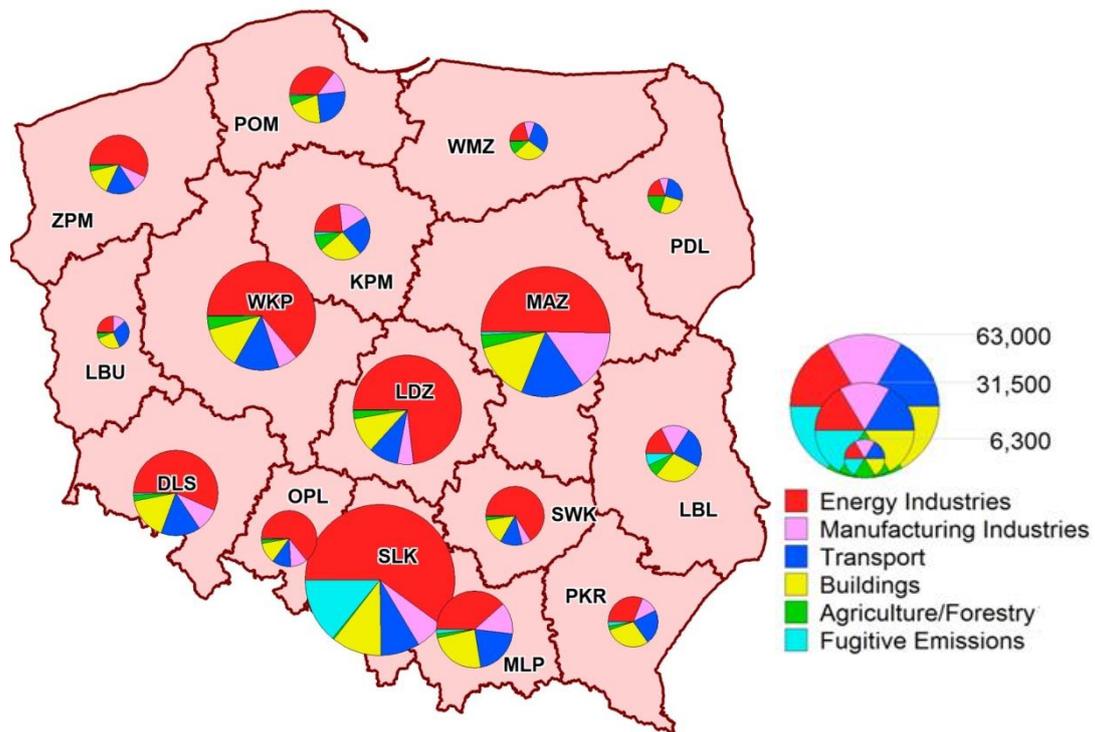


Figure 4. The total GHG emissions in the Energy Sector by sub-sectors (Poland, Gg, CO₂-equivalent, 2010)

6. Uncertainty analysis

The variables and parameters, used in the GHG inventory, are characterized by some uncertainties. These uncertainties are associated with lack of our knowledge on emission processes, inaccurate measuring instruments, etc. Uncertainties of the spatial inventory results, using the above described approach, include three components: (1) the uncertainty of spatial localization of emission sources and sinks; (2) the uncertainty of the used input statistical data, calorific values of fuels, and emission factors; and (3) the uncertainty of spatial disaggregation of activity data to the level of elementary objects. The uncertainty is characterized by the lower and upper limits of 95% confidence intervals of analyzed variables.

The first of these uncertainty components is not taken into account in this study. This is because of the above-described approach to building digital maps of emission sources and sinks, where the uncertainty of localization of point- and line-type elementary objects is kept to the minimum. The uncertainty of localization of area-type (diffused) emission sources / sinks is defined by the uncertainty of the Corine Land Cover maps.

As to the second component, the uncertainty of input statistical data, the uncertainty of calorific values, and the uncertainty of emission factors, we used the appropriate data from [5, 10], and other estimations (e.g., [6] and others). For these variables we used the symmetric and asymmetric (lognormal) distributions.

As described above, the algorithms for disaggregation of activity data are based on certain indicators, that are mostly statistical data. Therefore, for the analysis of the total uncertainty, it was assumed that these uncertainties are described as for statistical data. For some categories of human activities, such as the residential sector, the uncertainty of disaggregated data was evaluated by comparison with similar data from other known sources [2].

Based on these input uncertainties we estimated the distributions using the Monte-Carlo method. The uncertainties of results, as 95% confidence intervals, were calculated, i.e. the expectation, as well as the lower and upper limits. For the point-type sources we estimated the uncertainty of the results separately for each source, and for the total emissions in the category to which they belong. We also analyzed the sensitivity of the total uncertainty to changes of the uncertainties of separate components, as the uncertainties of input statistical data, the uncertainty of calorific values, and the uncertainty of emission factors.

As the number of elementary objects for line- and area-type sources of emissions are large (typically tens of thousands, as in the residential sector or the agriculture sector), we also evaluated the uncertainty of results using the Monte-Carlo method, but at the regional level. We also investigated the sensitivity of the total uncertainty to changes of uncertainties of separate components, including the uncertainties of input activity data, the uncertainty of calorific values, and the uncertainty of emission factors.

7. Conclusions

The presented approach provides the high resolution of GHG spatial inventory in Poland with the use of point-, line-, and area-type emission sources / sinks. The spatial analysis is carried out at the level of these sources without using any additional grid. Consequently, the information on administrative assignment of corresponding emission sources (plants, settlements, etc.) is retained, and this, in turn, makes it possible to aggregate the final results even to the level of sub-municipalities without decreasing accuracy of results.

In principle, according to this approach an inventory is carried out by 'bottom-up' method, but also an element of 'top-down' assessment exists, since we disaggregate available statistical data to the level of elementary objects as point-, line-, or area sources. However, this approach makes it possible to fully use the available even partial information about the specific territorial emission or absorption processes.

The results of spatial inventory of GHG emissions / absorption demonstrate an unevenness of these processes in Poland. Such an unevenness is specific to each category of anthropogenic activity. The positive aspect is that the spatial inventory enables to display a real contribution of each even very small territory to the overall emission processes. What is more, the results presented in such a form show the emissions values, as well as their structural features. It is of interest to authorities to support well-grounded decision making.

Since the spatial analysis takes into account the territorial specificity of many parameters that affect emissions or removals of greenhouse gases (e.g. the differentiated characteristics of the fossil fuel used in the energy sector, the climatic conditions and the energy sources availability in the residential sector, the species and age composition of forests and many others), the total inventory results for the region/country as a whole is more precise comparing with the national inventories without spatial components and regional specifics.

The results were achieved under GESAPU project financed by EU.

References

- [1] Andres R.J., Marland G., Fung I., and Matthews E. (1996) A $1^{\circ} \times 1^{\circ}$ distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cycles*, 10(3), 419-429.
- [2] Bank Danych Lokalnych (Local Data Bank), GUS, Warsaw, Poland, Available at: <http://stat.gov.pl/bdl>
- [3] Boychuk Kh., Bun R. (2014) Regional spatial cadastres of GHG emissions in Energy sector: Accounting for uncertainty, *Climatic Change*, 2014, 124(3), 561-574.
- [4] Corine Land Cover 2006 raster data, Available at: <http://www.eea.europa.eu/>
- [5] Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, Penman Jim, Dina Kruger, Ian Galbally et al.; IPCC, 2001.
- [6] Hamal Kh (2009) Geoinformation technology for spatial analysis of greenhouse gas emissions in Energy sector. Dissertation, Lviv Polytechnic National University.
- [7] IPCC (2006) IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K (eds)
- [8] Oda T., and Maksyutov S. (2011) A very high-resolution (1 km \times 1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights, *Atmos. Chem. Phys.*, 11, 543-556.
- [9] Olivier J.G.J., Van Aardenne J.A., Dentener F., Pagliari V., Ganzeveld L.N., and Peters J.A. (2005). Recent trends in global greenhouse gas emissions: regional trends 1970-2000 and spatial distribution of key sources in 2000, *Env. Sc.*, 2 (2-3), 81-99.

- [10] Poland's National Inventory Report 2012, KOBIZE, Warsaw, 2012, 358 p., Available online at: http://unfccc.int/national_reports
- [11] Rayner P. J., Raupach M.R., Paget M., Peylin P., and Koffi E. (2010) A new global gridded data set of CO₂ emissions from fossil fuel combustion: Methodology and evaluation, *J. Geophys. Res.*, 115, D19306, 11 p.

Spatial Greenhouse Gas (GHG) inventory and uncertainty analysis: A case study for electricity generation in Poland and Ukraine

Petro Topylko¹, Mariia Halushchak¹, Rostyslav Bun^{1,2}, Tomohiro Oda^{3,4},
Myroslava Lesiv⁵, Olha Danylo^{1,5}

¹ Lviv Polytechnic National University, St.Bandery, 12, Lviv, 79013, Ukraine, mail:
petrotopylko@gmail.com;

² Academy of Business in Dąbrowa Górnicza, Poland;

³ Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt,
MD, USA;

⁴ Goddard Earth Sciences Technology and Research, Universities Space Research Association,
Columbia, MD USA;

⁵ International Institute for Applied Systems Analysis, Laxenburg, Austria;

Abstract

Spatial inventory of greenhouse gas (GHG) emissions allows to identify emission changes in space. In this study we have analyzed the specificity of territorial distribution of GHG emission sources for Poland and Ukraine. Mathematical models and geoinformation technology for spatial analysis of GHG emissions from fuel consumption by power and combined heat and power plants have been improved by taking into account uncertainty of input parameters and specific factors for every separate electricity/heat generating companies. We have updated the input digital maps of emission point sources. Based on it, we have developed a spatial GHG emission distribution for 2012. The uncertainties of GHG emissions in CO₂-equivalent for the power plants which we consider in our study are asymmetric and the upper bounds of 95% confidence intervals do not exceed 20,3%.

Keywords: mathematical modeling, GHG emission inventory, electricity/heat production, uncertainty

1. Introduction

In the last decades, rapid technological progress has caused considerable impact on the environment, in particular on the climatic condition of the planet [16]. Therefore, the problem of global warming caused by the increase of greenhouse gas (GHG) concentration in atmosphere is very relevant. Inventory of GHGs is an essential tool for monitoring the practical implementation of commitments to reduce or stabilize emissions. Emission inventories are often provided with spatial distributions. Some examples of such inventories are published in [1, 5].

Electricity generation, is often the most prominent GHG emission category that is closely related to our economic activities. It corresponds to the subsector “1.A.1.a Public Electricity and Heat Production” in the IPCC methodology [15]. According to the IPCC methodology, enterprises producing electricity and heat (rather – their funnels) are emission point sources.

The high-resolution spatial GHG inventory provides localization of point sources not only for electricity generation, but also for many other categories of human activity. In most cases GHG emissions are calculated in a determinative way. GHG emissions from electricity generation are the largest emitters on the analyzed territory [5, 6]. The uncertainties of results of GHG inventories from electricity production have significant impact on the uncertainty of the inventory results not only for this category but also on the summary results of all categories of human activities [4].

The first results of spatial analysis of GHGs emissions caused by fossil fuel fired power and/or heat plants in a few Polish regions were presented in [13]. The study described the mathematical model for spatial analysis of GHGs, and the experiments that were carried out for all Poland. A similar study for Ukraine has been presented in [11]. It should be noted that those studies [11, 13] didn't take into account the uncertainties of the input data (activity data and emission factors). Uncertainty of inventory results must be estimated. The first results of uncertainty analysis of GHG emission inventory have been presented in [12], but the authors did not consider specific parameters for a single company in their estimations (e.g., limits of uncertainty of emission coefficients).

This paper focuses on analysis of uncertainty of GHG emissions in Poland and Ukraine from heat/power plants. We have recalculated facility level CO₂ emissions from national fuel consumption data [8, 10, 14] and carried out the spatial analysis of GHG emissions from electricity generation for Poland and Ukraine for 2012. We have updated the maps of point source emissions for the year 2012 [11, 13]. These maps include geolocation of all enterprises that produce electricity or combined electricity and heat, with maximal power more than 20 MW.

2. Electricity and heat generation in Poland and Ukraine

Electricity generation is one of the strategical categories in Poland and Ukraine, accounting more than 40% of total GHG emissions. Particularly, the percentage of coal in overall fuel consumption is very high. During burning, coal is characterized by the largest emission coefficients (emissions per ton of fuel). Below we specify the differences and similarities of electricity and heat generation in Poland and Ukraine.

2.1 Poland

Electricity or combined electricity and heat producing plants in Poland are categorized into two types: (1) public power/heat plants and (2) autoproducing power/heat plants. Power/heat plants of first type produce electricity for general purposes. Autoproducing power/heat plants, according to the source classification by IPCC [15], are included in another category of energy sector: "1.A.2.Manufacturing Industries and Construction". However, most of these autoproducing plants produce energy/heat not only for their own use but for residential consumers as well. So, it is difficult to identify to which category this plant should be included. Such division of plants is a part of the statistical reports about amount of burned fossil fuel.

Despite the development of renewable energy sources, almost 90% of electricity in Poland is produced by power/heat plants that are located close to the miners of fuel raw materials in Silesia, in the central regions and in Pomoria. The production of electricity based on coal is about 62% of the total energy production. Another important energy resource is brown coal that covers 30% of the energy production [3, 5, 7, 14]. Plants that use brown coal for electricity production are located close to the miners of that fuel [5, 6].

2.2 Ukraine

The specificity of the location of electricity generation plants in Ukraine is similar to Poland. Most power plants are concentrated in the industrial regions of the country and in the places of fuel mining – it is the eastern part of Ukraine. Some power plants are located in the strategic places, for example, Burshtyn power plant is located at

the intersection of the power lines that connect Ukraine with Hungary, Romania and Slovakia. Another power plant Dobrotvir is located not far from Poland and is connected to Polish electrical network.

Outdated equipment (boilers, generators, filters etc.) that is still utilized in many power plants increases the amount of GHG emission. It should be noted that statistical data of fossil fuel used for electricity generation by private companies is not available. Instead, there is available information about amount of generated power, fuel type and fuel consumption per unit of energy produced.

For spatial GHG inventory it is required to have information about amount of fuel consumption or amount of energy output for each individual plant, appropriate emission factors, technical power characteristics, fuel chemical composition, etc. It is difficult to fulfill these requirements due to the lack of available data. For example, in Ukraine the information about the activities of power plants in most cases is not available. In this study, a detailed research of technical parameters of power/heat plants activity (emission factors, technical power characteristics, fuel chemical composition) are done only for a small amount of enterprises.

3. Maps of electricity and heat generating companies

Unlike emissions from diffused sources (e.g. residential and commercial), it is often difficult to model emission spatial distributions for the intense point sources such as electricity and heat plants. Thus, those emissions need to be mapped using the specific location information (geographical coordinates of smoke stacks). We developed a map of point sources for the year 2012 [11,13]. The map for the year 2012 was developed in following steps: (1) collect and compile information about companies that produce electricity (we used all available data about electricity generation branch in Poland and Ukraine), (2) identify geographical coordinates (latitude and longitude) of power/heat companies by using Google Earth high resolution imagery (TM)) and (3) create digital maps using the MapInfo GIS environment. As a result, we developed a database of electricity generating plants that includes the next information: unique identification number; name of electricity generation plant, city, region;geographical coordinates (latitude and longitude);power of plant, MW.

4. Spatial GHG inventory for Poland and Ukraine

Using the point source map we developed, we developed a spatial GHG inventory for the category of electricity and heat production for Poland and Ukraine. We used the models for spatial GHG inventories presented in [11] and [13] for Poland and Ukraine, respectively. These models use specific emission factors for individual power plants and calorific values of fuel, the information about location of plant and its characteristics (if it is available).

The results of GHG emissions estimation are in the form of sets of numerical values of GHG emissions (carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O),) for certain year and for every point source. The thematic maps have been then created based on the results of GHG emission estimates by sources. Figures 1 and Figure 2 illustrate these thematic maps for Poland and Ukraine, respectively. The emission leaders are industrialized regions, where the main production facilities of the country take place. For example, in Ukraine it is Donetsk region, Luhansk region,

Dnipropetrovsk region and Kharkiv region; and in Poland it is Silesia voivodeship, Lodz voivodeship, Mazovia voivodeship and Wielkopolska voivodeship.

The results of spatial GHG inventory of Ukraine and Poland enable to adequately assess the current situation of the industry in terms of GHG emissions and to accept the relevant decisions for GHG emissions reduction.

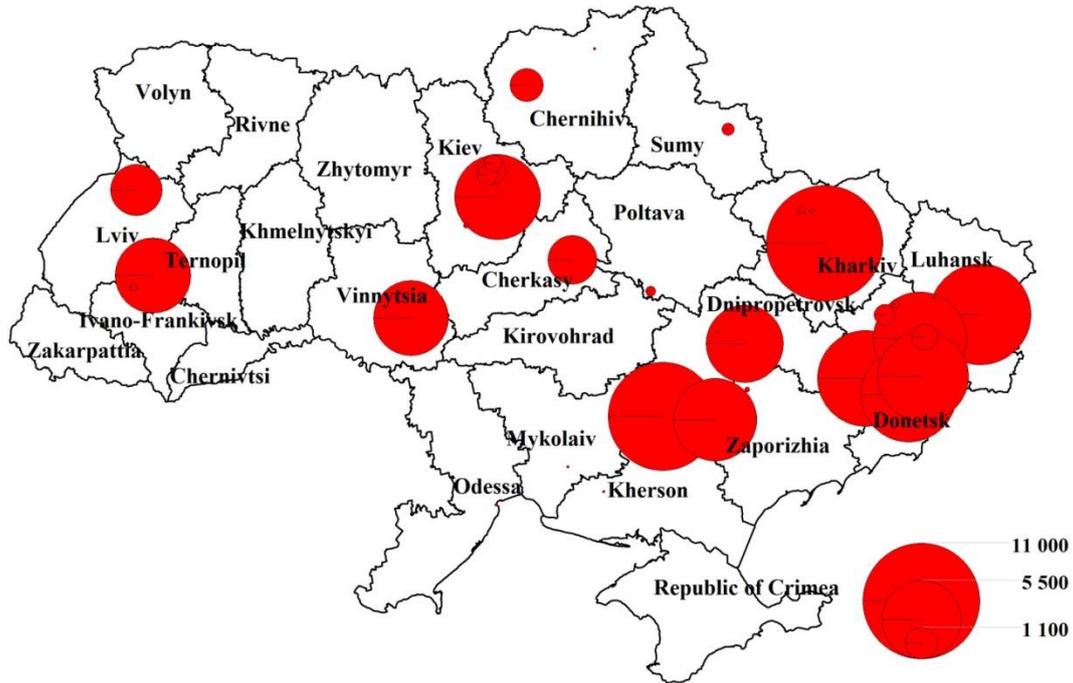


Figure 1. Map of GHG emissions from electricity production in Ukraine, CO₂ equivalent (2012, 10³ t)

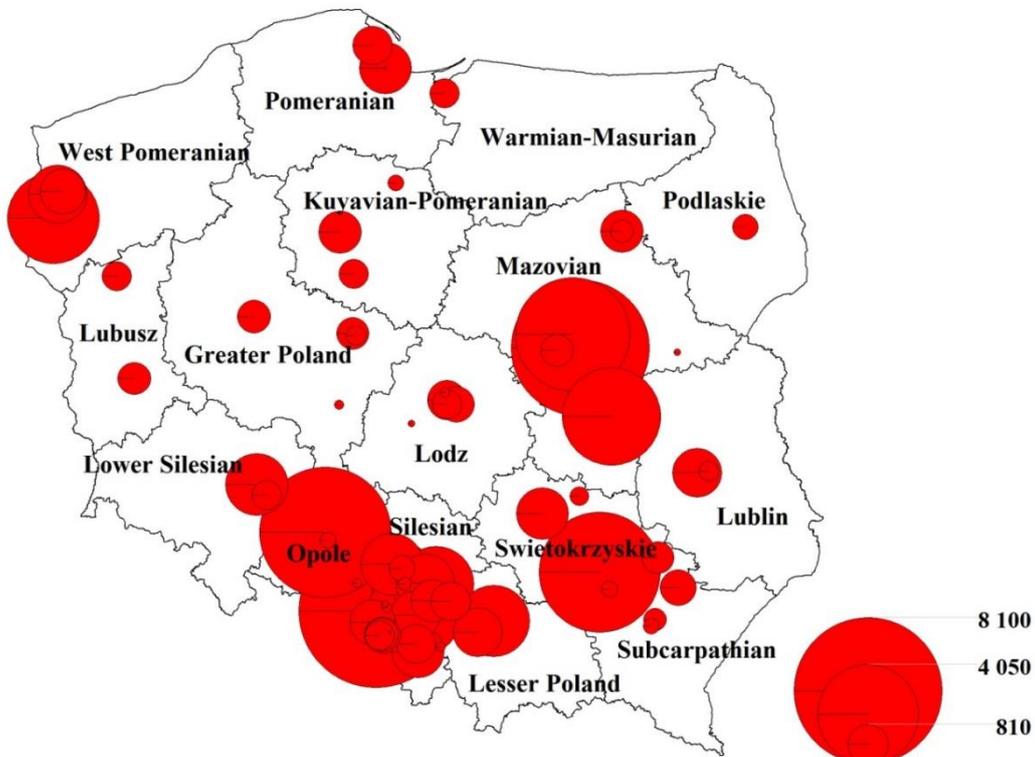


Figure 2. Map of GHG emissions from electricity production in Poland, CO₂ equivalent (2012, 10³ t)

6. Uncertainty analysis of the results of GHG emission inventories

The mathematical models of spatial inventory of GHG emissions include a number of parameters, which are characterized by uncertainty that has to be considered during the estimation of GHG emissions, e.g. uncertainty of emission factors and uncertainty of statistical data on consumption of fossil fuel. Uncertainties of the spatial inventory results depend on uncertainty of activity data, calorific values, emission factors etc. We are estimated such uncertainties by applying the Monte-Carlo method that is suitable for detailed evaluation of uncertainty "from category to category", particularly, if there are parameters with large uncertainties with distribution different from the normal. The density distribution is described by the complex functions and there are correlations between some input data, emission factors, etc.

The uncertainty of emission factors of major greenhouse gases for 1.A.1.a "Public electricity and heat production" category are characterized by normal (CO_2) distribution and log-normal (CH_4 , N_2O) distribution [2]. In Poland's National Inventory Report [8] it is published that emission factors of the major GHGs are characterized by the normal distribution. Uncertainty of emission factors in Poland's energy sector for carbon dioxide (CO_2) is 3,4%, methane (CH_4) – 15,5%, nitrous oxide (N_2O) – 11,3% (upper and lower limits of 95% confidence interval, which is approximately equal to two-sigma).

Table 1. Results of uncertainty estimation for ten the most powerful electricity plants in Poland (2012)

Power/heat plant	CO₂, th. t (uncertainty, %)	CH₄, th. t (uncertainty, %)	N₂O, th. t (uncertainty, %)	Total emissions, th. t (uncertainty, %)
BOT Elektrownia Bełchatów SA	17535,9 (-13,8: +15,3)	0,16 (-18,1: +20,8)	0,24 (-17,1: +19,5)	17611,5 (-13,8: +15,3)
Elektrownia Patnów II	28624,0 (-13,8: +15,3)	0,26 (-18,0: +20,8)	0,39 (-17,1: +19,5)	28747,4 (-13,8: +15,3)
Elektrownia Rybnik SA	7862,2 (-17,6: +20,2)	0,08 (-21,0: +24,8)	0,12 (-20,1: +23,7)	7901,8 (-17,6: +20,2)
BOT Elektrownia Turów SA	8317,4 (-13,8: +15,3)	0,07 (-18,0: +20,8)	0,11 (-17,1: +19,5)	8353,3 (-13,8: +15,3)
BOT Elektrownia Opole SA	6012,1 (-17,6: +20,2)	0,06 (-21,0: +24,8)	0,10 (-20,1: +23,7)	6042,4 (-17,6: +20,2)
Elektrownia Polaniec	5271,9 (-17,6: +20,2)	0,06 (-21,0: +24,8)	0,08 (-20,1: +23,7)	5298,5 (-17,6: +20,2)
Elektrownia Kozienice SA	3501,3 (-17,6: +20,3)	0,04 (-21,0: +24,9)	0,06 (-20,1: +23,7)	3519,0 (-17,6: +20,3)
Elektrociepłownia Siekierki SA	6511,1 (-17,6: +20,3)	0,07 (-21,0: +24,9)	0,10 (-20,1: +23,7)	6543,9 (-17,6: +20,3)
Elektrownia Dolna Odra SA	3117,5 (-17,6: +20,3)	0,03 (-21,0: +24,9)	0,05 (-20,2: +23,7)	3133,2 (-17,6: +20,3)
Elektrociepłownia Żerań SA	4746,0 (-17,6: +20,3)	0,05 (-21,0: +24,8)	0,08 (-20,2: +23,7)	4769,9 (-17,6: +20,3)

As an example, the results of uncertainty estimations for the 10 most powerful plants of Poland are shown in Table 1. The uncertainty range (calculated using the

Monte Carlo method) of emissions of Elektrownia Patnów II that uses brown coal for electricity production is from -13,8 to 15,3. These numbers are lower than the uncertainty range of GHG emission of Elektrownia Rybnik SA that consumes coal.

National Inventory Reports of GHG are published in Ukraine and in Poland. They include the information about results of GHG inventory of different sectors of human activity in accordance with IPCC methodology. Also, there are reports that present the detailed information about protection of environment for particular regions, e.g. the report on the state of the environment in the Lower Silesia. These reports potentially can be used for validation of the results presented in this study. We will carry out the comparison of the presented GHG emissions and the independent GHG emissions estimates in the nearest future.

Elektrownia Patnów II is the biggest plant in Poland, therefore we analyze sensitivity of the uncertainty of GHG emissions for this power plant to the uncertainty of input parameters (Figure 3). These results were calculated using the Monte Carlo method, and relative uncertainties of activity data, net calorific value, and CO₂ emission factor as input data. The results show that decreasing of the uncertainty of the net calorific value by 50% can decrease the total uncertainty for this plant from -13,8%:+15,3% to around 9%. As uncertainties of other parameters are rather low, the relative change of uncertainty of these parameters will not change overall uncertainty for this power plant significantly.

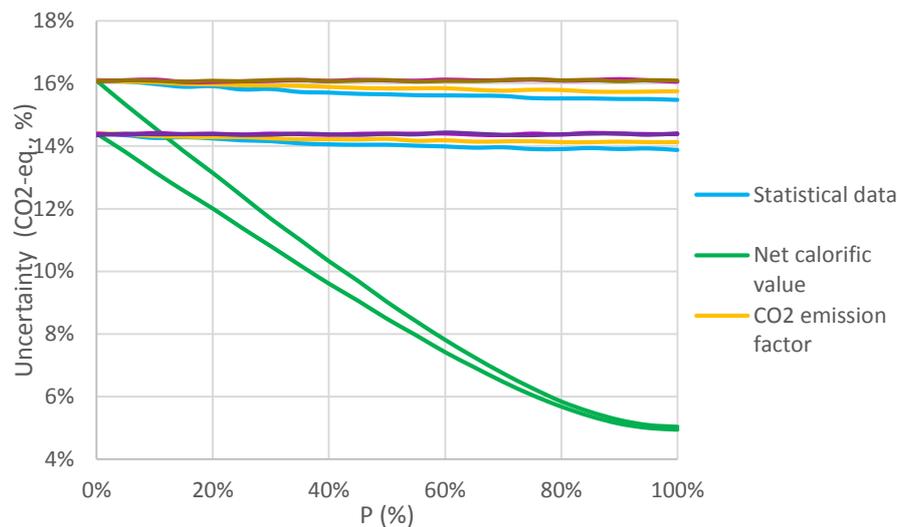


Figure 3. Dependence of total uncertainty of emission estimates for Elektrownia Patnów II to changes of uncertainty (on P %) of input parameters of inventory (the upper and lower limits of 95% confidence interval)

7. Conclusions

In this paper, we presented the results of spatial inventory of GHG emissions and their uncertainty analysis. The study of GHG emissions is based on the official statistics about fossil fuel consumption in Poland and Ukraine

The results of spatial inventory allow to identify power plants that produce the biggest quantity of GHG emissions and investigation of uncertainty of emission inventory results gives policymakers an effective tool for supporting decisions on strategic baselines of economic development and environmental policy. In Poland the leader in GHG emissions is Silesia voivodeship. It is the biggest industrial region

of country. The electricity and heat generating plants from the eastern part of Ukraine emit the most of GHGs in the country.

References

- [1] Danylo O. (2012) Spatial inventory of greenhouse gas emissions in the residential sector: a case-study for Poland and Ukraine, World with Reach: from Science to Policy: IIASA 40th Anniversary Conference, 24-26 October 2012, Vienna, Austria, Available online at: http://conference2012.iiasa.ac.at/poster_session.html
- [2] Hamal K., R. Bun, N. Shpak, O. Yaremchyshyn (2010) Spatial cadastres of GHG emissions: Accounting for uncertainty, The 3rd Intern. Workshop on Uncertainty in Greenhouse Gas Inventories : Proceedings, Lviv, LPNU, 81-90.
- [3] Horynski M., Pietrzyk W., Boguta A. (2012) A model of an energy efficient building automation system, *Econtechmod*, 1(1), 41-46.
- [4] Lesiv M., Bun A., Jonas M. (2014) Analysis of change in relative uncertainty in GHG emissions from stationary sources for the EU 15, *Climatic Change*. Springer, 124(3), 505-518.
- [5] Lesiv M., Bun R., Shpak N., Danylo O., Topylko P. (2012) Spatial analysis of GHG emissions in Eastern Polish regions: energy production and residential sector, *Ekontechmod*, 1(2), 17-23.
- [6] Lesiv M., Bun R., Topylko P. (2011) Geoinformation technologies and models for spatial analysis of GHG emissions: energy production in eastern Polish regions, *Proceedings of the International Workshop "Methods and Applications of Artificial Intelligence"*, 22-23 September 2011, Bielsko-Biała, CIM, 38-48.
- [7] Poland's National Inventory report 2012: Greenhouse Gas Inventory for 1988-2010, Warszawa, National Centre for Emission Management at the Institute of Environmental Protection - National Research Institute, 2012. Available at: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/6598.php.
- [8] Poland's National Inventory report 2014: Greenhouse Gas Inventory for 1988-2012, Warszawa, National Centre for Emission Management at the Institute of Environmental Protection - National Research Institute, 2014, Available at: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/8108.php.
- [9] Stan środowiska w województwie śląskim w 2012 roku. Available online at: <http://www.katowice.wios.gov.pl/monitoring/raporty/2012/raport2012.pdf>.
- [10] State statistics service of Ukraine. Available at: <http://www.ukrstat.gov.ua>
- [11] Topylko P., Bun R. (2013) Geoinformation technology for inventory of emissions of greenhouse gases in the electricity production sector of Ukraine, *Artificial intelligence (Donetsk)*, 4, 432-440.
- [12] Topylko P., Bun R., Striamets O., Danylo O. (2013) Uncertainty of greenhouse gases spatial inventory: power and heat production, *Proceedings of the 8th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT'2013)*, November 11-16, 2013, Lviv, 15-16.

- [13] Topylko P., Lesiv M., Bun R., Nahorski Z., Horabik J. (2013) Geoinformation technology for spatial inventory of greenhouse gas emissions: electricity and heat generation in Poland, *Econtechmod*, 2(2), 51-58.
- [14] Zużycie paliw i nośników energii w 2012 r., Główny Urząd Statystyczny, Warszawa, 2013. Available at: <http://stat.gov.pl/obszary-tematyczne/srodowisko-energia/energia/zuzycie-paliw-i-nosnikow-energii-w-2012-r-,6,7.html>
- [15] 2006 IPCC Guidelines for National Greenhouse Gas Inventories, H. S. Eggleston, L. Buendia, K. Miwa, T. Ngara, K. Tanabe, eds., IPCC, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan, 2006, 5 volumes. Available at: <http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>.
- [16] 2014 IPCC, *ClimateChange 2014: Synthesis Report of the Intergovernmental Panel on Climate Change*. Available at: <http://www.ipcc.ch/report/ar5/syr/>

Uncertainty analysis of GHG spatial inventory from the industrial activity: A case study for Poland

Nadiia Charkovska¹, Mariia Halushchak¹, Rostyslav Bun^{1,2}, Matthias Jonas³

¹Lviv Polytechnic National University, St.Bandery, 12, Lviv, 79013, Ukraine,
charkovska.n@gmail.com;

²Academy of Business in Dąbrowa Górnicza, Poland

³International Institute for Applied Systems Analysis, Laxenburg, Austria

Abstract

Taking into account the global climate change problem, an urgent task is greenhouse gas (GHG) emissions reduction, and associated with this problem the uncertainty analysis of input data and results of GHG inventory. The main purpose of this investigation is the assessment of inventory uncertainty at the level of separate company, region and even country in the industrial sector (including emission processes caused by chemical and physical transformation of materials, as well as by burning of fuels). We discuss the obtained results of spatial GHG inventory in the Industry sector in Poland, by usage of the bottom-up approach, based on IPCC guidelines, official statistics and digital maps of territories investigated. Monte-Carlo method was applied for estimation of inventory uncertainty in main categories of analyzed sector, taking into account the small and large variation of parameters in cases of symmetric and asymmetric distributions. We determined emission sources that have the greatest impact on overall uncertainty in the industrial sector, and evaluated the relative uncertainty depending on uncertainty in activity data and emission factors. The additional knowledge on spatial distribution of emissions and their structure, supports the processes of decisions making on emission reduction.

Keywords: GHG emission, spatial GHG inventory, uncertainty analysis, Monte-Carlo method.

1. Introduction

Nowadays climate change is one of the most urgent ecological problem. Systematic atmospheric measurements show that the concentration of carbon dioxide (CO₂), as the most important anthropogenic greenhouse gas (GHG), has increased more than 20% compared to 1958 year. Apart from the energy sector, a significant share in terms of greenhouse gas emissions belongs to industrial production.

The national inventories of GHG emissions are the key element of the global system of monitoring and control of climate change. For enhancement of the assessment accuracy of GHG emissions inventory it is necessary to improve the inventories by developing new mathematical and software tools. The development of mathematical models of emission processes for GHG inventory is an important task in estimation of emissions, since only for a small number of emission sources experts may make direct measurements.

The assessment of the uncertainty of GHG inventory at the country level as well as individual emission sources is an extremely important problem due to the fact that incorrect estimates may have a significant impact on the process of GHG trading. The results of GHG inventory have a special value only with estimates of uncertainties of the input data (activity data, and emission factors) and the output data (emissions) [5].

2. High resolution spatial GHG inventory

Traditional inventory at country level does not answer the question, where the biggest emitters are located. For this purpose we have developed a geoinformation approach to high resolution spatial inventory of GHG on the example of Poland. In the categories of spatial GHG inventory every large plant for the production of industrial materials is presented as a major point-type source of GHG emissions.

The case study covers all categories of industrial sector: cement, lime, ammonia, nitric acid, iron and sinter production etc. (categories within sectors 1A2, 2A, 2B, 2C according to IPCC Guidelines [1]). This approach basically includes the stages of creation the maps of emission sources, and emission calculation for each source. On the basis of official information on industrial companies localization it is possible to set locations of their production facilities using Google Earth (TM). Also using a digital map of land use (Corine Land Cover [4]) the map of industrial zones as area-type emission sources was created. Thus digital maps of point- and area-type sources of emissions in each category of economic activity were built.

We have developed special models of disaggregation of official statistical data (activity data) from the national level (or where possible from the regional level) to the level of separate plants or industrial areas, that we consider as the elementary objects of study. These models use a set of parameters as disaggregation indicators, including production capacity of companies, data on gross value added in the subregions, available data on the specific of technological processes, fuel used and many others.

Then, using created digital maps and mathematical models we carried out spatial inventory of emissions for each elementary object and got sets of geospatial data on CO₂, CH₄, N₂O emissions, and total emissions in CO₂ equivalent. Below, this approach is illustrated on the example of cement industry only.

The specificity of the main sources of GHG emissions in the industrial sector in Poland is their uneven territorial distributions. The spatial inventory reflects this peculiarity. Also the special feature of our spatial inventory is the high resolution of the obtained results. Maximum resolution is determined by the resolution of used digital maps of land use and does not exceed 100m.

3. Mathematical models and results

Cement industry suffered a significant development before and during the two periods of commitments under the EU ETS. In 2009-2011, the absolute CO₂ emissions from European cement industry decreased by 20-22% compared to the 2000-2005 years. Most emissions from the cement industry are caused by clinker production as an intermediate mineral in the cement production process [2].

Polish cement industry is widely developed in 7 of 16 voivodeships. The cement industry is presented by 11 cement production plants with full production cycle, 1 cement grinding plant, and 1 alumina cement production plant. The full production cycle means all stages of the cement production, in particular the processes of the clinker calcination and cement grinding [6]. The largest cement producers are Górażdże Cement S.A. (concern Heidelberg), Lafarge Cement S.A. (concern Lafarge), and Grupa Ożarów S.A. (concern CRH). The shares of these groups in total cement production are 26%, 21% and 17%, respectively [3].

The main three ways to reduce GHG emissions from the cement industry are the reduction of emissions caused by fuel combustion, the reduction of emissions due technological processes modernisation, and reduction of indirect emissions from

electricity consumption by improving the electric energy efficiency during the clinker and cement production. The reduction of emissions due technological processes modernisation is primarily connected with substitution of clinker by other minerals (a); with decreasing of carbon content in cement (b); with capture and storage / capture and disposal of carbon (c). The analysis showed that about 40-60% of the total emission reductions should be the option of carbon capture and storage of CO₂ [2].

We adapted the mathematical description of carbon dioxide emissions from cement production at the level of separate plant, as a single point source of emission [3], for GHG spatial inventory. According to this mathematical model the carbon dioxide emissions from a single point source is calculated as a product of the quantity of clinker produced, CaO content in clinker, and cement kiln dust losses by the formula:

$$E_{\text{Cement}}^{\text{CO}_2}(\zeta_n) = F_{\text{stat,clinker}}(\zeta_n) \cdot K_{\text{clinker}}^{\text{CO}_2}(\zeta_n) \cdot K_{\text{CKD}}, \quad (1)$$

$$\zeta_n \in \Xi_{\text{cement}}, \quad n = \overline{1, N_{\text{cement}}},$$

where $E_{\text{Cement}}^{\text{CO}_2}$ is the amount of annual carbon oxide emissions from the cement plant; $F_{\text{stat,clinker}}$ is the activity data on clinker production for the cement plant ζ_n ; $K_{\text{clinker}}^{\text{CO}_2}$ is the emission factor for clinker for the cement plant ζ_n ; K_{CKD} is the correction factor for losses of cement kiln dust (it was assumed that $K_{\text{CKD}} = 1.02$); Ξ_{cement} is the set of cement production plants; N_{cement} is the number of these plants.

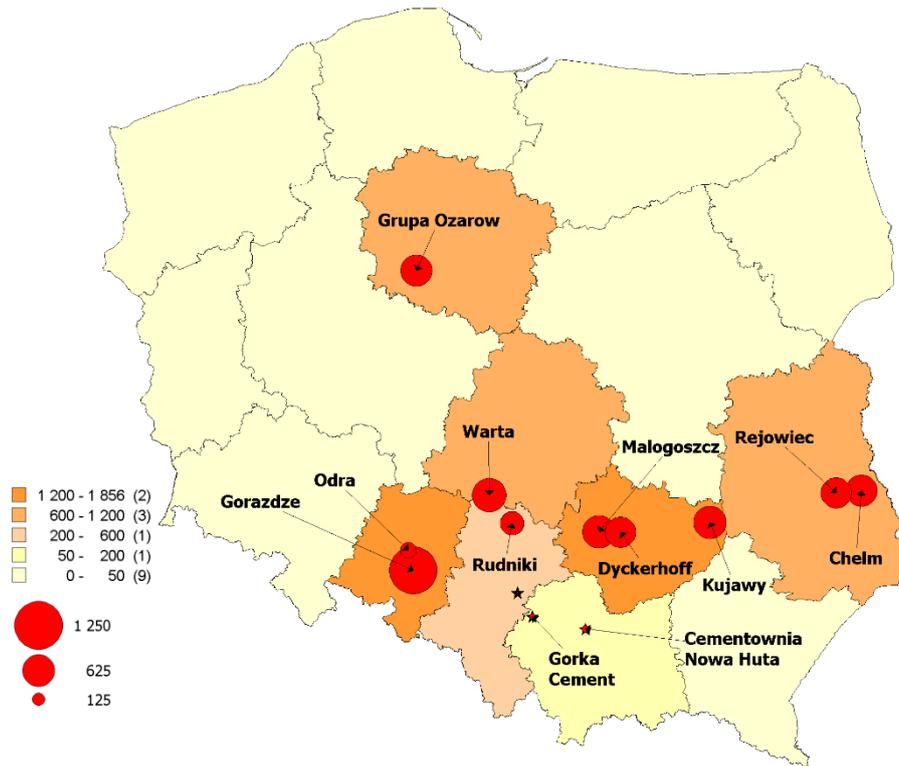


Figure 1. Results of the analysis: main sources of carbon dioxide emissions from cement production in Poland at production plants level, and region level (10³ tons; in 2010)

Data on industrial production were disaggregated from the country level (administrative regions level if possible) to the level of production companies by using specific indicators of disaggregation, for instance, the capacities of large plants as point-type sources of emissions. The above-described mathematical model of emission

processes from cement production (1) is implemented with the usage of geographic information systems (GIS) and developed software in MapBasic.

We formed a set of input geospatial data, and executed computations on GHG spatial inventory in cement production category in Poland (the results are presented in Figure 1). The similar spatial inventories are performed for other categories of economic activity of the Industrial sector, which are characterized by a significant level of emissions.

4. Input data for uncertainty analysis

Input data, which are used in our mathematical models for GHG spatial inventory, are associated with some uncertainties. These uncertainties reflect the lack of our knowledge about emission processes in each category of the industrial sector. Therefore, the statistical data on the output of major industrial goods, which are produced by corresponding plants, and the specific plant's GHG emission factors can be presented as random variables. One of the main methods for modelling GHG emissions taking into account the uncertainty of input data, is Monte Carlo method. Its advantage is the ability of using the information based on uncertainty of input parameters of mathematical models to estimate the level of uncertainty in GHG emissions for different companies, regions and the country as a whole. Today, there are good well-founded price incentives and financial requirements for the accurate accounting of the economic activity (production volumes) [7].

In this regard, the statistics on the results of economic activity tend to reduce uncertainties and decrease the correlation coefficient of data over time. Uncertainty analysis of the activity data in the industrial sector in Poland is carried within preparation of the national inventory reports on GHG emissions. According to estimates of the Polish experts in statistics, the uncertainty of statistical data for various categories of emission sources at the country level is in the range of 2-5% [8]. The national inventory reports show that the most accurate evaluations of emission factors for carbon dioxide together with their uncertainty ranges were obtained for the category "Cement production" (15%) and the category "Metal production" (10%). The uncertainty of methane emission factor for most sources in the industrial processes sector is 20% (normal distribution). In the analysis of uncertainties of GHG emissions from cement production at Polish factories, the following input data were used (see. Table 1): the uncertainty of volumes of clinker produced (symmetric, 2%), the uncertainty range for emission factors for carbon dioxide (symmetric, 15%) for all sources.

5. Results of uncertainty analysis

According to the study of the traditional inventory (not-spatial) the total uncertainty of GHG emission estimates in the industrial sector in Poland in general is 5.2% for carbon dioxide, 13.5% for methane, and 29.7% for nitrous oxide.

On the basis of the created set of geospatial data and the developed approach to the uncertainty analysis of GHG emissions the computing experiments were performed using Monte Carlo method for production of cement, lime, nitric acid, ammonia, iron and agglomerates in Poland (using activity data for 2010). The algorithm of realization of Monte Carlo method consists of 4 steps: setting probability distribution functions of each parameter of mathematical model separately for each cement production plant (1); generating pseudo-random data samples of statistical data and emission factor

accordingly to the density of the probability distribution (2); calculation with mathematical model (1) of the emission estimate using modeled at previous step the random values (one random values of emission is calculated for each emission source) (3); calculation of the total GHG emissions from all sources (4).

Based on the results of modelling in main categories, the uncertainty ranges mentioned above amounted to 7,4% for CO₂ emissions, 16,8% for CH₄ emissions, 35.5% for N₂O emissions (symmetric normal distribution is used). The verification of the correctness of realized mathematical and software tools was carried out using Polish national annual reports on greenhouse gases emission at the country level as a whole. The obtained results show a small overall uncertainty of inventory results of greenhouse gases in the production of basic metals, minerals and chemicals in 2010.

This should positively affect the total uncertainty regional or national total emissions for the all sectors, and give the authorities the opportunity to take into account this factor in the verification of the fulfilment of arrangements on reduction of GHG emissions.

Table 1. Input data for the uncertainty analysis for GHG emissions from cement production (Poland, 2010)

№	Name of plant	Volumes of clinker produced, 10 ³ tons/year	CO ₂ emission factor, t _{CO2} /t	CO ₂ emission, 10 ³ tons	Uncertainty range (lower), %	Uncertainty range (upper), %
1	Cementownia Góraźdże	2400	0,512	1228,82	-15,569	15,979
2	Cementownia Małogoszcz	1215	0,52	631,78	-15,655	15,981
3	Cementownia Kujawy	1215	0,52	631,82	-15,646	15,985
4	Grupa Ozarów	1144,4	0,529	605,38	-15,663	15,993
5	Cementownia Rejowiec	1065,6	0,529	563,72	-15,655	15,986
6	Cementownia Chełm	1137,5	0,529	601,73	-15,650	15,981
7	Cementownia Rudniki	682,5	0,529	361,02	-15,643	15,982
8	Dyckerhoff Polska	1050	0,529	555,43	-15,643	15,975
9	Cementownia Warta	1320	0,529	698,25	-15,659	15,989
10	Cementownia Odra	350	0,529	185,15	-15,653	15,982
11	Górka Cement	50	0,529	26,43	-15,654	15,986
12	Cementownia Nowa Huta	80	0,529	42,36	-15,652	15,978

The authorities should be interested in the reduction of uncertainty of inventory results, and thus in the reduction of uncertainty of its individual components. However, the reduction of uncertainty of emission estimates from certain human activities is an extremely complicated, lengthy and expensive process, it requires the significant investment in the research, measurement and refinement administrative measures. Of course, with gradually increasing of our knowledge about the nature of emissions the

respective uncertainties of individual components of the inventory process are reduced, but to solve this problem in global scale by a short time and with limited funds it is impossible.

Thus the problem of determining of the categories of economic activities, which are the most important in terms of sensitivity analysis, is quite interesting. It means that overall uncertainty of inventory results is the most sensitive to the changes in uncertainty of some input parameters. As an example, Figure 2 illustrates graphically a sensitivity of uncertainty of CO₂ emission estimates from cement production. The results show that the relative uncertainty for carbon dioxide emissions is the most dependent on the uncertainty of CO₂ emission factor. The uncertainty of total emissions little depends on improving knowledge about the activity data in cement industry. For example, the reduction of uncertainty ranges of CO₂ emission factor into 50% causes the decreasing of CO₂ emission uncertainty from 15,6% to 7,8%.

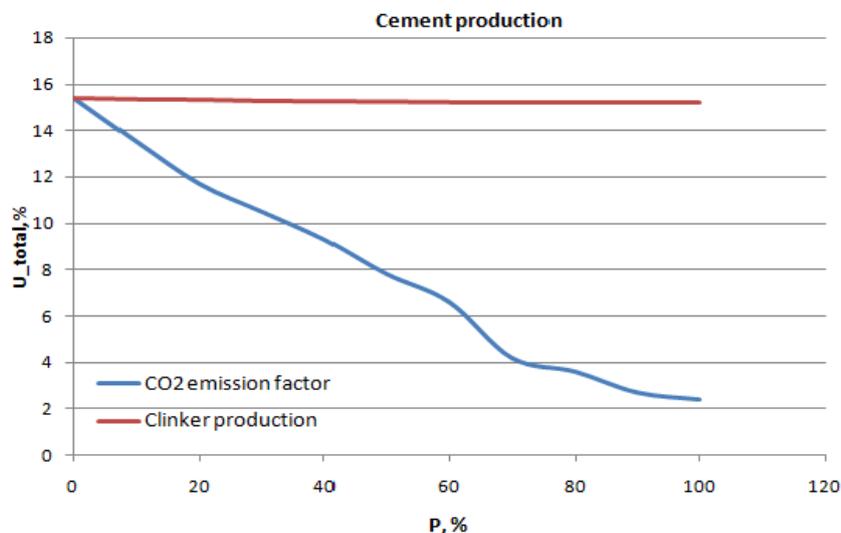


Figure 2. Dependence of total uncertainty of GHG inventory in cement production in Poland from decreasing uncertainty of input data into P percents

6. Conclusions

The obtained results of the mathematical modeling and the spatial analysis of GHG emissions in categories of the industry sector demonstrated the basically low uncertainties of emissions, especially emissions caused by production of cement by respective companies. It has a positive impact on the uncertainty of total regional or national emissions from all categories of economic activity. Thus it gives a possibility to authorities to take into account this factor in the verification of the performance of agreements on the reduction of GHG emissions. The relative uncertainty of carbon dioxide emissions highly depend on the uncertainty of CO₂ emission factor.

References

- [1] IPCC (2006) IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K (eds)

- [2] Carbon Control and Competitiveness Post 2020: The Cement Report Final Report February 2014, Available at: <http://www.climatestrategies.org/research/our-reports/category/61/384.html>
- [3] Charkovska N., Bun R., Nahorski Z., and Horabik J. (2012) Mathematical modeling and spatial analysis of emission processes in Polish industry sector: cement, lime and glass production, *Econtechmod*, 1(4), 17-22.
- [4] Corine Land Cover 2006 raster data, Available at: <http://www.eea.europa.eu/>
- [5] Hamal K., Bun R., Shpak N., and Yaremchyshyn O. (2010) Spatial cadastres of GHG emissions: Accounting for uncertainty, 3rd Intern. Workshop on Uncertainty in Greenhouse Gas Inventories: Proceedings, Lviv, LPNU, 81-90.
- [6] International Cement Review, Available online at: <http://www.cemnet.com>
- [7] Nahorski Z., and Jęda W. (2007) Processing national CO2 inventory emissions data and their total uncertainty estimates, *Water Air Soil Pollution: Focus*, 7(4-5), 513-527.
- [8] Poland's National Inventory Report 2012, KOBIZE, Warsaw, 2012, 358 p., Available online at: http://unfccc.int/national_reports

Spatial inventory of GHG emissions from fossil fuels extraction and processing: An uncertainty analysis

Mariia Halushchak¹, Rostyslav Bun^{1,2}, Matthias Jonas³, Petro Topylko¹

¹Lviv Polytechnic National University, St.Bandery, 12, Lviv, 79013, Ukraine,
mail: halushchak.m@gmail.com;

²Academy of Business in Dąbrowa Górnicza, Poland;

³ International Institute for Applied Systems Analysis, Laxenburg, Austria

Abstract

This article discusses bottom-up inventory analysis for greenhouse gas (GHG) emissions from fossil fuels extraction and processing in Poland. The approaches to modelling geo-referenced cadastres of emissions from fossil fuels extraction and processing are described as well as methods of uncertainty reduction using the knowledge on spatial greenhouse gas emissions distribution. The results of GHG emissions spatial inventory contain the information on geographical coordinates of emission sources. This information is useful for indication the largest emission sources. In this article we present the obtained results on spatial GHG inventory from fossil fuels extraction and processing in Poland, based on IPCC guidelines taking into account locations of emissions sources, official statistics and digital maps of territories investigated. Monte-Carlo method was applied for a detailed estimation of GHG emissions and results uncertainty in the main categories of analyzed sector.

Keywords: spatial GHG inventory, extraction and processing of fuels, fugitive greenhouse gas emissions, uncertainty

1. Introduction

Mankind has faced the most critical global environmental problem – climate changes. These changes are very likely due to increase of the concentration of anthropogenic greenhouse gases (GHG) in the atmosphere, especially for fossil fuel using in the energy sector and emissions in many other categories of anthropogenic activity. Spatial inventory allows us to determine the largest sources of emissions, and it is useful for planning the environmental protection measures on the regional level. Inventory of GHG emissions from fossil fuels extraction and processing in Poland was carried out only at the national level without identification of emissions sources location [3,11]. In this paper we present the mathematical models of emission processes of GHG emissions from fossil fuels extraction and processing at the level of separate emission sources.

The value of emissions we can't measure, only evaluate, that is why we talk about uncertainty of the results [5]. The assessment of the uncertainty of GHG inventory at the country level and level of individual emission sources is an important task. It play very big role in the trading of GHG.

2. Spatial GHG inventory

The spatial inventory includes determination of location emission sources and assessment emissions from these sources. Unlike the traditional inventory at the country level, the spatial one takes into account the location and specifics of each source of emission, which makes it possible to build more detailed emission inventories [4, 6, 10].

According to the IPCC Guidelines [9] we indicated emissions from extraction and processing oil and gas, and extraction and processing of solid fuels. Figure 1 presents the classification of emission categories in this sector. Here take place two types of emissions: the fugitive emissions, and emissions from fuel combustion [7-9].

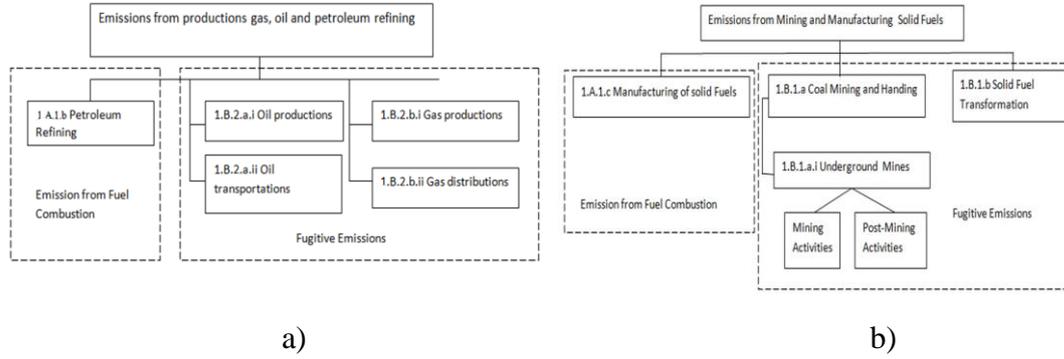


Figure 1. The classification of GHG emissions, which arise from extraction and processing: a) gas and oil; b) coal [8, 9]

For carrying out the GHG spatial inventory from fossil fuels extraction and processing industry, all places with extraction of coal, gas, and oil, coke plants and refineries are presented as point emission sources (their sizes are small in comparison to the size of the country and we can ignore area of them). The digital maps of point sources for emission categories under investigation were built.

We have developed the special models for disaggregation of official statistical data (activity data) from the national level (data at the level of mine or plant are confidential) to the level of separate sources. Then, using created digital maps and mathematical models we calculated emissions for each elementary object.

3. Mathematical model

For GHG spatial inventory we adapted the mathematical description of emission process in investigated area. As write before, we estimate separately fugitive GHG emissions and emissions from burning fuels. As an example, we present below the mathematical model of emissions from coal industry (the models for oil and gas industry are similar and depend on available of statistical data at corresponding level). According to this mathematical model the methane fugitive emissions we can calculate using formula:

$$E_{coal}^{gl}(\xi_n) = E_{coal,m}^{gl}(\xi_n) + E_{coal,p}^{gl}(\xi_n), \quad (1)$$

$$\xi_n \in \Xi_{coal}, n = \overline{1, N},$$

where $E_{coal}^{gl}(\xi_n)$ is the amount of annual fugitive emissions from ξ_n coal mine, $gl \in (CO_2, CH_4)$; $E_{coal,m}^{gl}(\xi_n)$ is the amount of fugitive emissions from mining process; $E_{coal,p}^{gl}(\xi_n)$ is the fugitive emissions which appear after mining process (keeping, transportation and other); Ξ_{coal} is the set of coal mines; N is the number of coal mines from Ξ_{coal} set.

Emissions from burning fuels at separate coke plant we calculate by formula:

$$E_{coke}^{g,f}(\eta_k) = D_{stat,coke}^f \cdot K_{coke}^f(\eta_k) \cdot K_{em,coke}^{g,f}(\eta_k), \quad (2)$$

$$\eta_k \in N_{coal}, \quad k = \overline{1, K},$$

where $E_{coke}^{g,f}(\eta_k)$ is the amount of annual emissions from burning fuels at η_k coke plant, $g \in \{CO_2, N_2O, CH_4\}$, f is the type of fuel, which are used at η_k coke plant; $D_{stat,coke}^f$ is the national data about using fuels in all coke plants; $K_{coke}^f(\eta_k)$ is the disaggregation factor, which depends on available of statistical data for η_k coke plant; $K_{em,coke}^{g,f}(\eta_k)$ is the emission factor for the f -th type of fuel, and different GHG; N_{coal} is the set of coke plants; K is the number of coke plants.

4. Results of spatial inventory

Coal is the main source of energy in Polish industry. That is why in the energy sector dominates the production of solid fuels, mostly coal. Poland ranks 8th place on the extraction of coal in the world (176 mln. t. in 2010). All coal mines are disposed in Silesian and Lublin voivodeships. There are 9 coke plants in Poland. They are located in Silesian (Śląskie) and Lesser Poland (Malopolskie) voivodeships. The largest metallurgical plants are in these voivodeships and coke is a main fuel in this industry.

Using a geographic information system the geoinformation technology has been developed, in which the above mentioned models (1) and (2) are used to estimate the emissions of GHG emissions from fossil fuels extraction and processing.

Statistical data have been collected, and an input geospatial database has been formed (separately for each source category) [1]. The database contains information on names and locations of emission sources, their production capacities, and the specific emission factors.

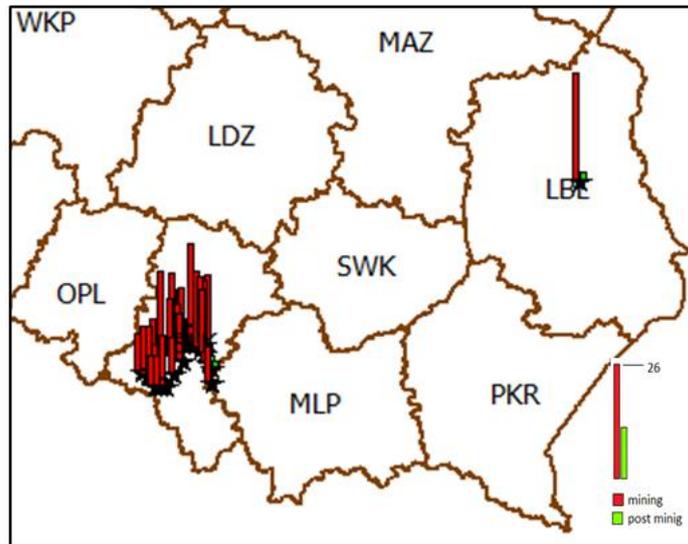


Figure 2. Fugitive emissions of CH_4 from coal mining for separate coal mines (th. t, Poland, 2010)

The results of inventory are obtained on the level of separate emission sources. In this paper we present the assessment of emissions in coal industry (see Figure 2 and

Figure 3). The similar spatial inventories are performed for oil and gas extraction and processing in Poland.

5. Uncertainty analysis

In the Polish national report on GHG emissions [11] the total uncertainty of GHG emission estimates is only in the energy industries sector (fuel combustion) without division by the types of anthropogenic activity. The simplified approach is used, which is based on the assumptions that every value is independent and probability distribution is symmetric. In general the uncertainty in the energy industries is 3,4% for carbon dioxide, 18,4% for methane, and 11,6% for nitrous oxide, uncertainty of fugitive emissions from fuels is 48,8% for methane (extraction of coal), and 5,4% for methane (extraction of gas and oil) [11].

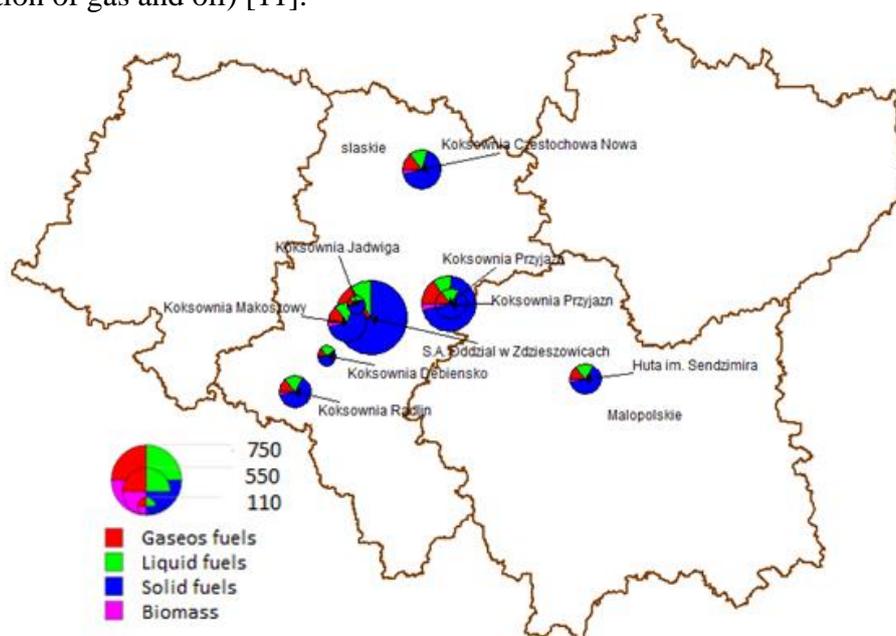


Figure 3. Result of GHG emissions spatial inventory from burning coal, oil, natural gas and biomass by type of fuel for separate coke plants (th. t.,CO₂-equivalent, Poland, 2010)

Uncertainty of emission factors for fuel combustion is low for carbon dioxide (or rather is considered a small) compared with other greenhouse gases, as the emission of this gas is mainly dependent on the carbon content in the fuel, which is very easy to identify. The uncertainty of CO₂ emissions is symmetric. The high level of uncertainty for methane and nitrous oxide can be explained by a lack of studies to establish national emission factors that take into account the specifics of individual processes and the lack of understanding of the formation of these emissions reductions [2,6].

Based on the results of modelling GHG emissions from burning coal in coke plants, the uncertainty is 3,76% for total emissions in CO₂ equivalent (symmetric normal distribution for CO₂ emission coefficient and lognormal for CH₄ and N₂O emission coefficient are used). The results of modelling of GHG emissions from burning coal and their uncertainties for separate coke plants are presented in Table 1. We also estimate uncertainties for separate coal mines. Table 2 presents results of modelling fugitive GHG emissions for ten the biggest coal mines (on data of 2010).

Table 1 The results of modelling GHG emissions and their uncertainties for separate coke plants

Name of coke plant	CO ₂ emissions, t; uncertainty, %	CH ₄ emissions, t; uncertainty, %	N ₂ O emissions, t; uncertainty, %	Total emissions, t; uncertainty, %
Coke plant Przyjaźń	464,408.5	2,756.6	603.8	466,722.0
	±3.687	-37.335..+49.308	-45.438..+68.080	-3.689..+3.767
Coke plant Jadwiga	50,018.4	968.5	65.0	50,154.675
	±3.689	-37.334..+49.308	-45.438..+68.080	-3.689..+3.767
Coke plant Dębieńsko	53,585.6	1,037.7	69.8	54,731.6
	±3,686	-37.335..+49.308	-45.438..+68.080	-3.687..+3.767
Coke plant Radlin	133,964.0	2,594.2	174.2	136,328.9
	±3.687	-37.335..+49.309	-45.439..+68.080	-3.689..+3.767
Coke plant Przyjaźń	133,964.0	2,594.2	174.2	136,328.9
	±3.687	-37.335..+49.309	-45.438..+68.081	-3.687..+3.768
Coke plant Częstochowa Nowa	232,204.2	4,496.6	301.9	236,936.8
	±3.687	-37.335..+49.308	-45.438..+68.080	-3.687..+3.767
Coke plant Makoszowy	206,434.0	4,150.7	268.4	210,996.3
	±3.689	-37.335..+49.308	-45.439..+68.080	-3.689..+3.767
S.A. Oddział w Zdźszowicach	722,518.0	1,4527.5	939.5	724,486.3
	±3.689	-37.335..+49.308	-45.439..+68.080	-3.689..+3.768
Ironworks im. Sendzimira	137,623.0	2,767.1	178.9	139,997.9
	±3.689	-37.335..+49.308	-45.438..+68.080	-3.689..+3.767

Table 2 The results of modelling GHG emissions and their uncertainties for the main coal mines

Name of coal mine	Volumes of coal extraction; 10 ³ tons/year	CH ₄ emission factor; t _{CO2} /t	CH ₄ fugitive emissions, Gg	Uncertainty, %
KWK Murcki Staszic	3.875	4.90	18.977	48.49
KWK Mysłowice-Wesoła	3.229	4.91	19.029	48.49
KWK Wujek	4.982	4.91	24.466	48.49
Oddział KWK Jankowice	2.759	4.91	13.547	48.49
Oddział KWK Knurów-Szczygłowice	3.792	4.91	18.622	48.49
Oddział KWK Sońnica-Makoszowy	3.285	4.91	16.13	48.49
Oddział KWK Ziemowit	4.097	4.91	19.912	48.49
Oddział KWK Piast	4.613	4.87	22.423	48.49
KWK Wieczorek	3.405	4.9	16.548	48.49
KWK Bogdanka	5.351	4.91	26.011	48.49

The total uncertainty of emissions depends on uncertainties of all input parameters of emission model, such as uncertainty of statistical data, uncertainty of CO₂, CH₄ and N₂O emission coefficients. Figure 4 graphically shows results of sensitivity analysis of

the total uncertainty of emissions estimates to improvement of uncertainty of input parameters on P percent. Results demonstrate that relative uncertainty of total emissions in CO₂-eqv. largely depends on uncertainty of statistical data and uncertainty of CO₂ emission coefficient. Uncertainty of total emissions stays almost unchangeable with the change of uncertainty of N₂O and CH₄ emission coefficients. For example, the reduction of uncertainty ranges of CO₂ emission factor into 50% causes the decreasing of total emission uncertainty from 3.76% to 2.63%.

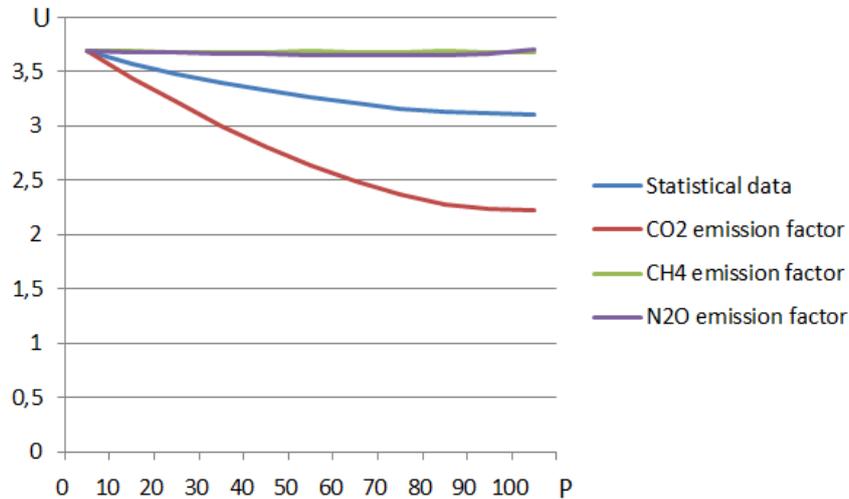


Figure 4. Dependence of total uncertainty of GHG inventory from burning coal in the coke plants of Poland (U) from decreasing uncertainty of input data into P percent (Monte Carlo simulations).

6. Conclusions

The mathematical models of GHG emissions resulting from fossil fuels extraction and processing were adapted to spatial inventory. The specialized geoinformation technology for spatial assessment of GHG emissions, which is based on elaborated mathematical models and uses the created geospatial database of input data was developed. The fugitive GHG emissions, that arise from extraction coal, gas and oil, as well as emissions from burning coal, oil, natural gas and biomass in the coke plants and refineries, and the fugitive emissions that arise during coking coal and processing oil, were examined.

The digital maps of the locations of extraction coal, gas and oil, the locations of coke plants and refineries in Poland were created, the layers with geospatial data about the structure of GHG emissions in the fossil fuels extraction and processing industry in Poland, taking into account specific emission factors for these objects, were formed. Based on performed numerical experiments the geospatial database and digital map of GHG emissions in Poland were obtained. The results of the inventory of greenhouse gas emissions were visualized by digital maps.

Spatial inventory of GHG emissions is useful to the authorities for making informed decisions to reduce GHG emissions. Estimation of uncertainty of GHG inventories can be helpful to comply the obligations on emission reduction. The sensitivity analysis demonstrates, that relative uncertainty of total emissions in CO₂-equivalent largely depends on uncertainty of statistical data and uncertainty of CO₂ emission coefficient. At the same time the uncertainty of total emissions stays almost unchangeable with the change of uncertainty of N₂O and CH₄ emission coefficients.

References

- [1] Bank Danych Lokalnych (Local Data Bank), GUS, Warsaw, Poland, Available at: <http://stat.gov.pl/bdl>
- [2] Bun R., Hamal Kh., Gusti M., Bun A. (2007) Spatial GHG inventory on regional level: Accounting for uncertainty, Proc. of the 2nd Intern. Workshop on Uncertainty in Greenhouse Gas Inventories, Laxenburg, Austria, 27-32.
- [3] Gawlik L., Grzybek I. (2003) Results of research on inventory of methane fugitive emission from coal mining system, *Archiwum Ochrony Środowiska (Instytutu Podstaw Inżynierii Środowiska PAN)*, 29(1), 3-23.
- [4] GIS, Spatial Analysis and Modeling (2005), Eds. D. J. Maguire, M. Batty, M. Goodchild, Redlands, ESRI Press, 482.
- [5] Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, Penman Jim, Dina Kruger, Ian Galbally et al.; IPCC, 2001.
- [6] Hamal Kh. (2009) Geoinformation technology for spatial analysis of greenhouse gas emissions in Energy sector, Thesis for a candidate's degree, Lviv Polytechnic National University, Lviv, 246.
- [7] Harrison M.R., Shires T.M., Wessels J.K., Cowgill R.M. (1997) Methane Emission from the Natural Gas Industry. EPA/GOO/SR-96/080
- [8] IPCC/OECD/IEA Programme on National Gas Inventories (1996). Fugitive emissions from Oil and Natural Gas Activities. Revised 1996 IPCC Guidelines for National greenhouse Gas Inventories.
- [9] IPCC (2006) Guidelines for National Greenhouse Gas Inventories, H.S.Eggleston, L.Buendia, K.Miwa, T.Ngara, K.Tanabe, eds., IPCC, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan, 2006, 5 volumes. Available online at: www.ipcc-nggip.iges.or.jp/public/2006gl/index.html
- [10] Lesiv M. (2011). Mathematical modelling and spatial analysis of greenhouse gas emissions in regions bordering Ukraine, Theses for Ph.D degree on technical sciences, Lviv Polytechnic National University, Lviv, 195.
- [11] Poland's National Inventory Report 2012: Greenhouse Gas Inventory for 1988-2010 (2102) National Centre for Emission Management at the Institute of Environmental Protection, Warszawa

Carbon Emission Inventory Calculation and Analysis based on Coal Lifecycle

Xiangyang Xu^{1,3} Junlian Gao^{2,3}

¹ School of Management, China University of Mining and Technology, Beijing, China
xxy@cumtb.edu.cn

² School of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing, China
gaojunlian100@126.com

³ Centre for Resources and Environmental Policy, China University of Mining and Technology, Beijing, China

Abstract

Coal is the main source of GHG emission in China. To systematic and comprehensive estimate the GHG emission of coal lifecycle, according to the intergovernmental panel on climate change (IPCC) method, the paper using Monte Carlo simulation, calculates carbon emission in each stage of coal lifecycle, and quantifies the uncertainties in results. The results show that GHG emission in coal mining, transportation and consuming is 8526.67 million tons, 3328.35 million tons and 1356723.59 million tons respectively in 2011, and the whole GHG emission in coal lifecycle is 1367186.72 million tons. The uncertainty analysis shows that coal mining contains the greatest uncertainties, the uncertainty of GHG emission from the coal lifecycle in 2011 is (-5.09%, 4.85%).

Keywords: Carbon emission inventory, Monte Carlo, Uncertainty

1. Introduction

Coal has been the dominant source of energy used to fuel the rapid economic development of China in the past two decades. In 2011, China relies on coal for approximately 70% of its energy. However, during the production and consumption process of coal, it has also produced significant impact on its physical environment, of which, the GHG emissions are well known.

Currently many researches have done on GHG emission inventories, but most of these researches have done on city level, and seldom researches have done on GHG emission according to different industries. Shi Huading etc., put forward the basic principles of GHG inventory methods of power sector, and established GHG inventory methods framework on power sector according to national conditions in China. Zheng Shuang proposed methodology about coalbed methane GHG emissions inventories. To our best knowledge, the studies on GHG emission inventory of coal lifecycle is rarely found in the literature.

2. Research Method

To calculate the emission inventory, the method is comes from “2006 IPCC Guidelines for National Greenhouse Gas Inventories” which is a relatively complete theory of GHG emissions and emission estimation methods and processes. It has been adopted by many countries now and “IPCC National Greenhouse Gas Inventory Good Practice and Uncertainty Management”. As to perform quantitative uncertainty analysis, the paper adopted the Monte Carlo simulation.

3. Greenhouse gas inventories of China's coal industry

The coal lifecycle consists of mining, washing, transportation, conversion and utilization. The GHG emission from each stage will be introduced in this section.

3.1 GHG emission from coal mining

Methane will dissipation during the geological process of coal formation, until some residual methane in coal is released from coal mining. In 2011, the total production of the coal in China is 3.888 billion tons, of which, 2.545 billion tons was produced through underground mining and 1.343 billion tons was through open pit. According to the provision in "Good Practice", there is 5% error in coal production register production, therefore we set the mining production through underground and open pit as triangular distribution. For simplicity, we set the production increase 5% over the register production as the maximum production, and the production decrease 5% as the minimum production, and the register production find in publication is the most possible values. The low, average and high CH₄ emission factors for underground mining are 16.75kg/t, 12.06kg/t and 6.7 kg/t respectively using global average value. The low, average and high CH₄ emission factors for surface mining are 2.0kg/t, 1.2kg/t and 0.3kg/t respectively using global average value. Related parameters are shown in Table 1.

Equation for estimating CH₄ emissions from coal mining is as follows:

$$CH_4 \text{ emissions} = \sum_i A_i \times EF_i * GWP \quad (1)$$

A_i is the activity level of mining way i , EF_i is emission factor of transportation mode i . GWP is global warning potential, and the GWP of CO₂ is 1, the GWP of CH₄ is 21, the GWP of N₂O is 275. Using Crystal Ball software, through 4000 times Monte Carlo simulation calculation, the total CH₄ emission in coal mining is 85.2667 million tons CO₂ equivalent, 95% confidence interval is (4783.05, 12502.50).

Table 1 Coal mining methane emissions related parameters

	Underground mining	Surface mining
Activity level probable value (10 thousand tons)	254502.74	134321.23
Activity level maximum value (10 thousand tons)	267227.88	141037.29
Activity level minimum value (10 thousand tons)	241777.60	127605.17
Emission factor probably value (kg/t)	12.06	8.04
Emission factor maximum value (kg/t)	16.75	13.40
Emission factor minimum value (kg/t)	6.70	2.01

3.2 GHG emission from coal transportation

The coal transportation in China mainly relies on railway, highway and waterway. In China's coal railway transportation, there are two kinds of locomotive, diesel locomotive and electric locomotive, which consumption oil and electricity respectively. Electricity is secondary energy, so we only take diesel locomotive into consideration, which use diesel as the main fuel. Coal transportation through highway usually uses medium-duty truck or heavy trucks which can load more than 20 tons, and most of them use diesel. Waterway transportation also uses ships mainly driven by diesel locomotive, from barge to large-scale ocean cargo ship.

In China, 2270.26 million tons coal was transported by railway in 2011. Its average haul distance was 645 km. Thus the volume of the coal circular flow was 1464.33177 billion tons kilometer, and 53.6% was transported by diesel locomotive and 46.4% by electric locomotive. The circular flow volume of diesel locomotive was 784.874 billion tons kilometer. In the railway transportation, the diesel locomotive consume 26.8 kg diesel each ten thousand tons kilometer, and railway transportation totally consumed 2103463.10 tons diesel. In China 0.35 billion tons coal was transported by highway, 2011. Its average haul distance was 250 km. Thus the volume of the coal circular flow through highway transportation was 87.5 billion tons kilometer. The diesel consumed by highway transportation is 6 kg each hundred tons kilometer, and highway transportation totally consumed 5250000 tons diesel. In China 0.65 billion tons coal was transported by waterway, 2011. Its average haul distance was 1768.75 km. Thus the volume of the coal circular flow through waterway transportation was 1149.688 billion tons kilometer. The diesel consumed by waterway transportation is 21.5 kg each ten thousand tons kilometer, and waterway transportation totally consumed 2471828.13 tons diesel. All kinds of GHG emission factor using the default values in IPCC 2006. Coal transportation parameters of activity level and emission factor are shown in Table 2.

Table 2 Coal transportation parameters of activity level and emission factor

Item	Railway transportation	Highway transportation	Waterway transportation
Diesel activity level probable value (t)	2103463.09	5250000.00	2471828.13
Diesel activity level minimum value (t)	1998289.94	4987500.00	2348236.72
Diesel activity level maximum value (t)	2208636.24	5512500.00	2595419.53
CO ₂ emission factor default value (kg/t)	3211.92	3211.92	3211.92
CO ₂ emission factor minimum value (kg/t)	3146.90	3146.90	3146.90
CO ₂ emission factor maximum value (kg/t)	3242.26	3242.26	3242.26
CH ₄ emission factor default value (kg/t)	0.18	0.17	0.30
CH ₄ emission factor minimum value (kg/t)	0.07	0.07	0.15
CH ₄ emission factor maximum value (kg/t)	0.45	0.41	0.46
NO emission factor default value (kg/t)	1.24	0.17	0.09
NO emission factor minimum value (kg/t)	0.62	0.06	0.05
NO emission factor maximum value (kg/t)	3.72	0.52	0.21

Table 3 GHG emission from coal transport (10 thousand CO₂ equivalent)

	Calculated value	Minimum value	Maximum value
Total CO ₂ emission	3142.88	2989.98	3300.89
Total CH ₄ emission	5.28	2.55	8.09
Total NO emission	168.25	69.17	297.38
Total emission	3328.35	3096.73	3574.06

According to IPCC (2007) recommendation, the equation for estimating CH₄ emissions from coal transportation is as follows:

$$E = \sum_i \sum_j A_i * EF_{ij} * GWP_{ij} \quad (2)$$

In the equation, E is GHG emission of coal, A_i is activity level of fuel i (t); EF_{ij} is emission factor of GHG j from fuel i (t/t fuel), GWP_{ij} is global warming potential of GHG j from fuel i . Using Crystal Ball software, the total GHG emission in coal transportation is calculated through 4000 times Monte Carlo simulation. The results are shown in Table 3.

3.3 Carbon emission inventory from the coal consumption

In the carbon emission inventory, Coal combustion is the main source of greenhouse gases. According to the availability of emission factors, we adopted the reference method recommended in the "Guide" and the latest emission factors announced by the National Bureau of Statistics. Set the data provided by the Statistical Yearbook as the most probable value in the triangular distribution. Set the date increase 5% over the probable value as maximum and -5% as minimum. The coal activity data is shown in Table 4.

Table 4 Coal consumption data in triangle distribution of 2011(10,000 tons)

Source category	Probable value	Minimum	Maximum
Raw coal	346799.81	329459.82	364139.80
Washed coal	49101.65	46646.57	51556.73
Briquette	1019.20	968.24	1070.16
Coke	38163.29	36255.13	40071.45
Other washed coal	11962.94	11364.79	12561.09

According to the recommendation of IPCC (2006), we use the emission factors in Table 5 provided by Chinese Academy of Engineering as the basis of calculation.

Table 5 Coal emission factors

Source category	Raw coal	Washed coal	Other washed coal	Briquette	Coke
CO ₂ emission factor defaults (kg CO ₂ /t)	2.009	2.531	1.004	1.689	3.044
minimum	1.9086	2.4045	0.9538	1.6046	2.8918
maximum	2.10945	2.65755	1.0542	1.77345	3.1962
CH ₄ emission factor defaults (kg CH ₄ /t)	20.908	26.344	10.454	17.584	28.435
minimum	19.86	25.03	9.93	16.7	27.01
maximum	21.953	27.661	10.977	18.463	29.857
N ₂ O emission factor defaults (kgN ₂ O/t)	31.362	39.516	15.681	26.376	42.653
minimum	29.794	37.54	14.897	25.057	40.52
maximum	32.93	41.492	16.465	27.695	44.786

The formula of total greenhouse gas emissions from coal transportation is as follows:

$$E = \sum_i \sum_j A_i * EF_{ij} * GWP_{ij} \quad (3)$$

In the formula, E is greenhouse gases emissions of coal, A_i is the activity level of i -th coal (t); EF_{ij} is the i -th coal's emission factor of j -th greenhouse gas(t/t); GWP_{ij} is the i -th coal's global warming potential of j -th greenhouse gas(t/t).

Table 6 The greenhouse gas emissions from the coal consumption (10000tCO₂)

	Calculation	Minimum	Maximum
CO ₂ emissions	951467.08	884349.76	1023140.33
CH ₄ emissions	205.24	190.65	220.12
N ₂ O emissions	403022.78	375227.33	434125.40
Total emissions	1356723.59	1266139.84	1442211.43

Simulating 4000 times by using Crystal Ball software, we could get the greenhouse gases emissions from the coal consumption. The results are shown in Table 6.

3.4 Total carbon emissions of the coal industry

Total GHG emission of coal lifecycle is equivalent with the sum emissions of coal mining, transportation and consumption as shown in Figure 1.

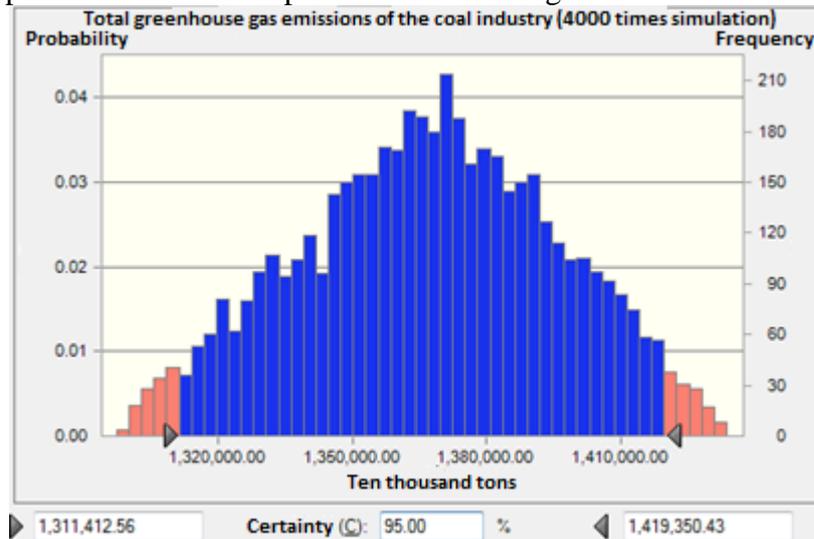


Figure. 1 Analog frequency view of total greenhouse gases emissions

It can be seen from Figure 2, total emissions of greenhouse gases from the coal industry in China is 13,671,867,200 tons in the year 2011, the 95% confidence interval is (1,297,537.98, 1,433,468.58).

4. Uncertainty analysis

Uncertainty estimate is an essential element of a complete emissions inventory. In the compile process, there are many uncertain factors in the estimation factors and activity levels, it can help to determine the direction of future efforts and guide decisions on methodological choice by uncertainty analysis. The “*National Greenhouse Gas Inventories Good Practice Guidance and Uncertainty Management*” formulated by the IPCC had united the methods of uncertainty quantification for each country. This section will analyze the uncertainty of China's coal emissions inventory in 2011.

Statistical values	Predictive value
Number of trials	5,000
Basic information	1,368,578.61
Mean	1,367,186.72
Median	1,367,834.03
Mode	---
Standard deviation	28,433.13
Variance	808,442,999.89
Skewness	-0.0806
Kurtosis	2.39
Coefficient of variation	0.0208
Min	1,297,537.98
Max	1,433,468.58
Average standard deviation	402.11

Figure 2 Simulation statistics of total greenhouse gases emissions

In this paper, we analyze the carbon emissions uncertainty for each source by the method of Bootstrap. When sampling from each group of model, sample distribution representing probability distribution model of the input of source category. We pick out 4000 random sample every time and calculate the average value. Therefore, by repeating 4000 times, we obtain 4000 average value on behalf of the input of model. The averages describe the distribution of uncertainty of the model input (emission factors and activity levels). Table 7 is the average of model input and the uncertainty of 95% confidence interval.

Table 7 Uncertainty Analysis of carbon emissions

Stages	Type	Averages (10,000 t CO ₂)	2.5% Quantile (10,000 tCO ₂)	97.5% Quantile (10,000 tCO ₂)	Uncertainty (%)
Coal mining	CH ₄	8526.67	4783.05	12502.50	(-43.90, 46.63)
	Total	8526.67	4783.05	12502.50	(-43.90, 46.63)
Coal transportation	CO ₂	3142.88	2989.98	3300.88	(-4.87,5.03)
	CH ₄	5.28	2.55	8.10	(-51.72,53.18)
	N ₂ O	168.25	69.17	297.38	(-58.89,76.75)
	Total	3328.35	3096.73	3574.06	(-6.96,7.38)
Coal consumption	CO ₂	951467.08	899598.69	1007308.23	(-5.45,5.87)
	CH ₄	205.24	193.92	216.80	(-5.52,5.63)
	N ₂ O	403022.78	381611.06	427324.76	(-5.31,6.03)
	Total	1356723.59	1286040.00	1422731.59	(-5.21,4.87)
Total emissions		1367186.72	1297537.98	1433468.58	(-5.09,4.85)

From the Table 7, we can see that the greenhouse gases inventory of coal mining has large uncertainty, Its' absolute value is greater than 40%,the main reason is the emission factor entered and activity emissions have large uncertainty, which spread to the total emissions of mining areas, it leads to the uncertainty of greenhouse emission is too large.

The uncertainty of CH₄ and N₂O from coal transportation is large, however the emissions uncertainty of greenhouse gases from coal transportation is low, it mainly due to the emission of CH₄ and N₂O is small, which has less impact on total uncertainty.

From the calculation, the total uncertainty of GHG emission inventory from coal lifecycle in China is (-5.09%, 4.85%) in 2011, which indicates the inventory can reflect GHG emissions from coal lifecycle accurately in 2011.

5. Conclusions

In the paper, we calculate the GHG emission inventory from coal lifecycle which includes coal mining, transportation and consumption in 2011 by using technique of Monte Carlo. The main conclusions are as follows:

- 1) The total GHG emission from the coal lifecycle in 2011 is 1367186.72 ten thousand tons CO₂, of which, coal mining, transportation and consumption is 8526.67, 3328.35 and 1356723.59 ten thousand tons CO₂ respectively.
- 2) The range of GHG emission uncertainty from coal lifecycle (-5.09%, 4.85%) in 2011. The coal mining, transportation and consumption is (-43.90%, 46.63%), (-6.96%, 7.38%) and (-5.21%, 4.87%) respectively.

References

- [1] National Bureau of Statistics of the People's Republic of China (2012). China Statistical Yearbook 2011. Statistical Press of China, Beijing, China.
- [2] Shi Huading, Qi Yongqing, Liang Haichao, Gao QingXian, Fu Jiafeng (2010). The Study on Calculation Methodology of Greenhouse Gas Emissions from Electric Power [J]. Advances in Climate Change Research, Beijing, China, 01: 40-46.
- [3] Zheng Shuang (2002). The Greenhouse Gas Emissions Inventory of Coalbed Methane[J]. China Coal, Beijing, China, 05:37-40.
- [4] Coal Information Research Institute (2012). China Coal Industry Yearbook 2011. Chinese Journal of Coal Industry Institute, Beijing, China.
- [5] IPCC (2007). IPCC National Greenhouse Gas Inventories Good Practice Guidance and Uncertainty Management.
- [6] IPCC (2007). 2007 IPCC Guidelines for National Greenhouse Gas Inventories Revised.
- [7] Chinese Ministry of Railways (2011). People's Republic of China Ministry of Railways 2011 Railway Statistical Bulletin.
- [8] Ministry of Transport Planning Division (2012), 2011 Highway and Waterway Transportation Industry Statistical Bulletin.
- [9] Ministry of Transport of People's Republic of China (2014). Development Statistics Bulletin of Transport Industry . <http://www.moc.gov.cn/zfxxgk/bnssj/zghgs/2014_05/t20140513_1618277.html> (May 13, 2014).

A metric for the prognostic outreach of scenarios: Learning from the past to establish a standard in applied systems analysis

Matthias Jonas, Elena Rovenskaya, Piotr Żebrowski

International Institute for Applied Systems Analysis,
Laxenburg, Austria

M. Jonas (✉): jonas@iiasa.ac.at

Abstract

Our study concerns retrospective learning, the characteristic feature of which is that prognostic uncertainty increases the more the further we look into the future. RL seeks to establish a metric for the outreach of prognostic scenarios. The purpose behind RL is to provide an easy-to-apply indicator, which informs non-experts about the time in the future at which a prognostic scenario ceases to be in accordance (for whatever reasons) with the system's past. Ideally, this indicator should be derived concomitantly with building a prognostic model. RL concerns the limitations of predictions and prognostic scenarios.

Keywords: Greenhouse gas emissions, emission inventories, emission scenarios, diagnostic uncertainty, prognostic uncertainty, learning

1. Introduction

Evaluating the performance of climate forecasts is becoming increasingly relevant [1]. At its heart this evaluation aims at judging the credibility of climate projections and quantifying the uncertainty in these projections [2–3]. In our study, which builds on (what we term) **retrospective learning** [RL], we take the opposite view.

RL seeks to establish a metric for the outreach of prognostic scenarios. The purpose behind RL is to provide an easy-to-apply indicator, which informs non-experts about the time in the future at which a prognostic scenario ceases to be in accordance (for whatever reasons) with the system's past. Ideally, this indicator should be derived concomitantly with building a prognostic model. In brief, RL concerns the limitations of predictions and prognostic scenarios.

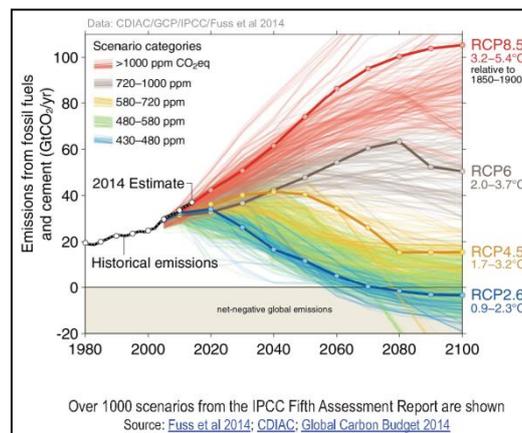


Figure 1. Historical and projected global fossil-fuel (CO₂) emissions, including emissions from cement production [4].

Figure 1 is a classical illustration of how quickly and strongly prognostic scenarios deviate from historical records. The figure shows historical and projected global CO₂ emissions resulting from fossil-fuel burning and cement production. From a purely intuitive perspective, only the highest emission scenarios appear to be in accordance with the historical record (until when?), but not the lower ones.

2. Motivation

From a theoretical point of view, we argue that the mathematical tools and techniques needed to quantify the outreach of prognostic scenarios based on learning from the past (that is, to apply RL) **are available**. However, the necessary epistemological insights to apply these tools and techniques properly, including outside their traditional context, are missing. The first statement (tools and techniques are available) is bold; while the second (knowledge to apply tools and techniques outside their traditional context is missing) is not new. Developing the first statement is subject to this paper. The second statement is at the core of empirical inference science, which is a maturing paradigm. Empirical inference science aims at complementing classical statistics in *Estimating dependencies on the basis of empirical data ... a central problem in applied analysis* [5: vii].

From a practical point of view, we argue that deriving the aforementioned indicator exhibits most interesting windfall profits: 1) We anticipate that generating the indicator while building a model will lead us onto new paths of constructing models and conducting systems analysis (i.e., towards a new standard of ‘good modeling’). 2) Our insights in RL will allow the chance of complying with—or the risk of exceeding—agreed global warming targets to be corrected. We conjecture that the risk of exceeding 2050 global warming targets ranging between 2 to 4 °C and greater is underestimated. We will return to these two issues at the end of our paper.

3. Terminology

We explain the difference between diagnostic and prognostic uncertainty, the two terms at the core of our paper, in Section 5.1 below. Their definitions will provide the basis for understanding the difference between learning in a diagnostic and prognostic context and the other terms (e.g., ‘prediction’ and ‘forecast’) that we use.

4. Status quo

Since their inception, climate treaty negotiations have set out to stabilize Earth’s climate by implementing mechanisms that reduce global greenhouse gas [GHG] emissions and lead to sustainable management of the atmosphere at a ‘safe’ steady-state level (assumed to hold for an increase in global average temperature of below 2 °C above preindustrial levels). In recent years, international climate policy has taken a step beyond achieving GHG concentration-related objectives by increasingly focusing on limiting temperature rise [6]. The idea of limiting cumulative global GHG emissions by adhering to a long-term global warming target was first discussed broadly and publicly by policymakers at the 2009 United Nations climate change conference in Copenhagen. It appears to be a promising and robust methodology [7–12] (cf. also Box 1). To comply with it, the emission reductions required from the fossil-fuel and land use/land-use change sector are daunting: 50%–85% below the 1990 global annual

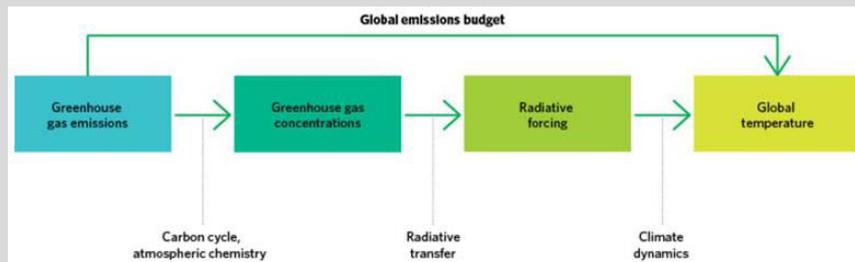
emissions, with even greater reductions for industrialized countries [13–15]. The underlying assumptions are equally daunting: terrestrial or oceanic sinks continuing to offset fossil-fuel and LUC emissions before achieving an emissions balance that goes beyond CO₂-C (i.e., CO₂-equivalents also including CH₄, N₂O, etc.), with no systemic surprises occurring during the transition process. In particular, the imperative followed for net emissions from LUC activities is that these will be reduced linearly to zero until 2050. That is, it is assumed that deforestation and other LU mismanagement will cease and that net emissions balance.

Box 1. Relationship between GHG emissions and global surface temperature [16: Fig. 3.2; 17].

The magnitude of an increase in global surface temperature is not determined by emissions in any one year, but by the concentration of GHGs in the atmosphere which, in turn, is the net outcome of total emissions and removals of GHGs to and from the atmosphere over an extended period.

Global emission budgets estimate the total amount of (net) GHG emissions that will result in a given temperature increase, within a probability range. This is why cumulative emissions (e.g., between today and 2050) are perceived as a good predictor for this temperature increase (e.g., in 2050 and beyond). That is, the emissions budget approach allows linking cumulative emissions of GHGs directly to temperature, without determining atmospheric concentrations of GHGs and their radiative forcing as intermediary observables (see figure below). The relationship between cumulative emissions and temperature is expressed as a probability, to reflect uncertainty of the climate response to a given amount of GHG emissions.

While global emission budgets identify the overall limit on global emissions, they do not prescribe the timing of peak emissions or the rate at which emissions must be reduced, so long as the overall budget is not breached. There will be a number of trajectories that could lead to the budgeted level of cumulative emissions and the related (but uncertain because trajectory-dependent) temperature increase over time. Because the emissions budget is ultimately fixed, however, delays in reducing emissions must be compensated with more rapid GHG emission reductions in future years.



In their study [15] Jonas *et al.* discuss diagnostic (retrospective: looking back in time) and prognostic (prospective: looking forward in time) uncertainty in an emissions-temperature-uncertainty [ETU] framework that allows any country to understand its national and near-term mitigation and adaptation efforts in a globally consistent and long-term context (worldwide coverage; warming range of 2–4 °C). To achieve this understanding, national linear emission target paths were established (from 1990 to 2050 or, alternatively, from 2000 to 2050) that are consistently embedded globally. In this systems context, cumulative emissions until 2050 are constrained and globally binding but are uncertain (i.e., they can be estimated only imprecisely); and whether or not compliance with an agreed temperature target in 2050 and beyond will be achieved is also uncertain. In a nutshell, the ETU framework can be used to monitor

a country's performance—past as well as prospective achievements—in complying with a future warming target in a quantified uncertainty-risk context (cf. Box 2). The authors' objective, in particular, was to understand the relevance of diagnostic and prognostic uncertainty in this global emissions-temperature setting and across temporal scales. Although the mode of bridging uncertainty across temporal scales still relies on discrete points in time ('today' and 2050) and is not yet continuous, the authors' study provides a valuable first step toward that objective.

Box 2. Output features of the ETU framework [15; adapted].

The output of the ETU framework provides national linear target paths for emissions, which are consistently embedded globally,

- for two temporal (predictor) regimes: 1990–2050 and 2000–2050;
- for CO₂ and all six Kyoto GHGs (cumulative);
- for individual spheres: technosphere and land use / land-use change;
- for three 2050 temperature targets: 2, 3 and 4 °C; and
- which allow monitoring Austria's performance—past (with and without embodied emissions) as well as projected achievements—in complying with these warming targets;

while accounting for both **diagnostic uncertainty** (which relates to the risk that true GHG emissions are greater than inventoried emission estimates reported in a specified year) and **prognostic uncertainty** (which relates to the risk that an agreed 2050 temperature target is exceeded).

5. Diagnostic versus prognostic uncertainty and learning in a diagnostic versus prognostic context

5.1 What is the difference between diagnostic and prognostic uncertainty and why do we consider them independent?

Jonas *et al.* [15] explain the difference between diagnostic and prognostic uncertainty in a temporal ('today'-versus-future) GHG emissions context:

Diagnostic uncertainty, our ability to estimate current emissions, stays with us also in the future. Assuming that compliance with an agreed emissions target is met in a target year allows prognostic uncertainty to be eliminated entirely. How this target was reached is irrelevant; only our real diagnostic capabilities of estimating emissions in the target year matter. This is how experts proceeded, e.g., when they evaluated ex ante the impact of uncertainty in the case of compliance with the Kyoto Protocol ... in 2008–2012, the Protocol's commitment period ...

Emissions accounting in a target year can involve constant, increased or decreased uncertainty compared with the start (reference) year, depending on whether or not our knowledge of emission-generating activities and emission factors becomes

more precise. The typical approach to date has been to assume that, in relative terms, our knowledge of uncertainty in the target year will be the same as it was in the start year.

However, uncertainty under a prognostic scenario always increases with time [conservative systems view]. The further we look into the future, the greater the uncertainty. This important difference suggests that diagnostic and prognostic uncertainty are independent. This differs from how prognostic modelers usually argue. A prevalent approach is to realize a number of scenarios and grasp prognostic uncertainty by means of the spread in these scenarios over time—which increases with increasing uncertainty in the starting conditions built into their models. However, this approach nullifies diagnostic uncertainty once a target (future) is reached.

The notion of a conservative systems view is central to RL, meaning that a system cannot exhibit surprises in the future that it has not experienced during its ‘one-reality’ past.

This difference between diagnostic and prognostic uncertainty is not only theoretical. It becomes relevant in the next section.

5.2 What do we understand by learning in a diagnostic and prognostic context?

Learning under diagnostic conditions requires the ‘measuring’ of differences or deviations. Here we follow Marland *et al.* [18], who discuss this issue in the context of emissions accounting and uncertainty:

Estimates of uncertainty have traditionally been expert judgments based on the data input to the calculations. But for CO₂ emissions from fossil fuels, there are actually at least four approaches that one can take to gain some insight into the full uncertainty of emissions estimates: comparison of estimates made by independent methods, comparison of estimates from multiple sources, evolution over time of estimates from a single source, and, soon (we hope), modeling against remotely sensed data.

With respect to the evolution of estimates over time (3rd approach), the authors state:

Many of the countries and organizations that make estimates of CO₂ emissions provide annual updates in which they add another year of data to the time series and revise the estimates for earlier years. Revisions may reflect revised or more complete energy data and more complete and detailed understanding of the emissions processes and emissions coefficients. In short, we expect revisions to reflect learning and a convergence toward more complete and accurate estimates.

Retrospective learning, in turn, is about the limitations of looking (**projecting**) into the future and may be best explained in contrast to retrospective forecasting. Retrospective forecasting strives for the most appropriate (best) forecast by minimizing the difference between forecast (**prediction**) and actual outcome, while the characteristics of the data record—here quantified by its dynamics and diagnostic uncertainty (random error)—are assumed **not** to change when inter- or extrapolating the historical data record. By way of contrast, retrospective learning seeks to capture the characteristic feature of prognostic uncertainty, namely, that **prognostic uncertainty increases the more the further we look into the future**,¹ while it is

¹ As a matter of fact, the confidence band of an (e.g.) linear regression also increases, but for mathematical rather than physical reasons, and it does so backward **and** forward in time.

assumed in the context of this proposal that the data record's memory is contained (as above) in its dynamics, not in its uncertainty.

Figure 2 attempts to visualize the fundamental difference between prediction and an advanced mode of learning from the past, the latter allowing the increase in prognostic uncertainty with time to be grasped. The mode of RL that we intend to explore builds on representation of the available data by way of two components: i) a Taylor (or equivalent) polynomial which captures the signal's predominant (lower-order) dynamics (**learning phase 1**); and ii) a linearly increasing 'uncertainty (learning) wedge',² which comprises the signal's higher-order dynamics and the uncertainty underlying the signal—or only the data record's higher-order dynamics if the data record is accurate and precise (**learning phase 2**). We expect this two-component split into lower-order dynamics and uncertainty wedge to be systems-dependent and unsharp, the latter resulting from uncertainty. In a nutshell, Figure 2 indicates that we seek to balance three things: the 'right' order of the dynamics and both the 'right' extension and the 'right' opening of the uncertainty wedge. It is this balance that must hold during the **testing phase**. The historical data held back for this phase have not been used before, that is, during learning phases 1 and 2, which is why we refer to this part of the data record as "historical future".

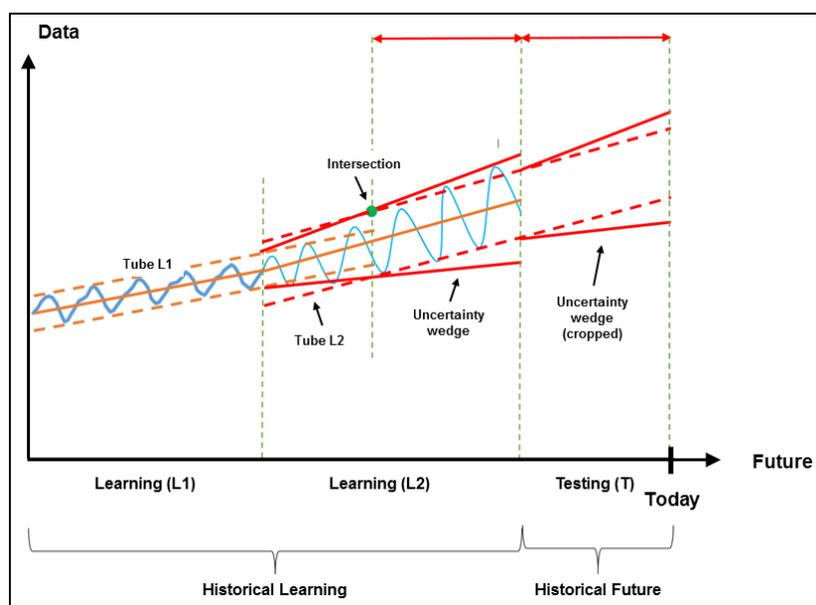


Figure 2. Illustrating the different steps of RL with the help of a simple (periodic, increasing, and periodically increasing) function.

6. Methodology—just one approach

Assume the following situation, namely, that we have more than one historical data record available, each **accurate and precise** (which can be easily relaxed to 'accurate and imprecise'), and that we have learned from the past (i.e., from an RL exercise):

- that each historical data record exhibits (but not necessarily) a linear dynamics;

² "Linear" meaning linear relative to the dynamics, which is why we also speak of **linear RL** (sufficient in the context of this study).

- that each data record's uncertainty (learning) wedge unfolds linearly into the future (up to what time in the future, however, is still unknown); and
- that our historical data records exhibit linear interdependencies. That is, in the case of an emissions-concentration-temperature [E-C-T] system, we mean linear interdependencies of a serial sort (fully sufficient in the context of this study): $T = T(C)$, $C = C(E)$, and $E = E(t)$; with T denoting global surface temperature, C atmospheric CO₂ concentration, E CO₂ emissions into the atmosphere, and t time. As a matter of fact, as individual time series these are exponential (posing no difficulties to treating their interdependencies in a similar way).

To facilitate understanding the philosophy behind the methodology, we consider two cases:

6.1 Serial interdependence $E \rightarrow C \rightarrow T$

Starting from $E = E(t)$, i.e.

$$E = m_{Et} t \quad [E] = \frac{PgC}{y}; [m_{Et}] = \frac{PgC}{y^2}; [t] = y$$

with m denoting the signal's (here) linear dynamics and Et indicating that we are in the E-t plane; and

$$E_u = f_{Et,u} m_{Et} t \quad [f_{Et,u}] = 1$$

$$E_l = f_{Et,l} m_{Et} t \quad [f_{Et,l}] = 1$$

with the constants $f_{Et,u}$ and $f_{Et,l}$ indicating the upper [u] and lower [l] borders of the uncertainty wedge. The difference between upper and lower border at any time is given by $\Delta E = \Delta f_{Et} m_{Et} t$.

On the other hand, the difference between upper and lower border can be perceived as error in E, which suggests that use is made of the law of error propagation:

$$\sigma_E^2 = \left(\frac{\partial E}{\partial m_{Et}} \right)^2 \sigma_{m_{Et}}^2 + \left(\frac{\partial E}{\partial t} \right)^2 \sigma_t^2 .$$

Assuming time to be known exactly (i.e., $\sigma_t = 0$):

$$\sigma_E = \sigma_{m_{Et}} t ; \quad [\sigma_E] = \frac{PgC}{y}; [\sigma_{m_{Et}}] = \frac{PgC}{y^2}$$

that is, the error in E is given by the error in the slope m_{Et} , the signal's dynamics. Alternatively:

$$\frac{\sigma_E}{E} = \frac{\sigma_{m_{Et}}}{m_{Et}} .$$

Requesting $\Delta E := 2\sigma_E$, one finds via comparison

$$\Delta f_{Et} m_{Et} t = 2\sigma_{m_{Et}} t \quad \text{or} \quad \Delta f_{Et} = \frac{\Delta E}{E} = 2 \frac{\sigma_E}{E} = 2 \frac{\sigma_{m_{Et}}}{m_{Et}} .$$

In a nutshell, an accurate-precise system has been merged with classical statistics, meaning (here) that we grasp the historical future of our data record with the help of a straight line, the slope of which is uncertain. Another point warranting attention is that the law of error propagation is approximate and can only be applied under conditions that guarantee the validity of partial derivatives. In particular, if $\Delta f_{Et} = \Delta f_{Et}(t)$, these conditions could be violated quickly with increasing t .

One can proceed similarly for $C = C(t)$, i.e.,

$$C = m_{Ct} t \qquad [C] = \text{ppmv}; [m_{Ct}] = \frac{\text{ppmv}}{y}$$

... (here not repeated). Alternatively, instead of analyzing $E = E(t)$ and $C = C(t)$ individually, one can also look at the linearly interdependent case $C = C(E)$, i.e.,

$$C = m_{CE} E = m_{CE} m_{Et} t = m_{Ct} t ; \qquad m_{Ct} = m_{CE} m_{Et} ; [m_{CE}] = \text{ppmv} \frac{y}{\text{PgC}}$$

or, to generalize further, at the linearly interdependent case $T = T(C) = T(C(E))$, i.e.,

$$\begin{aligned} T &= m_{TC} C = m_{TC} m_{CE} E \\ &= m_{TC} m_{CE} m_{Et} t = m_{Tt} t . \end{aligned} \qquad m_{Tt} = m_{TC} m_{CE} m_{Et} ; [m_{TC}] = \frac{^{\circ}\text{C}}{\text{ppmv}} ; [m_{Tt}] = \frac{^{\circ}\text{C}}{y}$$

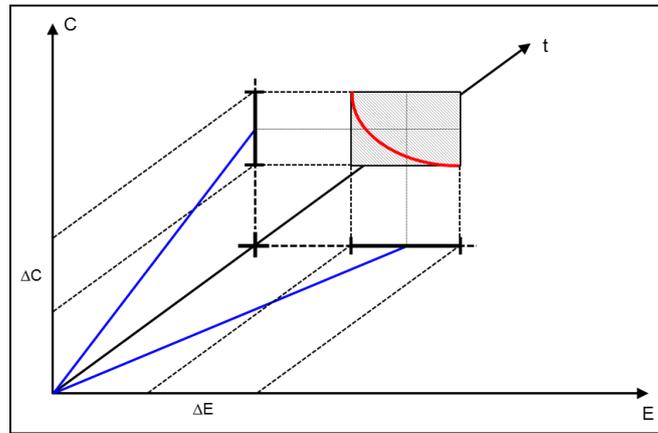


Figure 3. Graphical illustration of learning in the C-E space: independent versus linearly interdependent case. In the latter case, learning does **not** happen in a space which is spanned by a 2-dimensional square ($\Delta E \times \Delta C$), but along a 1-dimensional space (red curve) belonging to a curved uncertainty wedge.

The above cases can be conveniently summarized:

$$\Delta E = \Delta f_{Et} m_{Et} t = \Delta f_{Et} E$$

$$\Delta C = \Delta f_{Ct} m_{Ct} t = \Delta f_{CT} C = \sqrt{\Delta f_{CE}^2 + \Delta f_{Et}^2} C$$

$$\Delta T = \Delta f_{Tt} m_{Tt} t = \Delta f_{Tt} T = \sqrt{\Delta f_{TC}^2 + \Delta f_{CE}^2 + \Delta f_{Et}^2} T$$

... and so on. The interpretation is as follows: From analyzing (e.g.) the second equation, which allows $\Delta f_{Ct} = \sqrt{\Delta f_{CE}^2 + \Delta f_{Et}^2}$ to be extracted, it becomes obvious that the learning on the left side (Δf_{Ct}) is determined by the learning on the right side ($\sqrt{\Delta f_{CE}^2 + \Delta f_{Et}^2}$).

The resulting equation describes a second-order cone:

$$\Delta f_{CE}^2 + \Delta f_{Et}^2 - \Delta f_{Ct}^2 = 0 \quad \leftrightarrow \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0 \quad (\text{cf. also Fig. 3}).$$

The basic idea behind the above procedure is to grasp the learning (Δf -terms) with the help of the error-propagation approach, the mathematics of which is well-established and easy to apply (even concomitantly with building a prognostic model). In the case of linearly interdependent variables (here $C = C(E)$), the learning does **not** happen in a space which is spanned by a (here) 2-dimensional square ($\Delta E \times \Delta C$), but along a 1-dimensional curve belonging to a curved uncertainty wedge. It appears that this reduction to the 1-dimensional space is also preserved in the case of more than two linearly interdependent variables. But it would be premature to praise this as a major step forward in reducing uncertainty. We still do **not** have any knowledge on the outreach of the curved uncertainty wedge (which needs to be determined as indicated in Fig. 2).

6.2 Serial-parallel interdependence

$$\begin{array}{l} \mathbf{E}_1 \rightarrow \mathbf{C}_1 \\ \mathbf{E}_2 \rightarrow \mathbf{C}_2 \end{array} \rightarrow \mathbf{T}$$

Here, we do not derive the analytical expression for Δf_{Tt} which describes the learning. Deriving this expression is easy and straightforward. In contrast, another insight is much more important, namely, the analytical expression for Δf_{Tt} also holds for a system, where the second emissions source (E_2) has been replaced by a sink (R:

removal): $\begin{array}{l} \mathbf{E} \rightarrow \mathbf{C}_1 \\ \mathbf{R} \rightarrow \mathbf{C}_2 \end{array} \rightarrow \mathbf{T}$; meaning that the learning does not change while the two

systems differ: $C = C_1 + C_2$ versus $C = C_1 - C_2$. That is, a sink reduces a source but their uncertainties still add up.

It is this game changer that has not so far been considered by prognostic modelers: a shortfall with far-reaching consequences, notably, when determining the risk of exceeding an agreed global temperature target in the future.

7. Summary and preliminary outlook

The purpose of our paper is to present a particular methodology to tackle retrospective learning, the characteristic feature of which is that prognostic uncertainty increases the more the further we look into the future. Alternative methodologies are conceivable. We currently consider the discussion of necessary assumptions and, if

need be, simplifying assumptions (still) more important than immersing ourselves in numerical exercises.

So far, we see two important consequences emerging:

- The objective (i) to generate a metric / indicator to inform non-experts about the limitations of the predictive outreach of a prognostic scenario; and (ii) to demonstrate that this metric / indicator can be generated even concomitantly with building a prognostic model is within reach. We conjecture that the latter, in particular, will lead us, in the case of success, onto new paths of constructing models and conducting systems analysis—that is, towards a new standard of ‘good modeling’.
- RL informs us that, from an uncertainty perspective, emission sources and sinks need to be separated—which is not done in estimating the risk of exceeding an agreed global warming target in 2050. This very risk can be determined by using multi-model emission scenarios like those in Figure 1 in connection with emission-climate change models (where “climate change” is quantified by changes in global surface temperature). The cumulative emissions of these scenarios are used as a predictor for the expected global temperature increase in the future (cf. Box 1). However, the crux of this exercise is that it starts—erroneously—from net emissions. (Take Fig. 1 above, for example: removals eventually outpace emissions and net emissions even become negative.) From an uncertainty perspective, preferring **net emissions** to **emissions minus removals** runs counter to the law of error propagation which informs us that **a sink reduces a source but their uncertainties still add up**. This shortfall has far-reaching consequences. The correct approach would have been to deal with cumulated emissions and removals **individually** to determine their combined risk of exceeding the agreed temperature target. RL allows exactly this to be done: RL overcomes this shortfall and allows the effect of learning about emissions and removals individually to be grasped.

This is why we argue that understanding and grasping RL is of fundamental and global relevance.

References

- [1] Challinor, A.J., and C. Ferro (eds.) (2015). Managing uncertainty in predictions of climate and its impacts. *Clim. Change*, **132**(1–2), 367.
- [2] Otto, F.E.L., C.A.T. Ferro, T.E. Fricker and E.B. Suckling (2015). On judging the credibility of climate predictions. *Clim. Change*, **132**(1–2), 47–60, doi 10.1007/s10584-013-0813-5.
- [3] Lopez, A., E.B. Suckling, F.E.L. Otto, A. Lorenz, D. Rowlands and M.R. Allen (2015). Towards a typology for constrained climate model forecasts. *Clim. Change*, **132**(1–2), 15–29, doi 10.1007/s10584-014-1292-z.
- [4] GCP (2014). Global Carbon Budget 2014. PDF presentation file made available by the Global Carbon Project, Canberra, Australia, http://www.globalcarbonproject.org/carbonbudget/14/files/GCP_budget_2014_lowres_v1.02.pdf (accessed 08 September 2015).

- [5] Vapnik, V. (2006). Estimation of Dependencies Based on Empirical Data. Empirical Inference Science. Springer-Science+Business Media, New York, NY, USA, p. 98.
- [6] Rogelj, J., W. Hare, J. Lowe, D.P. van Vuuren, K. Riahi, B. Matthews, T. Hanaoka, K. Jiang, and M. Meinshausen (2011). Emission pathways consistent with a 2 °C global temperature limit. *Nature Clim. Change*, **1**, 413–418, doi 10.1038/nclimate1258.
- [7] Allen, M.R., D.J. Frame, C. Huntingford, C.D. Jones, J.A. Lowe, M. Meinshausen and N. Meinshausen (2009). Warming caused by cumulative carbon emissions toward the trillionth tonne. *Nature*, **458**(7242), 1163–1166, doi: 10.1038/nature08019.
- [8] Matthews, H.D., N.P. Gillet, P.A. Stott and K. Zickfeld, 2009: The proportionality of global warming to cumulative carbon emissions. *Nature*, **459**(7248), 829–833, doi 10.1038/nature08047.
- [9] Meinshausen, M., N. Meinshausen, W. Hare, S.C.B. Raper, K. Frieler, R. Knutti, D.J. Frame and M.R. Allen (2009). Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature*, **458**(7242), 1158–1162, doi: 10.1038/nature08017.
- [10] WBGU, 2009: Solving the climate dilemma: The budget approach. Special Report, German Advisory Council on Global Change (WBGU), Berlin, Germany [ISBN: 978-3-936191-27-1], p. 55. Available at: http://www.wbgu.de/fileadmin/templates/dateien/veroeffentlichungen/sondergutachten/sn2009/wbgu_sn_2009_en.pdf (accessed 26 January 2015).
- [11] Zickfeld, K., M. Eby, H.D. Matthews and A.J. Weaver(2009). Setting cumulative emissions targets to reduce the risk of dangerous climate change. *PNAS*, **106**(38), 16129–16134, doi: 10.1073/pnas.0805800106.
- [12] Raupach, M.R., J.G. Canadell, P. Ciais, P. Friedlingstein, P.J. Rayner and C.M. Trudinger (2011). The relationship between peak warming and cumulative CO₂ emissions, and its use to quantify vulnerabilities in the carbon-climate-human system. *Tellus*, **63B**(2): 145–164, doi 10.1111/j.1600-0889.2010.00521.x.
- [13] Fisher, B.S., and N. Nakicenovic (lead authors) (2007). Issues related to mitigation in the long-term context. In: *Climate Change 2007: Mitigation of Climate Change* [B. Metz, O. Davidson, P. Bosch, R. Dave and L. Meyer (eds.)]. Cambridge University Press, Cambridge, UK, 169–250.
- [14] Jonas, M., G. Marland, W. Winiwarter, T. White, Z. Nahorski, R. Bun and S. Nilsson (2010). Benefits of dealing with uncertainty in greenhouse gas inventories: Introduction. *Clim. Change*, **103**(1–2), 3–18, doi: 10.1007/s10584-010-9922-6.
- [15] Jonas, M., G. Marland, V. Krey, F. Wagner and Z. Nahorski (2014). Uncertainty in an emissions-constrained world. *Clim. Change*, **124**(3), 459–476, doi: 10.1007/s10584-014-1103-6.
- [16] Climate Change Authority (2012). Targets and Progress Review. Australian Government, <http://www.climatechangeauthority.gov.au/reviews/targets-and-progress-review-3> (accessed 10 September 2015).
- [17] Raupach, M.R., I.N. Harman and J.G. Canadell (2011). Global climate goals for temperature, concentrations, emissions and cumulative emissions. CAWCR Technical Report No. 042, Centre for Australian Weather and Climate Research,

Melbourne, Australia, http://www.cawcr.gov.au/publications/technicalreports/CTR_042.pdf (accessed 10 September 2015).

- [18] Marland, G., K. Hamal and M. Jonas (2009). How uncertain are estimates of CO₂ emissions? *J. Ind. Ecol.*, **13**(1), 4–7, doi: 10.1111/j.1530-9290.2009.00108.x.

Acronyms

C	concentrations
E	emissions
E-C-T	emissions-concentration-temperature
ETU	emissions-temperature-uncertainty
GHG	greenhouse gas
l	lower
RL	retrospective learning
T	temperature
u	upper

Assessing the improvement of greenhouse gases inventories: can we capture diagnostic learning?

Piotr Żebrowski, Matthias Jonas, Elena Rovenskaya

International Institute for Applied Systems Analysis, Advanced Systems Analysis Program,
Laxenburg, Austria
Piotr.Zebrowski@iiasa.ac.at

Abstract

Our study aims at modelling the diagnostic learning understood as a gradual improvement of the quality of greenhouse inventories. We quantify this improvement by the speed of convergence of consecutive revisions of emissions estimates to the most recently published ones, which we assume to be close to the true emissions values. On the example of Austria's National Inventory Reports we show that the diagnostic learning process exhibits exponential dynamics.

Keywords: Greenhouse gas emissions, revisions of emissions inventories, learning, uncertainty of emissions estimates

1. Introduction

Signatories to the United Nations Framework Convention on Climate Change (UNFCCC) are obliged to submit annual inventories of their greenhouse gas (GHG) emissions, together with revisions of historical emission estimates. Previous estimates are recalculated whenever errors in inventories were identified and corrected, new data sources were taken into account or new accounting methodologies were employed. Therefore, consecutive revisions of emissions estimates from previous years are thought to reflect the advancement of knowledge in constructing GHG inventories, while the most recent estimates are considered to be accurate. But can we detect this learning process in the historical data of GHG emissions reported to the UNFCCC?

In this work we show on the example of Austria's GHG inventories that indeed we can observe and model the improvement of GHG emission inventories.

2. Diagnostic learning

By diagnostic learning we call the process of gradual improvements in the quality of GHG emissions inventories. This "improvement of quality" we understand as the advancement of knowledge, which is reflected by the increase of accuracy (reduction of bias) and/or precision (reduction of standard deviation) of emissions estimates.

Several attempts have been made to grasp the diagnostic learning in a quantitative way. In Hamal's work [1] the notion of total uncertainty was used. The total uncertainty combines inaccuracy (in relative terms) between the most recent and the most initial estimates of emissions, with the imprecisions of these estimates (lack of precision). Another approach is to analyze the convergence of sequences of estimates reported in the consecutive National Inventory Reports as has been done in Nahorski *et. al.* [2].

In Marland *et. al.* [3] learning was understood as a convergence of revised estimates of emissions towards the more accurate ones. In our work we follow this line of thinking and investigate whether the consecutive revisions of estimates stabilize around certain

level (presumed to represent the accurate estimate), and if so, how fast this stabilization level is reached.

A common feature of the approaches presented in [1] and [2] is that is that both methods strived to capture the learning process from one report to another in the uniform way (ensemble approach). However, if one do not consider revised estimates of emissions occurred in different years separately, then the learning process one tries to describe is “contaminated” by structural changes in emissions in different years.

In contrast, our method grasp “pure” learning (unaffected by structural changes) because we model learning process in consecutive revisions of emissions estimates for only one fixed year of emissions at a time. But before we explain our approach to diagnostic learning in detail we first describe the data set we are analysing.

3. The data

We chose to present our method on the case example of Austria as for this country the temporal evolution of revised CO₂ emissions estimates is well pronounced. The data set we are working on has been compiled from the Austria’s National Inventory Reports (NIRs) submitted to the UNFCCC in the years 2003-2014.

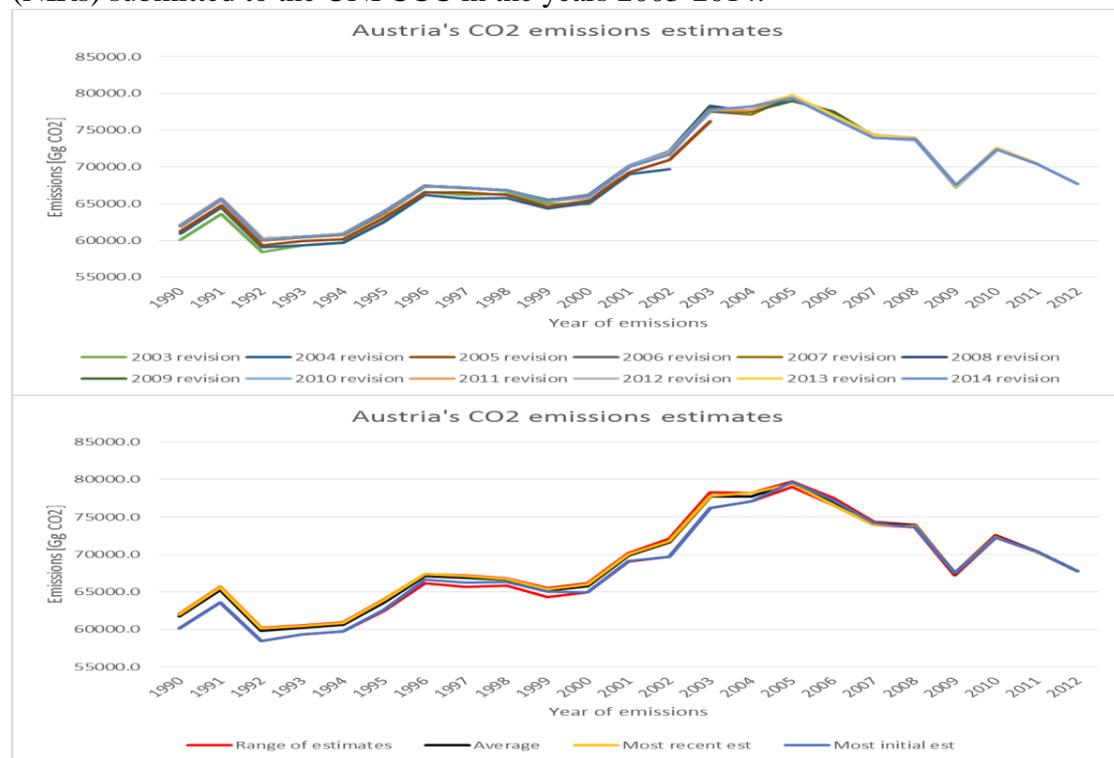


Figure 1. Revisions of the Austria’s CO₂ emissions estimates (top panel) and time evolution of ranges of Austria’s CO₂ emissions estimates (bottom panel).

It is important to note that our data set may be naturally divided into two parts. Part I contains revised estimates of emissions in the period 1990-2001. It may be organized into 11 sequences, with first one containing revised estimates of emissions occurred in 1990, second one contains estimates of emissions for 1991 and so on. Each of these sequences consists of estimates published in the years 2003-2014 and thus all of them have the equal length of 12. As a consequence, all most initial estimates in this part comes from the NIR published in 2003 (see Figure 2.). Part II containing the rest of the data (estimates of emissions occurred in the period 2002-2012) can be organized into

another 11 sequences in the same way as the part I. The only difference is that all these sequences are of different lengths. The first one, containing estimates of emissions in year 2002, is of the length 11 and the last one has only one element, that is the only available estimate of emissions in year 2012 published in the year 2014.

Figure 1. shows that revised emissions estimates differ slightly revision-wise but clearly follow the emissions path published most recently. However, for each year of emission the revised estimates may behave erratic and in general do not approach the most recent one (assumed to be correct) in the strictly monotonic way. It is also difficult to compare in absolute terms the changes in estimates that occurred in different years. Thus, in order to see a clearer picture, we assume the most recent estimates as a reference level and work with relative differences of estimates normalized estimates (i.e., values of estimates divided by the most recent ones).

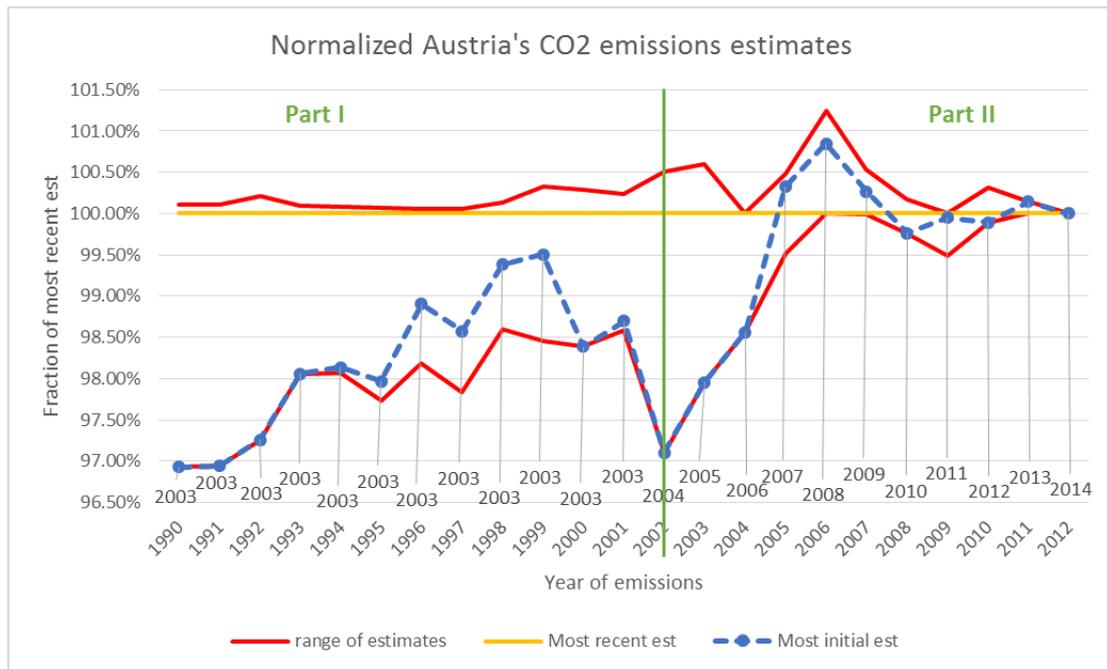


Figure 2. Austria’s CO₂ emissions estimates normalized by the most recent estimates. For the most initial estimates line the year of estimate’s publications is marked.

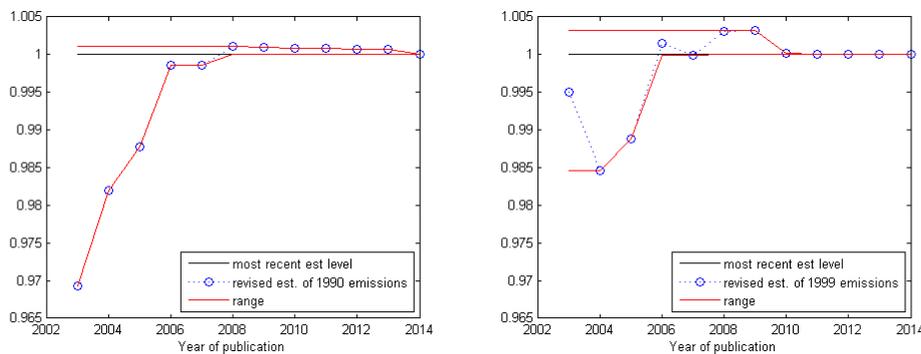


Figure 3. Time evolution of revised estimates of Austria’s CO₂ emissions in the year 1990 (left panel) and 1999 (right panel).

Figure 2. presents the data after normalization transformation and should be interpreted as follows: for example, initial estimate of the emissions in 1999 were lower

than the most recent one by 0.5% of the most recent estimate, and the range of all revisions spans from 98.5% to 100.3%. How these estimates change from revision to revision is shown on the Figure 3. (left).

After normalization of estimates we clearly see that typically the most initial estimates tend to underestimate emissions, but their subsequent corrections are not always in the direction of the most recent estimate. The range of revised estimates of emissions in a given year carries more information about the variability of these estimates and is a good proxy for the uncertainty of these estimates. Hence, we argue that analysis of the ranges of emissions estimates suits well the purpose of grasping diagnostic learning. Our methodology, which we describe in the following section, is based on this observation.

4. Model of diagnostic learning

4.1. Notation

Let $E_{n,y}$ denotes estimate of emissions in year n revised in year y , with $n = \{1990, \dots, 2012\}$ and $y = \{n+2, \dots, Y\}$, where $Y = 2014$ is the year of the last revision ($y \geq n+2$ reflects the fact that inventories are published with 2-year delay). We define

$$m_{n,y} = \min\{E_{n,y}, \dots, E_{n,Y}\} \text{ and } M_{n,y} = \max\{E_{n,y}, \dots, E_{n,Y}\}. \quad (1)$$

Then $m_{n,y}$ denotes the smallest of the estimates of emissions in year n that were published in years between y and Y . Similarly $M_{n,y}$ denotes the biggest of these estimates. How $m_{n,y}$ and $M_{n,y}$ (normalized by $E_{n,Y}$) change as $y \rightarrow Y$ can be seen on the Figure 3. (The lower and upper red lines on the left panel represents evolution of

$m_{1990,y}/E_{1990,2014}$ and $M_{1990,y}/E_{1990,2014}$ for y changing from 2003 to 2014. Similarly, the lower and upper red lines on the right panel correspond to $m_{1999,y}/E_{1999,2014}$ and $M_{1999,y}/E_{1999,2014}$, accordingly.)

4.2. Formal approach to diagnostic learning

As mentioned in section 2. we understand diagnostic learning as a convergence

$$E_{n,y} \rightarrow E_n \text{ as } y \rightarrow \infty, \quad (2)$$

where E_n is a true but unknown value of emissions that occurred in the year n . In practice, for each year of emissions n we observe only a few initial elements of sequence $E_{n,y}$ (at most 12 for the Part I of the data). However, if convergence (2) holds and the most recent estimate $E_{n,Y}$ is close to the true value E_n then we observe that revisions $E_{n,y}$ stabilize around level $E_{n,Y}$. Ten necessarily also $m_{n,y} \rightarrow E_{n,Y}$ and $M_{n,y} \rightarrow E_{n,Y}$ as $y \rightarrow Y$. Both these effects can be seen of Figure 3.

Let us fix a year of emissions n . As revised estimates $E_{n,y}$ may oscillate around level $E_{n,Y}$, the learning process is more apparent in the evolution of $m_{n,y}$ and $M_{n,y}$ since they converge monotonically to $E_{n,Y}$. The speed of this convergence may be interpreted as the rate of suppression of oscillations of estimates $E_{n,y}$ and the decrease of the difference $M_{n,y} - m_{n,y}$ may be regarded as the decrease of uncertainty.

Figure 2. reveals that the upper ranges of emission estimates $M_{n,y}$ do not vary much between different years of emissions n and in general are only slightly higher (typically by less than 0.5%) than the most recent estimates. Therefore it is the evolution of lower ends of the estimates ranges that reflects the diagnostic learning.

4.3. Modeling diagnostic learning in revisions of emission estimates for each year of emissions separately.

Figure 3. suggests that for each fixed year of emissions n lower end of estimate's range $m_{n,y}$ approaches level $E_{n,Y}$ more rapidly at the beginning and gradually slows down and stabilizes as $y \rightarrow Y$. Therefore, for each fixed n , it is natural to choose the exponential dynamics as a model of time evolution of $m_{n,y}$ as $y \rightarrow Y$, namely

$$\frac{m_{n,y}}{E_{n,Y}} = 1 - c_n e^{-\lambda_n (y-n-2)} \quad (3)$$

for $y = \{y_0, \dots, Y\}$, where $y_0 \geq n + 2$ is the year of publication of the most initial available estimate of emissions in the year n . The parameter λ_n is interpreted as the learning rate in the period between y_0 and Y .

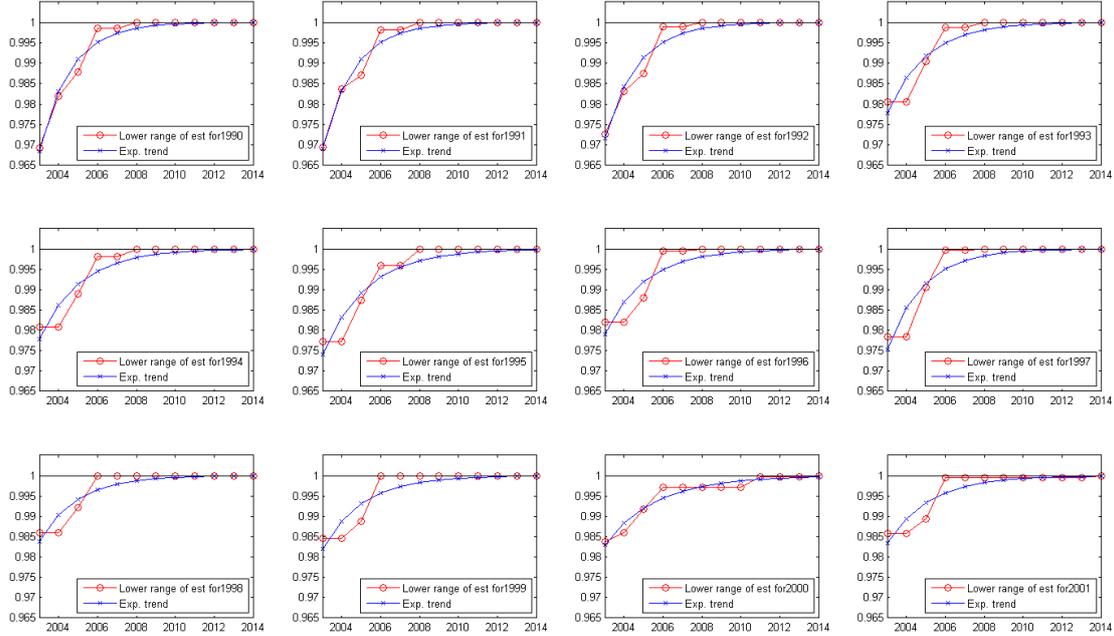


Figure 4. Exponential trends of evolution of lower ends of emissions estimates ranges.

Table 1. Values of learning obtained via formula (3)

Year of emissions n	1990	1991	1992	1993	1994	1995
Learning rate λ_n	0.6251	0.6210	0.5954	0.5043	0.4722	0.4425
Year of emissions n	1996	1997	1998	1999	2000	2001
Learning rate λ_n	0.4815	0.5405	0.5147	0.4786	0.3761	0.4497

We apply the model (3) to each sequence in the Part I of the data (that is to revisions of emissions for years $n = 1990, \dots, 2001$, that were published between years $y_0=2003$ and $Y = 2014$). The reason for this choice is two-fold. Firstly, all sequences in Part I are of equal length which ensures that the of the trend fit is comparable for all considered

samples. Second reason is that all these sequences reflects learning process in the same period, namely the years 2003-2014. On Figure 4. we present the trends in learning obtained from model (3) for the years of emissions covered by the Part I of the data, while Table 1. contains the corresponding learning rates.

4.4. Modeling diagnostic learning revision-wise

In the previous section we applied model (3) to grasp learning in the revised emissions estimates for each one year of emissions covered by the Part I of the data at a time. However, if our model of diagnostic learning is a correct one, we should be able to observe exponential trend in overall improvement of inventories from one revision to another. We perform such consistency check using the Part II of the data.

We suspect that the structural changes in emissions cause only minor differences between learning rates λ_n for different years of emissions n (see Hamal [1]). Therefore it is reasonable to assume that the learning process from revision to revision is uniform for estimates of emissions across all covered years of emissions n .

We calculate the average $\bar{\lambda} = 0.5085$ of all learning rates λ_n given in the Table 1. and interpret it as the approximate uniform learning rate of this overall improvement of all CO₂ inventories published in the period 2003-2014.

Provided this hypothesis is true we should then be able to use our model (3) to grasp diagnostic learning in the Part II of the data. Observe that the lower red line in the Part II of the Figure 2. corresponds to normalized lower ends of the ranges of emissions estimates $m_{n,n+2} / E_{n,Y}$ for the years of emission covered by Part II of the data ($n = 2002, \dots, 2012$). Now, if all $m_{n,y} / E_{n,Y}$ behave uniformly for all n as $y \rightarrow Y$, then model (3) and our assumptions yield that

$$\frac{m_{n,n+2}}{E_{n,Y}} = 1 - \bar{c} e^{-\bar{\lambda}(n-n_0)} \quad (4)$$

for all $n = \{n_0 = 2002, \dots, 2014\}$, where parameter $\bar{\lambda}$ is the uniform learning rate describing overall improvement of emissions inventories revision-wise. Thus, if our model is a correct one, value of $\bar{\lambda}$ calculated as the average of learning rates λ_n given in the Table 1. and the value of $\bar{\lambda}$ obtained directly from the Part II of the data via model (4) should match. As Table 2. shows, this indeed is the case.

Table 2. Two independent estimates of the uniform learning rate $\bar{\lambda}$

$\bar{\lambda}$ as the average of λ_n from Table 1.	0.5085
$\bar{\lambda}$ derived from equation (4)	0.4948

This close agreement of independent estimates of uniform learning rate $\bar{\lambda}$ strongly indicates that the model of diagnostic learning presented above is correct.

5. Further research plans.

When deriving equation (4) we assumed that the structural changes have negligible influence on how the learning rates λ_n changes with n . In reality averaging the learning rates over the years of emissions n both explicitly or implicitly (as we have done in case of estimation of $\bar{\lambda}$ with use of equation (4)) is susceptible to structural changes. In future we plan to factor structural changes into the methodology presented above.

References

- [1] Hamal, K. (2010). Reporting GHG Emissions: Change in Uncertainty and Its Relevance for Detection of Emission Changes. Interim Report IR-10-003, International Institute for Applied Systems Analysis, Laxenburg
- [2] Nahorski, Z., Jarnicka, J. Modeling Uncertainty Structure of Greenhouse Gases Inventories, unpublished manuscript.
- [3] Marland, G., Hamal, K., Jonas, M., (2009). How Uncertain Are Estimates of CO₂ Emissions? *Journal of Industrial Ecology*, 13(1) pp. 4-7.

A method for estimating time evolution of precision and accuracy of greenhouse gases inventories from revised reports

Jolanta Jarnicka¹, Zbigniew Nahorski¹

¹Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Jolanta.Jarnicka@ibspan.waw.pl, Zbigniew.Nahorski@ibspan.waw.pl

Abstract

In the paper a statistical method for estimating the evolution of GHG inventories is proposed. For estimation the revisions of inventories published in consecutive years are used. Data from the National Inventory Reports up to 2007, and then up to 2014 are analyzed. A parametric model and a procedure for estimating parameters are described, and examples of their applications are presented. As a result, statistically significant trajectories of standard deviations are obtained and clear improvement of inventory accuracy in time is observed.

Keywords: greenhouse gases inventories, uncertainty, modeling

1. Introduction

According to the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol, each of the cosignatories is obliged to provide annual data on greenhouse gas inventory. Each report contains data from a given year and revisions of past data, whenever required, but it also has to deal with uncertainties. Data for previous years are revised when more precise information is obtained. This means that revisions made in different years use different knowledge, and hence uncertainties in different revisions are incomparable. The question therefore arises, whether it is possible to compare and organize data on GHG emission, to get as much information about the unknown uncertainty as possible. Discussion on that problem can be found e.g. in [1], [2], [3], and many others.

The goal of that paper is an attempt to find an answer to that question, by proposing an alternative method of uncertainty assessment. This is done by analyzing how the uncertainty of the inventory reports changes over the consecutive yearly revisions. Based on the data from the National Inventory Reports up to 2007 and then the data up to 2014, in view of revisions of inventories published in consecutive years, we observed significant improvement of inventory accuracies in time. The results obtained indicate a clear learning effect in inventory calculation.

In Section 2 we present the idea of interpreting the data and propose a parametric model, that describes the uncertainty structure in the inventory reports. Section 3 contains the results of fitting the model to data from the National Inventory Reports for Austria. Conclusions are given in Section 4.

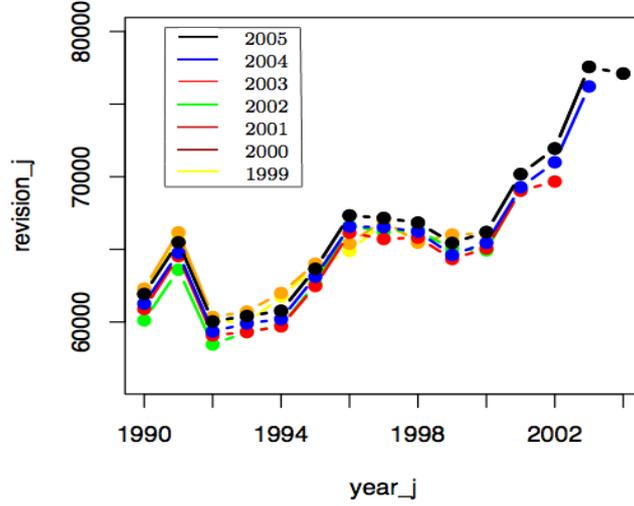
2. Presentation of the model

We analyze data from the national inventory reports. Let E^n – denote the inventory data for the country i , in the year n revised in the year y , and let Y – denote the last year, when the revision is made. For a given country i all the inventory data form a table, in which each row contains consecutive revisions of the data for a given year (Table 1).

Table 1. Indexing the data

$$\begin{array}{cccccccc}
 \vdots & \vdots \\
 \dots & E_{y,i}^{n-1} & E_{y,i}^n & E_{y,i}^{n+1} & \dots & E_{y,i}^y & 0 & \dots & 0 \\
 \vdots & \vdots \\
 \dots & E_{Y,i}^{n-1} & E_{Y,i}^n & E_{Y,i}^{n+1} & \dots & E_{Y,i}^y & E_{Y,i}^{y+1} & \dots & E_{Y,i}^Y
 \end{array}$$

We use the fact that, each revision data, for a given country, forms a realization of a stochastic process. These stochastic processes for a fixed country are different, but related. They form a bunch of stochastic processes. An example is given in Figure 1, presenting data from the National Inventory Reports for Austria. The data refers to CO₂ emissions in the years 1990-2005 (i.e. reported up to 2007), and revisions performed every year, from 1999-2005.


Figure 1. Revisions of the National Inventory Reports data on CO₂ emissions, 1999-2005, Austria.

For a given country i , we model any revision data to be composed of the “real” emission, which we call the “deterministic” fraction and a “stochastic” fraction, related to our lack of knowledge and imprecision of observation of the real emission. We assume that the uncertainty is related to the stochastic part of the model.

$$E_{Y,i}^n = D_{Y,i}^n + S_{Y,i}^n, \quad S_{Y,i}^n \sim \mathcal{N}(0, \sigma_{Y,i}),$$

where E – stands for the emission inventory, D – for its deterministic fraction, S – for the stochastic fraction, and n – is the year, for which the revised data were recalculated. Similarly, if $y_j < Y$,

$$E_{y_j,i}^n = D_{Y,i}^n + S_{y_j,i}^n, \quad \text{with } S_{y_j,i}^n \sim \mathcal{N}(m_{y_j,i}^n, \sigma_{y_j,i}^n),$$

where the mean values and standard deviations are of the form

$$m_{y_j,i}^n = a_i(Y - y_j), \quad \sigma_{y_j,i}^n = \sigma_{Y,i} + b_i f(Y - y_j), \quad b_i \neq 0,$$

and f is a function, such that

$$f(Y - y_j) > -\frac{\sigma_{Y,i}}{b_i}.$$

The parameters a_i and b_i , for a country i , associated with the stochastic fraction $S_{y_j,i}^n$ can be estimated from the data, together with $\sigma_{Y,i}$. Parameter a_i describes a shift in the accuracy of the inventory gathering, and b_i – a shift of the precision level. They both depend on the difference between the revision year y_j , and the most recent revision year Y , due to the learning. To find the deterministic fraction $D_{Y,i}^n$, the smoothing splines can be used, as presented in [4]. This approach, when applied to the most recently revised data $E_{Y,i}^n$ will give not only the estimate of the deterministic fraction, but also an estimate of the variance $\sigma_{Y,i}^2$.

Algorithm for a fixed country i

Fix i and consider all the inventory data $E_{y_j,i}^n$ in the year n for $n = 1, \dots, N_j$, revised in the year y_j , $j = 1, \dots, J$. For a fixed country i , the procedure can be described as follows.

1. For the most recently revised inventory data E_Y^n **calculate the smoothing spline Sp_Y and estimate the variance σ_Y^2** of the stochastic fraction S_Y^n .
2. **Subtract the spline Sp_Y , built on the data from the year Y , from all earlier revisions $E_{y_j}^n$, $y_j < Y$, calculating differences**

$$v_j^n = E_{y_j}^n - Sp_Y, \quad n = 1, \dots, N_j, \quad j = 1, \dots, J.$$

For some years the difference v does not exist, due to lack of revised inventories in this year. These years are skipped from the sequence of N_j data.

We consider the following **model**:

$$(1) \quad v_j^n \sim \mathcal{N}(m_j, \sigma_j), \quad n = 1, \dots, N_j, \quad j = 1, \dots, J,$$

where

$$(2) \quad m_j = a(Y - y_j), \quad \sigma_j = \sigma_Y - b(Y - y_j)^{c+1}, \quad b \neq 0.$$

Assume also that differences (1) are independent.

3. **Estimate parameters a , b , and c** (and hence m_j and σ_j , $j = 1, \dots, J$ in (2)) in the following three-step procedure.

Step 1. Estimate parameters α_j and β_j , $j = 1, \dots, J$ in the model

$$m_j = \alpha_j(Y - y_j), \quad \sigma_j = \hat{\sigma}_Y + \beta_j(Y - y_j), \quad \beta_j \neq 0,$$

using Maximum Likelihood estimators

$$\hat{\alpha}_j = \frac{1}{N_j(Y - y_j)} \sum_{n=1}^{N_j} v_j^n, \quad \hat{\beta}_j = \frac{\left(\sqrt{\frac{1}{N_j} \sum_{n=1}^{N_j} (v_j^n - \bar{v}_j)^2} - \hat{\sigma}_Y \right)}{(Y - y_j)},$$

$$\text{where } \bar{v}_j = \frac{1}{N_j} \sum_{n=1}^{N_j} v_j^n.$$

Step 2. Use $\hat{\alpha}_j, j = 1, \dots, J$, obtained in Step 1, to estimate parameter a in the first order autoregressive model

$$\alpha_{j-1} = \frac{1}{\tilde{a}}\alpha_j + \varepsilon_j, \quad |\tilde{a}| < 1, \tilde{a} \neq 0, \quad \text{where } \alpha_{J+1} := 0,$$

and ε_j are independent and $\varepsilon_j \sim N(0, \sigma)$. Estimator of the parameter a is then given by

Step 3. Use the sequence $\hat{\beta}_j, j = 1, \dots, J$, obtained in Step 1, to estimate parameters b and c in the regression model

$$(3) \quad \beta_j := -b(Y - y_j)^c, \quad j = 1, \dots, J, \quad \text{where } b < 0.$$

Since $\beta_j > 0, j = 1, \dots, J$, nonlinear model (3) can be converted into a linear one of the form

$$\ln \beta_j = \ln(-b) + c \ln(Y - y_j),$$

and the parameters $\tilde{b} := \ln(-b)$ and c can now be estimated using the Least Squares method.

3. Case study - NIR data for Austria

The data analyzed refers to CO₂ emissions in the years 1990 – 2005, and recalculations (revisions), performed every year, from 1999 to 2005. We start with building a smoothing spline Sp_Y for the most recently revised data – from the year $Y = 2005$ (Figure 2).

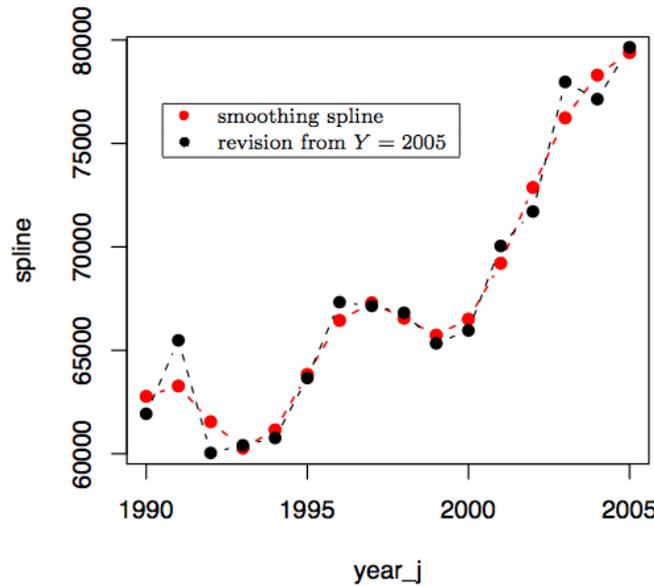


Figure 2. Smoothing spline and data on CO₂ emissions from Austrian NIR, $Y = 2005$

Then, we calculate the differences v_j between real emissions data, and the spline obtained, and estimate parameters a, b , and c in the model (1) - (2). Having obtained estimates for a, b , and c , we can find sequences of means and standard deviations. The results are presented in Figure 3 and Table 2.

Table 2. Estimates of parameters in the model (1) - (2).

j	1999	2000	2001	2002	2003	2004	parameters
m_j	8.62	7.18	5.74	4.31	2.87	1.44	$\hat{a} = 1.44$
σ_j	1119.33	1099.62	1074.95	1042.20	994.20	906.87	$\hat{b} = 1158.6, \hat{c} = -1.11$

Parameter	Estimate	Model fit
a	1.44	$\sigma^2 = 43669$
b	1158.6	St.error=0.056, t-test: p-value=0.000000023, $R^2 = 0.99$
c	-1.11	St.error = 0.045, t-test: p-value=0.0000152

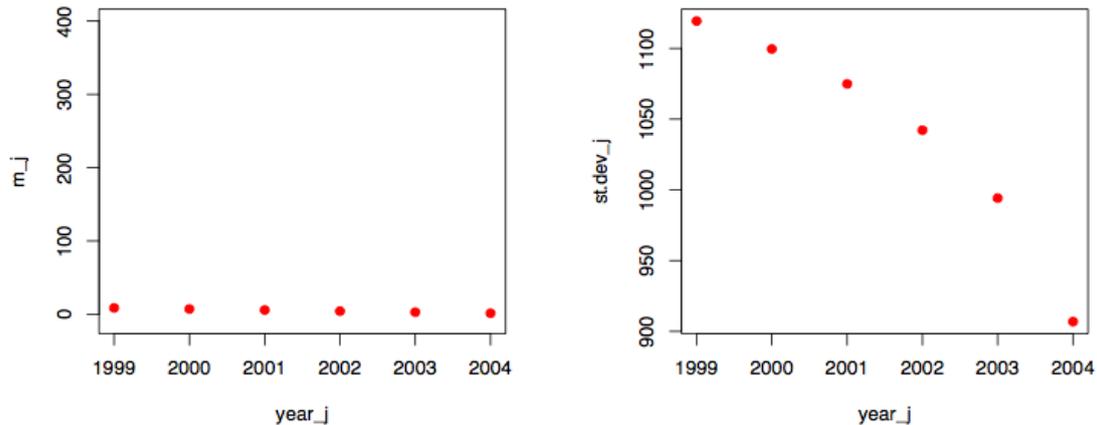


Figure 3. Estimates of m_j and σ_j , $j=1999..2004$, Austria.

The analysis conducted gives the information about the data considered. The mean values tend to zero and we may take into account the speed of that convergence. The sequence of standard deviations is strictly decreasing, indicating a reduction in the emissions.

The main result – the uncertainty assessment is the sequence of relative values of the form

$$\frac{\hat{\sigma}_j}{\hat{Sp}_j}, j = 1, \dots, J$$

depicted in Figure 4a), together with the relative uncertainties provided in the National Inventory Report. One can observe that the relative values are significantly smaller than the uncertainties published in the reports, which means that the assessment proposed is more accurate.

For comparison, in Figure 4b) we present the analysis, conducted for the Austrian National Inventory Reports data available up to the year 2014 (i.e. for $Y = 2012$). Also in this case, one can notice that uncertainty assessment obtained using the method proposed is significantly better than those published in the official reports.

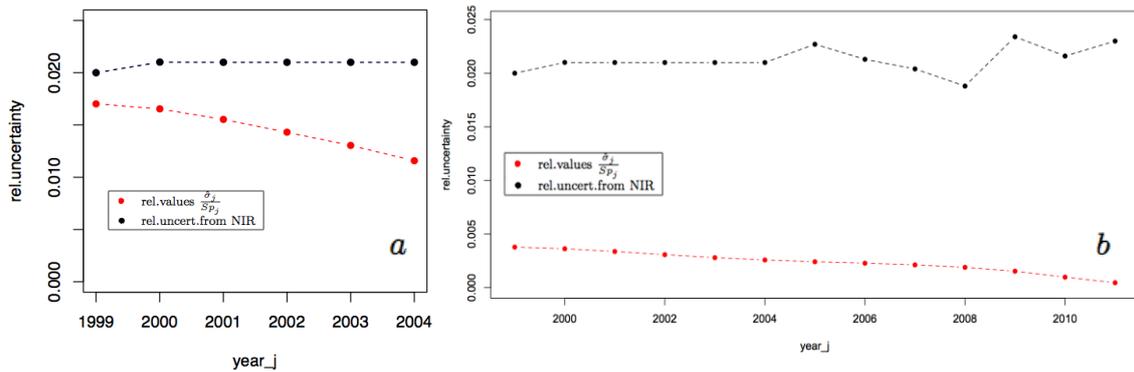


Figure 4. Comparison of relative uncertainty and relative values $\frac{\sigma_j}{Sp_j}$, NIR data on CO₂ emissions, Austria, a) $j = 1999 \dots 2004$, $Y = 2005$ b) $j = 1999 \dots 2011$, $Y = 2012$

4. Conclusions

The method proposed, proved to be a good tool for the uncertainty assessment. It is worth noting that it is based solely on the data, and works without any additional assumptions. It works well in practice (applied to the NIR data for EU countries). However, it is necessary to test larger data sets (as they become available), and other databases.

References

- [1] G. Marland (2008), Uncertainties in accounting for CO₂ from fossil fuels. *Journal of Industrial Ecology*, 12(2):136-139.
- [2] G. Marland, K. Hamal, M. Jonas (2009), How uncertain are estimates of CO₂ emissions? *Journal of Industrial Ecology*, 13(1):4-7.
- [3] K. Hamal (2010) Reporting GHG emissions: Change in uncertainty and its relevance for detection of emission changes. Interim Report IR-10-003. IIASA, Laxenburg.
- [4] Z. Nahorski, W. Jęda (2007), Processing national CO₂ inventory emission data and their total uncertainty estimates, *Water, Air, and Soil Pollution: Focus*, 7(4-5):513-527.

How uncertainty of air emission inventories impacts policy decisions at sub-national level. A Shift-Share Analysis undertaken in Piedmont (Italy).

Alessandra La Notte¹, Stefania Tonin¹, Silvio Nocera¹

¹Department of Design and Planning in Complex Environments, IUAV-University of Venice
lanotte@iuav.it

Abstract

Emission inventories are compiled at national and regional levels and used without taking uncertainty into account. We attempt to check whether and to what extent uncertainty related to emission inventories affect quantitative analysis used by policy makers to set strategies and implement actions at regional and sub-regional levels. We consider the regional air emission inventory of the Piedmont region in Italy. Uncertainty is calculated by adapting the insurance-based method. A hybrid accounting matrix is built and a Shift-Share analysis is undertaken for manufacturing and construction activities, and for the transport sector at regional and provincial levels. The procedure is repeated for data without uncertainty and data with uncertainty. Although in absolute terms total emissions are remarkably different, the outcomes of the Shift-Share Analysis vary among provinces: sometimes the messages are misleading when uncertainties are not included in the calculation; sometimes the differences are negligible. Some general conclusion can be drawn.

Keywords: air emission inventory, uncertainty, Shift-Share Analysis, hybrid environmental accounts

1. Introduction

Air emission inventories, and in particular Green-House Gas emissions, have always been thought as the primary source of information for the international Climate Change agreements and trading [1] and are usually compiled at national level. However, especially when developed at sub-national level, these datasets can be a precious source of information for policy makers at different administrative in accounting terms for descriptive analysis and for policy analysis [2]. Although few examples of air emission inventories used in policy analysis at subnational level already occur, uncertainty is never considered. In some cases uncertainty coefficients are not even available from the agencies and institutes responsible for the delivery of air emission inventories.

In this paper we are going to combine one particular technique to quantify uncertainty together with a hybrid environmental accounting framework and we are going to use a decomposition analysis tool to assess whether and to what extent estimates with and without uncertainty do affect the final message that policy makers use when planning strategy and actions for the territory they administer.

The case study we used to apply the methodology, accounting framework and the decomposition analysis is the Piedmont region and its provinces. Their air emission regional inventory is one of the best example existing in Italy. Their datasets are publicly available and the uncertainty coefficient are efficiently compiled by the functionaries in charge for the inventory.

After a brief description of the data, methodology and tool used (Section 2), the results are presented (Section 3) and some points for discussion raised (Section4). In

the conclusion (Section 5) we summarise the main findings of this first application through key messages.

2. Materials and methods

The Piedmont region is located in the North-Western part of Italy. In Piedmont automotive (the FIAT group and its induced activities) is the dominating compartment, followed by chemical, food, textile, clothing, electronics and editorial compartments. The tools at the basis of this application are the hybrid environmental accounts and the shift-share analysis that will be described in the following paragraphs.

2.1 Calculation of uncertainty and hybrid environmental accounts

The CORE INventory AIR emissions (CORINAIR) method is the framework supported by the European Environment Agency. It was adopted by the national environmental protection agency in compiling the national inventory. At regional level and specifically in Piedmont, the EMEP-CorinAIR inventory is compiled since the beginning of 2000s and the procedure has been greatly improved and updated since the first release. The regional inventory records data according to the SNAP (Selected Nomenclature for Air Pollution) classification. The inventory is composed by 11 macro sectors, 75 sectors and 430 activities for the following pollutants: methane (CH₄), carbon monoxide (CO), carbon dioxide (CO₂), nitrogen dioxide (N₂O), ammonia (NH₃), Volatile Organic Compounds (VOC), oxides of nitrogen (NO_x), sulphur dioxide (SO₂) and particulate matters (PM₁₀ and PM_{2.5}).

Uncertainty is compiled according to inventory guidelines. For each pollutant at the activity level it is possible to calculate uncertainty according to the following formula:

$$UNC_{ijk} = UnEF_{ijk} * UnAD_{i,k} \quad (1)$$

where

UNC_{ijk} = total uncertainty coefficient for the activity i, pollutant j, fuel k; UnEF = uncertainty coefficient assigned to Emission Factor for the activity i, pollutant j, fuel k; UnAD = uncertainty coefficient assigned to Activity Data for the activity i, fuel k

Marland et al. [3] borrow their approach to estimate uncertainty from the insurance industry by adding a charge called risk charge that represents the insurance for the insurer. Uncertainty is calculated using the approach suggested in EMEP Guidelines and this value is used as risk charge. The formula to calculate the uncertainty maximum limit becomes thus:

$$EUnc_{ijk} = (E_{ijk} * (1 - UNC_{ijk})) + E_{ijk} \quad (2)$$

where

EUnc_{ijk} = total emissions and uncertainty for the activity i, pollutant j, fuel k; E_{ijk} = emissions for the activity i, pollutant j, fuel k

It is possible to connect air emissions to their generating activity. The NAMEA-type accounting module allows to frame together economic data and emissions and can be compiled at local level [4]. The first step to undertake is to harmonize the SNAP classification system that is based on production processes with the NACE (*Nomenclature générale des Activités économiques dans les Communautés Européennes*) classification system that is based on economic sectors. For this application we choose to focus on the whole secondary sector (that includes all manufacturing activities and construction) and on the transport sector. We thus do not consider the primary sector (agriculture and forestry) the tertiary (services) and

households. Economic Data (local units and number of employees) are withdrawn from ASIA (register of active enterprises), and air emission data are withdrawn from EMEP-CorinAIR (in tonnes for all pollutants except CO₂ that is in 1,000 tonnes).

2.2 The Shift-Share Analysis

We apply decomposition analysis in order to investigate the mechanism that affects air emissions: the rationale for structural decomposition analysis is splitting an identity into its components. Changes in some variables are decomposed in changes in its determinants. The methodologies commonly used to decompose emissions trends are index decomposition analyses, input-output structural decomposition analysis and shift-share analysis [5].

The purpose of this application is to measure the role of the productive structure at the lower hierarchical level considered (in our case the provincial level) in explaining the emissions efficiency gap between this level (i.e. provincial) and the higher hierarchical level (in our case the regional level). Shift-share analysis in fact decomposes the source of change of the specified ‘dependent variable’ into provincial specific components (that constitutes the shift) and the portion that follows regional growth trends (that constitutes the share).

The question we aim to address is whether the gap between the considered province and the regional benchmark average depends on (lack of) environmental friendly technologies in the included economic sectors, and/or on a provincial specialization in sectors with higher/lower eco-efficiency.

We firstly calculate the intensity of emissions by considering the emission of each pollutant referred to the number of workers employed in each sector. This variable provides insights into the socio-environmental efficiency of the productive sectors, which is useful in order to plan a strategy to support environmental innovation at sector level. We then analyse the relative environmental efficiency of the provincial economic system with respect to the regional average, referring to the GHG pollutants and to the economic sectors included in the hybrid accounts.

The aggregate indicator of emission intensity is represented by ‘total emissions [E] on number of employees[Empl]’. The benchmark is represented by the regional value. We define the index of emissions intensity as X for the regional average ($X=E/Empl$), as X_{Pr} for the province ($X_{Pr}=E_{Pr}/Empl_{Pr}$) and as X^s for each sector for the province and $X^s_{Pr}=E^s_{Pr}/Empl^s_{Pr}$ for the region ($X^s=E^s/Empl^s$). We then define the share of sector value added as $P^s=Empl^s/Empl$ for the region and $P^s_{Pr}=Empl^s_{Pr}/Empl_{Pr}$ for the province.

$$X = \sum_s P^s X^s \quad (3)$$

$$X_{Pr} = \sum_s P^s_{Pr} X^s_{Pr} \quad (4)$$

The shift-share decomposition allows to identify three effects that explain the gaps in terms of aggregate emissions efficiency between the province and the region.

The first effect (‘structural’ or industry mix) is given by:

$$m_{Pr} = \sum_s (P^s_{Pr} - P^s) X^s \quad (5)$$

m_{Pr} assumes a positive (negative) value if the region is ‘specialized’ in sectors associated with lower (higher) environmental efficiency, given that the gap in value added sector shares is multiplied by the value of X of regional average (‘as if’ the province were characterized by average regional efficiency). The factor m_{Pr} assumes lower values if the province is specialized in (on average) more efficient sectors.

The second factor (‘differential’ or ‘efficiency’) is given by:

$$p_{Pr} = \sum_s (X_{Pr}^s - X^s) \quad (6)$$

P_{Pr} assumes a positive (negative) value if the region is less (more) efficient in terms of emissions (the shift between provincial and regional efficiency), under the assumption that ('as if') number of employees sector share were the same for the region and the province.

The third effect ('allocative component') is given by:

$$a_{Pr} = \sum_s (X_{Pr}^s - X^s) (P_{Pr}^s - P^s) \quad (7)$$

The a_{Pr} factor is positive (negative) if the province is specialized, relative to the regional benchmark, in sectors characterized by higher (lower) emission intensity.

Table 1. Data with and without uncertainty at regional and provincial level.

	CH ₄	CO	CO ₂	COV	N ₂ O	NH ₃	NO _x	PM ₁₀	PM _{2.5}	SO ₂
Secondary sector Piedmont										
REG (W/O)	107.657,1 7	6.089,22	15.564,0 2	27.109,2 0	2.276,79	1.209,42	21.830,4 6	785,44	558,87	8.281,56
REG (W)	203.505,1 7	9.375,98	23.672,2 1	46.169,0 5	3.001,98	2.103,48	30.070,7 1	1.129,7 2	832,51	11.996,8
Δ _Reg	0,89	0,54	0,52	0,70	0,32	0,74	0,38	0,44	0,49	0,45
Transport sector Piedmont										
REG (W/O)	288,00	11.215,42	3.902,60	2.518,18	121,09	31,99	29.859,3 6	4.692,4 0	1.460,8 7	48,97
REG (W)	488,11	19.219,08	6.628,90	4.302,59	207,52	54,28	50.369,4 2	8.155,4 0	2.485,6 3	86,29
Δ _Reg	0,69	0,71	0,70	0,71	0,71	0,70	0,69	0,74	0,70	0,76
Secondary sector Biella										
Prov (W/O)	2.959,16	73,29	177,06	255,45	6,29	81,35	364,90	32,50	21,31	232,42
Prov (W)	5.652,81	115,20	301,58	471,16	11,10	134,85	565,19	56,71	35,70	416,51
Δ _Prov	0,91	0,57	0,70	0,84	0,77	0,66	0,55	0,75	0,68	0,79
Transport sector Biella										
Prov (W/O)	7.56799	363,9551 4	121,4479	77,90136	3,38207	1,08423	846,3386	169,552	45,2134	0,771662
Prov (W)	12,86	626,26	207,63	133,04	5,79	1,85	1.437,20	295,62	77,29	1,32
Δ _Prov	0,70	0,72	0,71	0,71	0,71	0,71	0,70	0,74	0,71	0,71
Secondary sector Torino										
Pro(W/O)	61.832,97 116.244,1	2.262,16	6.590,38	9.204,56 15.437,0	218,08	270,97	7.304,90 10.138,7	183,44	146,92	718,78
Prov (W)	0	3.190,53	9.921,59	8	396,09	474,77	7	258,35	209,00	1.143,14
Δ _Prov	0,88	0,41	0,51	0,68	0,82	0,75	0,39	0,41	0,42	0,59
Transport sector Torino										
Pro(W/O)	143,9462	5323,606 6	1744,264	1220,169	55,0616	14,043	12821,56 21.785,4	1894,426	655,133 1.121,5	30,51204
Prov (W)	246,21	9.192,52	2.984,92	2.103,96	94,97	23,97	7	3.295,82	7	54,56
Δ _Prov	0,71	0,73	0,71	0,72	0,72	0,71	0,70	0,74	0,71	0,79
Secondary sector Vercelli										
Pro(W/O)	2.412,86	386,22	1.116,74	680,91	69,64	109,75	861,74	44,59	28,72	102,88
Prov (W)	4.475,64	694,15	1.593,49	1.048,54	120,05	197,48	1.154,17	53,10	37,14	167,17
Δ _Prov	0,85	0,80	0,43	0,54	0,72	0,80	0,34	0,19	0,29	0,63
Transport sector Vercelli										
Prov (W/O)	16,7377	641,8716	250,0528	140,6166	6,68947	2,03567	2071	326,973	89,5194	1,633588
Prov (W)	27,86	1.081,32	418,30	235,82	11,22	3,42	3.444,82	566,43	149,92	2,74
Δ _Prov	0,66	0,68	0,67	0,68	0,68	0,68	0,66	0,73	0,67	0,68

3. Results

The Piedmont region contains important industrial centers: from automobile to electronics, to mechanical, to food and beverage industries. Hybrid flow accounts have been compiled for all provinces of Piedmont and Shift-Share Analysis has been performed for regional-provincial cases. Results have been compared for estimates with and without uncertainties. Differences stand out in three cases that will be presented in details. The three provinces are Biella, Torino and Vercelli. Table 1 shows the estimates with (W) and without (W/O) uncertainties of these provinces compared with the regional level. It is important to consider the transport sector separately from the secondary sector. In the first case there are no remarkable differences between the regional and provincial deltas, while in the second case for some pollutants at provincial level shows remarkably higher differences.

In the province of Biella there are many factories working on spinning and weaving of wool and on other tissues. Table 1 shows that the difference in considering data with and without uncertainty for secondary sector diverges from the regional trend for the pollutants CO₂, N₂O, PM and SO₂. The shift-share analysis for Biella shows that when we consider uncertainty the productivity differential changes for two pollutants: NO_x and SO₂.

The second province we report is Torino: the most important province in the region from historical, economic and demographic points of view. The industrial sectors mostly developed in this province are the automobile industry and all its related industrial sectors, and electronics. Table 1 shows the estimates with (W) and without (W/O) uncertainties of this province compared to the regional level. Differently from the Biella province in Torino CO₂ and PM with and without uncertainty follow the regional trend. It is important to check on Table 1 that the only province of Torino generates almost half of CO₂ emissions due to traffic for the whole region. Clearly the province of Torino has a remarkable impact at regional level for some of the main pollutants due to traffic (i.e. CO₂ and PM). The shift-share analysis for Torino shows that the allocative component changes when data are computed with uncertainty: the main differences are recorder for the pollutants NH₃ and PM_{2.5}.

The last province we present is Vercelli, whose main economic activities are linked with the production of rice. Table 1 shows the estimates with (W) and without (W/O) uncertainties of this province compared with the regional level. This province shows different trends for many pollutants: in some cases the differences between data with and without uncertainties between the provincial and the regional levels are much higher (CO, N₂O, SO₂) and in some other cases are much lower (COV, PM₁₀, PM_{2.5}). The shift-share analysis for Vercelli shows that there are a lot of differences when data are computed with and without uncertainty. The structural component has a single critical pollutant: without uncertainty is NH₃ but with uncertainty is COV. The eco-efficiency component without uncertainty is favorable only for CH₄, COV and SO₂ while with uncertainty becomes favorable for all pollutants. The allocative component presents a more favorable condition without uncertainties where the only critical pollutants are CH₄, NH₃ and SO₂; with uncertainties in fact all pollutants become critical except CO₂ and COV.

4. Discussion and final remarks

Looking at the numbers, when considering the differences in absolute terms in most cases estimates with uncertainties double the initial estimates. However when using a

tool such as Shift-Share Analysis, doubled estimates do not dramatically affect the outcomes. In some cases differences can be found but not as striking as initially expected. As source of air emissions we consider all secondary sector and from the tertiary sector only transport. For all provinces we consider separately the secondary sector and transport when comparing the difference of estimates with and without uncertainty between the regional trend and the provincial trend (ref. Table 1). For all provinces the regional and provincial levels in secondary sectors show important differences for N₂O, PM and SO₂ (the one exception is the province of Torino). When the two sectors are summed some of these difference disappear: e.g. the difference (higher or lower) in PM between the regional and provincial trends disappear. Moreover, the territorial aggregation is impacted by some provinces which determine the weights of some pollutants (of course due to their generating activities) rather than others: Table 1 shows that it also applies in terms of uncertainties. For example the province of Turin has a considerable impact because it collects the major economic activities and host most of the population. In fact when comparing the regional and the provincial levels, for many pollutants that in other provinces shows important differences if considered with and without uncertainty, in the province of Torino the difference only emerge in two cases (N₂O and SO₂). In the province of Vercelli the economic activities and the number of inhabitants are less. This province records many differences in data with and without uncertainty compared to the regional level (ref. Table 1) and thus such a reality could not be represented by the regional level: it should be analyzed individually. If this difference shows up within a region like Piedmont, we can imagine the huge differences that would show at national level. In Italy for example the national level would never represent equally the Northern and the Southern parts: territorial policies, development and environmental policies not only must consider uncertainty but must also identify for the appropriate territories the appropriate administrative level.

However, the method we used to estimate uncertainty was applied in a very elementary way. In fact we did not make any difference among pollutants: we assumed that all measured data are underestimates. Some studies, e.g. [6], shows that some pollutants are prone to over estimates rather than underestimates and some other pollutants' estimates are fine. Moreover, we did not apply any refinement to the coefficient interval: we consider the worse hypothesis, i.e. the maximum possible applicable percentage of error.

Having set few statements in the previous section, we would like to conclude this paper with few remarks. Firstly, the calculation of uncertainty varies according to the administrative level considered. In our example we started from a sub-national level (the region) and further looked into a local context (the provinces). The message does thus amplify when the initial level is a nation or a macro-region. Secondly, by adding uncertainty to estimates can affect the message to policy makers, even if in some case less than expected when looking at the differences in absolute terms. Finally, a raw methodology, like the one we have applied in this paper, can help to identify which are the pollutants that require a deeper analysis. Considering the limits of time, budget and data, this kind of methodologies can work as sieve. To the identified pollutants and to the critical territorial contexts a more sophisticated approach should be applied in order to provide the policy maker with a correct and robust message.

References

- [1] Lieberman D., Jonas, M., Nahorski Z., Nilsson S. (2007). Accounting for Climate Change. Springer Netherlands.

- [2] La Notte A., Dalmazzone S. (2012). Feasibility and uses of the NAMEA-type framework applied at local scale: case studies in North-Western Italy, in V.Costantini, M.Mazzanti and A.Montini (Eds) Hybrid Economic-Environmental Accounts, Routledge studies in Ecological Economics, Taylor& Francis Group.
- [3] Marland, E., Cantrell, J. , Kiser, K., Marland, G., Shirley, K. (2014). Valuing uncertainty part I: The impact of uncertainty in GHG accounting. Carbon Management, 5(1), 35-42.
- [4] Dalmazzone S., La Notte A. (2013). Multi-scale environmental accounting: Methodological lessons from the application of NAMEA at sub-national levels, Journal of Environmental Management, 130, 405-416.
- [5] Esteban, J. (2000) Regional convergence in Europe and the industry mix: a Shift-share Analysis, Regional Science and Urban Economics, 3, 353-364.
- [6] Bergamaschi, P., Corazza, M., Karstens, U. , Athanassiadou, M. , Thompson, R.L., et al. (2015). Top-down estimates of European CH₄ and N₂O emissions based on four different inverse models. Atmospheric Chemistry and Physics, 15(2), 715-736.

Performance of global black carbon emission inventories in the Arctic

Ville-Veikko Paunu ^a and Kaarle Kupiainen ^a

^a *Finnish Environment Institute SYKE*
Email: ville-veikko.paunu@ymparisto.fi

Recent assessments indicate that short-lived climate pollutants, especially black carbon (BC) play a major role in the climate of the Arctic (AMAP, 2015). However, uncertainties remain in the impact assessments. One source of the uncertainties are the emission estimates of black carbon. Bond et al. (2004) presented a quantitative uncertainty estimation of emission inventories of carbonaceous aerosol. They found that major uncertainties are caused by insufficient information on emission parameters and major emitting sector activities. Global emission uncertainties for anthropogenic BC emissions were identified as 3.1 to 10 Tg/y (-30% to 120%) expressed as 95% confidence intervals. Sectors contributing most to the uncertainties were found to be Chinese coke making, residential wood combustion, industrial coal combustion, and on-road diesel. Regionally the largest uncertainties were estimated in Asian emissions.

While global BC emissions have an impact on the Arctic, pollutants emitted closer to the Arctic might have higher impact per emitted mass. Furthermore, BC is removed relatively quickly from the atmosphere, therefore having higher concentrations close to their sources. Thus, the spatial allocation of the emission estimates has an important effect on the climate impacts. However, this spatial dimension has previously been neglected in uncertainty assessments. An initial study, presented in AMAP 2015, showed that differences between global BC emission inventories were relatively large in higher latitudes. The differences were further analysed in this study.

We compared available spatially-distributed global BC emission datasets available from the ECCAD-GEIA website (<http://eccad.sedoo.fr>) and analysed differences in both emissions and their locations.

Some of the variation between the inventories was found to be due to different treatment of Arctic relevant source sectors. For example only some inventories included emissions from flaring in full extent, although the emissions are significant in the Arctic region. Notably the spatial representation of flaring in the oil and gas production areas close to and within the Arctic area was missing in most of the datasets. Another sector omitted in some inventories was international maritime transport. Inclusion of relevant emission sectors is a common improvement suggestion for all models.

There were significant differences between the spatial distributions of the different BC emission inventories and often the agreement between the spatial distributions was completely lacking. These differences also varied between source sectors. The differences indicated that the inventories use different spatial proxies for the emissions. We recommend that spatial proxies should be harmonized and important regions and source sectors for the Arctic area should be addressed as accurately as possible.

References

- [1] AMAP Impact of Tropospheric Ozone and Black Carbon on Arctic Climate, in press, technical report, 2015
- [2] Bond et al., *Journal of Geophysical Research* 109:D14203, 2004

Global anthropogenic particle number emissions and their size distributions

Pauli Paasonen^{1,2,3}, Kaarle Kupiainen^{2,3}, Zbigniew Klimont², Antoon Visschedijk⁴,
Hugo Denier van der Gon⁴ and Markus Amann²

¹University of Helsinki, Finland ²International Institute for Applied Systems Analysis (IIASA), Austria

³Finnish Environment Institute (SYKE), Finland ⁴TNO, the Netherlands

pauli.paasonen@helsinki.fi

Abstract

Aerosol particle number concentrations and size distributions affect our climate by determining the formation of cloud droplets and thus altering the cloud reflective properties. The aerosol-cloud interactions are one of the main uncertainties in estimating the future climate change. One of the weaknesses in current climate modelling is the description of number emissions and size distributions of particles. Here, we present the first global results of implementing particle number emission factors to GAINS emission scenario model and discuss the related uncertainties. The uncertainties for different source sectors vary significantly, causing a steep difference in total uncertainties in different parts of the world. The reason for these uncertainties is the scarcity of data on particle number size distributions for certain sources. The implemented particle number emission factors, however, are expected to be a significant improvement over previously applied particle number emissions estimates in climate modelling.

Keywords: particle number emissions, number size distribution, emission scenario model, aerosol-cloud interactions

1. Introduction

One of the main uncertainties in our understanding of the future climate change arises from the aerosol-cloud interactions [1]. One factor to these uncertainties is the inadequate description of aerosol number emissions from anthropogenic sources. The number of cloud droplets, which reflect solar radiation back to space, depends on the number concentrations of particles in cloud condensation nuclei -size range (CCN, diameters d_p close to or over $0.1 \mu\text{m}$). These particles are emitted to the atmosphere directly from anthropogenic sources or formed in atmosphere due to the growth of ultrafine particles (UFP, with diameters below $0.1 \mu\text{m}$), which may be of either biogenic or anthropogenic origin. As a source of CCN, the biogenic growth of UFP is roughly as significant a source of CCN as direct anthropogenic emissions [2]. On the other hand, UFP have severe adverse health effects, which are different to those of particulate mass [3]. Also the main anthropogenic sources of UFP, which typically dominate particle number concentrations (PN), are different to the main sources of particulate mass [4].

Here, we present the first results of the implementation of aerosol number emission factors (EF_{PN}) in the global emission scenario model GAINS (Greenhouse gas - Air pollutant Interactions and Synergies [5]) and discuss the related uncertainties.

2. Methods

The GAINS model (Greenhouse gas – Air pollutant Interactions and Synergies [5]) is an integrated assessment model, which brings together information on the sources

and impacts of air pollutant and greenhouse gas emissions and their interactions. GAINS combines data on economic development, the structure, control potential and costs of emission sources, the formation and dispersion of pollutants in the atmosphere and an assessment of environmental impacts of pollution. The political scenarios in GAINS allow for researchers and modellers to study the future global emissions and their spatial distribution and for decision makers to compare the costs and outcomes of regulations and investments on new technologies. GAINS describes the inter-relations between the effects and emissions of various pollutants (SO₂, NO_x, PM, NMVOC, NH₃, CO₂, CH₄, N₂O, F-gases) that contribute to these effects. GAINS assesses more than 1000 measures to control the emissions to the atmosphere for each of its 168 regions (mainly countries, of which some divided in regions, e.g. China consists of 32 regions, and some grouped, e.g. Middle East). In its optimization mode, GAINS identifies the least-cost balance of emission control measures across pollutants, economic sectors and countries that meet user-specified air quality and climate targets.

The annual emissions E in a country or a region i are calculated with

$$E_i = \sum_{jkm} E_{ijkm} = \sum_{jkm} A_{ijkm} X_{ijkm} EF_{ijkm}, \quad (1)$$

where the indices and symbols refer to

j	Source sector (e.g. domestic single house heating boilers)
k	Fuel (e.g. firewood, coal)
m	Abatement technology (e.g. pellet boilers, boilers with electrostatic precipitator)
A	Volume of annual activity (typically annual energy consumption in sector j with fuel k)
X	Share of abatement technology of the activity m (so that $\sum_m X_m = 1$)
EF	Emission factors for each sector-fuel-technology –combination (emissions per activity unit)

We have recently introduced aerosol particle number emission factors (EF_{PN}) with corresponding particle (number) size distributions (PSD) to GAINS [4]. For road transport PN_{EF} 's and PSDs are based on the latest version of TRANSPHORM database [6]. For the global analysis these were extended with separate PN_{EF} 's and PSDs for different fuel sulphur contents. For other sources, emission factors are obtained from the literature and from the emission inventory by TNO [7,8]. The implemented emission factors and size distributions represent the emissions of both primary and secondary particles immediately after cooling and dilution to the surrounding air. Particle size distributions in size range 3-1000 nm are described with 8 size bins, facilitating their application in air quality and climate modelling.

3. Results

Global PN emissions are dominated by emissions of UFP, which form close to 80% of the total global emissions (lower panel of Fig. 1).

Figure 1 shows, for year 2010, the shares of different sources in global continental total particle number emissions (PN_{tot}), in number emissions of ultrafine particles (UFP) and non-UFP ($d_p > 100$ nm), as well as in mass emissions of particles with $d_p < 1 \mu\text{m}$ (PM_1). The main source of UFP is road transport, representing 40 % of the total UFP emissions and thus being the largest contributor to total aerosol particle number emissions. Power production contributes to the UFP emissions with a 20 % share, while

residential combustion has a 17 % share. The shares of residential combustion and road transport in non-UFP number emissions are quite similar, roughly 30 % each, whereas the PM_{10} mass emissions are clearly dominated by residential combustion (> 50 %). The vast differences between the number and mass emission shares, especially from road transport and residential combustion, indicate the need for assessing the size segregated number emissions of aerosols in addition to mass emissions.

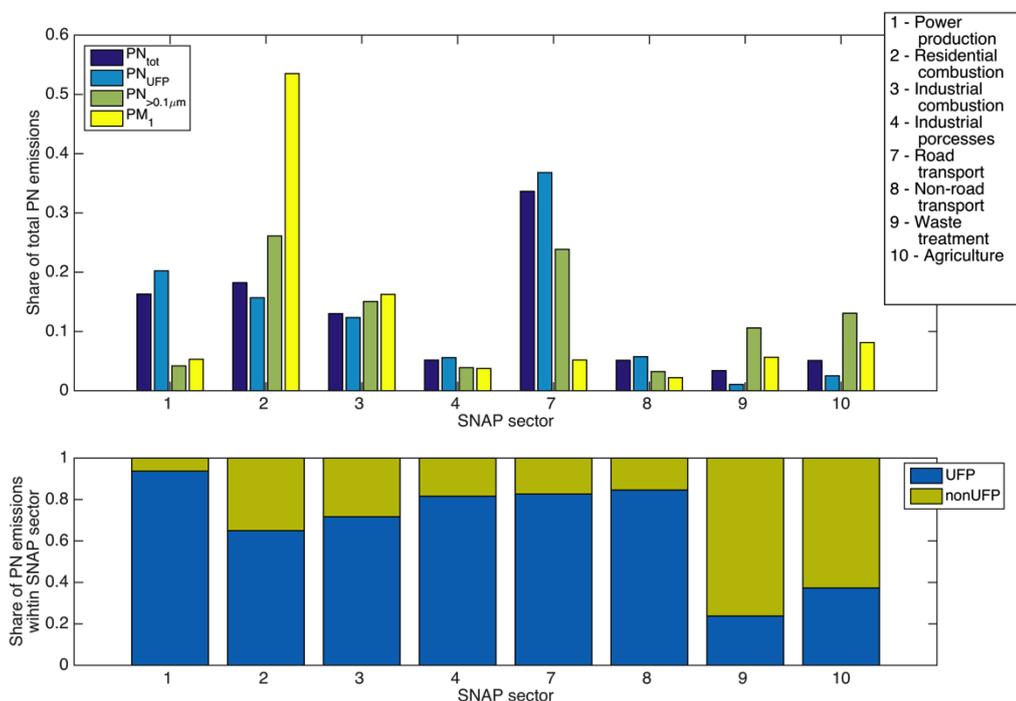


Figure 1. Upper panel: shares of different source sectors in aerosol number emissions of all (PN_{tot}), ultrafine (PN_{UFP}) and non-ultrafine (PN_{nonUFP}) particles and aerosol mass emissions of particles with diameters below $1 \mu m$ (PM_{10}) on 2010. Lower panel: shares of ultrafine and non-ultrafine particle in PN emissions for each SNAP-sector (see legend for clarification of SNAP sector codes).

The annual PN emissions and their estimated future trend in each source sector on different continents, with Eurasian continent divided to major countries and the rest of Europe and rest of Asia, are depicted in Figure 2. The future trend is based on the current legislation baseline scenario (ETP_CLE_v5) compiled in the ECLIPSE project [9]. In 2010, China emitted clearly the most aerosol particles due to high emissions from power production (especially from coke production), residential coal combustion and industrial combustion, followed by Asia (excl. China, India and Russia) and Europe (excl. Russia). In most parts of the world road transportation is the major source of particles.

The actions determined in current legislation are foreseen to decrease the PN emissions in China substantially due to decreases in emissions from coke production and residential coal combustion. This is partly related to the increase in the electricity network, replacing the coal fired cooking stoves. In Europe, North- and South-America and Australia due to the drastic decrease foreseen in traffic emissions due to improving particle emission abatement technologies, especially particle filters. On the contrary, especially in India and Russia, the increase in activities in industrial processes and combustion (the latter mainly in India) and combustion in gas pipeline compressors (in Russia) causes increases in total emissions. In Asia and Africa, the increase in road

transportation seems to overrule the benefits of improving emission abatement technologies, but also the emissions from other source sectors are estimated to increase. The global sum of continental anthropogenic emissions is predicted to decrease roughly by 15 % from 2010 to 2020 (from 1.5×10^{28} to 1.3×10^{28} particles/year), but expected to remain quite constant from 2020 to 2030.

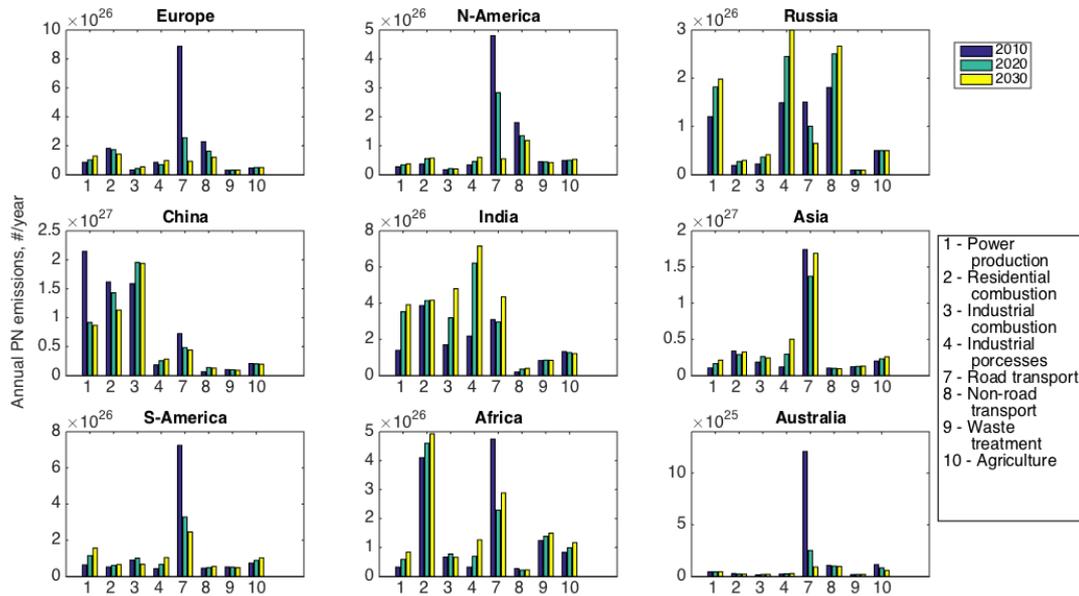


Figure 2. Contributions of different source sectors to particle number emissions in different parts of the world, from 2010 to 2030.

4. Uncertainties in the PN emissions

The uncertainties in the emissions arise from uncertainties specific for the different factors in Equation (1). Here we concentrate, however, only on the uncertainties related to the particle number emission factors EF_{PN} , because they can be estimated to be the main source of uncertainties in PN emissions due to the following reasons. The variation in particle numbers behaves typically in logarithmic scales and thus their concentrations and emissions can vary in orders of magnitude. Furthermore, e.g. Wang et al. [12] have shown that the emissions of the traditional pollutants (SO_2 , NO_x , $PM_{2.5}$ and PM_{10}) calculated with GAINS compare well with those based on measurements (for the studied pollutants the major relative error was found to be factor of 2.5), which suggests that the activities and shares of abatement technologies in Eq. (1) represent the reality reasonably well. Finally, since the emission factors for the traditional pollutants have been revised several times due to restrictions in their emissions and concentrations, but much less efforts have been put to determining particle number emissions, which are restricted only in the latest EURO standards for road traffic, it is quite obvious that the emission factors are the most uncertain part of the emission calculation for PN.

The uncertainty levels related to different sources vary significantly. The EF_{PN} and PSD applied here for road traffic have been determined in a long sequence of EU funded projects were compiled, revised and reviewed in the project TRANSPHORM. However, even in this source sector there remain uncertainties related e.g. to the varying

driving conditions, effects of varying ethanol concentration in fuels and differences between laboratory and real-world emissions. The road traffic emission factors can still be estimated to be among the best analysed ones together with the heating stoves, boilers and fireplaces fuelled with wood in Western countries, due to the various research articles on their $EF_{PN:s}$. On the other end, the emission factors for coke production plants [10] and coal combustion in residential (mainly cooking) stoves [11] are highly uncertain, because both are based on only one article. For these sources, among several others having a minor part in European emission, but potentially much larger in other parts of the world, the $EF_{PN:s}$ for different technologies cannot be determined from the literature, and thus we have set the effects of emission abatement technologies on EF_{PN} to be similar to EF for particulate matter mass emissions. This most certainly decreases the reliability of the emissions from these sources when newer technologies become more popular. Also the biomass, i.e. wood, agricultural residues and dung, combustion in residential sector especially in India and Africa are not fully representative, due to the lack of references for the typical burning equipment and conditions.

Another factor for uncertainty is often weak presentation of the smallest particles, $d_P < 0.01 \mu\text{m}$, in the emission factors and PSDs. High sulphur contents in the fuel lead typically to high emissions of below $0.01 \mu\text{m}$ particles before or immediately after the emissions to atmosphere. However, not all the instruments applied for determining the emission factors detect these particles and sometimes they are also too volatile to be detected with the used technologies. Additional uncertainty related to these smallest particles yields from the lack of emission factors, apart from road traffic, for different fuel sulphur contents and technologies removing SO_2 from the exhaust e.g. in coal plants or industrial combustion.

The variations in source sector specific uncertainties in emission factors described above cause steep differences in the spatial distribution of the uncertainties. In the areas dominated by traffic emissions (see Figure 2), the total uncertainty is the smallest, whereas the emissions in China, India and Russia can be considered much higher.

5. Final remarks

The particle number and mass emissions are typically dominated by different sources sectors and individual sources contribute very differently to these measures. Thus, despite all the above mentioned uncertainties, the PN emissions in GAINS can be expected to describe better the real world emissions than those estimated by converting mass emissions to number emissions with source sector –specific size distributions and mass-to-number factors. For reducing these uncertainties, it is necessary to conduct particle number emission and size distribution measurements for the indicated source sectors.

Acknowledgements

This work was funded by the Academy of Finland Centre of Excellence (grants no. 1118615 and 272041) and European Commission 7th Framework projects ECLIPSE (Project no. 282688), PEGASOS (265148), TRANSPHORM (243406) and ‘Assessment of hemispheric air pollution on EU air policy’ (contract no. 07.0307/2011/605671/SER/C3). We thank Leonidas Ntziachristos and Ilias Vouitsis at Aristotle University of Thessaloniki (Greece) for help and assistance in applying the

emission factors for road transport sector and professor Qiang Zhang from Tsinghua University (Beijing, China) for the spatial distribution of Chinese power plants for 2000, 2005, and 2010.

References

- [1] Stocker, T.F., D. Qin, G.-K. Plattner, L.V. Alexander, et al., (2013): Technical Summary. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., 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.
- [2] Paasonen, P., Asmi, A., Petäjä, T., et al., (2013a): Warming-induced increase in aerosol number concentration likely to moderate climate change. *Nature Geosci.* 6, 438-442.
- [3] WHO. Review of Evidence on Health Aspects of Air Pollution – REVIHAAP project. WHO 2013.
- [4] Paasonen, P., Visschedijk, A., Kupiainen, K., et al., (2013b): Aerosol particle number emissions and size distributions: Implementation in the GAINS model and initial results. Interim Report, IIASA, Laxenburg, Austria, IR-13-020.
- [5] Amann, M., Bertok, I., Borken-Kleefeld, J., et al., (2011): Cost-effective control of air quality and greenhouse gases in Europe: modeling and policy applications. *Environmental Modelling and Software* 26, 1489–1501.
- [6] Vouitsis, I., Ntziachristos, L., Han, Z. (2013). Methodology for the quantification of road transport PM emissions, using emission factors or profiles. *TRANSPHORM Deliverable D1.1.2*.
- [7] Denier van der Gon, H., Visschedijk, A., Johansson, et al., (2009): Size-resolved Pan European Anthropogenic Particle Number Inventory, *EUCAARI Deliverable 141*.
- [8] Kulmala, M., Asmi, A., Lappalainen, H.K., et al., (2011): General overview: European Integrated project on Aerosol Cloud Climate and Air Quality interactions (EUCAARI) - integrating aerosol research from nano to global scales. *Atmospheric Chemistry and Physics* 11, 13061–13143.
- [9] Stohl, A., Aamaas, B., Amann, M., et al., (2015): Evaluating the climate and air quality impacts of short-lived pollutants, *Atmos. Chem. Phys. Discuss.*, 15, 15155-15241, doi:10.5194/acpd-15-15155-2015
- [10] Weitkamp, E.A., Lipsky, E.M., Pancras, P.J., et al. (2005): Fine particle emission profile for a large coke production facility based on highly time-resolved fence line measurements. *Atmospheric Environment* 39, 6719–6733.
- [11] Bond, T. C., Doherty, S. J., Fahey, et al. (2013): Bounding the role of black carbon in the climate system: A scientific assessment, *J. Geophys. Res. Atmos.*, 118, 5380–5552.
- [12] Wang, S., Xing, J., Chatani, S., et al. (2011): Verification of anthropogenic emissions of China by satellite and ground observations, *Atmospheric Environment*, 45, 35, 6347-6358.

Uncertainty in gridded CO₂ emissions estimates

Susannah Hogue¹, Eric Marland², Robert J. Andres³, Gregg Marland⁴, Charles Robison⁵, Dawn Woodard⁶

¹ Department of Mathematical Sciences, Appalachian State University, Boone NC, USA.
hogues@appstate.edu

² Department of Mathematical Sciences and Research Institute for Environment, Energy, and Economics, Appalachian State University, Boone NC, USA. marlandes@appstate.edu

³ Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge TN, USA.
andresrj@ornl.gov

⁴ Research Institute for Environment, Energy, and Economics, Appalachian State University, Boone NC, USA. marlandg@appstate.edu

⁵ Department of Geography and Planning, Appalachian State University, Boone NC, USA.
robisonch@appstate.edu

⁶ Department of Earth Systems Science, University of California, Irvine CA, USA.
dawnwoodard@gmail.com

Abstract

We are interested in the spatial distribution of fossil-fuel-related emissions of CO₂, but it is important to understand the uncertainty in emissions estimates. Uncertainty is introduced in the magnitude and location of large point sources, the magnitude and distribution of non-point sources, and from the use of proxy data to characterize emissions. For the U.S. we develop estimates of the contribution of each component. At 1 degree resolution, in most grid cells, the largest contribution to uncertainty comes from how well the distribution of the proxy (population density) represents the distribution of emissions. In other grid cells the magnitude and location of large point sources make the major contribution. Uncertainty is strongly scale-dependent with uncertainty increasing as grid size decreases. Uncertainty for one degree grid cells is typically on the order of +/- 150% but this is perhaps modest in a data set where emissions per grid cell vary over 8 orders of magnitude.

Keywords: U.S. CO₂ emissions, gridded emissions, large point sources, proxy data

1. Introduction

There is a wide range of interest (both geochemical and geopolitical) in geographically explicit inventories of the sources and sinks of the greenhouse gas CO₂. It is a challenge to estimate sources and sinks in a spatially-explicit context and to best characterize the location and magnitude of emissions and sinks we would like to estimate also the associated uncertainty. Current gridded inventories of emissions from fossil-fuel use and industrial processes rely heavily on related, proxy and re-purposed data. In the following analyses we refine and combine the components of uncertainty and discuss them in the context of the widely-used Carbon Dioxide Information Analysis Center [1] gridded inventory for fossil-fuel related emissions from the U.S. (see also [2]).

Few studies have explored the uncertainty of global-scale, grid-level emissions datasets. Rayner et al. [3] noted that “none of the pointwise fossil emission products available today include” estimates of uncertainty and then estimated that for their dataset “uncertainties can be as high as 50% at the pixel level”. They also pointed out, importantly, that uncertainties for nearby pixels are not independent because, for example, the uncertainty for any given grid space includes consideration that a large point source might be only slightly displaced and the total for the ensemble of cells is constrained by national data. Rayner et al. emphasize that “using the uncertainty of this

pointwise map alone in an inversion is a serious error since it assumes independence of errors.”

2. Methods and analysis

Large point sources make up a large percentage of anthropogenic carbon dioxide emissions for the U.S. and for other industrialized countries [4]. In 2010 one third of U.S. emissions were reported from only 311 sites of large point sources [5]. As soon as the first few latitude and longitude data points from these data sets were typed into Google Earth many of these point sources were not observed at their reported locations. We have to deal with both magnitude and locational uncertainty. Total uncertainty in emissions from any geographic grid space thus has to reflect uncertainty in small or areal sources and in both the magnitude and location of large point sources.

Woodard et al. [6] have developed one key component of what we need to quantify the spatially explicit uncertainty in gridded inventories of CO₂ emissions – an approach for dealing with the uncertainty in the locations of large point sources. Also many gridded inventories exist that document ground level sources of anthropogenic emissions of CO₂ for the U.S. and the globe and these inventories use a variety of top-down and bottom-up methods to geographically distribute emissions that are not attributed to large point sources. Each of the top-down inventories uses some sort of proxy, such as population density or satellite-observed nightlights, to help distribute emissions totals from a large (national or state) scale down to the level of grids as small as 0.1 degrees on a side. Some of the inventories use multiple proxies to take advantage of their differing characteristics. Using proxy data, while necessary, can result in the misallocation of emissions values both spatially and temporally. Hutchins et al. [7] show that the differences among existing data sets increase as grid size is decreased. To address these issues, we have taken the first steps toward calculating the total uncertainty for a CDIAC-like inventory for the U.S. at the 1-degree grid scale. Data here are estimates of annual emissions for the year 2009.

In the gridded CDIAC inventory [1] data on population density are used as a proxy for the spatial distribution of all emissions within a country. For this analysis, we have removed emissions from electric power plants from the country total prior to using the population proxy to distribute the remaining national emissions. The power plants, with magnitudes and locations from EPA’s eGRID dataset [5], were then added back to give total emissions in each grid cell. The emissions inventory discussed here is thus comprised of two components, power plant emissions from the eGRID dataset and all remaining national emissions distributed using population density as a proxy. These remaining emissions do contain some additional, large industrial sources of CO₂, but reporting to the EPA GHG Reporting Program [8] shows that in 2010, 73% of emissions from sources greater than 25,000 metric tons of CO₂ equivalent were from power plants. Comparable data on industrial sources are not available outside of the U.S. and for the purposes of this study we assume these industrial sources to be part of the areal sources of emissions (hereafter “non-point sources”).

There are thus 6 components of uncertainty that need to be combined for an estimate of total uncertainty for the cells in the modified CDIAC database:

- Uncertainty in total national emissions
- Magnitude uncertainty for large point sources
- Spatial uncertainty for large point sources
- Magnitude uncertainty of the population proxy

- Spatial uncertainty for the population proxy
- Uncertainty in using population density as a proxy for emissions

For this analysis all calculations are based on one-sigma uncertainty. In combining the component uncertainties we assume that, except for the uncertainty in the national total, each of the components are independent of each other and they are therefore combined in Euclidean fashion (the square root of the sum of the squares) on a grid cell by grid cell basis. The uncertainty in the national total is passed to all components equally across grid cells. In this analysis we use data for emissions in 2009 as published by CDIAC in 2013. Temporal uncertainty would need to be considered in developing a time series of emissions.

Uncertainty in total national U.S. emissions is estimated at 2.5% (one sigma). This value is based on comparisons with other inventories, U.S. EPA analyses, and literature research of the components which combine to calculate total national U.S. emissions. (see also Andres et al. [2]). According to the U.S. EPA [9] the 95% confidence interval for total U.S. CO₂ emissions from fossil fuel combustion is -2% to +5%. Instead of using this asymmetric value we take the symmetric value of $\pm 5\%$, and since this is two standard deviations about the mean and our computations are all based on one standard deviation, the estimated national error used in our computations is $\pm 2.5\%$.

A random sample of 500 large point sources from the US EPA eGRID data set was taken in order to find the exact locations of the power plant discharges. We used Google Earth satellite imagery to identify the point sources and to verify each latitude and longitude. With spatial information from Google Earth, the distance between the actual location and the reported location was computed for each point source in the sample. The maximum distance from the reported location to the observed location of a point source was approximately 106 km. The mean distance from the reported location for all of the sample point sources (excluding zero) was 1.97 km. The mean distance from the reported locations for all of the point sources in the sample was 0.84 km. The latter value was then used as the mean spatial uncertainty. The spatial uncertainty for the top 81 emitters was larger than for the random sample of points. It was found that 60% were farther than 1km from the reported location. The mean difference in location was 7.94 km and the maximum spatial difference was about 122 km.

The information gathered from the 500-item random sample suggested that the differences between discharge locations and eGRID reported locations might be attributed to: 1.) differences between the plant site and the exact location of the CO₂ discharge stack, 2.) use of default locations in the EPA database, such as the centroid of a county, when the initial report to EPA did not include plant coordinates, 3.) typographical errors, 4.) reporting the location of a company office or mailing address instead of the plant site, and 5.) dealing with the existence of multiple stacks on the same site.

The locations of power plants are not part of a continuous distribution and therefore most traditional statistical methods do not work well in dealing with the uncertainty in their emissions. The discrete, or binary, nature of the locations (a plant either is in a given grid space or it is not) spurred the creation of a new method for dealing with the likely locations and the uncertainty in the emissions from power plants and the development of a new statistic, PSUM = Point Source Uncertainty Measure, [6] which we treat as a standard deviation in the analyses here. The uncertainty results are scale dependent, as with any spatial uncertainty. For a given locational uncertainty, the larger the grid cell the greater the probability that the point source will actually be found

in the reported cell. Monte Carlo analyses were run using the magnitude of emissions and the reported location for each point source as well as the calculated mean spatial uncertainty and the size of the geographic grid cells. A resulting grid of simulated means effectively distributes the reported CO₂ emissions from a point source to surrounding cells based on the fraction of the total number of simulation executions that fell in each cell.

The magnitude uncertainty for emissions from large point sources is taken to be a constant $\pm 10.62\%$ (one sigma) [10]. This number was derived by comparing data collected on smokestack emissions of U.S. electric power plants with emissions calculated from fuel deliveries at the same plants. The value 10.62 was the mean of the difference of the two measurements.

LandScan is a recently developed global data set [11] that estimates the average locations where people actually are rather than where their home location is. LandScan was first produced in 1998 and data sets for 2000-2012 are now available. CDIAC has contemplated use of LandScan population data but has not yet made the conversion. The CDIAC gridded CO₂ emissions data set relies, however, on a 1984 population distribution data base from the Goddard Institute for Space Studies [12]. While this GISS data set allows an estimation of the spatial distribution of CO₂ emissions over a long time series, an additional contribution to uncertainty results from the changes in urbanization and population distribution that have occurred since 1984. Although the GISS data are used as the proxy for distributing emissions, we use LandScan characteristics here to illustrate the uncertainty that could be achieved for the post-2000 time period.

Magnitude uncertainty in LandScan for the U.S. was assumed to be comparable to the estimates of uncertainty derived by the U.S. Census Bureau at the same spatial scale [13]. LandScan does not currently have any published estimates for uncertainty but we assume that it is very low in the U.S. As in all of the data sets used here, the uncertainty will vary by country or region in a global analysis. The uncertainty estimate provided by the U.S. Census Bureau is 0.01%. Spatial uncertainty in LandScan was estimated by looking at the changes incurred as a result of small shifts in the cell boundaries. We took the LandScan data set and distributed CO₂ emissions proportional to the population density values associated with each grid cell. We then shifted the grid cells by 10 kilometers (approximately one tenth of a grid cell in the central U.S.) in each direction (N, S, E, and W) so that each grid cell contained successively one tenth of each of the four surrounding cells. This effectively creates a weighted sum in which the central cell emissions value is weighted by 90% and the cell that is shifted towards the center is weighted by the remaining 10%. A weighted sum was computed for each of the four shifts that occurred. The standard deviation for the resulting four weighted sums was then computed and stored as the uncertainty value within the central cell.

In order to characterize the uncertainty associated with using population density as a proxy for CO₂ emissions we started with the per capita emissions in each state [14] (with data on large point sources removed) and the mean number of grid spaces per state. We calculated the standard deviation in per capita emissions from non-point sources at the state level and took this as a measure of the variability in the relationship between population density and emissions density at that scale. We assumed that the variability among states provides a measure of the variability at the grid level within states. This provides enough information that we can back-calculate to estimate the standard deviation in emissions by grid cell attributed to the population proxy. Then we can use this standard deviation as the uncertainty estimate for using population density as a proxy for emissions at the grid cell level.

3. Discussion and conclusions

The combination of the six aspects of uncertainty produces a total uncertainty, by grid space, for a hypothetical, modified CDIAC dataset. Figure 1 shows these results as a percent of total emissions in each grid space. Recall that these numbers apply to a hypothetical data set – one in which 1.) the data on large point sources have been substituted for an equal quantity of emissions that were previously distributed according to population density and 2.) the data on population density have an uncertainty attributed to data from the U.S. Census Bureau at the same scale. In both cases we expect that the best achievable uncertainty will be much higher in many other countries where the data on population and large point sources have greater uncertainty.

The scale shown on Figure 1 is quite high. This was not unexpected because of how much uncertainty there is for the exact location and magnitude for CO₂ sources at this scale. Figure 1 shows that the uncertainty associated with the modified CDIAC data set is consistently around 150% of the emissions total for each grid space. Recall that these estimates of uncertainty are for a modified CDIAC data set where we have now isolated large point sources before using population density to distribute the remaining emissions from non-point sources. We have also treated the population density data as though they had been derived from Landsat, thus avoiding the shifts in population density that have occurred since construction of the Goddard Institute for Space Studies data set for 1984. Recall too that this uncertainty is for individual grid cells of 1 degree scale and is very scale dependent. There is strong correlation among grid spaces because of the spatial uncertainty about the exact placement of large point sources and because the national total is a defined constant.

Total uncertainty is the combination of all of the components, but we also learn something of the role that each component takes in forming the whole. With an understanding of the relative magnitude of each of the pieces, and the locational characteristics of where each component is large, we can try to target specific efforts to best reduce the total uncertainty. Table 1 provides summary statistics on the different components of uncertainty. Table 1 indicates that proxy uncertainty has the highest mean percentage of all the components. In particular a full 52% of grid cells have 90% of their uncertainty coming from proxy uncertainty. The implication here is that in the majority of grid cells, reduction of uncertainty can only be done by addressing uncertainty in the proxy relationship. This means that we must obtain a better understanding of the relationship between the proxies we use and the emissions they are meant to represent. Contributions from large point sources often dominate uncertainty for the grid cells where large point sources are present. And, the values here depend very much on the geographic scale. Uncertainty will increase for many reasons if the grid size is decreased without reducing the parameters of spatial uncertainty.

Our efforts to systematically estimate the uncertainty in a gridded data set of CO₂ emissions suggest that the uncertainty is quite high in the U.S. and it is probably higher in many countries where data on large point sources and the distribution of population are less well documented. Uncertainty will increase as the geographic scale is decreased. While data users need to appreciate the data uncertainty, the best data are probably suitable for many purposes. The analyses suggest that at 1-degree latitude/longitude resolution the current uncertainty (one standard deviation) by grid space in the U.S. is on the order of +/- 150%. Taking this analysis to a global scale will require additional analysis to characterize spatial uncertainty for each country or group of similar countries.

Acknowledgement:

Portions of this work were supported by NASA CMS grant NNH11ZDA001N-CMS.

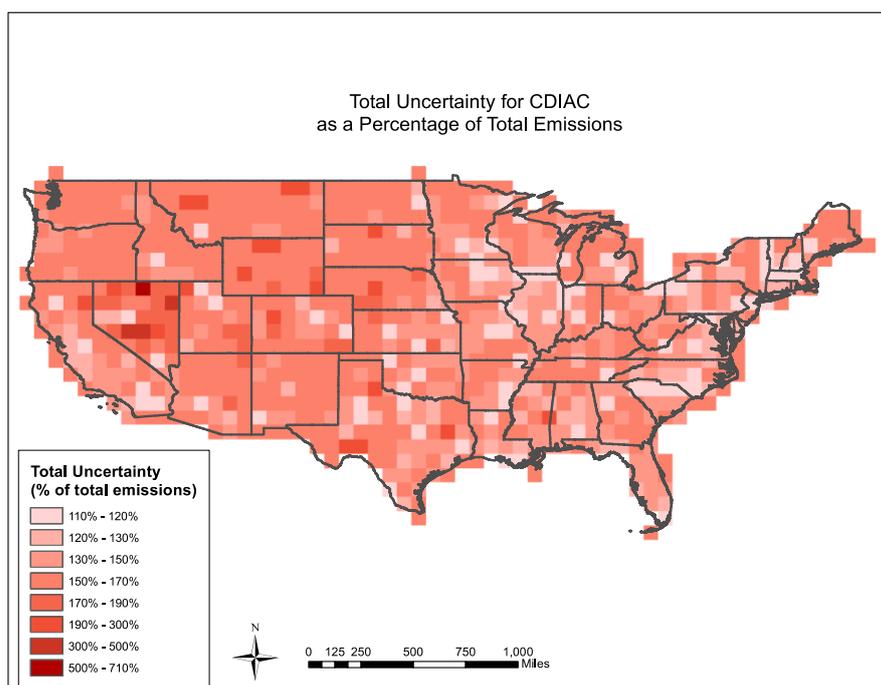


Figure 1. Uncertainty by grid space shown as a percentage of total emissions at 1-degree resolution. Areas shown with very high uncertainty are often a result of cities with abrupt changes in population density. Excluding these few areas of very high uncertainty we can see that the overall uncertainty in grid spaces is on the order of 1.5 times the total emissions. This is for a hypothetical, modified version of the CDIAC data set for 2009 (see text).

Table 1. By grid cell, a breakdown of each component of uncertainty with its summary statistics. Uncertainty in the national total is not included since it affects each of the grid cells equally. Note that the country borders create problems in that emissions may or may not occur even if a grid cell is designated as predominantly ocean, and some of the zero values lie along the eastern shoreline of the U.S. This is one of the challenges of cropping a global data set to a single country for a targeted analysis. Values are given as the percent uncertainty in a single grid cell.

	Min	1 st Quart.	Median	Mean	3 rd Quart.	Max
Magnitude Uncertainty, large point sources	0.00	0.00	13.96	41.70	94.88	110.60
Spatial Uncertainty, large point sources	0.00	0.00	12.62	37.69	85.77	100.00
Magnitude Uncertainty, LandScan	0.00	0.00	0.01	0.01	0.01	0.01
Spatial Uncertainty, LandScan	0.00	0.09	0.55	10.34	2.93	673.00
Proxy Uncertainty	0.00	22.99	141.20	100.70	161.50	161.50
Total Uncertainty, Emissions Data at Grid Cell Level	112.9	138.3	153.6	154.2	166.4	712.9

(Quart.=Quartile)

References Cited

- [1] Andres, R.J., T.A. Boden, and G. Marland, 2013. Annual fossil-fuel CO₂ emissions: mass of emissions gridded by one degree latitude by one degree longitude. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/ffe.ndp058.2013
- [2] Andres R.J., T.A. Boden, and D. Higdon, 2014. A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. *Tellus B*, 66, 23616. doi:10.3402/tellusb.v66.23616.)
- [3] Rayner, P.J., M.R. Raupach, M. Paget, P. Peylin, and E. Koffi, 2010. A new global gridded data set of CO₂ emissions from fossil fuel combustion: methodologies and evaluation, *J. of Geophysical Research* v. 115, pp. D19306, doi: 10.1029/2009JD013439.
- [4] Singer, A., M. Branham, M. Hutchins, J. Welker, D. Woodard, C. Badurek, E. Marland, and G. Marland, 2014. "The role of CO₂ emissions from large point sources in emissions totals, responsibility, and policy". *Environmental Science and Policy*, 44:190-200.
- [5] U.S. EPA, 2014. Clean Energy: eGRID, ninth edition with 2010 data. U.S. Environmental Protection Agency, <http://www.epa.gov/cleanenergy/energy-resources/egrid/>, released February 24, 2014.
- [6] Woodard, D., M. Branham, G. Buckingham, S. Hogue, M. Hutchins, R. Gosky, G. Marland, and E. Marland, 2015. A spatial uncertainty metric for anthropogenic CO₂ emissions. *Greenhouse Gas Measurement and Management*, 4:1-22. doi:10.1080/20430779.2014.1000793
- [7] Hutchins, M.G.H., J.D.C. Colby, G. Marland, and E. Marland, 2015. A comparison of five high-resolution spatially-explicit fossil fuel carbon dioxide emissions inventories, in review.
- [8] U.S. EPA, 2013. GHGRP 2011: Reported Data, U.S. Environmental Protection Agency, Greenhouse Gas Reporting Program. <http://www.epa.gov/ghgreporting/>
- [9] U.S. EPA, 2015. U.S. Greenhouse Gas Inventory Report: 1990-2013, U.S. Environmental Protection Agency, <http://epa.gov/climatechange/ghgemissions/usinventoryreport.html>
- [10] Quick, J.C., 2014. Carbon dioxide emission tallies for 210 U.S. coal-fired power plants: a comparison of two accounting methods. *J. Air and Waste Management Assoc.* 64:73-79.
- [11] Oak Ridge National Laboratory, 2014. LandScan, 2014. <http://web.ornl.gov/sci/landscan/>, retrieved June 10, 2015.
- [12] Andres, R.J., G. Marland, I. Fung, and E. Matthews, 1996. A 1° x 1° distribution of carbon dioxide emissions from fossil fuel combustion and cement manufacture, 1950-1990. *Global Biogeochemical Cycles*10(3): 419-429.
- [13] U.S. Census Bureau, 2011. Population distribution and change: 2000 to 2010: Census Briefs. <http://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf>. retrieved June 10, 2015
- [14] U.S. DOE/EIA, 2014. Environment: State-level energy-related carbon dioxide emissions 2000-2011 U.S. Department of Energy, Energy Information Administration <http://www.eia.gov/environment/emissions/state/analysis/>

Uncertainty associated with fossil fuel carbon dioxide (CO₂) gridded emission datasets

Tomohiro Oda^{1, 2, *}, Lesley Ott¹, Petro Topylko³, Mariia Halushchak³, Rostyslav Bun^{3, 4}, Myroslava Lesiv⁵, Olha Danylo^{3, 5}, Joanna Horabik-Pyzel⁶

¹ Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

² Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD, USA

³ Lviv Polytechnic National University, St. Bandery, 12, Lviv, 79013

⁴ Academy of Business in Dąbrowa Górnicza, Poland

⁵ International Institute for Applied Systems Analysis, Laxenburg, Austria

⁶ Systems Research Institute of Polish Academy of Sciences, Warsaw, Poland

**Tomohiro.Oda@nasa.gov*

Abstract

CO₂ emissions from fossil fuel combustion (FFCO₂) serves as a reference in carbon budget analysis and thus needs to be accurately quantified. FFCO₂ estimates from different emission inventories often agree well at global and national level, however their subnational emission spatial distributions are unique and subject to uncertainty in the proxy data used for disaggregation of country emissions. In this study, we attempt to assess the uncertainty associated with emission spatial distributions in gridded FFCO₂ emission inventories. We compared emission distributions from four gridded inventories at a 1 × 1 degree resolution and used the differences as a proxy for uncertainty. The calculated uncertainties typically range from 30% to 200% and inversely correlated with the emission magnitude. We also discuss limitations of our approach and possible difficulties when implemented at a higher spatial resolution.

Keywords: Emission inventory, carbon dioxide, fossil fuel emissions, uncertainty, atmospheric inversion

1. Introduction

CO₂ emission from fossil fuel combustion (FFCO₂) serves as a reference in carbon budget analysis where carbon uptake by natural processes is typically the biggest unknown. The uncertainty associated with global total FFCO₂ was estimated as 8% (2 sigma) by the work by Andres and co-authors [1] and FFCO₂ estimates from difference emission inventories often agree well at global and country level [2].

Disaggregation of country emissions is a common method to develop a gridded emission inventory. Emission spatial distributions are estimated using spatial proxy data such as population density/counts [e.g. 3, 4] and satellite-observed nightlights [e.g. 5, 6] for diffused sources, geographical locations of point sources (e.g. power plant, cement production facilities and steel furnaces) [e.g. 6, 7, 8] and line sources such as road and railroad networks, aircraft and ship tracks [e.g. 7], and combinations of those.

While the uncertainties for country total emissions are thought to be small (e.g. 5% for US), the emission disaggregation step can introduce significant errors in emissions estimates at higher spatial resolutions. The errors introduced will be propagated through atmospheric transport model simulations and subsequent budget analyses. Thus, it is critical to quantify and characterize the uncertainties (errors) associated with the spatial

distributions of fossil fuel emissions. In this study, we present a new approach to assess an uncertainty associated with spatial distributions in gridded FFCO₂ emission inventories.

2. Method

In disaggregation of country emissions, emission magnitude at grid cells can be achieved by multiplying the mass of the total emission to normalized spatial proxy data. In this example, a single emission sector is assumed along with the use of single proxy data for simplification

$$E_{i,j} = M_{Total} \times W_{i,j} \quad (1)$$

$$W_{i,j} = P_{i,j} / \sum_{i=1}^m \sum_{j=1}^n P_{i,j} \quad (2)$$

where $E_{i,j}$ is the emission magnitude of grid (i,j) , M_{Total} is the total mass of emissions for the domain of interest, $W_{i,j}$ is the weight given as a normalized proxy value of $P_{i,j}$. With a rule of combined uncertainty, the percent uncertainty of $E_{i,j}$ can be calculated as a combination of percentage uncertainties for (a) total emission mass and (b) weight at grid cell

$$\delta E_{i,j}/E_{i,j} = \sqrt{(\delta M_{Total}/M_{Total})^2 + (\delta W_{i,j}/W_{i,j})^2} \quad (3)$$

The uncertainty for the total mass is often available, however the challenge is to estimate the uncertainty for the second term based on spatial disaggregation. One could use the uncertainty estimates for the proxy data to estimate the second term in the root in the equation (3). A limitation of such approach is the inability of accounting for emissions that are not represented by proxy data used. The spatial distributions are unique and subject to the proxy data used for disaggregation of national emissions. Thus, the uncertainty assessment is specific to a particular disaggregation method and do not reflect the fact that some of the emission features might not be addressed by the underlying methodology.

In this study, we compared emission distributions from four emission inventories that are based on different disaggregation methods and used the differences to estimate the second term. We normalized the emission datasets to the same global total and calculated the mean and standard deviation at grid cells as follows:

$$\delta W_{i,j}/W_{i,j} \sim SD_{i,j}/Mean_{i,j} \quad (4)$$

We used four emission datasets used in a recent atmospheric inversion intercomparison work by Peylin and co-authors [9]: Emission dataset developed by Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORLN) [3] (hereafter CDIAC), the Open source Data Inventory for Anthropogenic CO₂ (ODIAC) [6] and two versions of Emission Database for Atmospheric Research (EDGAR) (v4.2 and Fast Track) [7]. Those emission datasets share some of underlying country level data and/or spatial proxy data, but the authors believe that their methods have produced distinct emission spatial distributions and can be considered to be reasonably different from each other. In common atmospheric CO₂ inversion, FFCO₂ is assumed to be the flux quantity with the least uncertainty and, unlike natural fluxes, is not optimized [e.g. 10]. The differences we would see in

different gridded FFCO₂ emission inventories are possible sources of uncertainty associated that could propagate into flux inverse estimates.

3. Results and discussions

Figure 1 shows spatial distributions of four different gridded emission datasets (CDIAC, ODIAC, EDGAR/v4.2, EDGAR/FastTrack). Here only emissions over land (normalized to the same total) were presented. We aggregated four different gridded inventories to a common 1 x 1 degree domain. The major patterns we can see in the global distributions are driven by country emissions estimates that have very good agreement in general, especially for top emitting countries. The differences seen in subnational emission distributions are largely attributable to the differences among disaggregation methods. In CDIAC and ODIAC for instance, areas with no emission are spreading over northern high latitudes and some desert areas such as Africa and the center part of Australia. This can be explained by the fact that CDIAC and ODIAC do not have an explicit representation of emissions from road network and its spatial distributions, while the two versions of EDGAR do. If uncertainty associated with spatial distributions is assessed just using a single gridded emission inventory, the emission distribution discrepancy like we found between CDIAC/ODIAC and EDGAR would not be addressed.

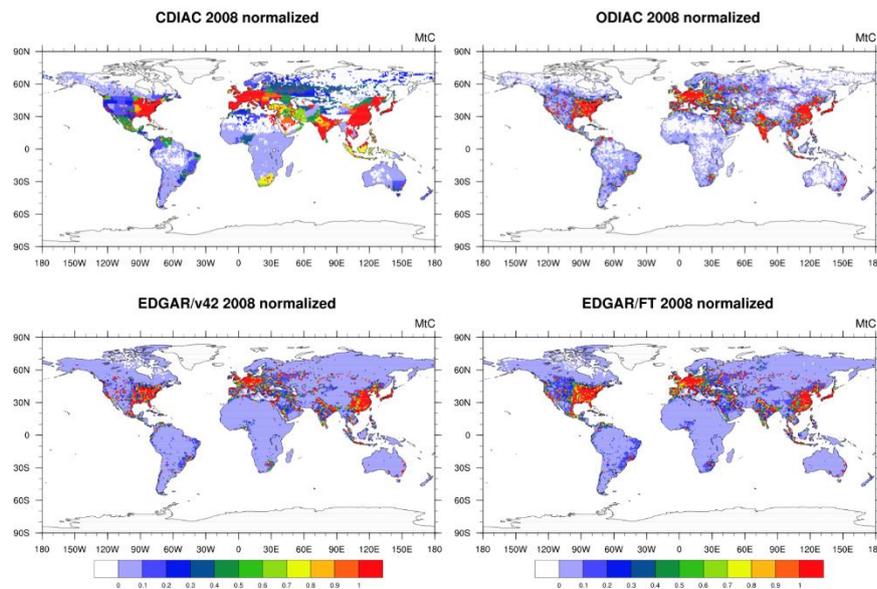


Figure 1. Spatial distributions of fossil fuel emissions from four different emission inventories (CDIAC, ODIAC, EDGAR v4.2 and EDGAR Fast track). Emission fields for the year 2008 were aggregated to a common 1×1 degree resolution and then global total are scaled to the same total as CDIAC. Values are given in the unit of mega tonne Carbon per year.

Our uncertainty estimate associated with emission spatial distributions is shown in Figure 2. The values in the map were calculated as standard deviation of emission values at grid cell from four different gridded inventories (normalized) divided by mean of the four, as briefly described in the section 2. The calculation was implemented at a common 1 x 1 degree resolution. The calculated uncertainties typically range from 30% to 200%. The uncertainty tends to be lower over areas with intense emissions and higher over the areas with relatively low emissions. This seems to be qualitatively reasonable

if we take emissions from road network as an example. The final uncertainty (uncertainty for emission estimates and spatial distribution) can be achieved by combining the uncertainty with the uncertainty for the global total emissions using a root-square fashion.

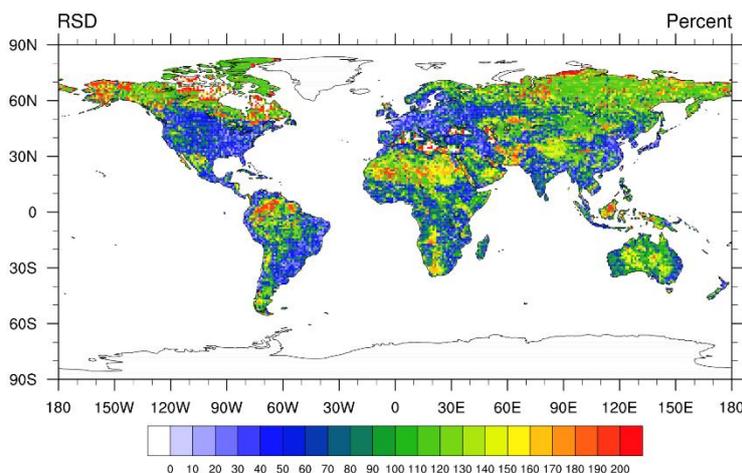


Figure 2. Estimates of uncertainty associated with the spatial disaggregation of country emissions. Values are given in the unit of percentage (%).

The uncertainty estimates from this approach are directly applicable to the studies such as atmospheric inversions where FFCO₂ is not optimized and, as a result, is assumed to be perfect. Our approach might be limited by the fact that the variety of disaggregation methods (hence, emission spatial distributions) is not rich enough to implement this type of analysis, although we attempted selected four emission datasets that contain substantial differences. We also acknowledge that the use of multiple emission dataset for estimating the spatial uncertainty does not assure it addresses all the possible error sources. This approach might not work if implemented a much higher spatial resolution where distinct spatial patterns of sector emissions become more visible. The proxy data is as it says “proxy” and in fact is not explicitly representing unique dynamics of human activities. At a high spatial resolution, geolocation information of sources would become a key to achieve accurate emission spatial distributions [e.g. 6]. Currently, it seems less common to collect geolocation information of sources as we regularly do for activity data. Collecting activity data with geolocation information for example would allow us to precisely map emissions and greatly improve emission spatial distributions even at a high spatial resolution.

4. Summary

We present a method to estimate the uncertainty associated with disaggregation of national emissions. We compared four different gridded inventories and used the differences as a proxy for the spatial uncertainty. The calculated uncertainty typically range from 30~200% and are inversely correlated with emission magnitudes. This seems to be qualitatively reasonable considering the fact intense emissions are relatively easy to identify and weak emissions are often difficult to place. We also discussed some of the methodological limitations we have identified.

Demands for an emission inventory gridded at a high spatial resolution has been increased as observational data have become rich and modelling capabilities have been improved to facilitate higher resolution transport and inversion modeling. The emission

inventories face a difficulty in achieving accurate spatial distributions at increasingly high resolutions. Collection of additional information associated with emission sources (e.g. geolocation) could greatly help us to accurately map emissions and assess uncertainties associated the resulting emission spatial distributions.

Acknowledgement

Tomohiro Oda is a recipient of the NASA Grant #NNX14AM76G.

References

- [1] Andres, R. J., Boden, T. A. & Higdon, D. A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. *Tellus B Chem. Phys. Meteorol.* 66, 23616 (2014)
- [2] Andres, R. J., Boden, T. A., Bréon, F.-M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., and Treanton, K.: A synthesis of carbon dioxide emissions from fossil-fuel combustion, *Biogeosciences*, 9, 1845-1871, doi:10.5194/bg-9-1845-2012, 2012.
- [3] Andres, R. J., G. Marland, I. Fung, and E. Matthews (1996), A $1^\circ \times 1^\circ$ distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990, *Global Biogeochem. Cycles*, 10(3), 419–429, doi:10.1029/96GB01523.
- [4] Olivier, J.G.J., Van Aardenne, J.A., Dentener, F., Pagliari, V., Ganzeveld, L.N. and J.A.H.W. Peters (2005). Recent trends in global greenhouse gas emissions: regional trends 1970-2000 and spatial distribution of key sources in 2000. *Env. Sc.*, 2 (2-3), 81-99. DOI: 10.1080/15693430500400345.
- [5] Rayner, P. J., M. R. Raupach, M. Paget, P. Peylin, and E. Koffi (2010), A new global gridded data set of CO₂ emissions from fossil fuel combustion: Methodology and evaluation, *J. Geophys. Res.*, 115, D19306, doi:10.1029/2009JD013439.
- [6] Oda, T. and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights, *Atmos. Chem. Phys.*, 11, 543-556, doi:10.5194/acp-11-543-2011, 2011.
- [7] Janssens-Maenhout G, Dentener F, Van Aardenne J, Monni S, Pagliari V, Orlandini L, Klimont Z, Kurokawa J, Akimoto H, Ohara, T, Wankmueller R, Battye B, Grano D; Zuber A, Keating T. EDGAR-HTAP: a Harmonized Gridded Air Pollution Emission Dataset Based on National Inventories. Ispra (Italy): European Commission Publications Office; 2012. JRC68434, EUR report No EUR 25 299 - 2012, ISBN 978-92-79-23122-0, ISSN 1831-9424
- [8] Asefi-Najafabady, S., P. J. Rayner, K. R. Gurney, A. McRobert, Y. Song, K. Coltin, J. Huang, C. Elvidge, and K. Baugh (2014), A multiyear, global gridded fossil fuel CO₂ emission data product: Evaluation and analysis of results, *J. Geophys. Res. Atmos.*, 119, 10,213–10,231, doi:10.1002/2013JD021296
- [9] Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an

ensemble of atmospheric CO₂ inversions, *Biogeosciences*, 10, 6699-6720, doi:10.5194/bg-10-6699-2013, 2013.

- [10] Gurney, K. R., Y.-H. Chen, T. Maki, S. R. Kawa, A. Andrews, and Z. Zhu (2005), Sensitivity of atmospheric CO₂ inversions to seasonal and interannual variations in fossil fuel emissions, *J. Geophys. Res.*, 110, D10308, doi:10.1029/2004JD005373.

Remapping gridded data using Artificial Intelligence: real world challenges

Jörg Verstraete¹

¹ Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland
Jorg.Verstraete@ibspan.waw.pl

Abstract

Working with spatial data, regardless of the specific content (emission data, population data, land use data, etc.), requires dealing with gridded datasets. A grid is a commonly used representation method for data, where a value of interest is associated with each cell of the grid. While very adequate for representing and analysing data, combining data from different sources implies working with different grids, and this is more complicated. In this article, we present some preliminary findings of applying a novel approach to map spatial data onto a different grid. The approach simulates intelligent reasoning through the use of artificial intelligence and employs additional knowledge to help create a high quality remapping. We also present the difficulties in applying this methodology in real world applications.

Keywords: map overlay problem, grid remapping, spatial operations, artificial intelligence

1. Introduction

Data regarding atmospheric emissions are one example of data that carries a spatial dependency. For research purposes, e.g. to investigate the exposure of a population or to correlate data, it is often necessary to combine data from different sources. As data tend to come from different sources, the grids on which they are defined can be incompatible: different size of cells, different orientation or a combination. This is called the *map overlay problem* and it occurs when the data of a grid cell in one grid needs to be correlated or even compared to data that is presented on a different, incompatible grid. The simplest solution is to use areal weighting, which allows to remap one grid onto a different grid, using the amount of overlap of the cells as weights used to redistribute the data. This however implicitly assumes that the data are uniformly distributed within each single grid cell. While this assumption may hold for some data, or even for some cells, it is not always a valid assumption. The current approaches are ignorant to the fact that many other data and knowledge are available; some of this data may be known to exhibit a correlation to the data that we need to remap. In [1], we presented the first concept of an artificial intelligent system that is able to perform the remapping of one grid onto another grid, using this additional information to improve the spatial distribution of the modeled data. Following the first concept, several implementations were made and experiments were performed. Here, we present our findings regarding the challenges ahead when needing to apply this method on real world data.

In the next section, the representation of spatial data is shortly introduced, while an introduction to the artificial intelligent system used it in Section 3. The challenges with real world data are presented in Section 4, followed by the conclusion.

2. Spatial data

2.1 Spatial data representation

Spatial data can be represented in one of two commonly used models: feature based or field based [2,3]. In a feature based model, basic geometric objects are used to represent real world objects: lines are used to represent roads, polygons represent areas, etc. In a field based approach, a numeric value that carries a spatial component (e.g. emission values) are modeled over a region of interest. This can be achieved using triangular networks (commonly used for e.g. altitudes) or grids. In the application of modelling emissions, grids are more common. In a grid, the region of interest is partitioned in a number of grid cells, that completely cover the region of interest; if all cells have the same shape and size, the grid is considered regular. The cell of a grid is considered the smallest possible unit. For a grid that represents e.g. emission values, the grid provides no information regarding the distribution of the emission within each cell: the emission can be concentrated in one part of the cell, uniformly spread over the cell, or can have any other distribution. This causes problems when incompatible grids – grids that have ill aligned grid cells – as there is no easy mapping from one grid onto another grid. Examples of incompatible grids are shown in Figure 1.

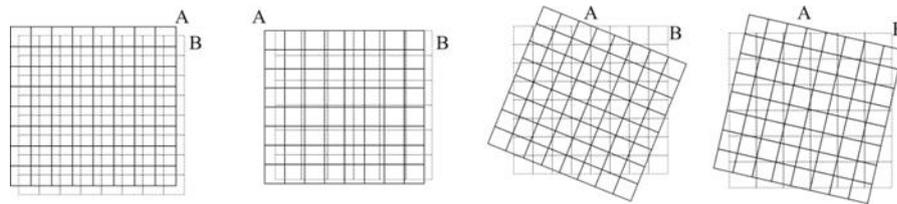


Figure 1. Examples of incompatible grids: shifted, different cell size, rotated or a combination.

2.2 Grid remapping algorithms

The map overlay problem occurs when one grid needs to be remapped onto another grid. Several approaches exist in literature, but all of them make either an implicit or an explicit assumption regarding the underlying distribution. For more details we refer to [4], and briefly describe the most common methods below.

The easiest and most commonly used algorithm for grid remapping is areal weighing. This approach implicitly assumes a uniform distribution of the data in each grid cell individually. The calculation to remap one grid onto another is very easy, as it suffices to consider the relative amount a cell of one grid overlaps with a cell of the other grid. This is illustrated on Figure 2. While effective in many cases, the method fails when the assumption does not hold.

Spatial smoothing is a second approach. Here, the modeled data is considered as a third dimension, which is subsequently smoothed and resampled. This is illustrated on Figure 2. The implicit assumption here is that the data is smooth of the entire grid and the performance of the method depends on the accuracy of this assumption.

The last method mentioned is spatial regression, where a priori statistical assumptions on the distribution of the data are used to control the grid remapping. The application of this method requires expert knowledge and quite complicated calculations. The distribution of the data is explicitly assumed here, but such knowledge may not be available.

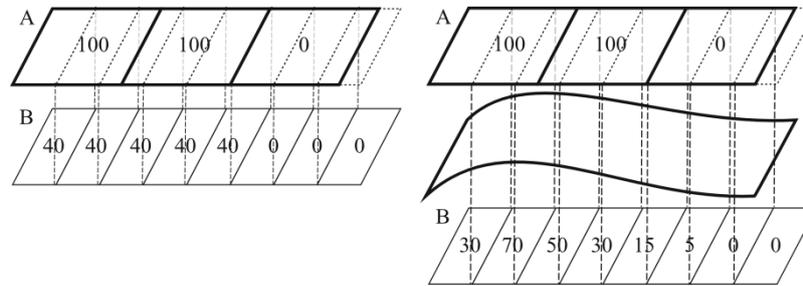


Figure 2. Illustrations of areal weighing (left) and spatial smoothing (right).

3. Artificial intelligent method and spatial grid remapping

3.1 Short introduction to fuzzy set theory

Fuzzy set theory is an extension to set theory, presented by Zadeh in [5]. In a fuzzy set, each of the elements carries a value from the interval $[0,1]$, this is the membership grade μ . This value can have one of three interpretations [6]: as a degree of membership – in which case it expresses “how much” the element belongs to the set, as a degree of certainty – in which case it reflects how certain it is the element belongs to the set, or finally as a degree of possibility – to indicate how possible it is the element belongs to the set. As such, a fuzzy set is defined by means of a traditional set and an associated membership function, which maps each element to its membership degree. Many applications of fuzzy sets exist [7], but for the application here we consider the possibilities of representing imprecise values and linguistic terms. An imprecise value (e.g. approximately 50) can be represented by a fuzzy set which has a membership grade 1 for the element 50, and decreasing membership grades as values are further from 50, as indicated on Figure 3. A linguistic term (such as “small”) can be represented as illustrated on Figure 3: 0 is considered small with degree 1, larger numbers have decreasing membership grades and numbers above 50 are not considered small (they have membership grade 0). Examples for the terms “medium” and “large” are also on Figure 3. The definitions of course depend on the domain and application and the provided fuzzy sets are just an example.

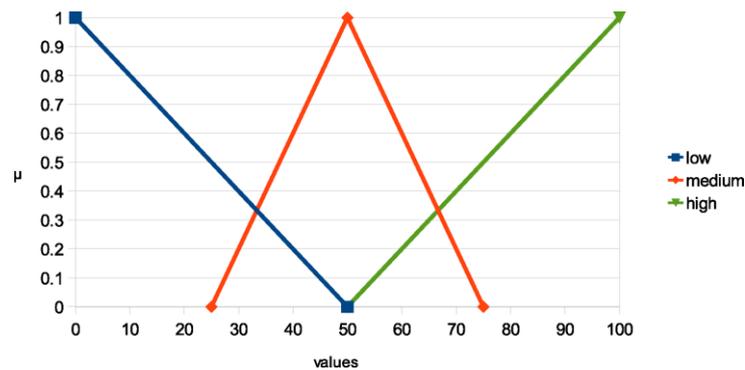


Figure 3. Examples of fuzzy sets used to represent low, medium and high numbers (on a scale from 0 to 100).

3.2 Fuzzy rulebase systems

Artificial Intelligence is a term that covers many approaches; for the presented algorithm, a rulebase system [8] is considered. Fuzzy systems have proven their effectiveness in control and applications can be found in many household appliances. A rulebase consists of a number of rules, for example:

if x is small then y is small
 if x is medium then y is medium
 if x is high then y is high

Here, x is an input parameter, which is a normal number, high/medium/low are linguistic terms represented by fuzzy sets, and y is an output parameter. All the rules are evaluated, so a given x can be both high and medium at the same time (e.g. using the definitions on Figure 3), but each to a different extent. Each rule results in a fuzzy set for y ; the outputs of all rules are aggregated and defuzzified to yield the final result. For more details on this method, we refer to [8].

3.3 Concept of grid remapping using a fuzzy rulebase

The use of a fuzzy rulebase system to perform grid remapping requires the creation of the rulebase. The first problem is: what are the parameters used in the rulebase (x in the above example)? Several parameters can be considered as mentioned in [9], but in general parameters are what allow the additional data to be used. One example for a parameter is the amount of overlap of the auxiliary grid with the input cell under consideration. The second problem is: to define low/medium/high, it is necessary to find limits for the parameter. A lower limit could be the value of the grid cells of the auxiliary grid that are fully contained by the cell, whereas an upper value could be the total value of the grid cells of the auxiliary grid that intersect the cell.

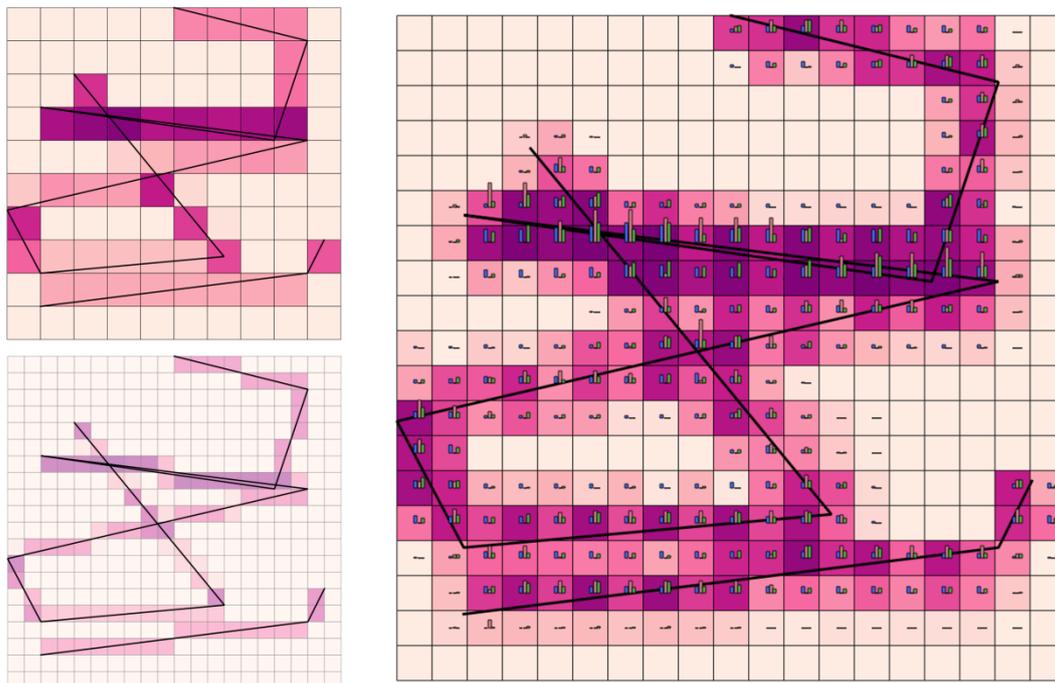


Figure 4. Example of the remapping algorithm: input (top-left), auxiliary data (bottom-left) and result (right). The line pattern shows the underlying distribution, the bar charts in the result cells show - from left to right - the result obtained through areal weighting, the ideal result and the result obtained with the presented approach.

Once it is known which parameters can be used, an appropriate rulebase can be constructed. To determine the result of the remapping, it suffices to apply the rulebase

for every output, calculating the parameters and evaluating the rulebase to yield a fuzzy result for the output cell. This concept is explained in more detail in [1].

An example is shown on Figure 4. The example is still artificial, but should highlight problems that also can occur on real world data. The dark lines are the underlying lines that contain data and from which the grids were defined. The shades in the result reflect the values obtained with the presented method. Compared to areal weighting, it is clear the method is able to identify the 2 nearly horizontal lines near the bottom, whereas areal weighting sees them as one big region. The presented method tends to assign lower values to cells that are located further away from the ideal line, which is also desirable. On the two nearly horizontal lines near the bottom, an alternating pattern is visible, from left to right, which is an artifact introduced by the method. The next section aims to explain the origin of these and other problems and ties them in to real world situations.

4. Challenges related to applying the rulebase system

Various prototype implementations and proofs of concept have proven that the methodology can work. However, initial attempts at applying the prototype implementations on real world data have revealed some issues that still need resolving. First, there are problems of a more technical nature, described in the next 2 subsections; next there are problems related to the real world data itself, described in the subsequent 2 subsections. The first two problems can still be resolved using artificially generated examples, but the latter two would benefit from real world data.

4.1 Mathematical precision

The first problem relates to the way numbers are handled on a computer system. Coordinates of grid cells are represented by floating point values, which have a limited precision on a computer system. The consequences of this are very well explained in [10, Chapter 4], and in particular they pose problems for parallel or near parallel lines, which is the case for the grid cells. The presented algorithm highly depends on correctly assessing the overlap between intersections and calculating intersection areas. Initial test data suffered less from such problems, as the coordinates tended to be more artificial.

An example of what happens when values get incorrectly rounded can be seen on Figure 5, where the thick lines indicate locations where the intersection is incorrectly identified. Such errors can lead to wrong limits for the parameters. One work around for this was recently developed and is presented in [11], a more general workaround has been developed but still needs to be verified.

4.2 Parameter definitions and their limits

The application of the rulebase requires both parameter values, lower limits and upper limits. All three of them have equal importance, as a poorly defined lower and upper limit can make the parameters useless. In [9], we presented some suggestions for parameters. The example on Figure 4 was performed using a single parameter that relates to overlap of the auxiliary data. While the data does concentrate more towards the lines, the end result somewhat reflects the auxiliary grid: this explains the fluctuations in data that should be constant. Use of multiple parameters, even relating to the same data, should neutralize this effect, while still providing a result that shows

a better distribution. Research in additional parameters is currently the next phase of the research.

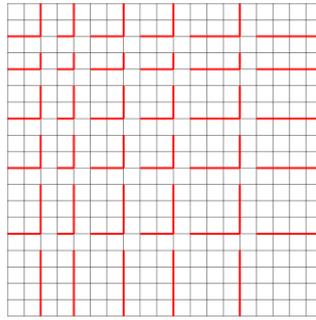


Figure 5. Example of a grid where rounding of the coordinates cause incorrect identification of intersection: the thicker lines are lines where neighbouring gridcells are incorrectly identified as intersecting.

4.3 Relative grid distributions

This topic is somewhat related to the previous topic, but there is a difference. Independent of the parameter definitions is the impact of the relative positions of the different grids. The parameter currently considered in the example is completely useless if the grid onto which the data is remapped exactly divides each grid cell. The reason for this is that in this situation, the parameter value, its lower limit and its upper limit will all be equal, and therefore no evaluation can be made by the rulebase system. The example on Figure 4 shows a fluctuating pattern along the near-horizontal lines: the values are alternating higher and lower. This is not desirable, as their values ought to be the same or at least similar: neither to input data nor the auxiliary data indicate this alternating pattern. The reason for its occurrence is the combination of the parameter that was used and the relative grid position. This is a second example that shows an effect that can happen, if the combination of the grid distributions and the parameters is not ideal. Solving this is both tied to solving the problems mentioned in 4.2, but also in making sure the auxiliary supplied data is indeed useful and of good quality for the considered problem.

4.4 Availability and quality of the data

The last aspect relates to the data itself. The assumption is made that auxiliary data are available, which for many research will be the case. The correlation between the data to be remapped and the auxiliary data should be known from prior research and not discovered on the data set at hand. The data should also be of good quality: grid obtained from down-sampling an existing grid might appear to be of higher precision, but internally is not. Regardless of the parameters implemented in the system, the use of such grids may provide unsatisfactory results.

5. Conclusions

The article shortly describes a novel approach to remap gridded spatial data and lists the challenges in bringing the approach from theory to practise. The biggest challenges are listed, along with the ideas that will be pursued to solve them.

References

- [1] Verstraete J. (2014) Solving the map overlay problem with a fuzzy approach. *Climatic Change*, pp. 1–14.
- [2] Rigaux P., M. Scholl, and A. Voisard (2002) *Spatial databases with applications to GIS*. Morgan Kaufman Publishers.
- [3] Shekhar S. and S. Chawla (2003) *Spatial databases: a tour*. Pearson Educations.
- [4] Gotway C. A. and L. J. Young (2002) Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458) pp. 632–648.
- [5] Zadeh L. A. (1965) Fuzzy sets. *Information and Control*, 8 pp. 338–353.
- [6] Dubois D. and H. Prade. (1999) The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90 pp. 141–150.
- [7] Zimmerman H.J. (1999) *Practical applications of fuzzy technologies*. Kluwer Academic Publishers.
- [8] Mendel J. M. (2001) *Uncertain rule-based fuzzy logic systems, Introduction and new directions*. Prentice Hall.
- [9] Verstraete J. Parameters to use a fuzzy rulebase approach to remap gridded spatial data. (2013) *Proceedings of the 2013 Joint IFSA World Congress NAFIPS Annual Meeting (IFSA/NAFIPS)*, pp. 1519–1524.
- [10] Hoffman C. M. (1992) *Geometric and Solid Modeling*. Chapter 4. <https://www.cs.purdue.edu/homes/cmh/distribution/books/geo.html>
- [11] Verstraete J. (2015) Dealing with rounding errors in geometry processing. (accepted for publication in the *Proceedings of FQAS*, October 2015)

Integration of multi-source information in disaggregation of spatial emission data

Joanna Horabik-Pyzel¹, Zbigniew Nahorski¹

¹ Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
Joanna.Horabik@ibspan.waw.pl, Zbigniew.Nahorski@ibspan.waw.pl,

Abstract

Quantification of CO₂ emissions at fine spatial scales is advantageous for many environmental, physical, and socio-economic analyzes; in principle it can be easily integrated with other data in gridded format. It is especially important for better assessment of carbon cycle and climate change. Some possibilities exist for incorporation of the additional knowledge to improve the results and their uncertainty. There is, for example, a constant progress in assessment of local emissions from observations done in the atmosphere. This information can possibly help in improving disaggregated emission estimates. This paper discusses these questions and outline possibility of using these additional knowledge for improving estimation of emissions in fine scales.

Keywords: Greenhouse gases emissions, spatially resolved data, disaggregation, integration of multi-model results

1. Introduction

This paper is meant as a discussion paper which attempts to overview the research results that can be used in improving spatial gridded GHG estimates at a fine resolution by using existing or possible to obtain information coming from different sources. This problem is intrinsically connected with uncertainties of the used information, as it is quite intuitively evident that a more sure information should be more credited in integration of knowledge than a less sure one, and therefore the former should be more weighted in the final result than the latter.

Quantification of GHG emissions at fine spatial scales is advantageous for many environmental, physical, and socio-economic analyzes; in principle, it can be easily integrated with other data in a gridded format. This is especially important for improved assessment of carbon and other chemical component cycles and climate change. To better understand the transport of different pollutants atmospheric dispersion models are used. This way influence of emissions can be confronted with the atmospheric concentration measurements. In this modelling two factors are considered to be mainly responsible for modelling errors: emission accuracy and meteorology. This is the reason for a battle for high accuracy gridded emission estimates.

In some applications, of particular importance are estimations of fossil fuel CO₂ fluxes, which are used to quantitatively estimate CO₂ sources and sinks, see e.g. [5]. A few institutions gather data on emissions from fossil fuels at national levels, like the US Department of Energy Carbon Dioxide Information Analysis Center (CDIAC) [6, 50]; the International Energy Agency (IEA) [35]. IPCC gathers data from national GHG inventories within the Kyoto Protocol agreement and its continuation [36]. British Petroleum company compiles energy statistics [7] that can be conveniently used for estimation of national CO₂ emissions. These datasets have been used for estimating global sources and sinks on a regional (e.g. continental) scale, [2, 11, 29, 59, 63, 64]. Their resolution is, however, too small to be directly useful for very fine emission grids.

Much less data are available on emissions in the below-national scales. Some countries publish data for provinces, but their scale is still too rough compared with other contemporary studies, like those presented in the sequel. Spatial disaggregation of emissions introduces additional uncertainty to a developed inventory. This is why the option of using additional information to reduce this uncertainty is of great interest.

A common approach to disaggregation of emissions is a usage of proxy data, which are most often areas of fine grids or population therein. However, independent estimates of GHG fluxes, like inverse modeling or eddy covariance, provide opportunity to incorporate additional knowledge and provide more comprehensive spatial quantification of carbon budget. Evidently, there is a mismatch between distinct approaches to estimation of fluxes. In general, there are two kinds of estimates: it can be either accounting of emissions (bottom-up) or by measuring concentrations of CO₂ and inferring about original emission fluxes (top-down). Merging such datasets is a challenging task due to incomplete accounting and uncertainties underlying each of the methods. Moreover, one should take into account various spatial scales and different scarcity of data. This paper outlines several methods, discusses advantages and limitations of using them for improving inventory emission estimates in fine scales, and review methods used for combining uncertain data sets, highlighting the issues related to spatial dimension of the task.

2. Disaggregation based on proxy data

2.1. Basic research stream

Disaggregation methods for obtaining high-resolution emissions typically use proxy data available in finer scales. The most straightforward approach to estimate data in a fine scale is to disaggregate national emissions proportionally to gridded population information, see e.g. [1, 55, 67] or in some cases proportionally to the area. Another proxy data are the satellite observations of nighttime lights [16, 17]. Direct use of these proxy data does not allow for very fine resolutions, as emissions from some sources, like power plants, do not correlate well with proxies. That is why Oda and Maksyutov [53] extracted emissions from point sources before disaggregating the non-point emissions proportionally to the nightlight distribution, and integrated them again to obtain 1km × 1km emission data. Rayner et al. [61] used a modified Kaya identity, in which emissions are modeled as a product of population density, per capita economic activity, energy intensity of economy, and carbon intensity of energy to predict emissions from several sectors, namely energy, manufacturing, transport (broken to land, sea, and air emissions).

A very high resolution of emission cadasters (2km × 2km grid) was obtained for Poland within the 7th FP GESAPU project [24]. It resulted from a detailed analysis of information from various sources, published by governmental and research agencies, as well as energy or industry plants (e.g. taking part in emission trading scheme); the analysis was followed by disaggregation on activity levels and precise modeling, see [8, 9]. At present, this approach to disaggregation seems to provide the best results. Nevertheless, the relevant procedure, like gathering data from numerous sources and publications or individual disaggregation of multitude of variables, requires immense input of human work, so this approach is as far suitable rather only for regions of a country or a few countries.

The GESAPU approach allows for rather straightforward assessment of uncertainty of disaggregated data, following the IPCC guidelines Tier 1 (error propagation) or Tier

2 (Monte Carlo method) methodology ([37, 38]. Again, Bun et al. [9] and accompanying papers ([12, 13, 30, 71] provide details of the analysis. As well, it is based on individual examination of all sources and sinks. Hogue et al. [31] discuss problems of doing similar analysis using existing global databases. Oda et al. [54] proposes a method of calculating uncertainty parameters. This method will be mentioned in the sequel.

2.2. Extensions

An intrinsic possibility is to use more than one proxy data for disaggregation. This can be done using linear regression function. The problem is to estimate its parameters, as no data for the fine resolution grid are given usually. Using data from other regions is questionable. Ghosh et al. [25] calculated the correlation between the nighttime lights and the Vulcan data [28] compiled for USA with the resolution $10\text{km} \times 10\text{km}$, and then used this correlation to calculate disaggregated emissions for other countries, but this approach did not provide satisfactory results. An ad-hoc method was proposed to improve this approach.

A method to use regression function, in a more general context of spatially autocorrelated data, was proposed in [32]. Its idea is to estimate the regression function parameters for the coarse grid and use them in the regression function for the fine grid. This method works well when the coarse and fine grid cell areas differ not more than a few to a dozen times ([32, 34]). But its range of applicability depends very much on the similarity of correlations on different area scales.

This methods enables automatic calculation of uncertainty distribution arising from statistical inference, see [33].

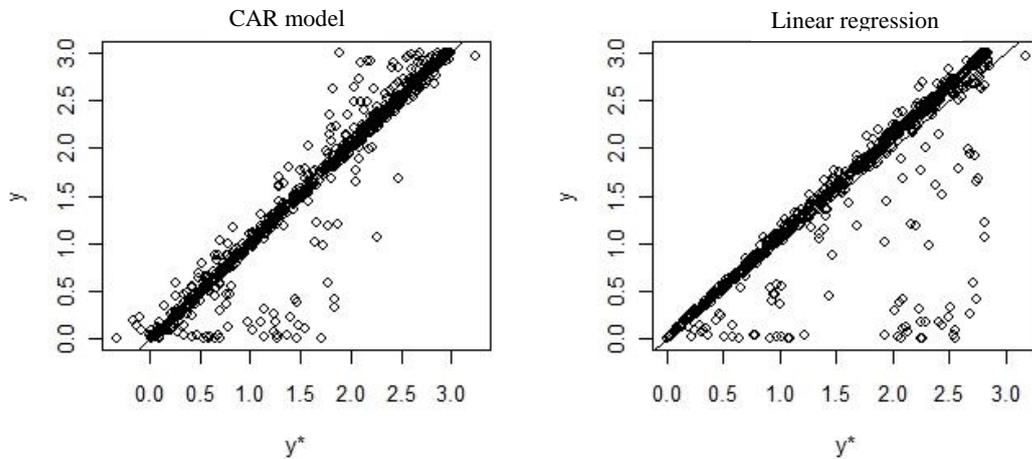


Figure 1. Predicted (y^*) versus observed (y) values

Figure 1 presents scatterplots of predicted values versus observations for original ammonia data in $5\text{km} \times 5\text{km}$ grid as well as the values disaggregated from $10\text{km} \times 10\text{km}$ grid. For the disaggregation from $10\text{km} \times 10\text{km}$ to $5\text{km} \times 5\text{km}$ grids, the mean square error (MSE) was 0.064 (conditional autocorrelation (CAR) model) and 0.186 (regression). Although introduction of spatial dependence evidently improved accuracy of prediction, the linear regression method gives pretty good match.

A problem with using regression function is in estimating zero emissions, i.e. emissions (or activities) for the cells where they do not exist. This problem, known also under the name semicontinuous variables, clumped-at-zero or zero-inflated data, can also happen in the disaggregation proportional to one proxy variable, but it is more

acute for the multivariable case. In this problem a variable has a continuous probability distribution for positive values and a nonzero probability mass at zero. Simple cutting off of the negative values is perhaps operative, but scientifically not well justified. Min & Agresti [52] discuss this problem and review methods of dealing with it. They will be not discussed here.

3. Other sources of information

3.1. Atmospheric observations

Two kind of observations can be of value to constrain inventory estimates and perhaps improve their accuracies. One is measurement of the specific gas concentration or mixing ratio. An example can be found in [75]. The problem is to partition the estimated atmospheric load obtained possibly from subtracting a background concentration, to emission sources. This is typically done using the inversion methods to estimate CO₂ fluxes. For this, the Bayes estimator is generally applied ([18, 68]). To use the inversion method, the function which relates emission with concentration (footprints) is needed. It is typically computed using the atmospheric dispersion models and is finally of the linear form

$$\mathbf{y}_{\text{obs}} = \mathbf{H}\mathbf{x} + \boldsymbol{\psi} \quad (1)$$

where \mathbf{y}_{obs} is an m -vector of the measured atmospheric concentrations (mixing ratios) in the receptor points, in space and time, above the background value, \mathbf{x} is an n -vector of fluxes (emissions) from sources in the region considered, and \mathbf{H} is the matrix that relates emissions in sources to the measurements. The elements of the $m \times n$ matrix \mathbf{H} are computed using a transport model. It is assumed that they are constant in the considered time period, which may be a rough approximation. $\boldsymbol{\psi}$ is an m -vector of uncertainties of the relation (10); it is modeled as a random variable with the Gaussian distribution

$$p(\boldsymbol{\psi}) = [(2\pi)^m \det \mathbf{C}_y]^{-1} \exp \left\{ -\frac{1}{2} \boldsymbol{\psi}^T \mathbf{C}_y^{-1} \boldsymbol{\psi} \right\} \quad (2)$$

The real fluxes are unknown but it is assumed that uncertain information on fluxes $\mathbf{x}_{\text{prior}}$ is given, so that

$$\mathbf{x} = \mathbf{x}_{\text{prior}} + \boldsymbol{\vartheta} \quad (3)$$

where again, the uncertainty is modeled as a random vector with the Gaussian distribution, independent on $p(\boldsymbol{\psi})$,

$$p(\boldsymbol{\vartheta}) = [(2\pi)^m \det \mathbf{C}_x]^{-1} \exp \left\{ -\frac{1}{2} \boldsymbol{\vartheta}^T \mathbf{C}_x^{-1} \boldsymbol{\vartheta} \right\} \quad (4)$$

Using the Bayes theory the conditional probability $p(\mathbf{x}|\mathbf{y}_{\text{obs}})$ is given by

$$p(\mathbf{x}|\mathbf{y}_{\text{obs}}) = \frac{p(\mathbf{y}_{\text{obs}}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y}_{\text{obs}})} \quad (5)$$

It is proportional to

$$p(\mathbf{x}|\mathbf{y}_{\text{obs}}) \sim \exp \left\{ -\frac{1}{2} \left[(\mathbf{y}_{\text{obs}} - \mathbf{H}\mathbf{x})^T \mathbf{C}_y^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{H}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_{\text{prior}})^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_{\text{prior}}) \right] \right\} \quad (6)$$

After some manipulations the value $\hat{\mathbf{x}}$ which maximizes the above conditional probability is obtained which gives the Bayes estimator of the fluxes

$$\hat{\mathbf{x}} = \mathbf{x}_{\text{prior}} + (\mathbf{H}^T \mathbf{C}_y^{-1} \mathbf{H} + \mathbf{C}_x^{-1})^{-1} \mathbf{H}^T \mathbf{C}_y^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{H}\mathbf{x}_{\text{prior}}) \quad (7)$$

The statistical uncertainty of the Bayesian estimator can be calculated as a covariance matrix

$$\hat{\mathbf{C}}_x = (\mathbf{H}^T \mathbf{C}_y^{-1} \mathbf{H} + \mathbf{C}_x^{-1})^{-1} = \mathbf{C}_x - \mathbf{C}_x \mathbf{H}^T (\mathbf{H} \mathbf{C}_x \mathbf{H}^T + \mathbf{C}_y)^{-1} \mathbf{H} \mathbf{C}_x \quad (8)$$

The estimate $\hat{\mathbf{x}}$ is the sum of the prior estimate plus a correction, which depends on the deviation of observations from their predicted values. This correction improves the initial estimate of fluxes (e.g. obtained from disaggregation of the inventory estimates). The expression (8) informs us that the errors of the improved estimates (the values on the diagonal of $\hat{\mathbf{C}}_x$) are not bigger (and very likely smaller) than the errors of the a priori estimate.

To use the above expressions, one has to know estimates of the covariance matrices $\hat{\mathbf{C}}_x$ and \mathbf{C}_y . This issue is discussed in numerous papers, e.g. [45, 57, 63]. Various methods of finding appropriate values have been proposed; very often diagonal matrices have been used. Exponential decay of covariance values, both in space and/or time, has been found to match the reality better. Michalak et al. [51] develop a maximum likelihood method for estimating the covariance parameters. The likelihood function is formulated and the Cramér-Rao bound is derived.

The idea to use the likelihood function approach has been also used in the so-called geostatistical inverse modelling [27]. In this setting, instead of using prior information, emissions are modelled as linear combinations of trends. More advanced modelling of the fluxes has been proposed in the so-called assimilation data method proposed by Kaminski et al. [39], and then used e.g. in [60]. In this method, a more thorough model of emissions from the biosphere is included.

The above expressions have been used mostly in flux inversion studies. Ciais et al. [14] provide various comments on practical applications of this sort of methods. Peylin et al. [57] use them for estimating monthly European CO₂ fluxes and report 60% reduction of errors. Rivier et al. [63] apply them for estimating monthly fluxes of CO₂ from the biosphere and ocean for the global and European scale. The Bayesian estimate errors are reduced therein by 76% for the western and southern Europe, and by 56% for the central Europe. Lauvaux et al. [45] give inversion results for a 300km × 300km region in the South-West of France near Bordeaux with the 8km × 8km resolution of CO₂ fluxes, reporting about 50% error reduction. Continuous measurements were taken in two towers, and two aircrafts measuring CO₂ were used. Thompson et al. [70] estimated the N₂O fluxes in the western and central Europe. With only one in-situ measurement point used for inversion, they obtained between 30% and 60% error reduction for Germany.

The idea of atmospheric inversion methods is very general, and it can be used for improving estimates given any additional information in a suitable form. Atmospheric measurements are rather rare in space, so it may be difficult to obtain significant improvement for a very fine spatial grid for large areas. However, using local measurements and fine gridded a priori data the good resolutions can be achieved. For example, Gałkowski [21] obtained this way emission estimates with the resolution of few kilometers using measurements performed the stations in Kraków, and at Kasprowy Wierch located on a mountain in Polish Tatras (1989 m a.s.l.) some 100 km south of Kraków. This resolution is only a few times coarser than the very fine resolutions of 1-2 km, which have been obtained by using inventory data and disaggregation based on proxies.

Atmospheric inversion methods seem nowadays to be the most important approaches used to constrain estimates of emission fluxes from the biosphere.

3.2. Measurements of tracers

Measurement of tracers connected with emissions helps to identify better the fluxes. The most important tracer is ¹⁴C isotope. The ¹⁴C isotope is produced by cosmic

radiation in the upper atmosphere, and then it is transported down and absorbed by living organisms. The ^{14}C isotope decays in time of a few hundred years (its half-life equals approximately 5700 years), while the fossil fuels come from organisms which lived million to hundred million years ago. Intensive burning of the fossil fuels dilutes the atmospheric concentration of the ^{14}C isotope [66]. This way (a lack of) ^{14}C isotope may be used as a tracer of fossil fuel originated CO_2 emissions, and the rate of dilution can be used to assess local/regional/global emissions of fossil fuel CO_2 .

The ^{14}C isotope has not been the only tracer of CO_2 emissions considered. Also, SF_6 and CO have been investigated [22, 47, 72], but ^{14}C has been found to be the most useful and directly available. Lopez et al. [49] used additional tracers of CO , NO_x , and $^{13}\text{CO}_2$, besides that of $^{14}\text{CO}_2$, to estimate relative fossil fuel (from liquid and gas combustion) and biosphere fossil fuel (from biofuels, human and plant respiration) CO_2 in Paris, and got good agreement.

Estimation of the fossil fuel CO_2 basically comes from two mass balance equations, for CO_2 and ^{14}C (or $^{14}\text{CO}_2$), which are presented in the concentration form (or, more often, in the mixing ratio form; the mixing ratio s is defined as $s = \rho_c/\rho_a$, where ρ_c is a CO_2 density and ρ_a is the air density)

$$\text{CO}_2^{\text{obs}} = \text{CO}_2^{\text{bg}} + \text{CO}_2^{\text{ff}} + \text{CO}_2^{\text{bio}} + \text{CO}_2^{\text{other}} \quad (9)$$

$$^{14}\text{C}^{\text{obs}} = ^{14}\text{C}^{\text{bg}} + ^{14}\text{C}^{\text{ff}} + ^{14}\text{C}^{\text{bio}} + ^{14}\text{C}^{\text{other}} \quad (10)$$

where the superscripts stand for, respectively, the observed ($^{\text{obs}}$) mixing ratio, background ($^{\text{bg}}$) mixing ratio – without the local fossil fuel emission, fossil fuel ($^{\text{ff}}$) mixing ratio, biosphere (photosynthesis and heterotrophic respiration) component ($^{\text{bio}}$), and other components, like those coming from burning of biomass, nuclear industry or ocean ($^{\text{other}}$). The ^{14}C isotope is typically measured as a relative difference between the (^{13}C corrected) sample and absolute rate [40, 65]

$$\Delta^{14}\text{C} = \frac{\left(\frac{^{14}\text{C}}{\text{C}}\right)_{\text{obs}} - \left(\frac{^{14}\text{C}}{\text{C}}\right)_{\text{abs}}}{\left(\frac{^{14}\text{C}}{\text{C}}\right)_{\text{abs}}} \quad (11)$$

where the absolute ($_{\text{abs}}$) value is the absolute radiocarbon standard ($1.176 \cdot 10^{-12} \text{ mol}^{14}\text{C}/\text{molC}$), related to oxalic acid activity. Equation (11) is usually expressed in per mill (‰) and written as

$$\Delta^{14}\text{C} = \left[\frac{\left(\frac{^{14}\text{C}}{\text{C}}\right)_{\text{obs}}}{\left(\frac{^{14}\text{C}}{\text{C}}\right)_{\text{abs}}} - 1 \right] \cdot 1000 [\text{‰}] \quad (12)$$

After some transformation the following final equation can be obtained

$$\Delta^{14}\text{C}^{\text{obs}} \text{CO}_2^{\text{obs}} = \Delta^{14}\text{C}^{\text{bg}} \text{CO}_2^{\text{bg}} + \Delta^{14}\text{C}^{\text{ff}} \text{CO}_2^{\text{ff}} + \Delta^{14}\text{C}^{\text{bio}} \text{CO}_2^{\text{bio}} + \Delta^{14}\text{C}^{\text{other}} \text{CO}_2^{\text{other}} \quad (13)$$

From (9), the concentration of one component can be calculated and inserted to (13). The choice of the eliminated component depends in principle on possibility of measuring the values in the equations, and the case considered. For example, having eliminated CO_2^{obs} , the equation for CO_2^{ff} is found as follows

$$\text{CO}_2^{\text{ff}} = \frac{(\Delta^{14}\text{C}^{\text{obs}} - \Delta^{14}\text{C}^{\text{bg}}) \cdot \text{CO}_2^{\text{bg}}}{\Delta^{14}\text{C}^{\text{ff}} - \Delta^{14}\text{C}^{\text{obs}}} + \frac{(\Delta^{14}\text{C}^{\text{obs}} - \Delta^{14}\text{C}^{\text{bio}}) \cdot \text{CO}_2^{\text{bio}}}{\Delta^{14}\text{C}^{\text{ff}} - \Delta^{14}\text{C}^{\text{obs}}} + \frac{(\Delta^{14}\text{C}^{\text{obs}} - \Delta^{14}\text{C}^{\text{other}}) \cdot \text{CO}_2^{\text{other}}}{\Delta^{14}\text{C}^{\text{ff}} - \Delta^{14}\text{C}^{\text{obs}}}$$

As the concentration (and mixing ratio) of ^{14}C in the fossil fuel CO_2 is equal to 0, then from (11) we have $\Delta^{14}\text{C}^{\text{ff}} = -1000$. It is often assumed that $\text{CO}_2^{\text{other}} = 0$, particularly when a site is far from other sources. Other assumptions may be appropriate for the area

considered, as this methodology can be applied to the studies of different scales, ranging from the global ones to small-scale.

Various authors discuss assumptions and assess underlying uncertainties. Turnbull et al. [73] present a systematic discussion and quantify uncertainties using modelling and the above equations.

Another important aspect is the choice of location for measurements of background values. Since the measurements are usually taken for long time periods, the background values are taken from observations at high-altitude sites. In Europe, commonly used background observations come from the High Alpine Research station Jungfraujoch at 3450 m a.s.l. in the Swiss Alps. Other local sites considered in Europe are the Vermunt station in Austria (1800 m a.s.l.) and the Schauinsland in Germany (1205 m a.s.l.). In Poland, there is an observation site at Kasprowy Wierch (1989 m a.s.l.) in the High Tatra Mountains, which can be used as a regional reference station [44]. Turnbull et al. [73] estimate differences of 1-3 ‰ due to the choice of a background site.

The resolution of CO_2^{ff} determination depends, first of all, on a spatial distribution of $\Delta^{14}\text{C}^{\text{ff}}$ measurements. The $\Delta^{14}\text{C}^{\text{ff}}$ measuring observation stations are rather scarce. For instance, in 2008 there were only 10 measurement sites in Europe [56]. Much better spatial resolution can be obtained using measurements in plant materials, like corn leaves, rice, grape wine ethanol, grass, tree leaves, and tree rings. Most of them allow only for annual estimation, so measurements have to be done for many years to get longer time series. Only wine ethanol and tree rings enable historical records. This way Palstra et al. [56] was able to measure ^{14}C in 165 different wines from 32 different regions in 9 different European countries. The measurements were compared with those obtained from a regional atmospheric transport model, predicting fossil fuel CO_2 with the resolution $55\text{km} \times 55\text{km}$, with a good compatibility. Riley et al. [62] used measurements from winter annual grasses collected at 128 sites across California, USA, to model transport of fossil fuel CO_2 by using a regional transport model with the resolution $36\text{km} \times 36\text{km}$. These resolutions are still not high enough to be directly useful in very fine gridded cells and need to be disaggregated for this or used for improving estimates in coarser grid.

3.3. Direct local measurements of fluxes

The fluxes can be also measured. The fluxes from big chimneys are actually estimated with quite good accuracy. But also fluxes coming from the biosphere or urban environment can be measured using several methods. Observations from the flux towers are taken above the plant canopies, and use the so-called eddy covariance method. The basic idea of the eddy covariance can be found in [10]. Foken & Wichura [19] discuss the connected errors. Other possible measurements use chamber system, see an example in [75], to measure fluxes coming from the soil.

The flux tower observations could be a perfect way to provide very high resolution emission fluxes from the biosphere both in space and time provided that a net of flux towers is dense enough. Unfortunately, flux towers are rather scarce. Even over the large area of USA and Canada, only 36 flux tower observations are reported [58]. Their use can be therefore rather considered in the future, when more flux towers are constructed. At present, they are used mainly for an assessment of biosphere emission models, see [3] or [58].

When using local flux measurements, particularly coming from the soil, some problems may arise with high spatial variability of the obtained results. For example, Gałkowski [21] reports three times difference between measurements of nitrous oxide fluxes from the soil in a distance of few meters.

To be useful in estimation of areal data, the point measurements have to be interpolated. Geostatistical methods, like Kriging [15, 23], are usually applied. It is, however, suitable only for homogeneous fields. Bayesian melding [20] is a method that can cope with usual inhomogeneity of pollution fields, see e.g. [48]. These methods have been developed to include time dependence, [4, 46].

4. Combining multi-model estimates

There are many different ways to combine results of models. Three groups are distinguished in this paper, although the authors do not claim that they cover all published methods.

Averaging. Simple averages, see e.g. Oda (2015), can be quite efficient. Weighted averages

$$\bar{x} = \frac{\sum_i w_i x_i}{\sum_i w_i}$$

perform usually better, but the weights w_i have to be defined in them. Usual method of determining weights from the historical differences between the model output and observations cannot be used, as observations of real emission values do not exist. The situation is a bit similar to combining models in climate projections, see [41, 69] for review of the methods used there. As the models use projections for the future, both the discrepancies of the model output from the ensemble mean in the current time and in the future are considered there. For example, in reliability ensemble average method proposed in [26] the weights are calculated as the product of two terms inversely proportional to the absolute values of these discrepancies. Adaptation to emission models could use only one discrepancy. The calculations can follow iteratively.

Bayesian approach. The Bayesian methodology proposed in the atmospheric inversion may be applied for combining model results, which give independent information complementing each other. In this case, the following function has to be minimized

$$J = (\mathbf{y}_{\text{obs}} - \mathbf{x})^T \mathbf{C}_y^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{x}) + (\mathbf{x} - \mathbf{x}_{\text{prior}})^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_{\text{prior}}) \quad (14)$$

where \mathbf{y}_{obs} is the vector of the emission estimates from the first model and $\mathbf{x}_{\text{prior}}$ is the vector of estimates from the second one. The solution of the minimization problem is

$$\hat{\mathbf{x}} = \mathbf{x}_{\text{prior}} + (\mathbf{C}_y^{-1} + \mathbf{C}_x^{-1})^{-1} \mathbf{C}_y^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{x}_{\text{prior}}) = \mathbf{x}_{\text{prior}} + \mathbf{C}_x (\mathbf{C}_y + \mathbf{C}_x)^{-1} (\mathbf{y}_{\text{obs}} - \mathbf{x}_{\text{prior}}) \quad (15)$$

and the estimate of the improved estimate covariance matrix takes the form

$$\hat{\mathbf{C}}_x = (\mathbf{C}_y^{-1} + \mathbf{C}_x^{-1})^{-1} = \mathbf{C}_x - \mathbf{C}_x (\mathbf{C}_x + \mathbf{C}_y)^{-1} \mathbf{C}_x \quad (16)$$

Particularly simple computations are obtained for diagonal covariance matrices \mathbf{C}_x and \mathbf{C}_y . In this case the above formulae read

$$\hat{x}_i = x_{i,\text{prior}} + \frac{c_{ii,x}}{c_{ii,x} + c_{ii,y}} (y_{i,\text{obs}} - x_{i,\text{prior}}), \quad i = 1, \dots, n \quad (17)$$

$$\hat{c}_{ii,x} = \frac{1}{\frac{1}{c_{ii,x}} + \frac{1}{c_{ii,y}}} = \frac{c_{ii,x} c_{ii,y}}{c_{ii,x} + c_{ii,y}}, \quad i = 1, \dots, n \quad (18)$$

It is readily seen that $\hat{c}_{ii,x} \leq c_{ii,x}$ and $\hat{c}_{ii,x} \leq c_{ii,y}$. This procedure can be generalized to more than two models.

Joint probability distribution approach. This approach has been proposed by Kryazhinsky [42]. In his approach no weighting is used. He operates on the joint probability distribution (multivariable distribution) obtained as the product of the distributions of individual distributions under assumption of their independence, as in the bivariate case

$$p(x, y) = p(x)p(y)$$

although the method can be applied to a higher multivariate distribution as well. In opposition to the two previously mentioned ones, this approach has been as yet not extensively evaluated in practical applications. It was used in [43] for combining estimates of net primary production of forest obtained from two models.

5. Final remarks

A preliminary review of possibilities of using additional knowledge to improve fine gridded estimates of GHG emissions is presented. Besides the mentioned above, there may be still more information that can help in better estimation of gridded emissions, but not dealt with in the paper. Different constraints on local emission (like lack of specific sources in the cells) can be possibly used in obtaining better accuracy. There may be, for example, independent emission assessments done on part of considered regions. Also some common sense knowledge can be used. One of important problems connected with integration of different knowledge is mismatch of the grids used in different studies, often spotted in real applications. These problems are discussed [74], who also presents an approach based on artificial intelligence methods to solve them. Many practical difficulties are also pointed to in [31].

A basic question is how much all additional knowledge can improve the estimates obtained by using proxy variables. This probably will depend on specific case. Not much can be probably expected in the case of industrial emissions, particularly of the carbon dioxide gas. However, for such emissions like nitrous oxide, N₂O, from the biosphere, which is very poorly estimated by present techniques, introduction of, say, atmospheric inversion methods can perhaps give a considerable improvement. These questions can be solved only by investigations of specific cases. This makes the area for interesting research projects.

Acknowledgements

Joanna Horabik-Pyzel was supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing; project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

References

- [1] Andres R.J., Marland G., Fung I., Matthews E. (1996) A $1^0 \times 1^0$ distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990. *Global Biogeochemical Cycles*, 10:419-429.
- [2] Baker D.F., Law R.M., Gurney K.R., Rayner P., Peylin P., Denning A.S., Bousquet P., Bruhwiler L., Chen Y.H., Ciais P., Fung I.Y., Heimann M., John J., Maki T., Maksyutov S., Masarie K., Prather M., Pak B., Taguchi S., Zhu Z. (2006) TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO₂ fluxes, 1988-2003. *Global Biogeochemical Cycles*, 20(1), GB10002, DOI: 10.1029/2004GB002439.
- [3] Baldocchi D., Meyers T. (1998) On using eco-physiological, micrometeorological and biochemical theory to evaluate carbon dioxide, water vapour and trace gas

- fluxes over vegetation: a perspective. *Agricultural and Forest Meteorology*, 90:1-25.
- [4] Banerjee S., Carlin B., Gelfand A. (2003) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC. London, Boca Raton.
- [5] Battle M., Bender M.L., Tans P.P., White J.W.C., Ellis J.T., Conway T., Francey R.J. (2000) Global carbon sinks and their variability inferred from atmospheric O₂ and δ¹³C. *Science*, 287:2467-2470.
- [6] Boden T.A., Marland G., Andres R.J. (2010) Global, regional, and national fossil-fuel CO₂ emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, USA. DOI: 10.3334/CDIAC/00001_V2010.
- [7] BP (2014) Statistical Review of World Energy. London. <http://www.bp.com/content/dam/bp/pdf/Energy-economics/statistical-review-2014/BP-statistical-review-of-world-energy-2014-full-report.pdf>.
- [8] Boychuk K., Bun R. (2014) Regional spatial inventories (cadastres) of GHG emissions in the energy sector: Accounting for uncertainty. *Climatic Change* 124:561-574.
- [9] Bun R., Nahorski Z., Horabik-Pyzel J., Danylo O., Charkovska N., Topylko P., Halushchak M., Lesiv M., Striamets O. (2015) High resolution spatial inventory of GHG emissions from stationary and mobile sources in Poland: summarized results and uncertainty analysis. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [10] Burba G., Anderson D. (2007) Introduction to the Eddy Covariance method: general guidelines and conventional workflow. LI-COR Biosciences, <http://www.instrumentalia.com.ar/pdf/Invernadero.pdf>
- [11] Canadell J.G., Le Quéré C., Raupach M.R., Field C.B., Buitenhuis E.T., Ciais P., Conway T.J., Gillett N.P., Houghton R.A., Marland G. (2007) Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences*, 104(47):18866-18870, DOI: 10.1073/pnas.0702737104.
- [12] Charkovska N., Bun R., Danylo O., Horabik-Pyzel J. (2015) Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [13] Charkovska N., Halushchak M., Bun R. (2015) Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [14] Ciais P., Rayner P., Chevallier F., Bousquet P., Logan M., Peylin P., Ramonet M. (2010) Atmospheric inversion for estimating CO₂ fluxes: methods and perspectives. *Climatic Change*, 103(1-2):69-92.
- [15] Cressie N.A.C. (1993) *Statistics for Spatial Data*, Wiley, New York.
- [16] Doll C.N.H., Muller J.P., Elvidge C.D. (2000) Nighttime imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions. *Ambio*, 29:157-162.
- [17] Elvidge C.D., Baugh K.E., Kihn E.A., Koehl H.W., Davis E.R., Davis C.W. (1997) Relations between satellite observed visible near-infrared emissions, population, economic activity and power consumption. *Remote Sensing*, 18:1373-1379.

- [18] Enting I.G. (2002) *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press, New York.
- [19] Foken T., Wichura B. (1996) Tools for quality assessment of surface-based flux measurements. *Agricultural and Forest Meteorology*, 78:83-105.
- [20] Fuentes M., Raftery A. (2005) Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* 62:36-45.
- [21] Gałkowski M. (2015) Temporal and Spatial Variability of Nitrous Oxide in the Atmosphere over Małopolska Region: Determination of Loads and Emissions. PhD dissertation. AGH University of Science and Technology, Kraków, Poland.
- [22] Gamnitzer U., Karstens U., Kromer B., Neubert R.E.M., Meijer H.A.J., Schroeder H., Levin I. (2006) Carbon monoxide: A quantitative tracer for fossil fuel CO₂? *Journal of Geophysical Research*, 111, D22302, DOI: 10.1029/2005JD006966.
- [23] Gelfand A.E., Diggle P.J., Fuentes M., Guttorp P., Eds. (2010) *Handbook of Spatial Statistics*, Chapman & Hall/CRC.
- [24] GESAPU Report Summary (2015). CORDIS. Community Research and Development Information Service. European Commission, http://cordis.europa.eu/result/rcn/157618_en.html
- [25] Ghosh T., Elvidge C.D., Sutton P.C., Baugh K.E., Ziskin D., Tuttle B.T. (2010) Creating a global grid of distributed fossil fuel CO₂ emissions from nighttime satellite imagery. *Energies*, 3:1895-1913, DOI: 10.3390/en312.
- [26] Giorgi F., Mearns L. (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'reliability ensemble method' (REA) method. *Journal of Climate*, 15:1141-1158, DOI: 10.1175/1520-0442(2002)015,1141:COAURA>2.0.CO;2.
- [27] Gourdj S.M., Mueller K.L., Schaefer K., Michalak A. (2008) Global monthly averaged CO₂ fluxes recovered using a geostatistical inverse modelling approach: 2. Results including auxiliary environmental data. *Journal of Geophysical Research*, 113, D21115, DOI: 10.1029/JD009733.
- [28] Gurney K.R., Mendoza D.L., Zhou Y., Fischer M.L., Miller C.C., Geethakumar S., de la rue du Can S. (2009) *Environmental Science & Technology*, 43 (14), 5535-5541
- [29] Gurney K.R., Rachel L.M., Denning A.S., Rayner P.J., Baker D., Bousquet P., Bruhwiler L., Chen Y.H., Ciais P., Fan S., Fung I.Y., Gloor M., Heimann M., Higuchi K., John J., Maki T., Maksyutov S., Masarie K., Peylin P., Prather M., Pak B.C., Randerson J., Sarmiento J., Taguchi S., Takahashi T., Yuen C.W. (2002) Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature*, 415:626-630.
- [30] Halushchak M., Bum R., Topylko P. (2015) Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [31] Hogue S., Marland E., Andres R.J., Marland G., Robinson Ch., Woodard D. (2015) Uncertainty in gridded CO₂ emissions estimates. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.

- [32] Horabik J., Nahorski Z. (2014) Improving resolution of a spatial air pollution inventory with a statistical inference approach. *Climatic Change*, 124:575-589.
- [33] Horabik J., Nahorski Z. (2014) The Cramér-Rao lower bound for the estimated parameters in a spatial disaggregation model for areal data. In: D. Filev, J. Jabłkowski, J. Kacprzyk, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szykarczyk, S. Zadrozny: *Intelligent Systems 2014*, Springer International Publishing, pp. 661-668.
- [34] Horabik-Pyzel J, Charkovska N., Danylo O., Nahorski Z., Bun R. (2015) Conditionally autoregressive model for spatial disaggregation of activity data in GHG inventory: Application for agriculture sector in Poland. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [35] IEA (2007) CO₂ emissions from fuel combustion: 1971-2005. International Energy Agency, Paris, France.
- [36] IPCC (1996) Revised IPCC 1996 guidelines for national greenhouse gas inventories. Technical Report IPCC/OECD/IEA, Paris. <http://www.ipcc-nggip.iges.or.jp/public/gl/invsl.html>.
- [37] IPCC (2001) Penman J., Kruger D., Galbally I. et al. (2001) Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories,; IPCC.
- [38] IPCC (2006) Eggleston H.S., Buendia L., Miwa K., Ngara T., Tanabe K. (eds) IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme,
- [39] Kaminski T., Knorr W., Rayner P. J., Heimann M. (2002) Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Global Biogeochemical Cycles*, 16(4):1066, doi:10.1029/2001GB001463, 2002
- [40] Karlen I., Olsson I.U., Kilburg P., Kilici S. (1968) Absolute determination of the activity of two ¹⁴C dating standards. *Arkiv Geophysik*, 4:465-471.
- [41] Knutti R., Furrer R., Tebaldi C., Cermak J., Meehl G.A. (2010) Challenges in combining projections from multiple models. *Journal of Climate*, 23:2739-2756.
- [42] Kryazhimskiy A. (2013) Posterior integration of independent stochastic estimates. Interim Report IR-13-006, IIASA, Laxenburg, Austria. Available at: http://www.iiasa.ac.at/publication/more_IR-13-006.php.
- [43] Kryazhimskiy A, Rovenskaya E, Shvidenko A, Gusti M, Shchepashchenko D, Veshchinskaya V (2015). Towards harmonizing competing models: Russian forests' net primary production case study. *Technological Forecasting and Social Change*, 98:245-254.
- [44] Kuc T., Rozanski K., Zimnoch M., Necki J., Chmura L., Jelen D. (2007) Two decades of regular observations of ¹⁴CO₂ and ¹³CO₂ content in atmosphere carbon dioxide in Central Europe: Long-term changes of regional anthropogenic fossil CO₂ emissions. *Radiocarbon*, 49(2):807-816.
- [45] Lauvaux T., Uliasz M., Sarrat C., Chevallier F., Bousquet P., Lac C., Davis K.J., Ciais P., Denning A.S., Rayner P.J. (2008) Mesoscale inversion: first results from the CERES campaign with synthetic data. *Atmospheric Chemistry and Physics*, 8:3459-3471.

- [46] Le N.D., Zidek J. (2006) *Statistical Analysis of Environmental Space-Time Process*. Springer, Berlin.
- [47] Levin I., Karstens U. (2007) Inferring high-resolution fossil fuel CO₂ records at continental sites from combined ¹⁴CO₂ and CO observations. *Tellus, Ser. B*, 59:245-250.
- [48] Liu Z., Le N. D., Zidek J. V. (2011) An empirical assessment of Bayesian melding for mapping ozone pollution. *Environmetrics*, 22:340-353.
- [49] Lopez M., Schmidt M., Delmotte M., Colomb A., Gros V., Janssen C., Lehman S.J., Mondelain D., Perrussel O., Ramonet M., Xueref-Remy I., Bousquet P. (2013) CO, NO_x and ¹³CO₂ as tracers for fossil fuel CO₂: results from a pilot study in Paris during winter 2010. *Atmospheric Chemistry and Physics*, 13:7343–7358, DOI:10.5194/acp-13-7343-2013
- [50] Marland G., Boden T.A., Andres R.J. (2008) Global, regional, and national fossil fuel CO₂ emissions. In: *Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.
- [51] Michalak A., Hirsch A., Bruhwiler L., Gurney K.R., Peters W., Tans P.P. (2005) Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversion. *Journal of Geophysical Research*, 110, D24107, DOI: 10.1029/2005JD005970.
- [52] Min, Y. and Agresti, A. (2002). Modeling nonnegative data with clumping at zero: A survey. *Journal of the Iranian Statistical Society*, 1:7–33.
- [53] Oda T., Maksyutov S. (2011) A very high-resolution (1 km × 1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmospheric Chemistry and Physics*, 11:543-556, DOI: 10.5194/acp-11-543-2011.
- [54] Oda T., Ott L., Topylko P., Halushchak M., Bun R., Lesiv M., Danylo O., Horabik-Pyzel J. (2015) Uncertainty associated with fossil fuel carbon dioxide (CO₂) gridded emission datasets. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [55] Olivier J.G.J., Aardenne J.A.V., Dentener F.J., Pagliari V., Ganzeveld L.N., Peters J.A.H.W. (2005) Recent trends in global greenhouse gas emissions: Regional trends 1970-2000 and spatial distribution of key sources in 2000. *Environmental Sciences*, 2(2-3):81-99, DOI: 10.1080/15693430500400345.
- [56] Palstra S.W., Karstens U., Streurman H.J., Meijer H.A.J. (2008) Wine ¹⁴C as a tracer for fossil fuel CO₂ emissions in Europe: Measurements and model comparison. *Journal of Geophysical Research*, 113, D21305, DOI: 10.1029/2008JD010282.
- [57] Peylin P., Rayner P.J., Bousquet P., Carouge C., Hourdin F., Heinrich P., Ciais P., AEROCARB contributors (2005) Daily CO₂ flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology. *Atmospheric Chemistry and Physics*, 5:3173-3186.
- [58] Raczka B.M., Davis K.J., Hutzinger D., Neilson R.P., Poulter B., Richardson A.D., Xiao J., Baker I., Ciais P., Keenan T.F., Law B., Post W.M., Ricciuto D., Schaefer

- K., Tian H., Tomelleri E., Verbeeck H., Viovy N. (2013) Evaluation of continental carbon cycle simulations with North American flux tower observations. *Ecological Monographs*, 83(4):531-556.
- [59] Raupach M.R., Marland G., Ciais P., Le Quéré C., Canadell J.G., Klepper G., Field C.B. (2007) Global and regional drivers of accelerating CO₂ emissions. *Proceedings of the National Academy of Sciences*, 104(24):10288-10293, DOI: 10.1073/pnas.0700609104.
- [60] Rayner P.J., Scholze M., Knorr W., Kaminski T., Giering R., Widmann H. (2005) Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, 19, GB2026, DOI: 10.1029/2004GB002254.
- [61] Rayner P.J., Raupach M.R., Paget M., Peylin P., Koffi E. (2010) A new global gridded dataset of CO₂ emissions from fossil fuel combustion: Methodology and evaluation. *Journal of Geophysical Research*, 115, D19306, DOI: 10.1029/2009JD013439.
- [62] Riley W.G., Hsueh D.Y., Randerson J.T., Fischer M.L., Hatch J., Pataki D.E., Wang W., Goulden M.L. (2008) Where do fossil fuel carbon dioxide emissions from California go? An analysis based on radiocarbon observations and an atmospheric transport model. *Journal of Geophysical Research*, 113, G04002, DOI: 10.1029/2007JG000625.
- [63] Rivier L., Peylin P., Ciais P., Gloor M., Rödenbeck C., Geels C., Karstens U., Bousquet P., Brandt J., Heimann M., Aerocarb experimentalists (2010) European CO₂ fluxes from atmospheric inversions using regional and global transport models. *Climatic Change*, 103(1-2):93-115.
- [64] Stephens B.B., Gurney K.R., Tans P.P., Sweeney C., Peters W., Bruhwiler L., Ciais P., Ramonet M., Bousquet P., Nakazawa T., Aoki S., Inoue T.M.G., Vinnichenko N., Lloyd J., Jordan A., Heimann M., Shibistova O., Langefelds R.L., Steele L.P., Francey R.J., Denning A.S. (2007) Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science*, 316:1732-1735.
- [65] Stuiver M., Polach H.A. (1977) Discussion. Reporting of ¹⁴C data. *Radiocarbon*, 19(3):355-363.
- [66] Suess H.E. (1955) Radiocarbon concentration in modern wood. *Science*, 122:415-417, DOI: 10.1126/science.122.3166.415-a.
- [67] Tans P.P., Fung I.Y., Takahashi T. (1990) Observational constraints on the global atmospheric CO₂ budget. *Science*, 247:1431-1438.
- [68] Tarantola A. (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia.
- [69] Tebaldi C., Knutti R. (2007) The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A*, 365:2053-2075, DOI: 10.1098/rsta.2007.2076.
- [70] Thompson R.L., Gerbig C., Rödenbeck C. (2011) A Bayesian inversion estimate of N₂O emissions for western and central Europe and the assessment of aggregation error. *Atmospheric Chemistry and Physics*, 11:3443-3458, DOI: 10.5194/acp-11-3443-2011.

- [71] Topylko P., Halushchak M., Bun R. (2015) Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [72] Turnbull J.C., Miller J.B., Lehman S.J., Tans P.P., Sparks R.J. (2006) Comparison of ¹⁴CO₂, CO and SF₆ as tracers for recently added fossil fuel CO₂ in the atmosphere and implications for biological CO₂ exchange. *Geophysical Research Letters*, 33, L01817, DOI: 10.1029/2005GL024213.
- [73] Turnbull J., Rayner P., Miller J., Naegler T. Ciais P., Cozic A. (2009) On the use of ¹⁴CO₂ as a tracer for fossil fuel CO₂: Quantifying uncertainties using an atmospheric transport model. *Journal of Geophysical Research*, 14:D22302, DOI: 10.1029/2009JD012308.
- [74] Verstraete J. (2015) Remapping gridded data using Artificial Intelligence: real world challenges. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.
- [75] Zimnoch M., Necki J, Chmura L., Jasek A., Galkowski M., Kuc T., Gorczyca Z., Bartyzel J., Rozanski K. (2015) Urban metabolism: atmospheric loads and fluxes of major greenhouse gases (CO₂, CH₄) in Krakow, southern Poland. Short paper. 4th Int. Workshop Uncertainty in Atmospheric Emissions. Kraków, 7th – 9th October 2015.

Sensitivity of marginal abatement cost curves to variation of G4M parameters

Mykola Gusti^{1,2}, Nikolay Khabarov¹, Nicklas Forsell¹

¹ International Institute for Applied Systems Analysis, Austria

² Lviv Polytechnic National University, Ukraine

gusti@iiasa.ac.at

Abstract

Because of the G4M model non-linearity marginal abatement cost curves (MACCs) are sensitive to variation of the model parameters, irrespective of the fact that the same parameter variations are applied in both zero-CO₂ price and non-zero-CO₂ price runs. Since integrated assessment models in general are complex computer models with non-linearity one may expect all MACCs constructed using such models are sensitive to variation of the model parameters. The MACCs constructed using G4M are much more sensitive to parameter variation at a certain range of CO₂ prices, usually low CO₂ prices. The MACCs for total biomass CO₂ emissions constructed using G4M are most sensitive to variation of corruption coefficient (measuring efficiency of use of abatement costs) and, on the second place, to agriculture land price. Experts applying MACCs for policy analysis must be aware of uncertainty features of the MACCs as the uncertainty can influence the outcome of the analysis.

Keywords: G4M, marginal abatement cost curve, sensitivity, model parameters

1. Introduction

Marginal Abatement Cost Curve (MACC) relates potential of greenhouse gas (GHG) emissions reduction over a baseline and costs of the reduction. It is often used by research institutions and governments in a number of countries for analysis of mitigation policies. MACCs are constructed, in particular using integrated assessment models. MACCs provide information for analysis of such policy instruments as implementation of a CO₂ tax or a cap-and-trade system [1].

Experts employing MACCs for policy analysis must be aware of uncertainty in the MACCs as the uncertainty can influence the outcome of the analysis. For example, in case of a CO₂ tax implementation an uncertain MACC may give wrong information on possible reduction of CO₂ emissions resulting from the implemented tax; in case of introduction of a cap-and-trade system an uncertain MACC may misinform on carbon price that could result from a certain volume of carbon allowances.

Global Forest Model (G4M) simulates afforestation, deforestation, forest management directed at sustainable wood production, response of the mentioned processes to CO₂ price incentives and respective CO₂ emissions. G4M is applied for development of MACCs including such mitigation options as enhanced afforestation, avoiding deforestation and forest management directed to both wood production and carbon sequestration [2].

This study is aimed at answering the questions: what is sensitivity of the MACCs to selected model parameters and how the parameter uncertainties can impact GHG abatement policies related to forest sector?

2. Method

We study sensitivity of MACCs to variation of three G4M parameters selected in the consultations among the project³ partners: corruption coefficient (cr), wood price (w) [USD/m³] and agriculture land price (l) [USD/ha]. The corruption coefficient measures the efficiency of incurred costs for abatement: $cr=1$ (highest efficiency) means that no abatement costs are consumed by corruption and $cr=0$ (lowest efficiency) means that all costs are consumed by corruption.

We run the G4M model for a number of CO₂ price scenarios: initial prices starting in 2020 (0, 1, 3, 5, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120 USD/tCO₂) and rising 5% per year (that results in CO₂ price range of 4-520 USD/ton CO₂ in 2050) using standard parameter values as in [2]. For the purposes of sensitivity analysis we vary the values cr , w , and l mentioned above: we decrease/increase them by 1, 2.5, 5, 10, 50 and 90% (only single parameter was changed during a run). For a year within the range 2020-2050 a MACC is defined as a difference of biomass CO₂ emissions at zero CO₂ price and a non-zero CO₂ price. The emissions include afforestation, deforestation and forest management components. The parameter deviation was applied to all CO₂ price runs thus serving as a bias for MACC. For the run we used population and GDP following SSP2 scenario (<https://secure.iiasa.ac.at/web-apps/ene/SspDb/dsd?Action=htmlpage&page=about>), wood demand, regional agriculture land prices and wood prices were estimated by GLOBIOM model (<http://www.globiom.org/>) under assumption of bioenergy demand of 50PJ/year.

We calculated 12 MACC variations for each parameter. The results get the following notations: $crpV$, $crmV$, wpV , wmV , lpV and lmV , where p means an increase of a parameter, m means a decrease of a parameter and V means 1, 2.5, 5, 10, 50 or 90% change of the parameter. Because of limited space of the paper we present detailed analysis of MACC sensitivities to 10, 50 and 90% variation of the parameters globally as well as summary information for Brazil, Indonesia and Mexico for the year 2030.

3. Results

At 1-5% variation of the parameters deviation of the global MACC curve follows the shape of respective deviations at 10% variation of the parameters but the amplitude is smaller. At 10% variation of the parameters the global MACC curve is most sensitive when CO₂ price is 5 USD/tCO₂ (Figure 1-3). At this CO₂ price a decrease of the corruption coefficient (means more abatement costs are consumed by corruption) makes the highest impact on the MACC – the efficiency of abatement costs decreases by 230 MtCO₂/year. Increase of the corruption coefficient (means less abatement costs are consumed by corruption) has a slightly smaller effect on the MACC – the abatement increases by 229 MtCO₂/year. Agriculture land price variation influences the MACC considerably – decrease of the land price yields 172 MtCO₂/year higher abatement while increase of the land price decreases the abatement by 122 MtCO₂/year. Global MACC's deviation from a baseline (all parameters cr , w , l unchanged) diminishes with the increasing CO₂ price slower than the countries' MACCs considered in the study. The variation of corruption coefficient makes the maximum impact on the global MACC across the 3 parameters at CO₂ prices 1-30 and 80 USD/tCO₂. The variation of wood price makes the maximum impact on MACC across the 3 parameters at 40-60

³ "Options Market and Risk-Reduction Tools for REDD+" funded by the Norwegian Agency for Development Cooperation under agreement number QZA-0464 QZA-13/0074.

and 100-120 USD/tCO₂. Wood price reaches its maximum impact on MACC at 15 USD/tCO₂.

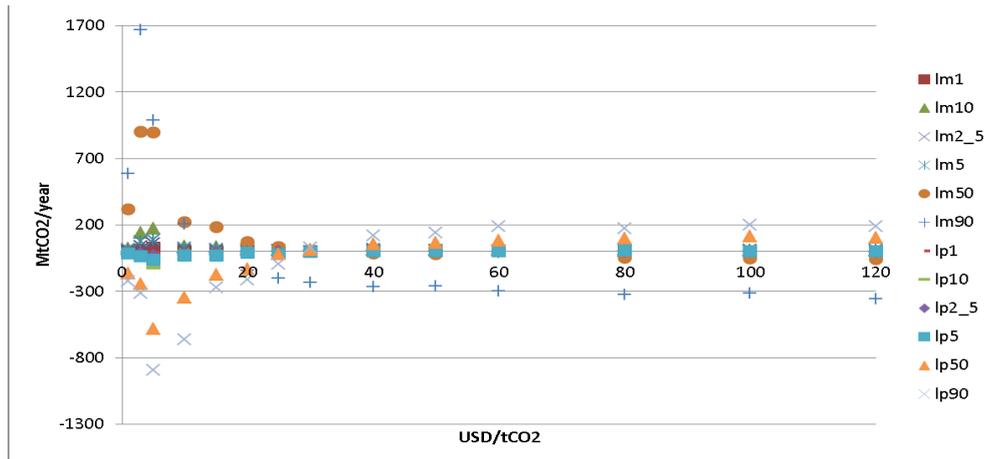


Figure 1. Sensitivity of MACC for total biomass CO₂ emissions to deviations of agriculture land price globally in 2030.

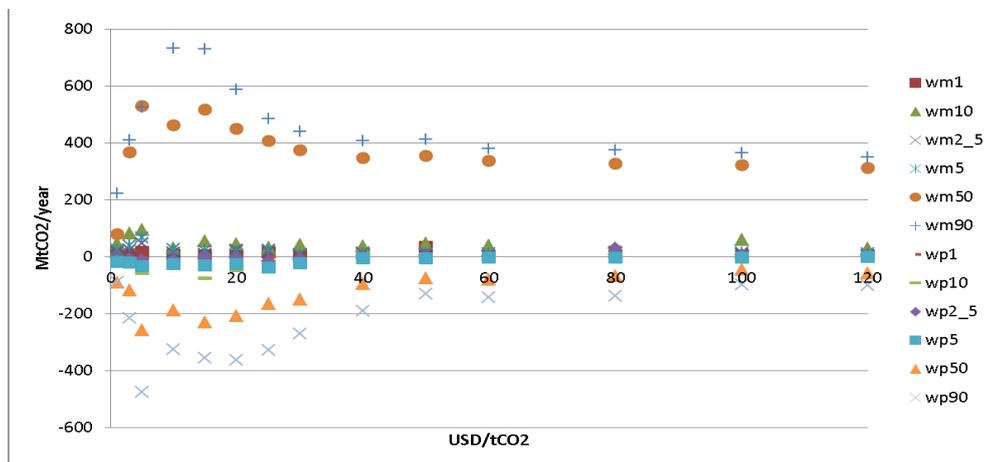


Figure 2. Sensitivity of MACC for total biomass CO₂ emissions to deviations of wood price globally in 2030.

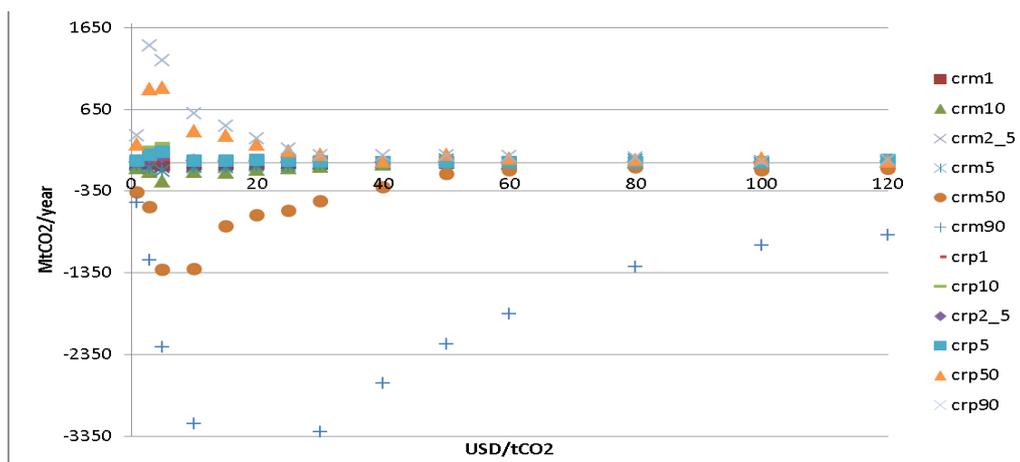


Figure 3. Sensitivity of MACC for total biomass CO₂ emissions to deviations of corruption coefficient globally in 2030.

At 50% variation of the parameters the global MACC curve is most sensitive when CO₂ price is 5 USD/tCO₂ (Figure 1-3). At this CO₂ price decrease of the corruption coefficient (means more abatement costs are consumed by corruption) causes deviation of the MACC by -1,310 MtCO₂/year. The effect of the corruption coefficient variation diminishes by 15 times to -75 MtCO₂/year at 120 USD/tCO₂. The corruption coefficient has the largest impact on MACC across the parameters at 1 and 5-30 USD/tCO₂. Wood price has a considerable effect on MACC at all CO₂ prices with maximal value of 531 MtCO₂/year at 5 USD/tCO₂ and has the largest effect across the parameters at 40-120 USD/tCO₂. Agriculture land price makes maximum impact on MACC (903 MtCO₂/year) at 3 USD/tCO₂ when it overcomes the effect of the other parameters.

At 90% variation of the parameters the MACC curve is most sensitive at 20 USD/tCO₂ (Figure 1-3). At this CO₂ price decrease of the corruption coefficient (means more abatement costs are consumed by corruption) causes deviation of the MACC by -3,477 MtCO₂/year. The effect of the corruption coefficient variation diminishes slowly and has the highest impact on MACC across the parameters at 5-120 USD/tCO₂. Agriculture land price has lower impact with maximum at 5 USD/tCO₂ (1,699 MtCO₂/year), it exceeds the impact of the other parameters at 1 and 3 USD/tCO₂. Wood price reaches its maximum impact on MACC (733 MtCO₂/year) at 10 USD/tCO₂.

The corruption coefficient has the largest impact on the MACC at all levels of the parameter changes. With increasing the amplitude of the parameter variation the maximum impact shifts from 5 USD/tCO₂ (at 10 and 50% variation) to 20 USD/tCO₂ (at 90% variation). Wood price has relatively even impact at all CO₂ prices, while agriculture land price has two picks – higher at low CO₂ prices and lower at high CO₂ prices. Increase of the parameter variation amplitude to 90% defuses the CO₂ price at which individual parameters cause maximum deviation of MACC.

For Brazil and Mexico similarly as in the global case considered above the corruption coefficient has the largest impact on the MACC at all levels of parameter variations. With increasing the amplitude of the parameter variation the maximum impact shifts from lower CO₂ prices to higher: in Brazil – from 5 USD/tCO₂ (57 MtCO₂/year) at 10% corruption coefficient increase to 10 USD/tCO₂ (-443 MtCO₂/year) at 50% corruption coefficient decrease and to 15 USD/tCO₂ (-644 MtCO₂/year) at 90% corruption coefficient decrease; in Mexico – from 10 USD/tCO₂ (-9 MtCO₂/year) at 10% corruption coefficient decrease to 15 USD/tCO₂ (-51 MtCO₂/year) at 50% corruption coefficient decrease and to 25 USD/tCO₂ (-70 MtCO₂/year) at 90% corruption coefficient decrease.

In Indonesia the corruption coefficient does not make the overall maximum impact on the MACC, nevertheless the impact is considerable especially for CO₂ prices over 5 USD/tCO₂ and large variation of the parameter. With larger decrease of the corruption coefficient the maximum of its impact on MACC shifts to higher CO₂ prices. The corruption coefficient has the largest impact on MACC across the parameters at CO₂ prices greater than 5 USD/tCO₂ for 50 and 90% variation of the parameters. Agriculture land price decrease has the largest impact on MACC at all levels of the parameter changes. With increasing the amplitude of the parameter variation the maximum impact shifts from 5 USD/tCO₂ (-28 MtCO₂/year) at 10% increase of the agriculture land price to 3 USD/tCO₂ (177 MtCO₂/year) at 50% and to 1 USD/tCO₂ (306 MtCO₂/year) at 90% decrease of agriculture land price.

4. Discussion

The parameter deviation was applied to all CO₂ price runs thus serving a bias for MACC. In this case MACC deviation is caused by the model non-linearity across CO₂ prices, i.e. different sensitivity of the emissions to the same deviation of a parameter at zero and non-zero CO₂ prices. For the studied countries and globally the emission response to alteration of agriculture land price is very high at CO₂ prices 3-10 USD/tCO₂ symmetrically to negative and positive deviations of the parameter (see Figure 4 for the global case).

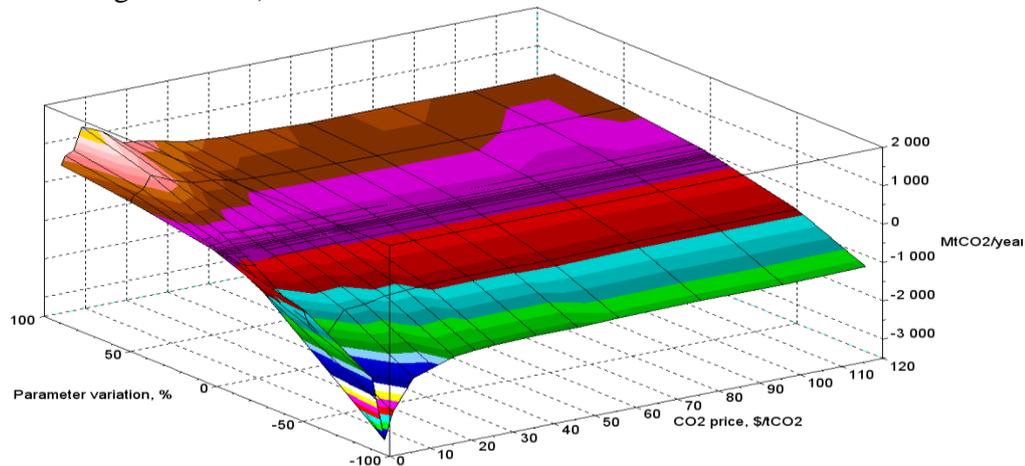


Figure 4. Sensitivity of total biomass CO₂ emissions to agriculture land price globally in 2030.

The emission response to wood price alteration has different shapes in the studied countries while the global case incorporates features of all countries. In Brazil the sensitivity is high at all CO₂ prices but at the prices 1-5 USD/tCO₂ the sensitivity changes its sign (with a maximum at 10 USD/tCO₂). The “anomaly “ is explained by the fact that at some CO₂ prices increase of wood price causes increase of deforestation rate because a part of deforested wood is sold that pushes switching from forestry to agriculture. This is the effect of an interplay between agricultural land price, CO₂ price, and wood price. The effect comes from the decision-making algorithm of G4M: conversion from forest to agriculture is based on the highest level of net present value (NPV) that can be achieved by one of these land use alternatives. In this case a higher wood price is not enough for economically sustainable forestry and (as a one-time profit from selling the wood) adds an incentive for moving to agriculture (deforestation) [2]. In Indonesia the emission response to wood price is variable over the CO₂ prices with maximum deviations around 3 and 60 USD/tCO₂. In Mexico the emission response to wood price is symmetrical by the sign of the parameter variation with maximum at 10-15 USD/tCO₂. For Mexico we see the same effect of increasing deforestation with increasing wood price at 5-10 USD/tCO₂. The global picture communicates similar message: for the carbon price about 10 USD/tCO₂ an increase of wood price increases deforestation as compared to a baseline corresponding to that carbon price (10 USD/tCO₂).

The emission response to variation of the corruption coefficient has a similar shape – with a sharp maximum deflection of the emissions at CO₂ prices 3-10 USD/tCO₂ when the corruption coefficient increases (see Figure 5 for the global case). When the corruption coefficient decreases the sensitivity is high at a wide range of CO₂ prices up to the whole range if the corruption coefficient decreases by 90%.

The G4M model is non-linear on and sensitive to the variation of the cr , w , and l model parameters. Existence of a range of CO₂ prices under which the MACCs are very much sensitive to the variation of model parameters, is, probably, model specific and connected with simulation of decision making and values of NPVs of the alternative land uses.

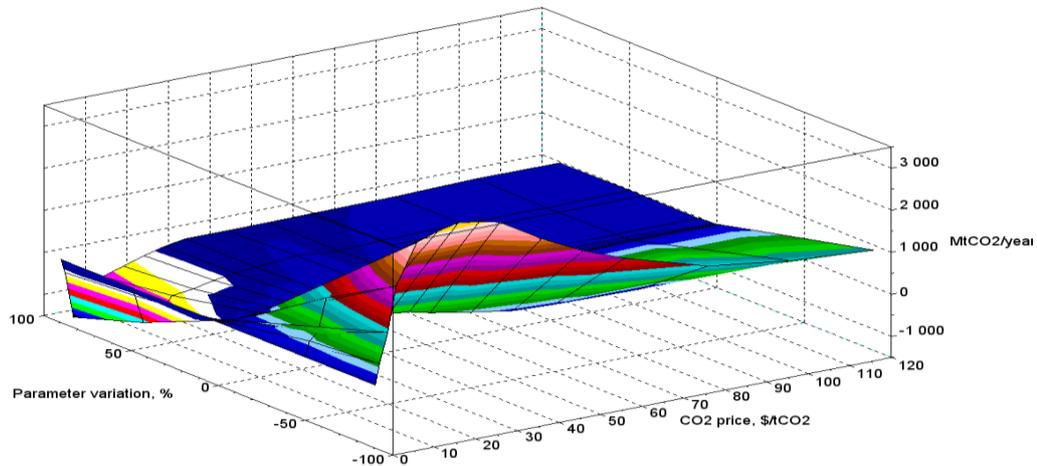


Figure 5. Sensitivity of total biomass CO₂ emissions to corruption coefficient globally in 2030 (greater coefficient means higher efficiency and less corruption).

4. Final remarks

Because of the G4M model non-linearity MACCs are sensitive to variation of the model parameters, irrespective of the fact that the same parameter variations are applied in both zero-CO₂ price and non-zero-CO₂ price runs. Since integrated assessment models in general are complex computer models with non-linearity one may expect all MACCs constructed using such models are sensitive to variation of the model parameters.

The MACCs constructed using G4M are much more sensitive to parameter variation at a certain range of CO₂ prices, usually low CO₂ prices.

The MACCs for total biomass CO₂ emissions constructed using G4M are most sensitive to variation of corruption coefficient (measuring efficiency of use of abatement costs) and, on the second place, to agriculture land price.

Experts applying MACCs for policy analysis must be aware of uncertainty features of the MACCs as the uncertainty can influence the outcome of the analysis, e.g. misinform on possible reduction of CO₂ emissions resulting from the implemented CO₂ tax or misinform on carbon price that could result from certain total carbon allowances in case of introduction of a cap-and-trade system.

Acknowledgements

The work has been carried out within the project "Options Market and Risk-Reduction Tools for REDD+" funded by the Norwegian Agency for Development Cooperation under agreement number QZA-0464 QZA-13/0074.

References

- [1] Kesicki, F. (2011) Marginal abatement cost curves for policy-making – expert-based vs model-derived curves. UCL Energy Institute, University College London. www.homepages.ucl.ac.uk/~ucft347/Kesicki_MACC.pdf. Accessed 14 September 2015
- [2] Gusti, M. & Kindermann, G. (2011). An approach to modeling landuse change and forest management on a global scale. In SIMULTECH-2011. Proc. of 1st Intern. Conf. on Simulation and Modeling Methodologies, Technologies and Applications, Noordwijkerhout, pp. 180–185

Multi-agent auction simulation of the GHG international emission permit trading

Piotr Pałka¹, Jarosław Stańczak², Weronika Radziszewska², Zbigniew Nahorski²

¹ Institute of Control and Computation Engineering, Warsaw University of Technology
Warsaw, Poland

P.Palka@ia.pw.edu.pl

² Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

*Zbigniew.Nahorski@ibspan.waw.pl, WeronikaRadziszewska@ibspan.waw.pl,
Jaroslaw.Stanczak@isbpan.waw.pl*

Abstract

This short paper presents elements of a simulation environment for negotiation of prices in GHG emission permit trade and results of simulations of international trades performed by programmable agents. Several market mechanisms and strategies used by programmable agents are discussed and applied in simulations. The results show convergence of the trading schemes to the equilibrium, depending on the case that consists of the trading mechanism and the strategies used by the agents. The simulation can be used for estimation of equilibrium price at market designing stage. It can be also used for simulation of proposed markets for uncertain emission inventories, which is envisaged for further studies.

Keywords: Greenhouse gases, emission permit trading, computer simulation, multi-agent systems

1. Introduction

Although the trading of GHG emission permits has been introduced in the Kyoto Protocol more than a dozen years ago and some experience has been already gathered, the markets are still unpredictable, particularly in assessment of the equilibrium price. Good estimate of the equilibrium price would allow the market planner to better organize the market and plan its parameters. Ermoliev et al. [8] proposed simulation of a bilateral trade using multi-agent systems for assessing the equilibrium price. A programmable agent would be under control of a party taking part in the trade and would use discreet private information to bid in a process of automatic negotiation. This simulation approach differs from the game-theoretic simulations, like those presented in [2, 3].

Recently, estimates of marginal abatement curves obtained by using simulation tools were published, like those calculated in GAINS [26], EPPA [16], and using bottom-up modeling [1]. This enables simulations of trades among parties. In this paper a few market mechanisms are considered to simulate trade among 16 regions of the world. The cost curves were adopted from [16]. Results obtained by using different negotiation methods and different price formation strategies by programmable agents are compared and discussed. This paper develops the earlier approach by Nahorski et al. [18], where also a much smaller group of 5 parties was considered, by considering more market mechanisms and many other negotiation strategies. Multi-agent trade simulations could be also used for checking designed market schemes before they are practically implemented.

The simulation approach could be also used for markets for enterprises. The main obstacle is knowledge of marginal abatement curves in such markets.

Another difficulty in organizing the GHG emission permit markets is high diversity of emission accuracy bounds among the trading countries. There is a couple of approaches to cope with this problem, see e.g. [9, 14, 15, 18, 22, 27]. Simulation of these approaches by using multi-agent tools is envisaged as continuation of the results presented in this paper.

2. The trading mechanisms used in simulations

A negotiation is a dialog between two or more parties, in order to resolve a conflict or to reach an agreement. A dialog is an exchange of communicates between two or more parties to reach their personal aim. The details of conducting the negotiation define most of the trading schemes. The automated agents negotiate parameters of the deal, as for example price, or conditions of delivery. The negotiation models can be divided into two categories: bilateral negotiations, which involve usually two parties (although multilateral negotiations are also possible), and auctions that by definition include multiple parties. There are many types of auctions which are in use, to mention the English Auction, the First-Price Sealed-Bid Auction, the Vickerey Auction, the Dutch Auction or Dutch Flower Auction, and the Continuous Double Auction. In the paper we consider the Dynamic Bilateral Negotiations and the following types of auctions: Continuous Double Auction, Sealed-Bid Auction, Sealed-Bid Reverse Auction, and Sealed-Bid Double Auction. For more detailed introduction to trade by programming agents see e.g. [12].

A market participant intending to buy a commodity (a buyer) places an offer; called a *bid*, together with the number of units and the price that the buyer is willing to pay for it. A participant intending to sell a unit of a commodity (a seller) places an offer; now called an *ask*, which includes the number of units and the price the seller wants. The market clears whenever the price of a bid is equal or greater than the price of the ask. The paired offers are removed from the market, and all other offers remain unchanged. The clearing price in every trade mechanism is set in the middle of the lowest accepted buying price (lowest accepted bid) and the highest accepted selling price (highest accepted ask). Every offer consists of the offered price and the offered number of permits.

2.1 Continuous Double Auction (CDA)

The Continuous Double Auction (CDA) is one of the market mechanisms frequently used in the stock market and also in their computer simulations. This type of market consists of three entities: the sellers, the buyers, and the market operator (the broker) that manages the trade: orders the bids and asks, and arranges transactions if prices of asks are lower or equal than prices of bids. The broker also records important market events and outstanding offers. The current lowest ask is called the outstanding ask, and the highest bid is called the outstanding bid, both these values are important during formulation of the offer price.

The buyers and sellers in the market are expected to behave rationally: their bids and asks should be profitable, and the ask-bid spread should be reduced in time to enable the market prices to evolve toward the equilibrium price.

There are variants of CDA markets, which depend on specifics of particular markets or traded goods. The CO₂ emission permits market also requires some modifications of

the general schema. The most important is division of CDA trade into cycles. Market participants may give their offers simultaneously, but only one of each kind in one cycle. Having collected all offers, the cycle terminates. Only these transactions are executed which are profitable for both participants. Offers unused in one cycle could be valid in a limited number of cycles. But it is rather favorable to make a new offer in the consecutive cycle, taking into account new market events. Cycles in auctions are not identified with any real periods of time and time is not crucial in this kind of market. Surpluses and shortages from one cycle can be sold/bought in subsequent ones.

Due to changes of marginal prices of market participants caused by conducted transactions (see Fig. 3), limitations are imposed on the number of transferred permits in one transaction to avoid big perturbations of prices in consecutive cycles.

In another variant, the transactions are concluded immediately after an ask and a bid matches have been detected. Also this variant is considered in the simulations.

2.2 Dynamic Bilateral Transactions (DBT)

In the bilateral trading, agents split into pairs and a single negotiation process occurs inside any pair. The splitting process is performed randomly, it occurs after termination of the running negotiation process, and is repeated iteratively. Established pairs conduct bilateral contracts depending on their expected profits. Each negotiation process may lead to an agreement or not.

2.3 Sealed-bid Auction\Reverse Auction (SA\SRA)

In the sealed-bid auction mechanism there are two roles in the trade: the auction operator, and the bidders. The operator calls for the auction to sell/buy a number of permits, possibly specifying the minimum/maximum unit price. Responding, a bidder gives its preferred unit price. The operator collects all the bids, and selects the winning one, with the highest/lowest unit price. In the simulations, the operator role is chosen randomly among the agents, while the remaining are the bidders.

2.4 Sealed-bid Double Auction (SDA)

In the double auction there are three roles: buyers, sellers, and the operator. The operator calls for the auction, and the sellers put the asks, and the buyers put the bids. Single clearing uses a clearing price that is not greater than prices of accepted bids, and not lower than prices of accepted asks. The clearing can also consist of more than two offers. Then, the clearing price should be set to satisfy as many asks and bids as possible.

3. Strategies used in the simulations

Vytelingum et al. [25] define a strategy of an agent as a set of atomic actions (that the agent can do), which were chosen based on the history of the market states and on the agent states. In a real situation it is very unlikely that an agent has information about all historic states of the market and especially about all parameters of the market. That is why real strategies are operating with limited number of variables, considering limited computational and sensory resources.

Strategies can be divided to those which use only current information and those which take into account also the history of the market states [17]. Among the former one there are the Frank or Truth Telling strategy, Pure Simple strategy, Kaplan strategy [23], Zero Intelligence strategy, and Preist and van Tol strategy [20].

Among the strategies that consider history, GD strategy proposed in [10] estimates so-called belief function from gathered information, which helps to form the proposed price. Other, complicated strategies with multi-level learning, are the Adaptive-Aggressive strategy and FL strategy [13]. The latter uses fuzzy logic reasoning. In this paper we use few chosen strategies described below.

3.1 Frank (F) or Truth Telling (TT) strategy

In the Frank strategy all agents bid according to its current marginal prices. As they do not behave strategically, the strategy is called Frank or Truth Telling.

In a similar simple strategy, described in [4] and called Pure Simple strategy, agents bid a constant 10% below the value of private evaluation. This strategy under the name Gamer was also played in the Santa Fe tournament [21], where it reached a similar, very low place to the TT and ZI strategies.

3.2 Zero-intelligence plus (ZIP) strategy

Zero intelligence (ZI) strategy was proposed by [11]. A ZI trader simply submits a random offer drawn from a uniform distribution.

Zero-intelligence plus (ZIP8) strategy, described in [5], bases on the auction history. 8 means the number of parameters passed to the strategy. It was later extended to 60 parameter strategy [6], which is however not discussed here. Every agent has the private price limit λ_i . For the seller it is the minimal value for which he is willing to sell one permission, and for the buyer it is the maximal value for which he is willing to buy one permission. At any time t , agent i calculates the price using its real-valued profit-margin $\mu_i(t)$.

$$p_i(t) = \lambda_i (1 + \mu_i(t)) - \text{for sellers} \quad (1)$$

$$p_i(t) = \lambda_i (1 - \mu_i(t)) - \text{for buyers} \quad (2)$$

The ZIP8 strategy assumes the constant recalculation of the real-valued profit margin. It is first drawn from the uniform distribution according to parameters:

$$\mu_i \sim U(\mu_{min}, \mu_{max})$$

In the course on the auction, the real-valued profit margin changes its value. To calculate it, the following is assumed. In the course of the auction, an agent can be either **greedy** or **careless**, and this property changes during the auction. A greedy agent wants to draw a largest possible profit from every transaction, neglecting the inherent risk. Careless agent is more aggressive and eager to enter into transaction, caring not so much about the profit.

An agent chooses at any time its behaviour (greedy or careless). It calculates its own possible offer from equations (1) – (2), and then checks the relation between calculated and the previous offer submitted to the market q_{j-1} . If the last offer has been rejected, an agent becomes careless, otherwise:

- if the agent is a seller and the last selling offer was greater than it had been calculated, it becomes greedy, otherwise it becomes careless,
- if the agent is buyer and the last buying was greater than it had been calculated, it becomes careless, otherwise it becomes greedy.

Agent calculates its interim offer according to the following equation:

$$\tau_j = R_i q_{j-1} + A_i$$

The parameter q_{j-1} is the previous offer submitted to the market, A_i and R_i are parameters that are drawn from the uniform distribution according to the following rules.

If the agent is careless:

$$A_i \sim U(1 - C_R, 1), \quad R_i \sim U(-C_A, 0)$$

If the agent is greedy:

$$A_i \sim U(1, 1 + C_R), \quad R_i \sim U(0, C_A)$$

where C_A and C_R are the parameters of the strategy.

Then, an agent calculates the interim profit margin:

$$d_j = 1 - \frac{\tau_j}{\lambda_j}$$

and subtracts the margin used in the previous negotiation round:

$$\delta_j = d_j - \mu_{j-1}$$

Next, the profit margin is modified using the Widrow-Hoff delta rule, that is the new margin is calculated as

$$\mu_j = \mu_{j-1} + \Delta_j$$

where Δ_j is calculated using the following rules with the learning rate β_i :

- if the agent is a seller:
 - if it is careless and $\delta_j \leq 0$ or it is greedy and $\delta_j > 0$, then:

$$\Delta_j = \beta_i(d_j - \mu_{j-1})$$
 - if it is careless and $\delta_j > 0$ or it is greedy and $\delta_j \leq 0$, then:

$$\Delta_j = \beta_i(\mu_{j-1} - d_j)$$
- if the agent is a buyer:
 - if it is careless and $\delta_j > 0$ or it is greedy and $\delta_j \leq 0$, then:

$$\Delta_j = \beta_i(\mu_{j-1} - d_j)$$
 - if it is careless and $\delta_j \leq 0$ or it is greedy and $\delta_j > 0$, then:

$$\Delta_j = \beta_i(d_j - \mu_{j-1})$$

The learning rate β_i is drawn from the uniform distribution:

$$\beta_i \sim U(\beta_{min}, \beta_{max}).$$

Now, we can calculate the new profit margin:

$$\mu_j = \mu_{j-1} - 1 + \Gamma_j$$

where Γ_j is the updating parameter, calculated using the equation:

$$\Gamma_j = \gamma \Gamma_{j-1} + (1 - \gamma) \Delta_j$$

where γ is the profit margin momentum coefficient. It is set by drawing from the uniform distribution:

$$\gamma \sim U(\gamma_{min}, \gamma_{max})$$

3.3 Adaptive-aggressive (AA) strategy

The Adaptive-Aggressive strategy of price formulation has been presented in [24]. This is a rather complicated model of the price formulation, considering market events, marginal costs of contractors and estimates of the market equilibrium with elements of short- and long-term learning.

The market participants are divided into intra-marginal and extra-marginal traders, depending on their limit price (marginal costs) $\lambda_j(t)$ in moment t and its relation to an estimate of the market equilibrium price $\hat{p}^*(t)$, that is the moving average of last N transaction prices p_i .

$$\hat{p}^*(t) = \frac{\sum_{i=t-N}^t w_i p_i}{N}, \quad \sum_{i=t-N+1}^t w_i = 1, \quad w_{i-1} = \rho w_i, \quad \rho = 0.9 \quad (3)$$

where:

p_i – transaction prices,

N – time horizon of the equilibrium estimation,

w_i – weights,

ρ – the forgetting factor (its value is set by the trading party).

Each of these groups is naturally divided into buyers and sellers (this allocation of participants changes during the market activity). Thus, four different causes of price formulation are distinguished:

- for an intra-marginal buyer $\lambda_j(t) > \hat{p}^*(t)$,
- for an intra-marginal seller $\lambda_j(t) < \hat{p}^*(t)$,
- for an extra-marginal buyer $\lambda_j(t) < \hat{p}^*(t)$,
- for an extra-marginal seller $\lambda_j(t) > \hat{p}^*(t)$.

So, the intra-marginal buyers and sellers are in good position for trading, while the extra-marginal ones in much worse. However, in the presented market functions of traders can be easily changed in consecutive iterations.

The price formation in the AA strategy requires a lot of information about the market, and the formulae used for bidding and asking are different for buyers and sellers:

- for a buyer

$$bid_j(t) = \begin{cases} o_{bid}(t) + \frac{\min(\lambda_j(t), o_{ask}^+) - o_{bid}(t)}{\eta} & \text{- the first round} \\ o_{bid}(t) + \frac{\tau_j(t) - o_{bid}(t)}{\eta} & \text{- other rounds} \end{cases} \quad (4)$$

$$o_{ask}^+ = (1 + \zeta_r) o_{ask}(0) + \zeta_a, \quad o_{ask}(0) = MAX_{ASK}$$

where:

MAX_{ASK} – the maximum price allowed on the market,

$\lambda_j(t)$ – the marginal (secret) price of the bidder,

$o_{bid}(t)$ – the current outstanding bid,

$\tau_j(t)$ – the target price (described later),

η – a correction factor, $1 \leq \eta \leq \infty$ (suggested value is 3),

ζ_a, ζ_r – modification factors (suggested values are 0.01 and 0.02, respectively).

- for a seller

$$ask_j(t) = \begin{cases} o_{ask}(t) + \frac{o_{ask}(t) - \max(\lambda_j(t), o_{bid})}{\eta} & \text{- the first round} \\ o_{ask}(t) - \frac{o_{ask}(t) - \tau_j(t)}{\eta} & \text{- other rounds} \end{cases} \quad (5)$$

$$o_{bid} = (1 - \zeta_r) o_{bid}(0) - \zeta_a, o_{bid}(0) = 0$$

where $o_{ask}(t)$ is the current outstanding ask.

The target prices $\tau_j(t)$ are different for all four categories:

- for an intra-marginal buyer

$$\tau_j(t) = \begin{cases} \hat{p}^*(t) \left(1 - \frac{\exp(-r\theta) - 1}{\exp(\theta) - 1}\right) & \text{- for } -1 < r \leq 0 \\ \hat{p}^*(t) + \frac{(\lambda_j(t) - \hat{p}^*(t)) \exp(r\theta) - 1}{\exp(\theta) - 1} & \text{- for } 0 < r < 1 \end{cases} \quad (6)$$

- for an intra-marginal seller

$$\tau_j(t) = \begin{cases} \hat{p}^*(t) + (MAX_{ASK} - \hat{p}^*(t)) \frac{\exp(-r\theta) - 1}{\exp(\theta) - 1} & \text{- for } -1 < r \leq 0 \\ \lambda_j(t) + (\hat{p}^*(t) - \lambda_j(t)) \left(1 - \frac{\exp(r\theta) - 1}{\exp(\theta) - 1}\right) & \text{- for } 0 < r < 1 \end{cases} \quad (7)$$

- for an extra-marginal buyer

$$\tau_j(t) = \begin{cases} \lambda_j(t) \left(1 - \frac{\exp(-r\theta) - 1}{\exp(\theta) - 1}\right) & \text{- for } -1 < r \leq 0 \\ \lambda_j(t) & \text{- for } 0 < r < 1 \end{cases} \quad (8)$$

- for an extra-marginal seller

$$\tau_j(t) = \begin{cases} \lambda_j(t) + (MAX_{ASK} - \lambda_j) \frac{\exp(-r\theta) - 1}{\exp(\theta) - 1} & \text{- for } -1 < r \leq 0 \\ \lambda_j(t) & \text{- for } 0 < r < 1 \end{cases} \quad (9)$$

The aggressiveness of the trader is an element of the short-time learning strategy and is controlled by the parameter $r \in [-1, 1]$. A trader with the value r close to -1 is called a **completely passive**. It tries to buy at price near 0 and sell at price near MAX_{ASK} (with the maximum profit). A trader, for which $r = 0$ is called **active**. It tries to buy and sell at price close to $\hat{p}^*(t)$ (with a moderate profit). At last the trader with r close to 1 is called a **completely aggressive** and tries to buy and sell at its price $\lambda_j(t)$ (without profit).

The degree of aggressiveness is controlled using the Widroff-Hoff rule:

$$r_j(t+1) = r_j(t) + \beta_1 (\delta_j(t) - r_j(t)) \quad (10)$$

$$\delta_j(t) = (1 \pm \zeta_r) r_{j_{shout}} \pm \zeta_a \quad (11)$$

where:

β_1 – random variable with a uniform distribution on the interval $[0.2, 0.6]$,

$r_{j_{shout}}$ – a value of r_j at the currently last bid,

$\delta_j(t)$ – is obtained according to following rules:

if there was no transaction at time $t - 1$, $\delta_j(t)$ is determined as follows:

for purchase:

$$\text{if } \tau_j(t-1) \leq bid_j(t-1), \text{ then } \delta_j(t) = (1 + \zeta_r) r_{j_{shout}} + \zeta_a$$

-that means an increased aggressiveness of the buyer at the auction,

for sale:

if $\tau_j(t-1) \geq ask_j(t-1)$, then $\delta_j(t) = (1 + \zeta_r)r_{j_{shout}} + \zeta_a$

- that means an increased aggressiveness of the seller at the auction,

if there was a transaction at time $t-1$, then:

for purchase:

if in the session $t-1$ the buyer j bought emissions for the price $q_j(t-1)$,

then:

if $\tau_j(t-1) \geq q_j(t-1)$, then $\delta_j(t) = (1 - \zeta_r)r_{j_{shout}} - \zeta_a$

- that means a decreased aggressiveness of the buyer at the auction.

otherwise $\delta_j(t) = (1 + \zeta_r)r_{j_{shout}} + \zeta_a$

- that means an increased aggressiveness the buyer at the auction,

for sale:

if in session $t-1$ the seller j sold emissions for the price $q_j(t-1)$,

then:

if $\tau_j(t-1) \leq q_j(t-1)$, then $\delta_j(t) = (1 - \zeta_r)r_{j_{shout}} - \zeta_a$

- that means a decreased aggressiveness of the seller at the auction,

otherwise $\delta_j(t) = (1 + \zeta_r)r_{j_{shout}} + \zeta_a$

- that means an increased aggressiveness of the seller at the auction.

Suggested values are $\zeta_a = 0.01$, $\zeta_r = 0.02$.

Similarly the long-term learning rule also uses the Widroff-Hoff rule to update the value of θ :

$$\theta(t+1) = \theta(t) + \beta_2(\theta^*(\bar{\alpha}(t)) - \theta(t)) \quad (12)$$

$$\theta^*(\bar{\alpha}(t)) = (\theta_{max} - \theta_{min})(1 - \bar{\alpha}(t)\exp(\gamma(\bar{\alpha}(t) - 1))) + \theta_{min} \quad (13)$$

$$\bar{\alpha}(t) = (\alpha(t) - \alpha_{min})/(\alpha_{max} - \alpha_{min}) \quad (14)$$

$$\alpha(t) = \frac{\sqrt{\frac{1}{N} \sum_{i=t-N+1}^t (p_i - \bar{p}^*(t))^2}}{\bar{p}^*(t)} \quad (15)$$

where:

α_{min} , α_{max} – minimal and maximal value of factor α ,

θ_{min} , θ_{max} – given minimal and maximal values of parameter θ (suggested values -2 and 8, respectively).

γ – the function shaping factor (suggested value 2).

4. Simulations

4.1. Simulation system

The mathematical formulation of the market and organization of simulation was analogous to that described in [18] and is not repeated here. The differences are presented below.

To organize simulations, the maximal and minimal prices for the parties are needed. The maximal price can be taken as a reasonably high arbitrary value. The minimal value is, however, bound by the shadow price of the party. The shadow price is the derivative of the marginal abatement cost curve at the current emission value. The marginal abatement cost curves for reducing the emission of greenhouse gases have been developed on the basis of the data for 2010 published by [16] and its online supplementary documentation. Originally they were calculated using version 4 of the

MIT EPPA model [19]. To use them in the computer simulations, the cost curves were approximated by polynomials fitted by using the regression method. The data were given for 16 countries and regions in the world: USA, Canada (CAN), Japan (JPN), European Union (EUR), Australia and New Zealand (ANZ), Eastern Europe (EET), Former Soviet Union (FSU), India (IND), China (CHN), Indonesia (IDZ), East Asia (ASI), Mexico (MEX), Central and South America (LAM), Middle East (MES), Africa (AFR), Rest of World (ROW).

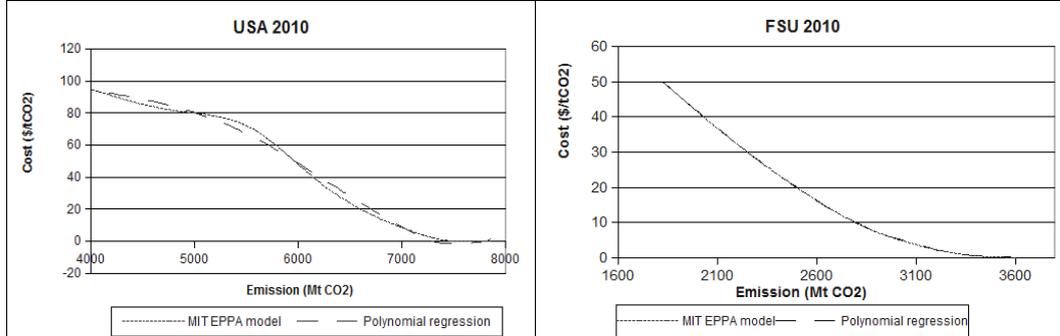
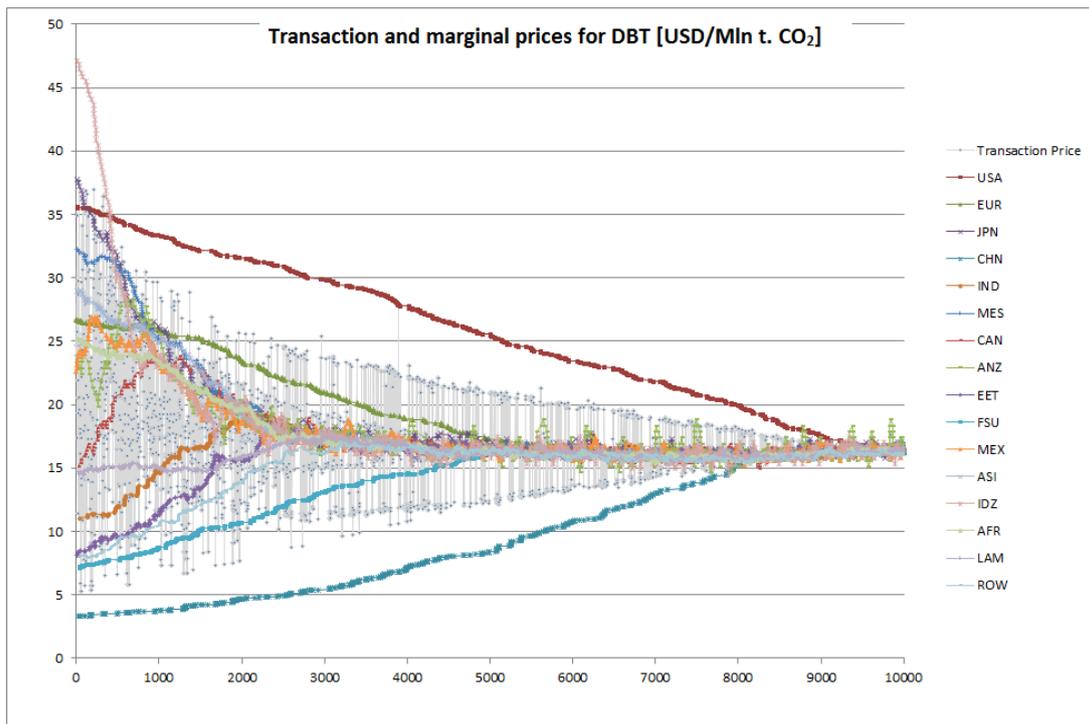
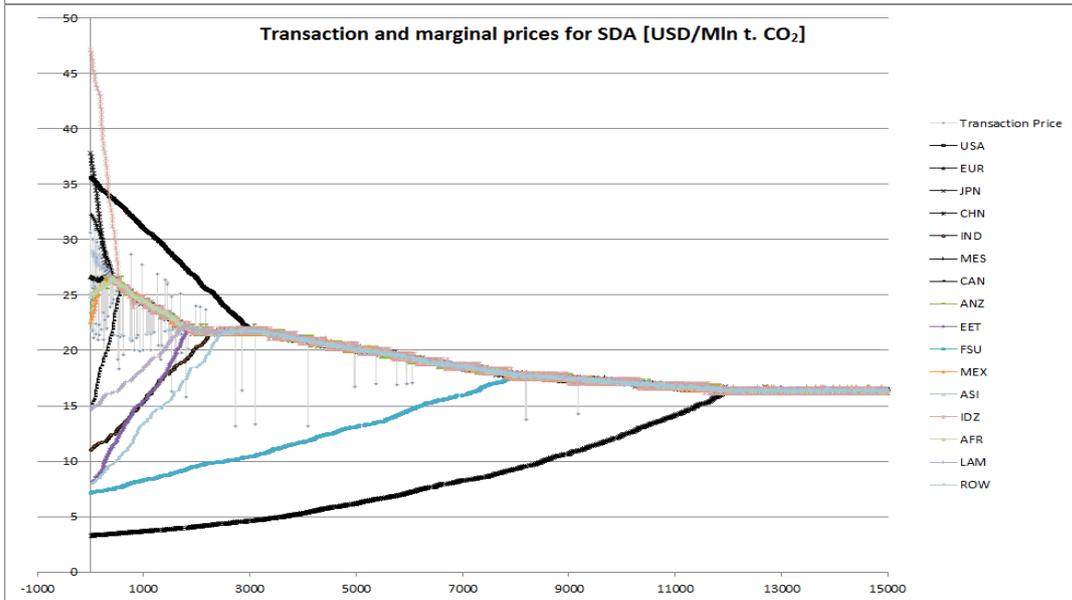
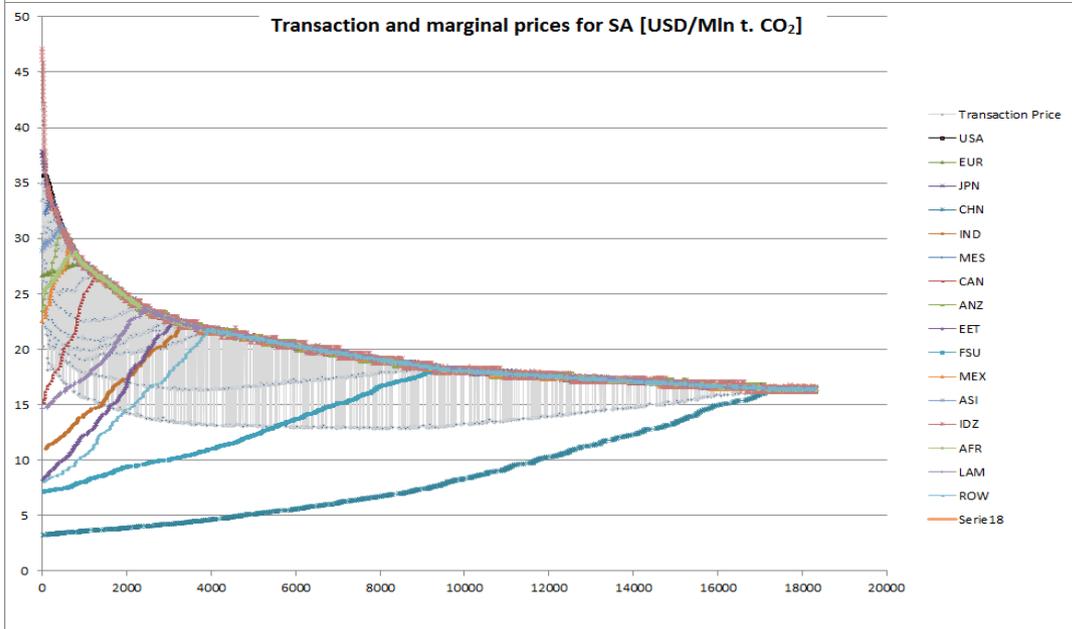
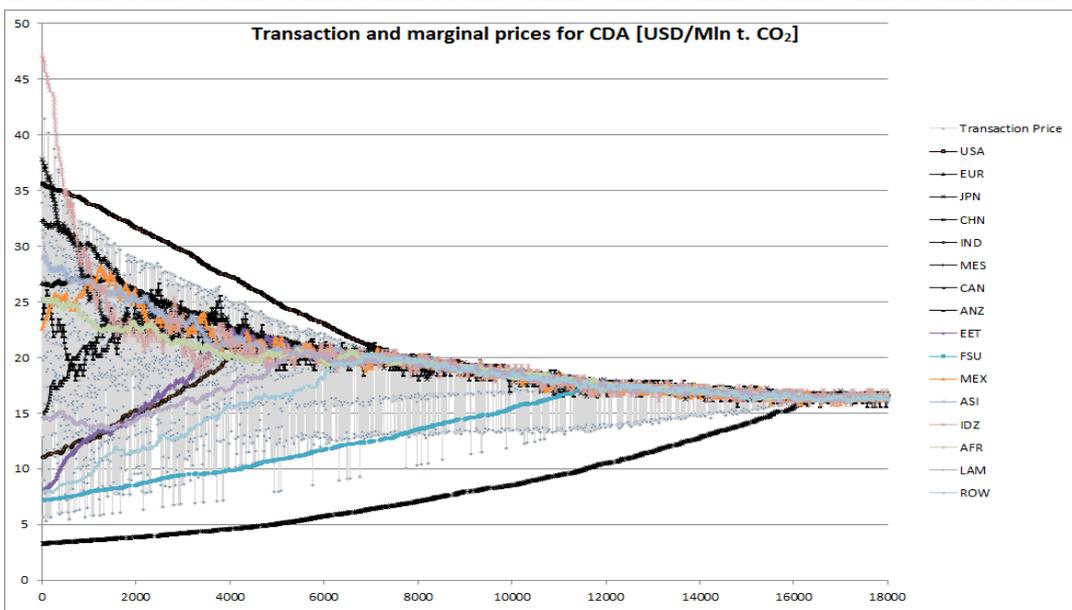


Figure 1. Exemplary cost curves and their approximations for USA and FSU in 2010.

Sample plots for the USA and FSU are shown in Figure 1. The approximated curve almost perfectly fits the original curve for FSU, while for USA the polynomial function is of too low order to fit the original data exactly. But for computational purposes this error is admissible.





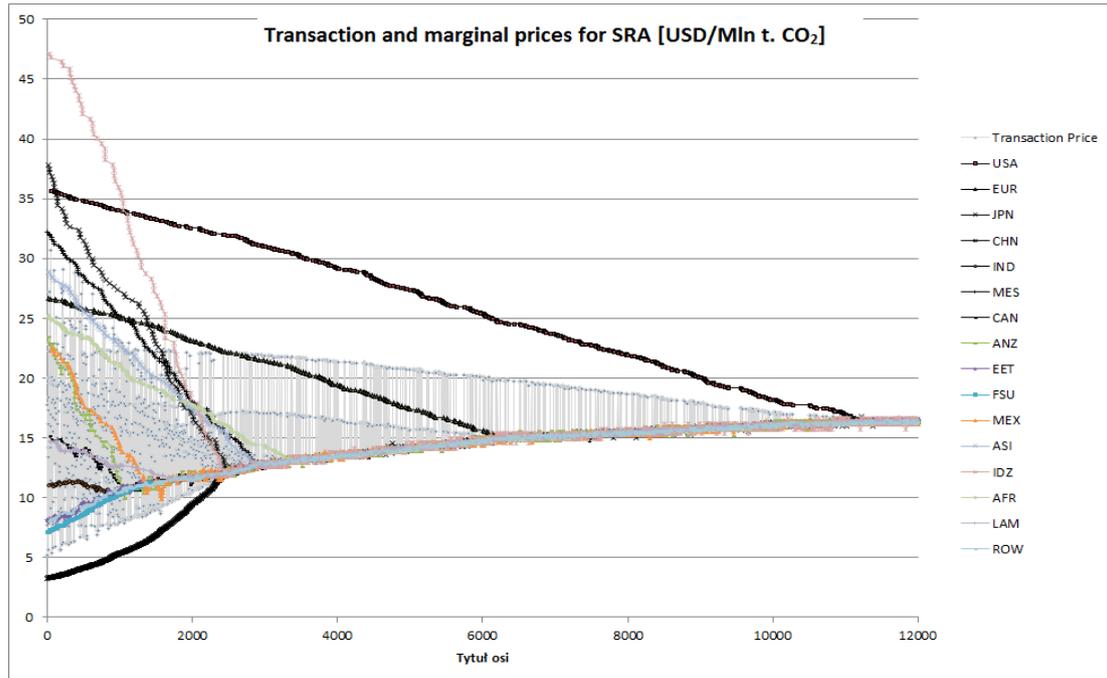
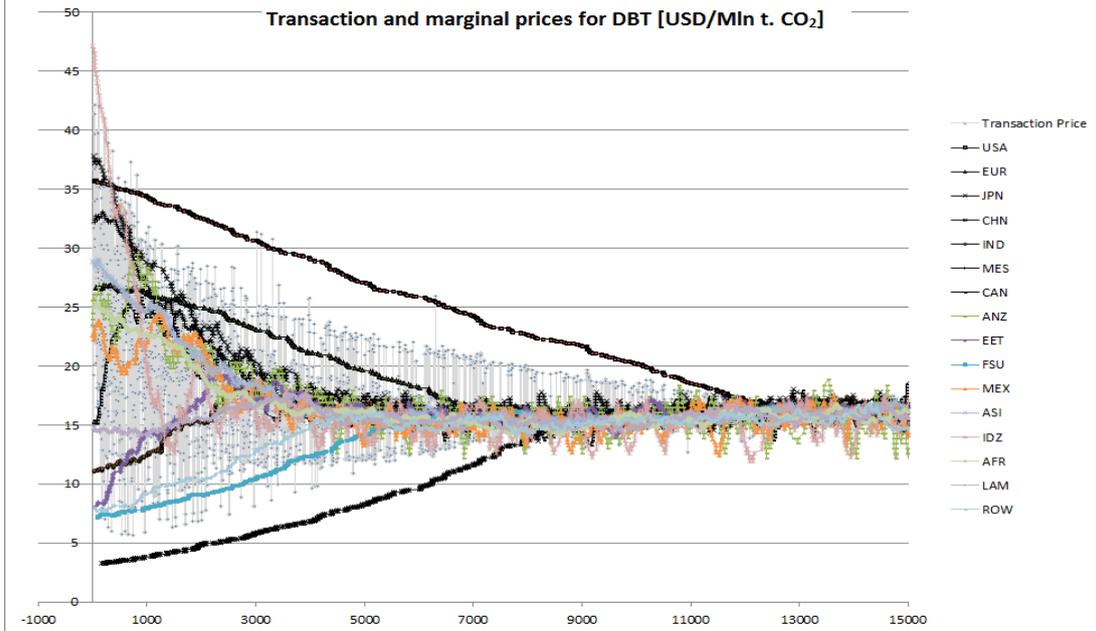
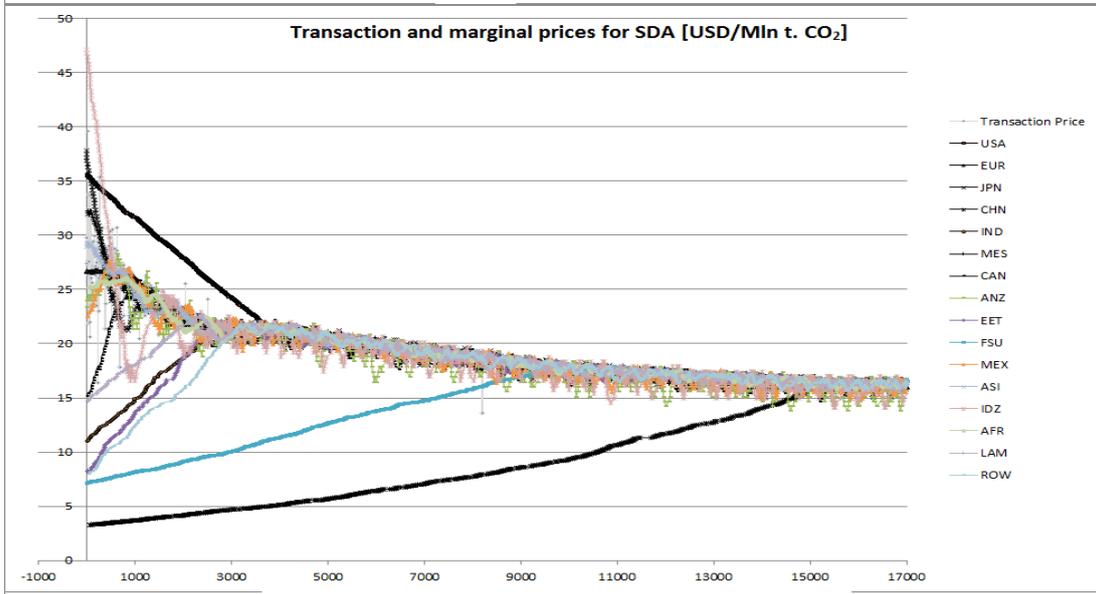
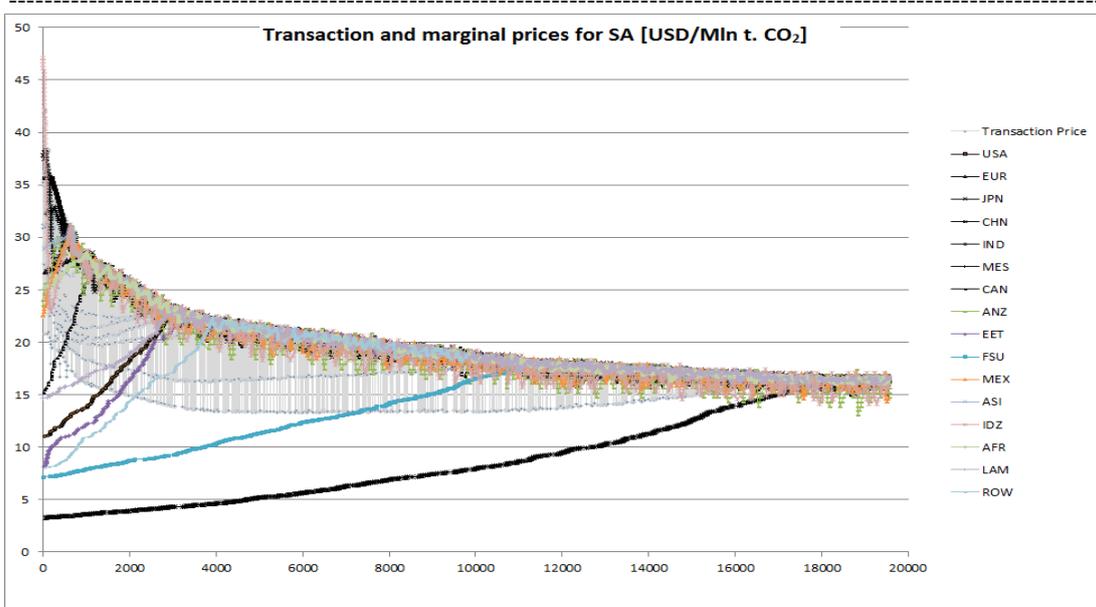


Figure 2 Trajectory of marginal and transaction prices in consecutive contracts, in [USD/MtC/y] for different trading mechanisms. In consecutive panels from the top: SA, DBT, SDA, CDA, SRA. Every party uses Frank strategy.

There are evident differences in parties trajectories for different cases that are connected with the random elements of the strategies. However, the following features can be observed for the Frank and ZIP strategies:

- For Continuous Double Auction and Dynamic Bilateral Trade, the plots of transaction prices are most spread out; the reason is high randomness of strategies used by parties which are concluding contracts, as in both mechanisms the decision of contract is made with less knowledge as compared with other auctions.
- For the Sealed-bid Auction the contracts are concluded one after the other by decreasing prices.
- For the Sealed-bid Reverse Auction, the contracts are concluded one after the other by increasing prices.
- For the Sealed-bid Double Auction, the plot of transaction prices is most condensed, as the contracts are concluded by choosing the most attractive offer among more than two competing ones.
- For the Frank strategy the plots of marginal prices are smoother than for the other strategies, what comes from the fact, that the parties do not use random strategies.
- In every simulation, the prices converge to a value about 16.4 [USD/MT CO₂]



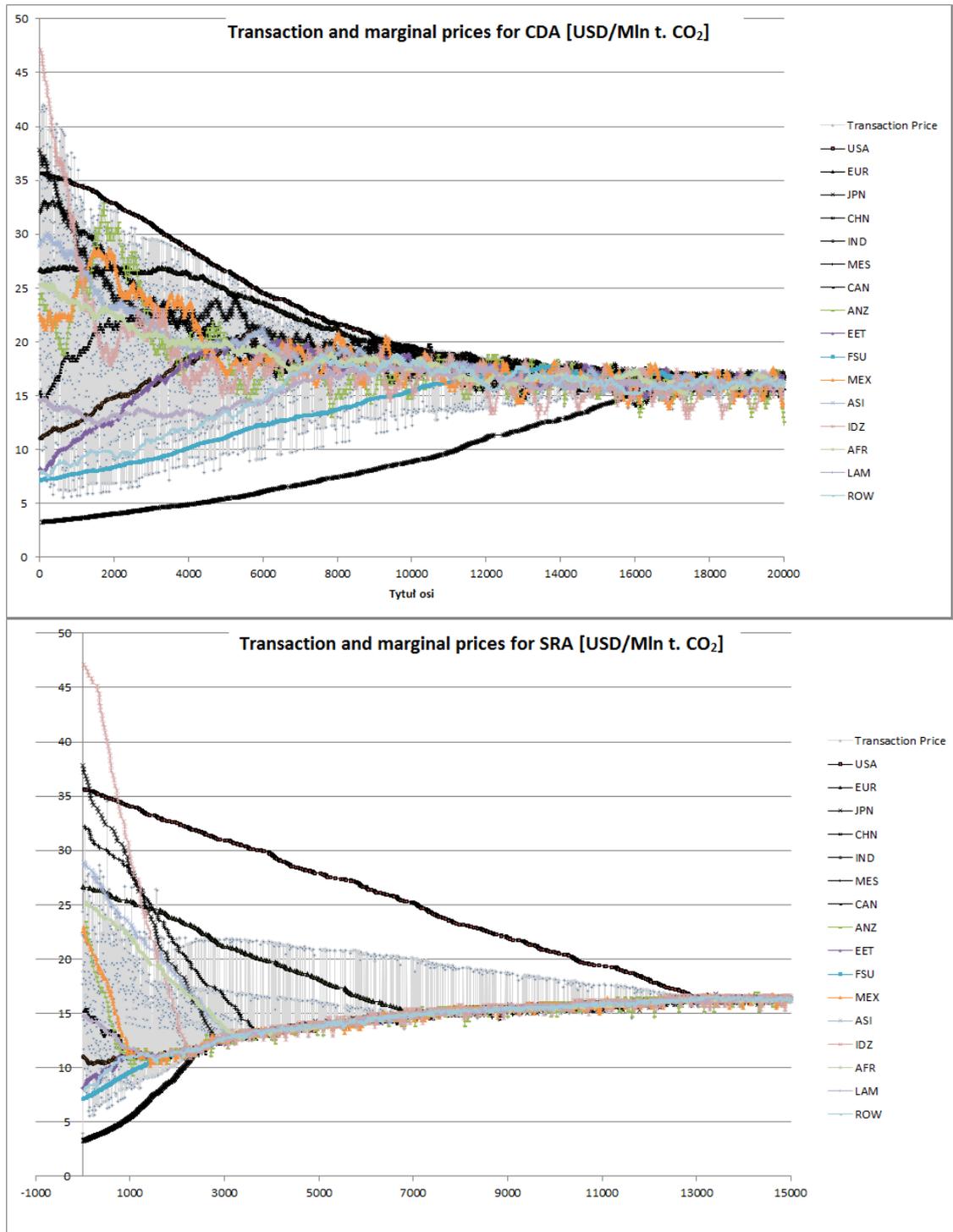


Figure 3 Trajectory of marginal and transaction prices in consecutive contracts, in [USD/MtC/y], for different trading mechanisms. In consecutive panels from the top : SA, DBT, SDA, CDA, SRA. Every party uses ZIP strategy.

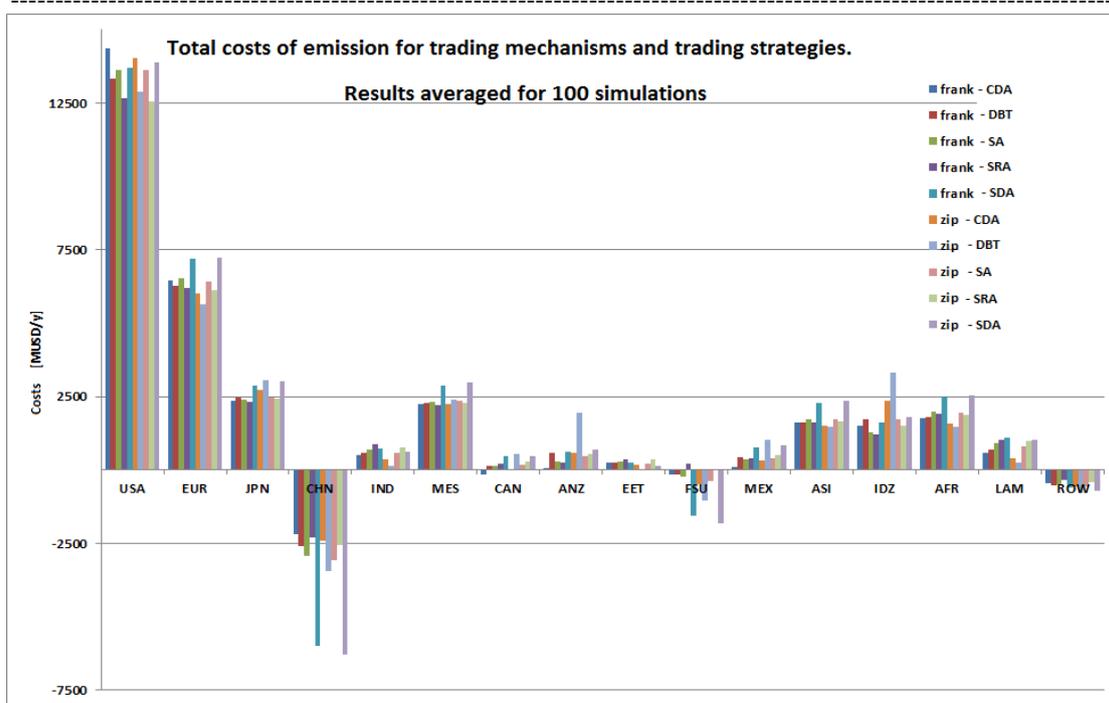


Figure 4. Total averaged costs of emission for trading mechanisms and trading strategies.

Analysing the averaged costs after 100 simulations we can note the following:

- USA mainly purchases the permits, so the difference among particular mechanisms is not significant; the lower costs can be observed for Sealed-bid Reverse Auctions (for both strategies), and for Dynamic Bilateral Trade with ZIP strategy.
- In turn, China (CHN) mostly sells the permits and benefits from the trade. The mechanisms that are most favourable for CHN is the Sealed-bid Double Auction.
- The same situation is for FSU, where this party benefits bet from the Sealed-bid Double Auctions, while for other mechanisms it benefits only marginally or even loses.
- EUR, ASI, AFR, and MES are the permits buyers, and similarly to CHN the best for them mechanism is the Sealed-bid Double Auction.
- For IND and LAM the lowest costs are for Continuous Double Auction and Dynamic Bilateral Trade with ZIP strategy.

The most interesting feature of the AA strategy for CDA is much shorter time of convergence of the transaction and marginal prices to the equilibrium. While the Frank strategy needed 1000-1600 iterations for convergence and ZIP strategy 1300-1600 iterations, the convergence for the AA strategy required about 300 iterations for transaction prices and about 500 iterations for the marginal prices. This is caused by very complicated strategy trying to elaborate as good offer as possible. The much

shorter negotiations are an important feature for applications, where agent representing real parties negotiate.

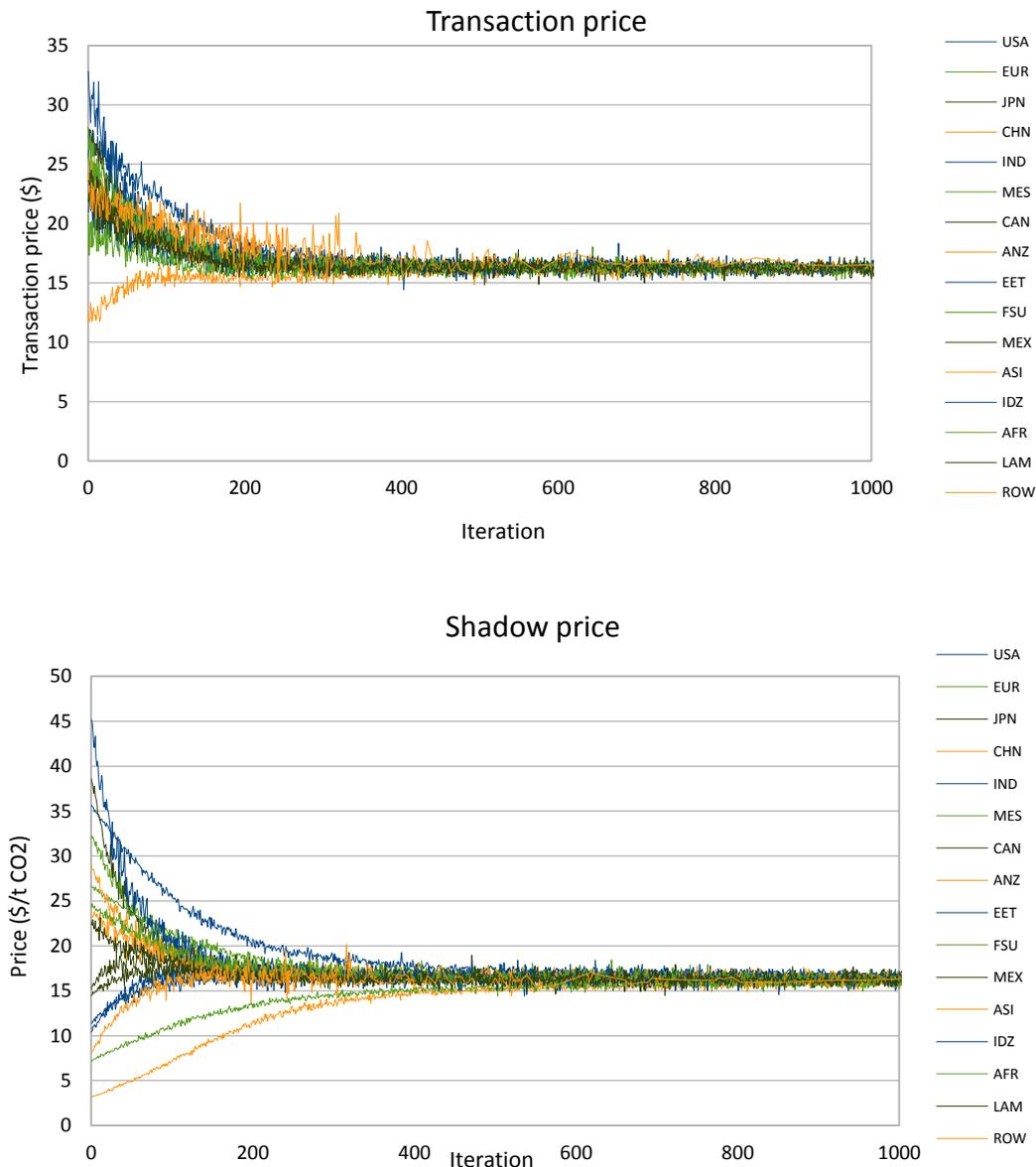


Figure 5. Averaged (100 iterations) values of transaction prices (upper panel) and shadow prices (lower panel) for 16 countries and regions for the Continuous Double Auction and AA strategy.

5. Conclusions

This paper presents the basic parts of a computer environment for simulation of emission permit trade using multi-agent framework. As far, different trading mechanisms, like bilateral trade and few types of auctions, and strategies which can be used by programmable agents, were reviewed, coded as computer subroutines, and described in the text. The results of simulated trade are presented. The case considered is a trade of GHG emission permits between 16 countries and regions of the world. The marginal costs for these parties were taken from [19].

Different simulated cases, each of which consists of a chosen pair of a trade mechanism and a strategy, showed differences in details, but similar general behaviour concerning convergence to the equilibrium and relative final total costs of the trading parties.

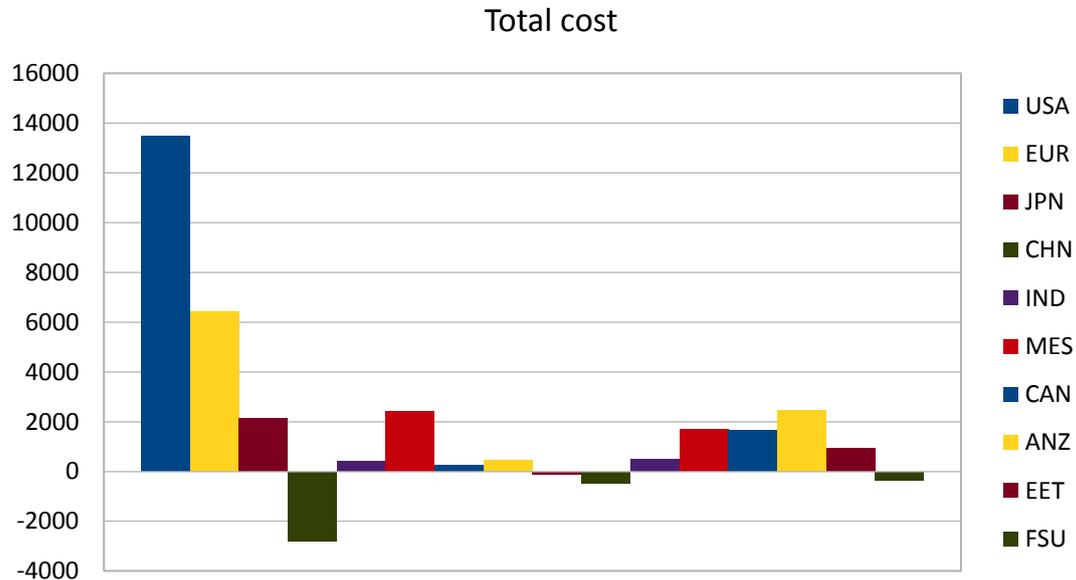


Figure 6. Averaged (100 iterations) values of total costs of CO₂ emission reduction for 16 countries and regions for the Continuous Double Auction and AA strategy.

The proposed simulation can be developed for estimating the equilibrium price in a designed market by implementing a simulated game, in which a programmable agent situated and operated by a playing party can use its secret information on marginal costs to take part in the game. Results of such a game would help to better design the market rules, not only by getting better information on the equilibrium price, but also on possible malfunctioning of the market.

The elaborated environment can be also developed for simulation of proposed markets for uncertain emissions. This kind of markets has not exist until now and can show unexpected features. The simulated experiments of such markets can demonstrate their strong and weak sides.

Acknowledgement

The research of Weronika Radziszewska was supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing. The project is financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund. The research of Piotr Pałka was supported by the research project “Coordination in cooperative multi-agent systems using reinforcement learning methods”.

References

- [1] Akimoto K., Sano F., Homma T., Wada K., Nagashima M., Oda J. (2012) Comparison of marginal abatement cost curves for 2020 and 2030: longer

- perspectives for effective global GHG emission reductions. *Sustainability Science*, 7:157-168, DOI: 10.1007/s11625-012-0165-5.
- [2] Babonneau F., Haurie A., Vielle M. (2013) A robust meta-game for climate negotiations. *Computational Management Science*, 10:299-329. DOI: 10.1007/s10287-013-0188-0.
- [3] Bernard A., Haurie A., Vielle M., Viguiier L. (2002) A two-level dynamic game of carbon emissions trading between Russia, China, and Annex B countries. NCCR-WP4 Working Paper 11, Swiss National Centre of Competence, University of Geneva & Paul Scherrer Institute. Available at <http://ecolu-info.unige.ch/~nccrwp4/>.
- [4] Cai K., Niu J., Parsons S. (2007) Using evolutionary game-theory to analyse the performance of trading strategies in a continuous double auction market. In: *Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning*, pp. 44-59.
- [5] Cliff D. (1997) Minimal-intelligence agents for bargaining behaviors in market-based environments. Technical Report, School of Cognitive and Computing Sciences, University of Sussex.
- [6] Cliff D. (2005) ZIP60: Further explorations in the evolutionary design of online auction market mechanisms. Technical Report, School of Cognitive and Computing Sciences, University of Sussex.
- [8] Ermoliev Y., Ermolieva T., Jonas M., Obersteiner M., Wagner F., Winiwarter W. (2015) Integrated model for robust emission trading under uncertainties: Cost-effectiveness and environmental safety. *Technological Forecasting & Social Change*, <http://dx.doi.org/10.1016/j.techfore.2015.01.003>.
- [9] Ermolieva T., Ermoliev Y., Jonas M., Obersteiner M., Wagner F., Winiwarter W. (2014) Uncertainty, cost-effectiveness and environmental safety of robust carbon trading: integrated approach. *Climatic Change*, 124(3):633-646, DOI: 10.1007/s10584-013-0824-2.
- [10] Gjerstad S., Dickhaut J. (1998) Price formation in double auctions. *Games and Economic Behaviour*, 22:1-29.
- [11] Gode D.K., Sunder S. (1993) Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119-137.
- [12] He M., Jennings N. R., Leung H. (2003a) On agent-mediated electronic commerce. *IEEE Trans. Knowledge and Data Engineering*, 15(4):985-1003.
- [13] He M., Leung H., Jennings N.R. (2003b) A fuzzy-logic based bidding strategy by autonomous agents in continuous double auctions. *IEEE Transactions on Knowledge and Data Engineering*, 15:1345-1363.
- [14] Marland E., Cantrell J., Kiser K., Marland G., Shirley K. (2014) Valuing uncertainty. Part I: The impact of uncertainty in GHG accounting. *Carbon Management*, 5(1):35-42.
- [15] Marland E., Cantrell J., Kiser K., Marland G., Shirley K. (2014) Valuing uncertainty. Part I: The impact of risk charges in dealing with time issues in lifecycle analysis and GHG accounting. *Carbon Management*, 5(1):43-53.

- [16] Morris J., Paltsev S., Reilly J. (2012) Marginal abatement costs and marginal welfare costs for greenhouse gas emissions reductions: results from the EPPA model. *Environmental Modeling & Assessment*, 17:325-336, DOI: 10.1007/s10666-011-9298-7.
- [17] Nahorski Z., Radziszewska W. (2012) Price formation strategies of programmable agents in continuous double auctions. In: Busłowicz M., Malinowski K. (Eds) *Advances in Control Theory and Automation. Committee of Automation and Robotics of the Polish Academy of Sciences*, 181-194.
- [18] Nahorski Z., Stańczak J., Pałka P. (2014) Simulation of an uncertain emission market for greenhouse gases using agent-based methods. *Climatic Change*, 124(3):647-662, DOI: 10.1007/s10584-013-1039-2.
- [19] Paltsev, S., Reilly, J., Jacoby, H., Eckaus, R., McFarland, J., Sarofim, M., Asadoorian, M., Babiker, M. (2005). The MIT Emissions Prediction and Policy Analysis (EPPA) model: Version 4, MIT Joint Program on the Science and Policy of Global Change Report 125, August. http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt125.pdf (accessed 30.06.2015).
- [20] Preist Ch., van Tol M. (2003) Adaptive agents in a persistent shout double auction. Technical Report HPL-2003-242, Hewlett Packard.
- [21] Rust J., Miller J.H., Palmer R. (1994) Characterizing effective trading strategies. Insights from a computerized double auction tournament. *Journal of Economic Dynamics and Control*, 18:61-96.
- [22] Shirley K., Cantrell J., Kiser K., Marland E., Marland G. (2014) Valuing uncertainty. Part II: The impact of risk charges in dealing with time issues in lifecycle analysis and GHG accounting. *Carbon Management*, 5(1):43-53.
- [23] Vytelingum P. (2006) The structure and behaviour of the continuous double auction. PhD Thesis, University of Southampton.
- [24] Vytelingum, P., Cliff, D., Jennings, N.R. (2008) Strategic bidding in continuous double auctions, 172(14):1700–1729.
- [25] Vytelingum P., Dash R.K., He M., Jennings N.R. (2005) A framework for designing strategies for trading agents. In: *IJCAI Workshop on Trading Agent Design and Analysis*, pp. 7-13.
- [26] Wagner F., Amman M., Borken-Kleefeld J., Cofala J., Höglund-Isaksson L., Purohit P., Rafaj P., Schöpp W., Winiwarter W. (2012) Sectoral marginal abatement cost curves: implications for mitigation pledges and air pollution co-benefits for Annex I countries. *Sustainability Science*, 7:169-184, DOI: 10.1007/s11625-012-0167-3.
- [27] Willet K., Caplanova A., Sivak R. (2015) Emission discharge permits, ambient concentrations, and a margin of safety: trading solutions with a computer-assisted smart market for air quality. *Environmental Modeling & Assessment*, 20:367-382, DOI: 10.1007/s10666-014-9436-0.

Urban metabolism: atmospheric loads and fluxes of major greenhouse gases (CO₂, CH₄) in Krakow, southern Poland

Mirosław Zimnoch¹, Jarosław Necki¹, Łukasz Chmura^{1,2}, Alina Jasek¹, Michał Galkowski¹, Tadeusz. Kuc¹, Zbigniew Gorczyca¹, Jakub Bartyzel¹, Kazimierz Rozanski¹

1 - AGH University of Science and Technology, Krakow, Poland

2 - Institute of Meteorology and Water Management, National Research Institute, IMGW-PIB
Branch of Krakow, Krakow, Poland
zimnoch@agh.edu.pl

Abstract

Emissions of carbon dioxide from fossil fuel burning constitute an important component of atmospheric carbon budget, on both global and regional scales. In urban settings across Europe anthropogenic CO₂ burden of local atmospheres is particularly visible. Methane is the second most important anthropogenic greenhouse gas. In southern Poland, apart of CH₄ emissions associated with agricultural sector and coal production, other sources of anthropogenic methane including leakages from urban gas supply networks and numerous landfills in the region are also abundant. We are presenting a brief overview of long-term investigations aimed at quantification of atmospheric loads and fluxes of CO₂ and CH₄ in Krakow agglomeration using atmospheric observations combined with ground-level measurements and modeling.

Keywords: carbon dioxide, methane, urban metabolism, greenhouse gases

1. Introduction

Urban centers play an important role in the global greenhouse gas emissions. More than 70% of anthropogenic sources of CO₂ are located in urban areas [1]. The European continent, with its dense transportation network, developed industrial infrastructure and high population density is currently responsible for more than 25 % of the global fossil fuel CO₂ emissions [2]. Urban agglomerations are also important source of methane. It has been estimated that aging natural gas distribution networks are leaking at the rate of up to 8% of the total gas consumption [3].

Here we present the results of long-term investigations aimed at quantification of atmospheric loads and fluxes of CO₂ and CH₄ in Krakow agglomeration using atmospheric observations combined with ground-level measurements and modeling. In the first part we demonstrate how measurements of atmospheric mixing ratios of carbon dioxide, combined with analyses of its carbon isotope composition (¹³C/¹²C and ¹⁴C/¹²C ratios) can be used to make partitioning of the local atmospheric CO₂ budget and quantifying the contribution of fossil fuel-derived CO₂ in the local atmosphere. In the second part, we present two independent methods of assessing surface fluxes of CO₂ and CH₄ originating from distributed sources.

Figure 1 shows schematic diagram of carbon dioxide and methane budget in the urban atmosphere. Typical for urban regions are generally elevated concentrations of CO₂ and CH₄ in the local atmosphere, when compared to remote, clean areas [4]. These local loads are caused by intense surface emissions of CO₂ and CH₄ from both point and distributed sources. When combined with diurnal variations in the intensity of vertical mixing of the lower atmosphere, they often lead to characteristic variations in

atmospheric concentrations of CO₂ and CH₄ in urban regions, with high levels of those gases recorded during the night and reduced concentrations during mid-day and afternoon hours.

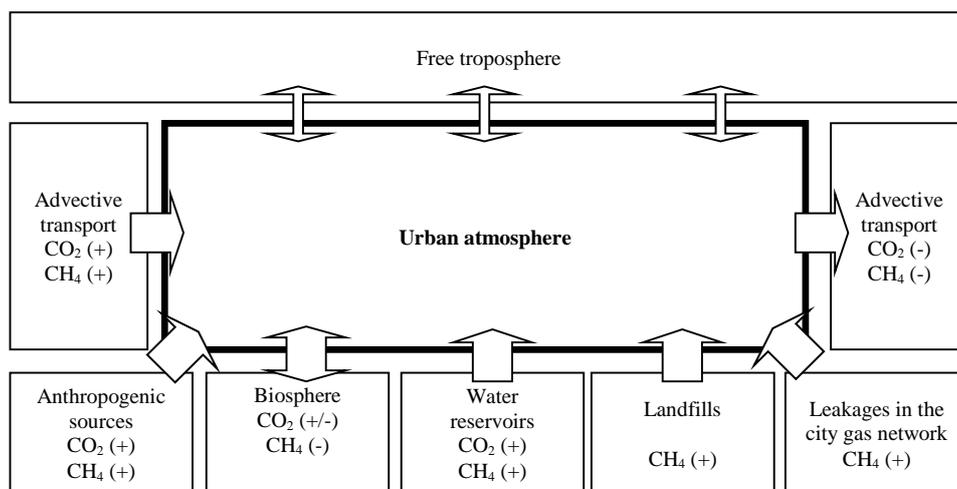


Figure 1. Schematic diagram of carbon dioxide and methane budget for urban areas. Sources and sinks of CO₂ and CH₄ are marked by '+' and '-' sign, respectively.

2. Site description

Krakow, the second largest city in Poland (approx. 800 000 inhabitants) is located in the Vistula river valley. Characteristic features of the local climate are significant percentage of calm periods (ca. 36%) and frequent temperature inversions, sometimes extending over several days. The average wind speed for the period 2005-2009 was around 3.3 m s⁻¹. West and south-west direction of surface winds prevail. Westerly circulation is generally connected with stronger winds (wind speeds above 4 m s⁻¹). Monthly air temperature at the site reveals a distinct seasonal cycle, with summer maximum (July-August) reaching ca. 20-25°C and winter minimum (January-February) between ca. -5 and 0°C.

Atmospheric measurement site (red dot on Figure 2) was situated on the University campus located in the western part of the city (50°04'N, 19°55'E, 220 m a.s.l.), bordering recreation and sports grounds. Air intake for atmospheric measurements was located on the roof of the Faculty building, ca. 20 meters above the local ground.

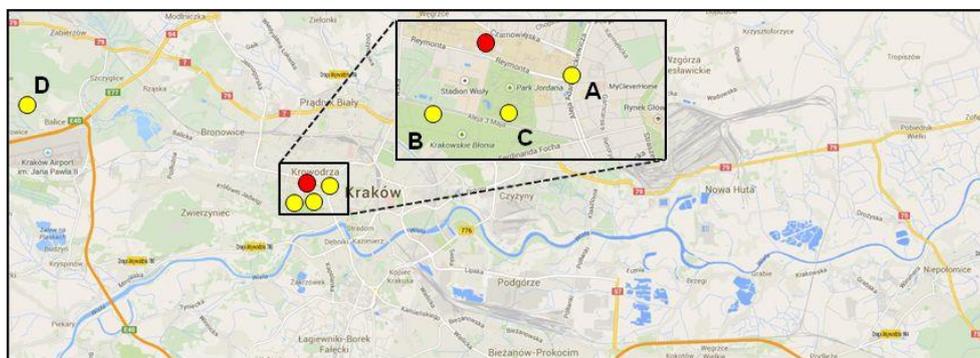


Figure 2 Location of the measurement points. Red dot represents location of atmospheric measurements, yellow points represent sites where measurements of soil fluxes of CO₂ and CH₄ were conducted.

Four sites located within the limits of Krakow agglomeration were chosen for measurements of soil fluxes of CO₂ and CH₄ (Figure 2): (i) site A (50°03'51"N, 19°55'26"E, 209 m a.s.l.) located in immediate vicinity of one of the city's major streets, (ii) site B (50°03'41"N, 19°54'08"E, 203 m a.s.l.) placed within urban meadow Blonia – 48 ha grassland recreation area in the city center with no car traffic in the direct neighborhood, (iii) site C (50°03'39"N, 19°55'08"E, 205 m a.s.l.) placed in the city park between sites A and B, and (iv) site D (50°05'24"N, 19°48'07"E, 247 m a.s.l.) located in the home garden on the outskirts of the city, in the upwind distance of ca. 11km west from other points. Site D was considered a reference site, not directly influenced by the city.

3. Methods

3.1 Atmospheric loads and partitioning of CO₂ budget

Regular measurements of atmospheric mixing ratios of CO₂ and CH₄ were performed using gas chromatograph HP6890 equipped with Shin Carbon ST column, Ni catalyst and FID detector [5]. Typical uncertainty of mixing ratio measurements was in the order of 0.1 μmol mol⁻¹ and 5 nmol mol⁻¹, for CO₂ and CH₄, respectively. In addition, during the period 2007-2009 several diurnal measurement campaigns aimed at collection of air samples for mixing ratio and isotopic analysis of CO₂ were performed. For each campaign 5 to 7 spot air samples were collected in 4-hour intervals. Carbon dioxide was cryogenically extracted from the collected air samples for further isotope analyses using IRMS and AMS techniques to determine the isotopic composition of carbon (δ¹³C and Δ¹⁴C) [6,7]. The measured ¹³C/¹²C ratios are expressed in delta notation on the VPDB scale [8], while the radiocarbon content is expressed in capital delta notation relative to Oxalic Acid standard and normalized for the year 1950 [9]. Overall uncertainty of isotope measurements was in the order of 0.2‰ for δ¹³C and 0.7 ‰ for Δ¹⁴C.

The measured atmospheric CO₂ mixing ratio at a given site (C_{meas}) consists of three components: (i) regional background component (C_{bg}) which provides the bulk of the atmospheric CO₂ load, (ii) biogenic component (C_{bio}), and (iii) fossil-fuel derived component (C_{foss}). When the carbon isotope composition of atmospheric CO₂ is measured in addition to atmospheric mixing ratio, the following mass and isotope balance equations can be formulated:

$$C_{meas} = C_{bg} + C_{bio} + C_{foss} \quad (1)$$

$$(\delta^{13}C_{meas} + 1)C_{meas} = (\delta^{13}C_{bg} + 1)C_{bg} + (\delta^{13}C_{bio} + 1)C_{bio} + (\delta^{13}C_{foss} + 1)C_{foss} \quad (2)$$

$$(\Delta^{14}C + 1)C_{meas} = (\Delta^{14}C_{bg} + 1)C_{bg} + (\Delta^{14}C_{bio} + 1)C_{bio} + (\Delta^{14}C_{foss} + 1)C_{foss} \quad (3)$$

Equations 1-3 enable partitioning of the measured CO₂ load in the local atmosphere into fossil-fuel derived, biogenic and regional background components (see [4] for details).

3.2 Direct measurements of surface fluxes

Soil CO₂ and CH₄ fluxes were measured using closed-dynamic chamber system coupled with Picarro G2101-i trace gas analyser. CO₂ concentration measurements inside the chamber were performed every second. The chamber method is a widely used

technique for measuring gas exchange between soil surface and the atmosphere [10,11,12].

The chamber system used in this work was designed and built in the Faculty of Physics and Applied Computer Science. The system consists of three stainless steel chambers of cylindrical shape, coupled with a collar hammered to the depth of ca. 5 cm into the soil and equipped with water seal to avoid gas leakages. The system was equipped with a flow regulator stabilizing the air flow through each chamber, pressure, temperature and relative humidity sensors installed in each chamber and a drying agent (magnesium perchlorate, $Mg(ClO_4)_2$) used for drying the air collected in the glass flasks for isotope analyses. In order to measure the rate of changes of CO_2 and CH_4 concentration inside the chamber, each chamber was sequentially connected to the analyzer during the measurement period. One path was equipped with flask ports allowing to take air samples for further analyses of isotopic composition of CO_2 using IRMS technique. The flux of CO_2 and CH_4 was calculated using the following expression:

$$f = \frac{p \cdot V \cdot \frac{dC}{dt}}{R \cdot T \cdot S} \quad (4)$$

where:

f - molar flux density ($mmol\ m^{-2}\ h^{-1}$),

p - atmospheric air pressure (Pa),

dC/dt - rate of concentration change of the given gas inside the chamber ($mmol\ mol^{-1}\ h^{-1}$),

R - universal gas constant ($kg\ m^2\ s^{-2}\ mol^{-1}\ K^{-1}$),

T - ambient air temperature (K)

V and S - chamber volume (m^3) and soil surface area covered by the chamber (m^2),

3.3 Indirect assessment of surface fluxes

During the day, when thermal convection operates in the lower atmosphere, trace gases emitted from the surface are dissolved in a large volume of well-mixed layer within the Planetary Boundary Layer (PBL), leading to relatively low concentrations of those gases observed close to the ground. Inversion of vertical temperature profile in the lower atmosphere during late afternoon leads to drastic reduction of the intensity of vertical mixing and subsequent accumulation of trace gases in near-ground atmosphere.

Diurnal changes in the dynamics of vertical mixing within PBL in combination with nocturnal boundary layer (NBL) height measurements can be applied for the assessment of surface fluxes of selected trace gases. The rate of nocturnal increase of atmospheric concentrations of trace gases is controlled by the mixing layer height according to mass balance equation:

$$H \frac{d \langle C \rangle}{dt} = F_{in} - F_{out} \quad (5)$$

where:

H - height of the mixing layer

$\langle C \rangle$ - mean concentration

F_{in} - surface flux

F_{out} - flux associated with removal processes (horizontal and vertical transport). For nights with low wind speed ($< 1 \text{ m s}^{-1}$) and the adopted frequency of measurements, this term can be neglected.

During stable atmospheric conditions with low wind speeds, a distinct vertical gradient of trace gas concentration is established within the PBL. As the measurements are performed close to the surface, at the height of ca. 20 m, a correction factor relating the average increase of the concentration within the PBL ($d\langle C \rangle/dt$) to the increase of this concentration observed close to the ground is required:

$$F_{in} = \frac{H}{k} \frac{dC_{surf}}{dt} \quad (6)$$

where:

H - mixing layer height

k - correction factor

C_{surf} - concentration of the measured trace gas at the adopted measurement height.

The correction factor k was calculated using numerical simulation of vertical profiles of atmospheric ^{222}Rn using EMEP atmospheric model [13]. Radon is a natural radioactive noble gas emitted from soils. The soil radon flux is mainly controlled by concentration of ^{226}Ra and in some extent by physical properties of the soil. This fact enables the use of ^{222}Rn in the assessment of surface fluxes of other trace gases such as CO_2 and CH_4 . The NBL height H was monitored using VHS sodar (Version 3) built by the Krakow Branch of the Institute of Meteorology and Water Management. Detailed description of the sodar system can be found in [14] whereas flux assessment methodology is described in [15].

4. Results and discussion

4.1 Atmospheric loads and partitioning of CO_2 budget

Figure 3a shows an example of atmospheric loads of CO_2 and CH_4 in Krakow during the month of June 2007. Shown are daily means of the respective mixing ratios. For comparison, Figure 3a contains also daily means of CO_2 and CH_4 recorded at Kasprowy Wierch station, ca. 100 km south of Krakow, located on top of the Kasprowy Wierch mountain in the Polish Tatras (1989 m a.s.l). This station can be considered a regional background station, free of local influences [5]. It is apparent from Figure 3a that urban atmosphere of Krakow is characterized by elevated concentrations of the measured trace gases. The excess concentrations with respect to Kasprowy Wierch can be substantial; for instance, at June 15th the daily mean CO_2 concentration in Krakow was around 462 ppm, to be compared with ca. 383 ppm at Kasprowy Wierch. Also methane revealed elevated concentration: ca. 2.33 ppm to be compared with 1.85 ppm at Kasprowy Wierch.

Figure 3b shows the results of event sampling performed on 20/21 June 2007. Shown are the values of three parameters: CO_2 mixing ratio, $\delta^{13}\text{C}$ and radiocarbon content of CO_2 . Distinct diurnal variability of CO_2 mixing ratios is observed, with the maximum (417 ppm) recorded late in the night and minimum (391 ppm) in the afternoon. Initial increase of CO_2 mixing ratio is accompanied by drop of $\delta^{13}\text{C}$ and increase of $\Delta^{14}\text{C}$. Later on, during the day 21th June significantly lower values of $\Delta^{14}\text{C}$ are recorded.

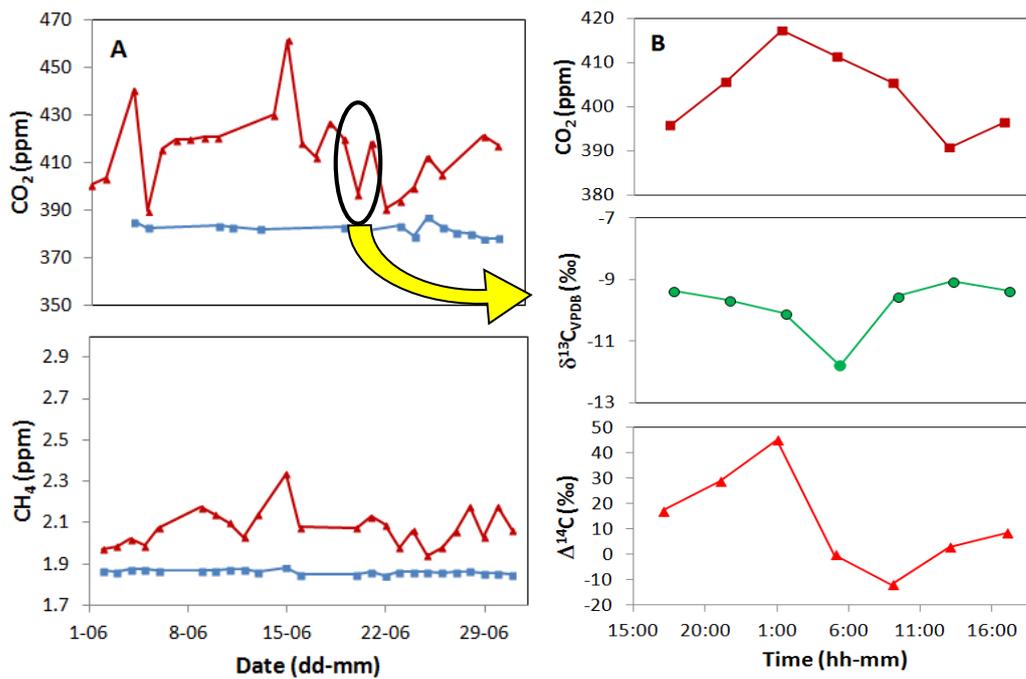


Figure 3. (A) Daily means of CO₂ and CH₄ mixing ratios recorded at Krakow and Kasprowy Wierch stations during June 2007. (B) concentration and carbon isotope composition of atmospheric CO₂ recorded during measurement campaign carried out between 20 and 21 June 2007 in Krakow.

The data shown in Figure 3b were used to calculate the budget of CO₂ in the local atmosphere of Krakow during 20/21 June 2007. Equations 1-3 were used to derive temporal evolution of individual components of this budget (fossil-fuel derived, biogenic and regional CO₂ background component). The results are presented in Figure 4. The calculated contributions indicate a dominant role of biogenic CO₂ emissions during night hours. During day time anthropogenic emissions dominate. The calculated background mixing ratios of CO₂ are in good agreement with the measurements performed at this time period at Kasprowy Wierch station [4].

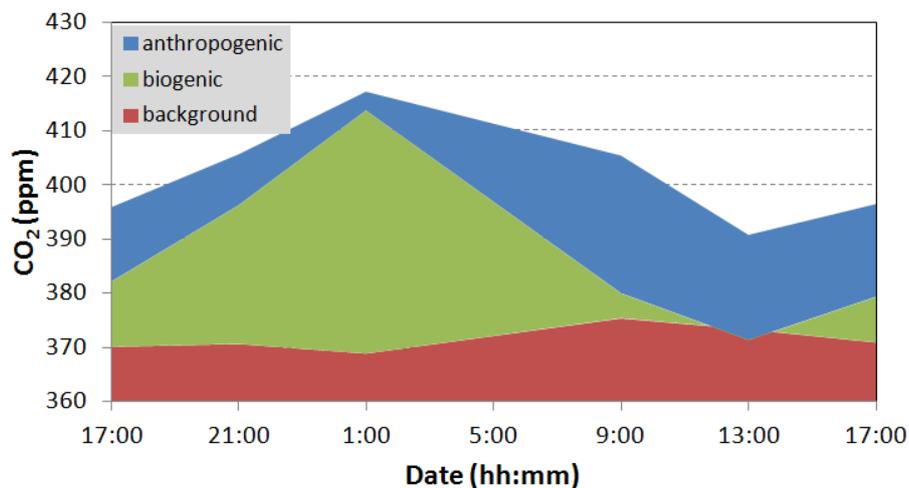


Figure 4. Partitioning of atmospheric CO₂ load over Krakow during 20/21 June 2007 into fossil-fuel derived, biogenic and regional CO₂ background component.

4.2 Direct measurements of surface fluxes

Figure 5 summarizes soil CO₂ fluxes measured at four different sites in Krakow between July 2009 and June 2013. All investigated sites reveal a strong seasonality of CO₂ emissions induced by seasonal vegetation cycle and seasonal variations of soil parameters (temperature, water content).

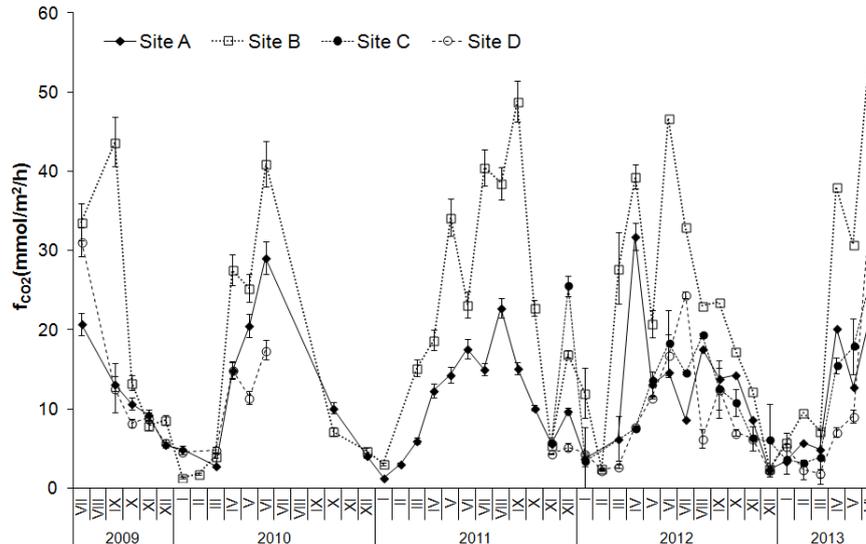


Figure 5. Seasonal variations of soil CO₂ fluxes measured at four selected ecosystem sites in Krakow agglomeration (A-D - cf. Figure 2), differing in degree of anthropogenic influence.

Minimum values of the CO₂ fluxes were recorded during winter months (December, January and February). Low soil temperatures limit respiration activity, while higher water contents reduce permeability of the soil during this season. Snow cover, if present, may further reduce transport of CO₂ between the soil and the atmosphere during winter months. Typical CO₂ fluxes measured during winter varied between approximately 1 and 5 mmol m⁻² h⁻¹ for all sites. During the vegetation period (April to October) the CO₂ fluxes increase significantly, reaching maximum values of 25-30 mmol m⁻² h⁻¹ at site A, C and D and 40-50 mmol m⁻² h⁻¹ at site B. Timing of the CO₂ flux maxima varies from year to year. The mean value of soil CO₂ flux calculated for entire observation period and all four sites reaches 16.2 mmol m⁻² h⁻¹ and is comparable to the mean local fossil fuel CO₂ flux, which is approximately 17.8 mmol m⁻² h⁻¹ (2008) [2]. Measurements of methane fluxes revealed small negative numbers (between -1 and -2 μmol m⁻² h⁻¹ for all urban sites and close to zero for the reference site D) indicating that soils in the study area are a weak sink of methane. This fact suggests that apparent source of methane leading to night-time maxima frequently observed in atmospheric CH₄ record available for Krakow station, has entirely anthropogenic origin.

4.3 Indirect assessment of surface fluxes

Calculations of surface night-time fluxes of CO₂ and CH₄ based on simplified atmospheric budget of these gases (cf. section 3.3) were performed for the period from May 2005 to May 2009, for nights with wind speed lower than 1m s⁻¹ (for details see [15]). The calculated mean monthly surface fluxes of CO₂ (Figure 6) reveal distinct

seasonality, with a minimum of ca $2 \text{ mmol m}^{-2} \text{ h}^{-1}$ occurring during winter time (December to March) and a maximum of ca. $20 \text{ mmol m}^{-2} \text{ h}^{-1}$ in summer (July and August). This seasonality is also seen in direct measurements of soil CO_2 fluxes (cf. Figure 5) and is modulated by the biospheric component, originating from soil respiration. Relative uncertainty of the monthly mean CO_2 fluxes derived from simplified atmospheric budget varies between ca. 1 and 87%.

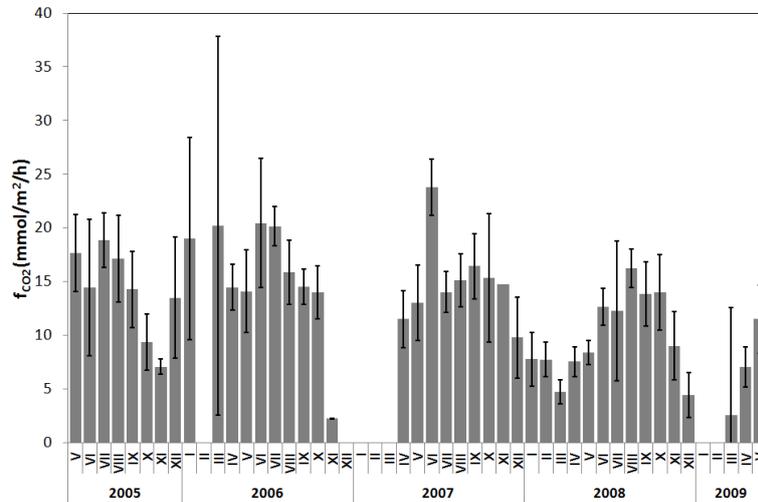


Figure 6. Monthly means of surface-averaged CO_2 flux for Krakow city derived from the atmospheric balance method.

Contrary to the CO_2 flux, the surface fluxes of CH_4 (Figure 7) do not reveal any distinct seasonality and, in contrast to chamber measurements, are distinctly positive. The monthly means of the calculated CH_4 flux scatter between ca. 50 and $200 \text{ } \mu\text{mol m}^{-2} \text{ h}^{-1}$, with relative uncertainty of the mean values varying between 10 and 227%. The mean CH_4 flux calculated for the whole analyzed period (May 2005 – December 2008) is equal $97.2 \pm 5.4 \text{ } \mu\text{mol m}^{-2} \text{ h}^{-1}$. The quoted uncertainty represents one standard deviation of the mean value. Multiplying the mean CH_4 flux by the surface area of the city (326.8 km^2), leads to the total mean annual emission of methane into the atmosphere in Krakow in the order of $(6.2 \pm 0.4) \times 10^6 \text{ m}^3 \text{ yr}^{-1}$. Leakages of the city gas network are thought to be the main source of this methane.

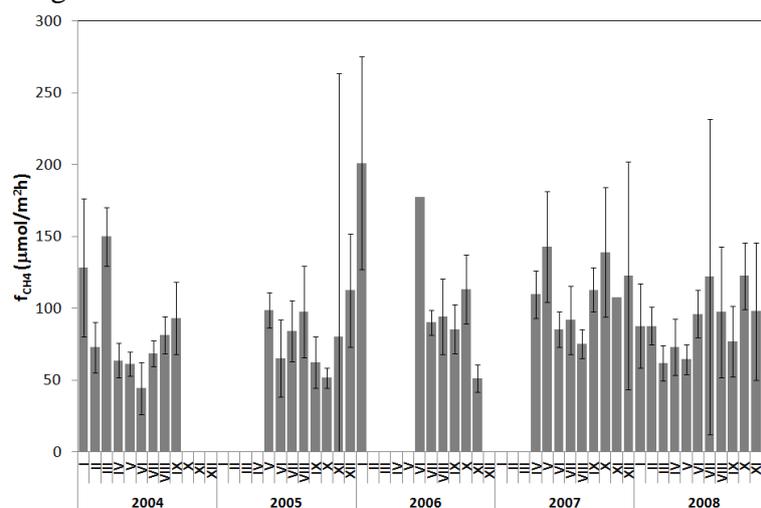


Figure 7. Monthly means of surface-averaged CH_4 flux for Krakow derived from the atmospheric balance method.

5. Summary

Urban centers are important elements of the global carbon cycle. Their importance will likely increase in the foreseeable future, with the continuing transformation of the world population from rural to urban. It is therefore important to gain a deeper understanding of carbon cycling in the urban environment in order to design appropriate strategies to control emissions of major greenhouse gases into the atmosphere.

An overview of long-term investigations aimed at quantification of atmospheric loads and fluxes of CO₂ and CH₄ in Krakow agglomeration using atmospheric observations combined with ground-level measurements and modeling, presented above, reveals the potential of such approach for quantifying atmospheric loads and surface fluxes of major greenhouse gases in complex urban environment, thus constituting an independent verification tool for greenhouse gas emissions reported in the framework of international agreements by signatory countries.

Acknowledgements

The work has been partially supported by EU projects EUROHYDROS and CARBOEUROPE, grants of the Ministry of Science and Higher Education (project nos. 2256/B/P01/2007/33 and 4132/B/T02/2008/43) and the statutory funds of the AGH University of Science and Technology (project no. 11.11.220.01).

References

- [1] Velasco E. and Roth M., 2010, Cities as net sources of CO₂: Review of atmospheric CO₂ exchange in urban environments measured by eddy covariance technique. *Geography Compass* 4/9, pp.1238-1259.
- [2] European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR), release version 4.2. <http://edgar.jrc.ec.europa.eu>, 2011.
- [3] Alvarez, R. A.; Pacala, S. W.; Windbrake, J. J.; Chameides, W. L.; Hamburg, S. P., Greater focus needed on methane leakage from natural gas infrastructure. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 109 (17), 6435-6440.
- [4] Mirosław Zimnoch, Dorota Jelen, Michał Galkowski, Tadeusz Kuc, Jarosław Necki, Łukasz Chmura, Zbigniew Gorczyca, Alina Jasek and Kazimierz Rozanski 2012. Partitioning of atmospheric carbon dioxide over central Europe: insights from combined measurements of CO₂ mixing ratios and their carbon isotope composition, *Isotopes in Environment and Health Studies* 48(3), 421-433.
- [5] Necki J.M., Schmidt M., Rozanski K., Zimnoch M., Korus A., Lasa J., Graul R., Levin I., 2003. Six-year record of atmospheric carbon dioxide and methane at a high-altitude mountain site in Poland. *Tellus*, 55B, 94-104.
- [6] Zimnoch M, Florkowski T, Necki J, Neubert R, 2004. Diurnal variability of delta C-13 and delta O-18 of atmospheric CO₂ in the urban atmosphere of Krakow, Poland, in *Environment and Health Studies* 40 (2): 129-143.
- [7] Goslar T. and Czernik J. 2000. Sample preparation in the Gliwice Radiocarbon Laboratory for AMS ¹⁴C dating of sediments. *Geochronometria* 18, 1-8.

- [8] Coplen T. 1996. New guidelines for reporting stable hydrogen, carbon and oxygen isotope-ratio data. *Geochim. Cosmochim. Acta*, 60, 3359-3360.
- [9] Mook W.G. and van der Plicht J. 1999. Reporting ^{14}C activities and concentrations. *Radiocarbon* 41(3), 227-239.
- [10] Norman, J. M., Kucharik, C. J., Gower, S. T., Baldocchi, D. D., Crill, P. M., Rayment et al., 1997. A comparison of six methods for measuring soil-surface carbon dioxide fluxes. *Journal of Geophysical Research*, 102(D24), 28771-28777.
- [11] Epron, D., Farque, L., Lucot, E., Badot, P.-M. 1999. Soil CO_2 efflux in a beech forest: dependence on soil temperature and soil water content. *Ann. For. Sci.*, 56, 221-226.
- [12] Chojnicki, B., Michalak, M., Acosta, M., Juszczyk, R., Jürgen, A., Drösler et al. 2010. Measurements of carbon dioxide fluxes by chamber method at the Rzecin Wetland ecosystem, Poland. *Polish Journal of Environmental Studies*, 19, 283-291.
- [13] Zimnoch M., Wach P., Chmura L., Gorczyca Z., Rozanski K., Godłowska J., Mazur J., Kozak K. and Jericevic A., 2014, Factors controlling temporal variability of near-ground atmospheric ^{222}Rn concentration over central Europe., *Atmos.Chem. Phys.* 13, pp.9567-9581.
- [14] Netzel, P., Stano, S. and Zarebski, M. 1995. Vertical doppler sodar VDS. *Wiad IMGW* 18(39), 3-4, 119-125.
- [15] Zimnoch M., Godłowska J., Necki J.M. and Rozanski K., 2010, Assessing surface fluxes of CO_2 and CH_4 in Urban environment: a reconnaissance study In Krakow, Southern Poland. *Tellus62B*, pp573-580.

Inventories of domestic heating sources and their emissions in urban areas – methods, results and uncertainty

Marek Rosicki¹, Magdalena Załupka²

ATMOTERM S.A.
Opole, Poland
¹rosicki@atmoterm.pl
²zalupka@atmoterm.pl

Abstract

Results of recent projects on domestic heating emission inventories in Poland are presented in the paper. Wider context information concerning the impact of domestic heating emissions on air quality in urban areas is provided based on the authors' experiences at the field of Air Quality Action Plans (AQAP). Past and current methods used for emission inventories are discussed including GIS analysis, questionnaires, interviews and calculations. The influence of Low Carbon Economy Plans (LCEP) on the domestic heating emission inventories is presented as well. Case studies for Kraków (AQAP) and Gdańsk (AQAP/LCEP) cities are described together with general conclusions on the role of precise inventories in the urban environmental and development programmes.

Keywords: air quality, emission inventory, domestic heating sources

1. Introduction

Past and recent investigations have demonstrated that urban air quality in many Polish cities depends on the emissions from domestic heating systems. PM10 and B(a)P parameters measured in the heating season are elevated comparing to the summer period (see Figure 1).

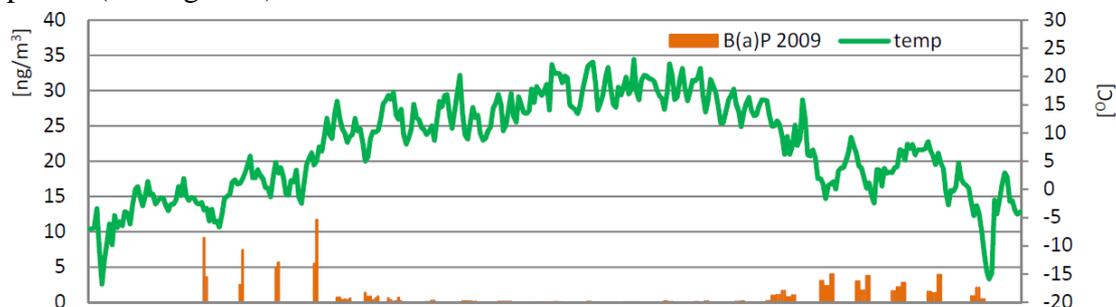


Figure 1. 2009 time series of B(a)P concentration in Poznan (Załupka et al., 2012).

During the JRC project (Junninen et al., 2009) the high contribution of coal fired domestic sources in air pollution in Krakow was proved by the chemical mass balance method.

Air quality modelling results (Lochno et al., 2013) also show that domestic heating sources are one of the main group of emission sources that influence the air quality in Krakow and other cities. An example of source apportionment modelling results is presented in Table 1.

Solid fuel fired domestic heating sources (SFFDH sources) has been indicated as a high priority issue in majority of Air Quality Action Plans (AQAP) prepared in Poland since 2005. Local and regional decision makers are facing the problem of SFFDH sources identification to set up the appropriate strategy of emissions reduction. In many cases heat energy balance calculation method is applied to provide information for

decision makers on the overall number of SFFDH sources. However some of city administrations decided to start deep investigation and launched bottom-up inventory projects.

Table 1. Contribution of different emission sources to PM10 and B(a)P annual concentration in Kraków for the areas where EU limit/target values are exceeded (Lochno et al., 2013)

	<i>Local domestic heating</i>	<i>Local transport</i>	<i>Local industry</i>	<i>Background and other sources</i>
<i>PM10</i>	42,3%	17,0%	21,0%	19,7%
<i>B(a)P</i>	48,7%	9,4%	3,5%	38,4%

2. Kraków case study

In May 2014 Kraków City administration began a new air quality project titled ‘MONIT-AIR: An integrated spatial data monitoring system for better air quality in Kraków’. As a part of this project a precise bottom-up inventory of existing SFFDH sources was started. The main reason to undertake this task was necessity to understand where the remaining SFFDH sources are located, are they concentrated or dispersed and what is the real scale of the solid fuels problem within the city. The city of Kraków was divided into seven areas for which inventory sub-projects were set up (Figure 3).

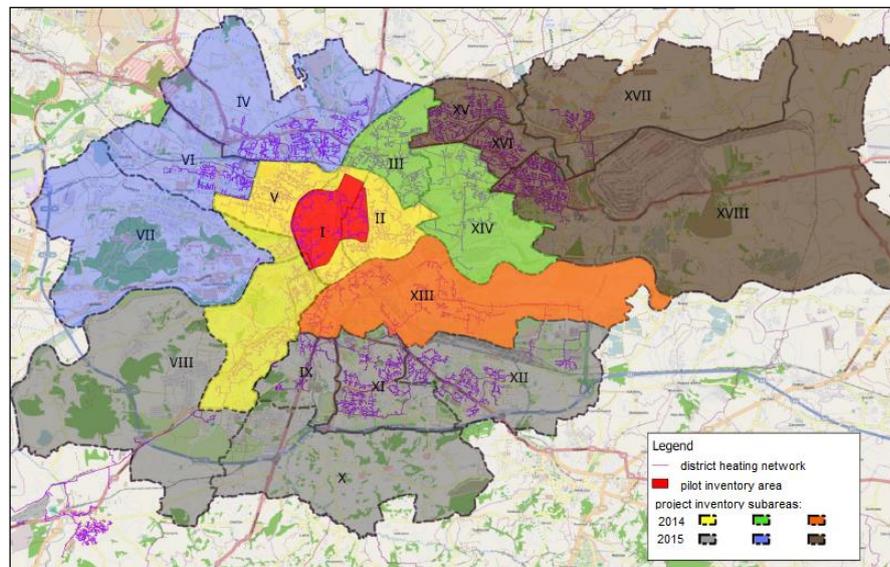


Figure 2. The areas of SFFDH inventory in Kraków.

After several months of field work the inventory is still in preparation. Before publishing the final results it is already possible to share some experiences on methodology and uncertainty issues.

The inventory sub-project starts from the analysis of existing data on non-solid-fuel based heating systems and networks within the selected area (district heating, gas and electricity networks). Basing on such analysis the set of buildings likely to contain SFFDH sources is being prepared. Then a questionnaire interview begins. Each

interviewer has assigned a subset of buildings where he asks questions on the type of heating, fuels used for heating and the number of heating sources. In the cases where the residents of building are absent or refuse to answer the questions there is a special procedure of filling in the questionnaire basing on interviewer observations. All information collected using the questionnaires has to be recorded in the database using an on-line IT application.

The interviewers are trained before starting the field work on every relevant aspects of their work. Although a special attention is paid during the trainings to the possible errors there are a lot of uncertainty sources that can be met in the inventory process. The most important are specified in Table 2.

Table 2. Factors influencing uncertainty of questionnaire-interview based inventory of SFFDH sources.

<i>Factor</i>	<i>Description</i>
Place of interview	The interviewers have no rights and obligation to go inside the building/apartment. There can be situations where the interview is carried out at the front of the building. The street noise and traffic can adversely affect the results of interview.
Attitude and knowledge of interviewees	Some interviewees can have negative attitude to the project and interview itself. In some cases the persons asked have not sufficient knowledge to answer the questions.
Skills of the interviewers	Although all the interviewers are trained before starting their project activities they cannot avoid some errors during the interviews and during entering data into the web IT application.
Time scale of the project	During the project run (18 months) changes in the heating systems can happen.

The final results of the inventory will be available through the urban GIS. This will provide decision makers and general public with precise information on the location of SFFDH sources and allow to develop detailed financial plan of solid fuel elimination. Further steps will include a transformation of the inventory into an IT system which will support city administration to monitor continuously domestic heating technology improvement and emissions changes. The main outcome of the inventory will be the numbers of SFFDH sources in the subareas. Using the specific emission factors it will be possible to convert the data into emission inventory and then use localised emission data for air quality modelling tasks. For PM10 emissions application of NFOŚiGW emission factor (380 g/GJ) is considered.

The SFFDH inventory project report in Kraków will include an uncertainty analysis of the received numbers of SFFDH sources. So far two uncertainty models have been considered: GUM based model and a counting model (W. Bich and F. Pennechi, 2012). For the GUM based model a symmetric rectangular distribution is assumed for every measurement process (each interviewer). The limit uncertainty Δ_{gi} is assessed individually for each interviewer 'i' and a standard uncertainty can be calculated as:

$$u_i = \frac{\Delta_{gi}}{\sqrt{3}}$$

Then a combined uncertainty can be determined using the GUM equation for uncorrelated input quantities. As a second method of uncertainty estimation a model proposed by Bich and Pennechi (2012) for counting processes has been considered. A preliminary uncertainty estimation gives results below 10% for the number of SFFDH sources in each investigated subarea. The counting model gives generally the lower uncertainty results than GUM based model.

3. Gdańsk case study

First time an inventory of domestic heating sources was made for the purposes of the Air Quality Action Plan (AQAP 2013). Calculated emissions for 2010 were prepared to use for air quality modelling. Calculations were based on heat demand analysis for the city buildings pertaining to the number of city residents and the contributions of different heating methods. This procedure requires to accept several assumptions. The first is heat demand per capita. Perfectly, such information comes from a heat supply plan. This plan for Gdańsk related then to the 2003 year. For that reason it was necessary to analyse changes from 2003 to 2010. It was done using statistical data, which are published by the Central Statistic Office of Poland every year. During the first inventory the following data presented by the Central Statistic Office were used:

- the number of residents in Gdańsk,
- the number and useable floor area of dwellings,
- the number of dwellings heated with natural gas,
- the sale of district network heat to residential buildings.

Because of the target – air quality modelling – an identification of emissions locations was very important. This is why a GIS analysis of heating and gas networks arrangement was necessary. After collating buildings layer and district heating network layer an emission-free areas were obtained. These are multi-family housing development built in the second half of the twentieth century. For the remaining areas emission density was differentiated depending on the type of buildings (scattered single-family or dense multi-family) and heating method. Emissions were calculated using EMEP/EEA factors proposed in the 2009 guidebook.

A next inventory of emissions from residential buildings in Gdańsk has been made for the purposes of Low Carbon Economy Plan (LCEP 2015, ongoing). In that case the carbon dioxide emissions were in the foreground. However at the same time emissions of air pollutants were also inventoried, i.e.: particulate matter (PM10 and PM2.5), benzo(a)pyrene, sulphur dioxide, nitrogen oxides. There are a few basic differences in LCEP inventory method comparing to the AQAP inventory:

- taking into account additional fuels - propane-butane and firewood,
- carrying out a field questionnaire-based survey on SFFDH sources in selected areas to verify the calculated number of heating sources fired by solid fuels.

Other elements which have an impact on the emissions in LCEP inventory include:

- a different base year – 2013,
- availability of an updated heat supply plan for 2012,
- an attempt of getting the actual data from district heat suppliers and fuel suppliers,

- a change of the EMEP/EEA emission factors in the new 2013 guidebook - they were reduced substantially, especially for gas and oil heating (see Figures 3 and 4).

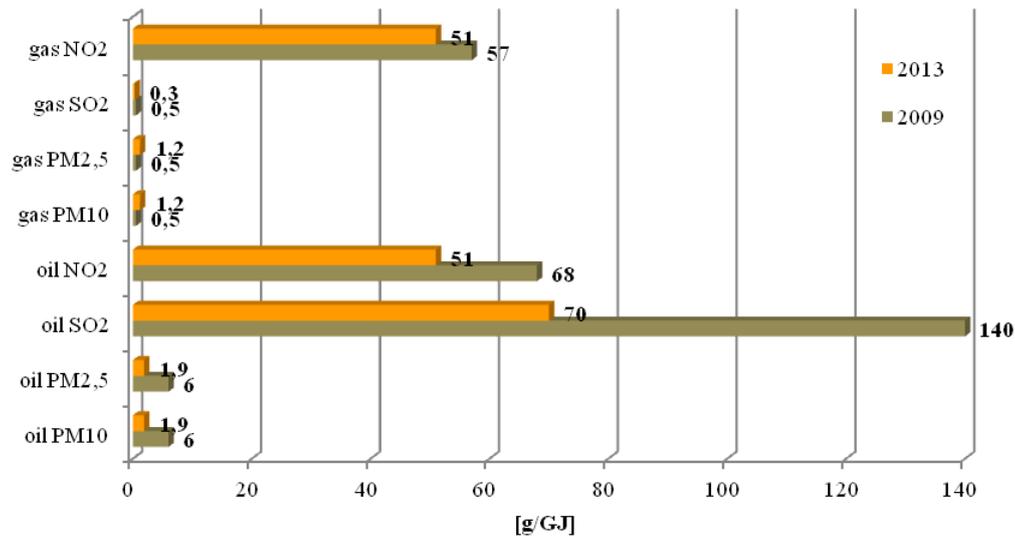


Figure 3. A comparison of EMEP/EEA emission factors for PM, SO₂ and NO₂.

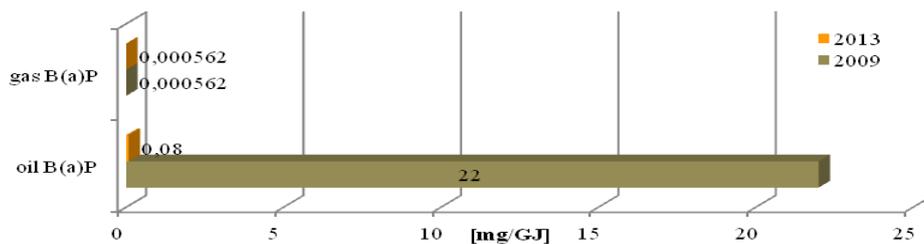


Figure 4. A comparison of EMEP/EEA emission factors for benzo(a)pyrene.

A comparison of the results for both inventories shows that the second one, more detailed, gave lower emissions. Experiences gained during both projects allow to indicate the following sources of uncertainty:

- Availability of data on fuel consumption is partly insufficient. Heat suppliers can provide data on heat consumption. Natural gas supplier can provide data on gas consumption. However the consumption of other fuels like coal, propane-butane, oil and firewood must be estimated.
- There is no accurate information on the age of the residential buildings and their thermal insulation which affects the estimation of heat demand.
- Applied EMEP/EEA emission factors may be not suited for the local conditions.
- In the course of the inventory projects human mistakes can happen (e.g. when data on the heat/gas consumption are collected).

4. Conclusions

Depending on the applied methodology approach and type of input data different results of emission inventory from domestic heating sources can be obtained. Factors

influencing uncertainty of such inventory include human mistakes, lack of raw data and emission factors errors. To achieve precise data for air quality modelling and management a critical data set concerns a number and location of solid-fuel-fired emissions sources. The project of bottom-up inventory of SFFDH sources in Kraków should demonstrate a lower uncertainty comparing with a traditional calculation based inventory. Further investigation and analysis on uncertainty of domestic heating sources inventories is necessary.

References

- [1] Załupka M., et al. (2012). Program ochrony powietrza w zakresie benzo-alfa-pirenu dla strefy aglomeracja poznańska (Air quality action plan in the range of benzo(a)pyrene for Poznan city), a study in Polish – Annex to the Resolution No. XXIX/566/12 of the Wielkopolska Regional Council.
- [2] Junninen H. et al. (2009). Quantifying the impact of residential heating on the urban air quality in a typical European coal combustion region. *Environmental Science and Technology* 43(20), pp. 7964-70.
- [3] Lochno A., et al. (2013). Program ochrony powietrza dla województwa małopolskiego (Air quality action plan for Małopolska region), a study in Polish – Annex No. 1 to the Resolution No. XLII/662/13 of the Małopolska Regional Council.
- [4] Evaluation of measurement data – Guide to the expression of uncertainty in measurement – JCGM 100:2008 (GUM 1995 with minor corrections). Bureau International des Poids et Mesures 2008.
- [5] W. Bich and F. Pennechi (2012). Uncertainty in measurements by counting. *Metrologia* 49 (2012), pp. 15-19.
- [6] Program ochrony powietrza dla strefy aglomeracji trójmiejskiej (Air quality action plan for Gdańsk-Gdynia-Sopot agglomeration), a study in Polish – Annex No. 1 to the Resolution No. 754/XXXV/13 of the Pomorskie Regional Council, 2013.
- [7] EMEP/EEA air pollutant emission inventory guidebook, 2009.
- [8] EMEP/EEA air pollutant emission inventory guidebook, 2013.

Accounting uncertainty for spatial modeling of greenhouse gas emissions in the residential sector: fuel combustion and heat production

Olha Danylo^{1,2}, Rostyslav Bun¹, Linda See², Petro Topylko¹, Xu Xiangyang³, Nadiia Charkovska¹, Przemysław Tymków⁴

¹Lviv Polytechnic National University, St.Bandery, 12, Lviv, 79013, Ukraine
olha.danylo@gmail.com;

²International Institute for Applied Systems Analysis, Laxenburg, A-2361, Austria;

³China University of Mining and Technology, Beijing, China;

⁴Wrocław University of Environmental and Life Sciences, Wrocław, Poland

Abstract

Energy consumption in households has a great potential for energy savings as well as for greenhouse gas emission reduction. As national inventory reports provide estimates at only a country or regional level, we have developed a new GIS approach that increases the resolution of emission inventories. We consider stationary emission sources, such as fossil fuel combustion and heat production for household energy needs that cover energy demand for cooking, water and space heating. We estimate the spatial emissions of greenhouse gases based on IPCC guidelines using official statistics on fuel consumption and spatial data about population density. The heating degree-day method was then used to determine the climatic conditions and spatial variability in energy demand. The results of the spatial inventory are obtained for settlements that are presented as area-type emission sources in a geospatial database. The uncertainties in the inventory results are estimated using a Monte Carlo method. The results show that uncertainties in greenhouse gas emissions at the regional level are significantly higher than at the country level although the uncertainty of emissions in CO₂-equivalent does not exceed 17.0%.

Keywords: Greenhouse gas emissions, spatial inventory, residential sector, heat production, uncertainty of inventory results.

1. Introduction

Climate change refers to a change in the long-term weather patterns that are specific to different regions of the world. Many researchers agree that the most likely reason for climate change is the increasing greenhouse gas emission concentration in the atmosphere [9]. Since the industrial revolution began, the emissions of greenhouse gases have increased by 40% compared with the pre-industrial period [16]. It is therefore important to draw attention to issues of climate change, directing our efforts towards reducing greenhouse gas emissions and mitigating the negative effects of processes that are already happening [9], [18]. However, these can only be achieved through joint efforts that involve as many countries as possible globally.

Greenhouse gas emissions can be reduced through more efficient usage of energy resources, transition to fuels with higher calorific values and lower emission coefficients or increasing the share of renewable energy. However, before investing in these changes, it is important to identify large emission sources, and therefore estimate the potential for emission reduction for those areas. Accounting uncertainty for spatial modeling of greenhouse gas emissions is an important part of such an analysis.

As the residential sector has a great potential for greenhouse gas (GHG) emission reduction compared to other sectors, especially in developing countries, we, therefore,

focus in this study on the analysis of emissions from direct fossil fuel use and heat production for household energy needs. We discuss mathematical models and algorithms that have been developed for spatial analysis of GHG emissions and present the method to estimate the influence of the uncertainty of individual inventory components on the total uncertainty of the inventory results. We apply the approach developed here using available statistical data for Poland and present the obtained results.

2. Spatial modeling of greenhouse gas emissions in the residential sector

Greenhouse gas emissions in the residential sector constitute a significant share in total emissions in most countries [20]. The emissions in this sector are caused by burning coal, natural and liquefied gas, wood and other fossil fuels in order to meet the energy needs of the households such as cooking, water and space heating. These needs can be covered from centralized energy supplies (e.g., district heating) or from “individual” energy sources (i.e. energy that is obtained through direct fossil fuel combustion).

The assessment of energy demand is not a new task. There are many global and regional models that analyze energy needs in the residential sector. Some models describe the household energy consumption on global scale for different regions of the world [4], [10], while other models are developed for a particular country [7], [13], administrative district or a city [11-12], [14-15], [19]. However, most of these approaches analyze energy demand and greenhouse gas emissions only on an aggregated level and do not perform an inventory for large areas at a high spatial resolution.

The aim of spatial modeling of GHGs is to obtain geographically distributed information on the amount of gases emitted from different sources [2], [3]. Therefore the inventory process consists of two main steps. First, we identify the emission sources for the sector of interest and then we allocate the amount of fossil fuels that were burned to cover certain energy needs in this sector. Based on this data we can estimate the GHG emissions for every emission source.

A spatial inventory of GHG emissions in the residential sector includes one additional step. As there are no statistical data on energy consumption and fossil fuel use at the level of settlements, statistical data must be disaggregated from the regional level to the level of elementary objects. Therefore, the development of the disaggregation algorithm is an important part of the inventory process.

2.1. Input data

Population density is the most essential input data for the estimation of the spatial distribution of GHG emissions in the residential sector as it is used in the creation of the geospatial database for the spatial GHG inventory. The territory of Poland was first clipped from the map of population density for the European Union and Croatia [8] and then the population estimation was adjusted using official statistical data. The objects of the population map are referred to as elementary, as they contain information about the size of the population at the lowest possible level.

As most of the energy in the residential sector in Poland is used for space heating, heating degree-days are, therefore, another important driving factor for the amount of energy that is used by households. Yearly heating degree-days were estimated as a sum of the daily deviation of the mean temperature from a heating base temperature but were

only calculated for those days when the mean temperature was lower than the base temperature [5].

The statistical data on the amount of fossil fuels used in the residential sector was taken from the Energy Statistics Yearbook published by the Central Statistical Office while other input information was taken from online database of official statistical data in Poland [1]. Net calorific values as well as emission coefficients were taken from [17]. In the case where country specific values are available, we use default coefficients from IPCC guidelines [6].

2.2 Estimation of greenhouse gas emissions

In many countries the statistical data about fossil fuels consumption are available only at a country or regional level. Therefore, to obtain the information about the amount of fossil fuel burned at the level of an elementary object, we need to disaggregate the fossil fuels. In the case of direct fossil fuel combustion, the amount of fuel is allocated proportionally to energy demand for cooking, water and space heating, taking into account access to energy sources, the percentage of living area equipped with central heating and the hot water supply.

The total amount of energy required to provide comfortable indoor temperature depends on the characteristics of the building (e.g. the number of floors, the year of construction), climatic features of the region (e.g. amount of heating degree-days), the intensity of building use (e.g. working-time or day of the week factor) and heat energy losses due to several factors such as entrance/exit and ventilation. Heating degree-days are determined based on the daily average temperatures taking into account the duration of the heating period. The total energy demand for space heating in the elementary object (i.e. city, town or village) is determined as the sum of the heat energy needed for space heating for all buildings in the settlement. It is estimated using average indicators of heat energy needed for heating one square meter of living space:

$$Q_h = k_{HDD} \cdot \sum_{year} Q_{h,year}^{sqm} \cdot LA_{year},$$

where $Q_{h,year}^{sqm}$ is the energy demand per square meter of living area constructed in *year* in the elementary object, k_{HDD} is the relative change in the amount of heating degree-days and LA_{year} is the living area per person (square meter) in buildings that are constructed in *year*.

District space heating is another source of energy for space heating. It is provided to households from many heat production locations, in most cases within urban areas. As it is complicated to identify the location of these point sources, we consider settlements as area sources of emissions. The statistical data on the amount of heat energy provided to households is available at the district level while the amount of fossil fuel that is used to generate this heat energy is accessible only at the regional level. Therefore, these data are disaggregated to the elementary object level proportional to the living area equipped with central heating.

As fossil fuels for centralized heat production and individual household energy needs were disaggregated to the level of elementary objects, we calculate the emissions of greenhouse gas G from burning fossil fuel i in elementary object n using the following formula:

$$E_{i,n}^G = M_{i,n} \cdot EF_{i,n}^G, n = \overline{1, N},$$

where $EF_{i,n}^G$ is the emission factor of G greenhouse gas and $M_{i,n}$ is the amount of fossil fuel.

At the level of elementary objects, the GHG emissions are determined as the sum of emissions from all sources within this geographical object.

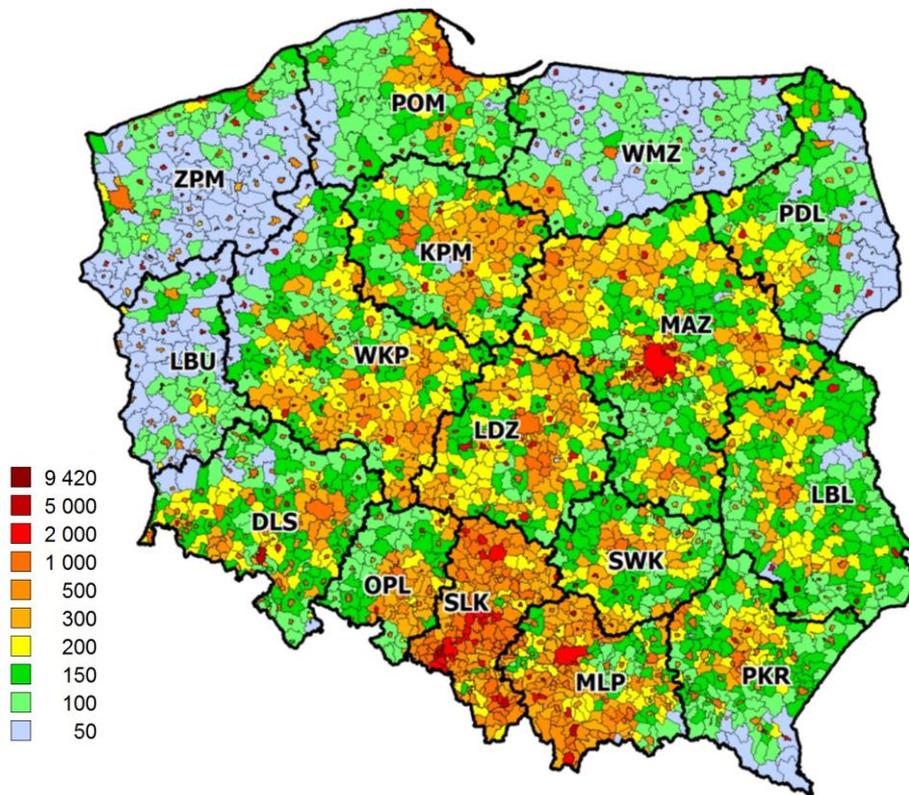


Figure 1. Specific greenhouse gas emissions from direct fossil fuel combustion in the residential sector in Poland at the level of municipalities (tons/km², CO₂-equivalent, 2010)

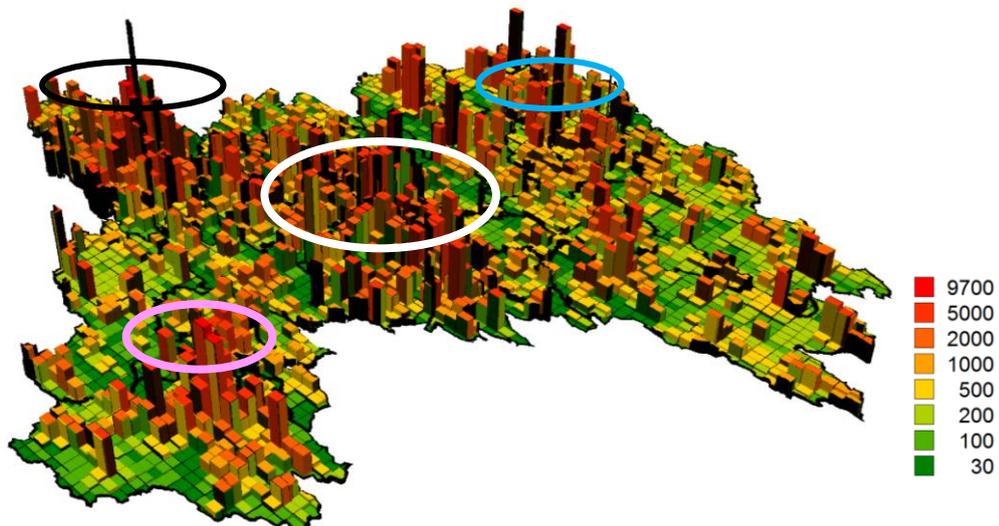


Figure 2. Prism-map of specific greenhouse gas emissions from direct fossil fuel combustion in the residential sector in the Silesia region at the level of elementary objects (tons/km², CO₂-equivalent, 2 x 2 km, 2010)

2.3. Results of the spatial inventory of greenhouse gas emissions

The largest emissions from direct fossil fuel combustion in the residential sector are in Masovian (MAZ - 14.4%), Silesian (SLK - 11.9%) and Greater Poland (WKP - 10.3%) regions as shown in Figure 1. Emissions in other regions do not differ significantly and range from 2.03% to 8.12%.

As the region of Silesia is the most urbanized area in Poland and it is a part of the Katowice agglomeration, accounting for more than 3 million people according to the Department of Economic and Social Affairs of the United Nations organization, the emissions in this region are, therefore, high when compared to other territories (Figure 1). Figure 2 presents specific emissions of greenhouse gases from direct fossil fuel combustion in the residential sector of Silesia. The highest specific emissions are in Katowice, Gliwice, Tychy (white circle), and Rybnik and Racibórz (black), Bielsko-Biala (pink) and Częstochowa (blue). The emissions from heat production are the highest in the Silesian region, but as emissions sources are very densely located (Figure 3), we sum up the emissions in this region together to compare with other regions that have much lower emission sources. GHG emissions from heat production in the Silesian region are much higher than in the Masovian region (2360 thousand tons in CO₂-eq., 16.87% of total emissions in this sector, Figure 3) and any other region although emissions in the Masovian region are also relatively high (1790 thousand tons, 12.8% of total emissions).

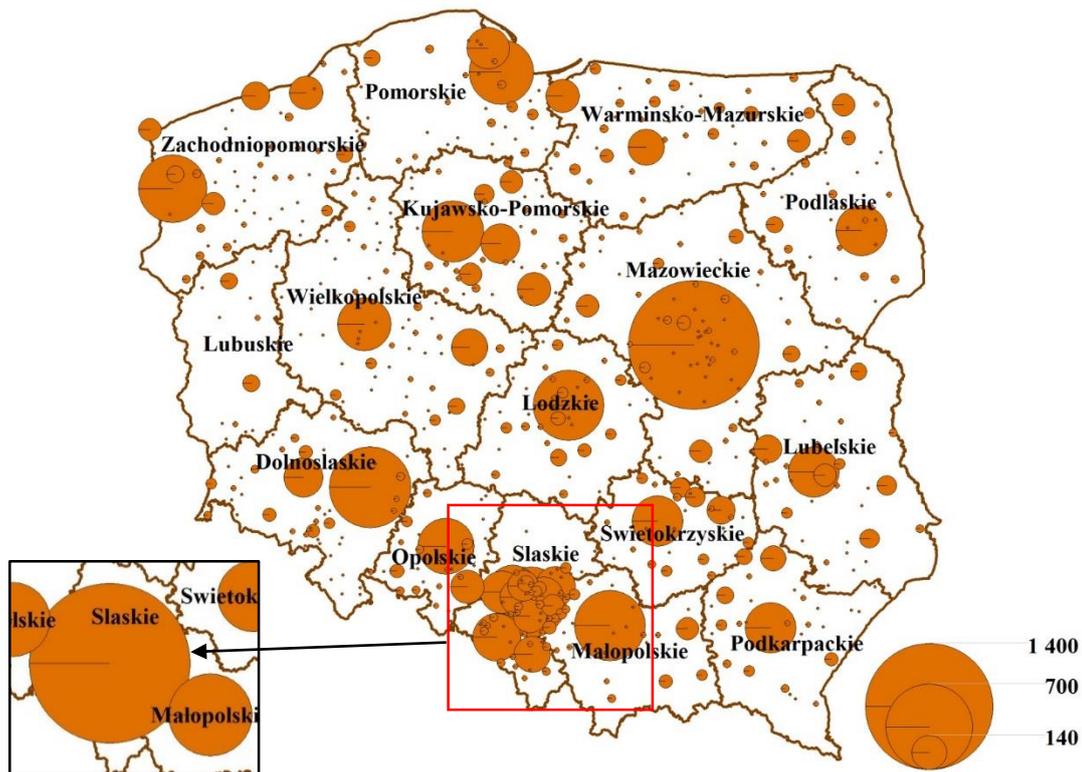


Figure 3. Greenhouse gas emissions from heat production in Poland (thousands tons, CO₂-equivalent, 2010)

3. Uncertainty analysis of inventory results

The obtained results of the spatial inventory are based on a set of input parameters that are not certain and are characterized by normal or log-normal distributions. The

uncertainty in the input values of the model can be combined using an error propagation approach, but only in the case when all values are normally distributed and uncertainties are small. In our case, some input values are log-normally distributed (e.g., CH₄ and N₂O emission coefficients). Therefore we used a Monte-Carlo method to sample random variables from these distributions using R. We considered a 95% confidence interval for estimating uncertainty in the emissions. We then performed an uncertainty analysis of the inventory results at the regional level, as most of the fossil fuels are disaggregated from this level downwards.

Table 1. Results of uncertainty estimation of greenhouse gas emissions from direct fossil fuel combustion in the residential sector in Poland (using data from 2010).

Voivodeship	CO₂, Gg (uncertainty, %)	CH₄, Gg (uncertainty, %)	N₂O, Gg (uncertainty, %)	Total emission Gg (uncertainty, %)
Lower Silesian	2635,8	5,4	0,03	2780,50
	(-12,9 : +14,9)	(-21,4 : +25,5)	(-19,7 : +23,2)	(-13,2 : +15,2)
Kuyavian-Pomeranian	1741,5	4,0	0,02	1848,54
	(-14,5 : +16,7)	(-21,5 : +25,5)	(-19,9 : +23,4)	(-14,7 : +16,9)
Lublin	1982,9	4,5	0,03	2103,56
	(-14,3 : +16,5)	(-21,5 : +25,6)	(-19,8 : +23,4)	(-14,5 : +16,8)
Lubusz	700,4	1,3	0,01	735,77
	(-11,8 : +13,6)	(-21,3 : +25,4)	(-19,3 : +22,7)	(-12,1 : +14,0)
Łódź	2451,2	5,8	0,03	2606,73
	(-15,0 : +17,3)	(-21,6 : +25,6)	(-20,0 : +23,6)	(-15,2 : +17,5)
Lesser Poland	3091,0	6,3	0,04	3258,20
	(-12,7 : +14,7)	(-21,4 : +25,5)	(-19,7 : +23,3)	(-13,0 : +15,0)
Mazovian	4966,4	9,2	0,05	5211,88
	(-11,6 : +13,4)	(-21,3 : +25,3)	(-19,3 : +22,7)	(-11,9 : +13,7)
Opole	893,8	2,1	0,01	948,65
	(-14,5 : +16,7)	(-21,5 : +25,6)	(-19,9 : +23,5)	(-14,7 : +17,0)
Subcarpathian	1889,5	3,9	0,02	1994,03
	(-13,0 : +15,0)	(-21,4 : +25,5)	(-19,8 : +23,4)	(-13,3 : +15,3)
Podlaskie	759,6	1,7	0,01	805,83
	(-14,3 : +16,5)	(-21,4 : +25,5)	(-19,8 : +23,2)	(-14,5 : +16,8)
Pomeranian	1478,1	2,8	0,02	1552,03
	(-11,7 : +13,5)	(-21,3 : +25,4)	(-19,2 : +22,7)	(-12,0 : +13,8)
Silesian	4591,5	10,1	0,06	4860,85
	(-13,8 : +15,9)	(-21,4 : +25,6)	(-19,9 : +23,4)	(-14,1 : +16,2)
Świętokrzyskie	1106,2	2,5	0,01	1174,20
	(-14,5 : +16,7)	(-21,5 : +25,6)	(-19,9 : +23,4)	(-14,7 : +17,0)
Warmian-Masurian	900,1	1,9	0,01	949,97
	(-13,0 : +15,0)	(-21,4 : +25,5)	(-19,5 : +23,0)	(-13,2 : +15,3)
Greater Poland	3013,4	5,9	0,04	3172,27
	(-12,4 : +14,3)	(-21,3 : +25,4)	(-19,5 : +22,9)	(-12,7 : +14,6)
West Pomeranian	1163,7	1,8	0,01	1210,98
	(-9,6 : +11,0)	(-21,0 : +25,1)	(-18,6 : +21,9)	(-9,9 : +11,3)

According to the IPCC classification, the residential sector is part of the 1A4b category in the energy sector. The uncertainty of CO₂ emissions in this sector at the country level is relatively low – around 3.0% (National inventory report of Poland, 2010), as the uncertainty of the activity data is 4% and the uncertainty of the emission

factor is 1% for liquid fuels and 2% for gaseous and solid fuels. Such low uncertainties in the input data can be explained by the high level of statistical data aggregation and overall accurate energy statistics at the country level.

However, in our spatial approach, we used data about fossil fuel consumption at the regional level, which is in units of weight. We first converted these data to energy units using national or default (IPCC guidelines) net calorific values, which we later disaggregated and then used for the emission estimations. By using regional statistics on fossil fuel combustion instead of country level data, we reduced the uncertainty due to disaggregation, but at the same time we introduced uncertainty in the net calorific values, which are rather high (e.g. 19.1% for coal). This leads to an increase in emission uncertainty at the regional level (Table 1) compared to the uncertainty in emissions at country level, as mentioned previously.

The statistical data on coal and liquid gas usage at the regional level were available in units of weight while the data on natural gas combustion are available in energy units. As a result, the uncertainty is slightly lower in those regions where natural gas constitutes a bigger share in the total energy consumption in the residential sector (e.g. Wielkopolskie voivodeship) compared to other regions.

4. Conclusions

Uncertainty analysis is an important part of the spatial inventory of greenhouse gas emissions as it gives an understanding of the ranges in which the emission estimates fall and which parameters of the model introduce the highest uncertainty into the model results. This information can be used to find possible ways to decrease uncertainties in the emission inventory when disaggregated to subnational level and to build more certain spatial cadasters of greenhouse gas emissions.

The spatial inventory of greenhouse gas emissions from direct fossil fuel combustion and heat production for residential sector energy needs was conducted using the official statistical data, national emission coefficients and net calorific values (or recommended by IPCC in the absence of national data). The obtained results show that uncertainties of greenhouse gas emissions at the regional level are significantly higher than at the country level; however, the uncertainty of emissions in CO₂-equivalent does not exceed 17.0%.

Acknowledgment

The study was partly conducted within the 7FP Marie Curie Actions IRSES project No. 247645 (GESAPU). Olha Danylo was a recipient of the Visegrad Fund Scholarship.

References

- [1] BDL: Bank Danych Lokalnych. [Online]. Available: <http://stat.gov.pl>. [Accessed: 21-Aug-2015].
- [2] Boychuk K. and Bun R., “Regional spatial inventories (cadastres) of GHG emissions in the Energy sector: Accounting for uncertainty,” *Clim. Change*, vol. 124, no. 3, pp. 561–574, Jun. 2014.
- [3] Bun R., Hamal K., Gusti M., and Bun A., “Spatial GHG inventory at the regional level: accounting for uncertainty,” in *Greenhouse Gas Inventories*, Jonas M., Nahorski Z., Nilsson S., and Whiter T., Eds. Springer Netherlands, 2011, pp. 227–244.

- [4] Daioglou V., van Ruijven B. J., and van Vuuren D. P., “Model projections for household energy use in developing countries,” *Energy*, vol. 37, no. 1, pp. 601–615, Jan. 2012.
- [5] Degree-Days - Theory and Application - TM41 2006. [S.l.]: CIBSE, 2006.
- [6] Eggleston S., Buendia L., Miwa K., Ngara T., and Tanabe K., “IPCC guidelines for national greenhouse gas inventories,” Inst. Glob. Environ. Strateg. Hayama Japan, 2006.
- [7] Ekholm T., Krey V., Pachauri S., and Riahi K., “Determinants of household energy consumption in India,” *Energy Policy*, vol. 38, no. 10, pp. 5696–5707, Oct. 2010.
- [8] Gallego F. J., “A population density grid of the European Union,” *Popul. Environ.*, vol. 31, no. 6, pp. 460–473, Jul. 2010.
- [9] Hannah L., “Chapter 2 - The Climate System and Climate Change,” in *Climate Change Biology (Second Edition)*, Hannah L., Ed. Boston: Academic Press, 2015, pp. 13–53.
- [10] Isaac M. and van Vuuren D. P., “Modeling global residential sector energy demand for heating and air conditioning in the context of climate change,” *Energy Policy*, vol. 37, no. 2, pp. 507–521, Feb. 2009.
- [11] Kannan R. and Strachan N., “Modelling the UK residential energy sector under long-term decarbonisation scenarios: Comparison between energy systems and sectoral modelling approaches,” *Appl. Energy*, vol. 86, no. 4, pp. 416–428, Apr. 2009.
- [12] Kavacic M., Summerfield A., Mumovic D., Stevanovic Z. M., Turanjanin V., and Stevanovic Z. Z., “Characteristics of indoor temperatures over winter for Belgrade urban dwellings: Indications of thermal comfort and space heating energy demand,” *Energy Build.*, vol. 47, pp. 506–514, Apr. 2012.
- [13] Kenisarin M. and Kenisarina K., “Energy saving potential in the residential sector of Uzbekistan,” *Energy*, vol. 32, no. 8, pp. 1319–1325, Aug. 2007.
- [14] Kialashaki A. and Reisel J. R., “Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks,” *Appl. Energy*, vol. 108, pp. 271–280, Aug. 2013.
- [15] Li C., He L., Cao Y., Xiao G., Zhang W., Liu X., Yu Z., Tan Y., and Zhou J., “Carbon emission reduction potential of rural energy in China,” *Renew. Sustain. Energy Rev.*, vol. 29, pp. 254–262, Jan. 2014.
- [16] Marland G., Boden T. A., Andres R. J., Brenkert A. L., and Johnston C. A., “Global, regional, and national fossil fuel CO₂ emissions,” *Trends Compend. Data Glob. Change*, pp. 34–43, 2003.
- [17] National Inventory Submissions 2014. [Online]. Available: http://unfccc.int/national_reports/annex_i_ghg_inventories/national_inventories_submissions/items/8108.php. [Accessed: 31-Jul-2015].
- [18] Sachs J. D., *The Age of Sustainable Development*. New York: Columbia University Press, 2015.
- [19] Sarak H., “The degree-day method to estimate the residential heating natural gas consumption in Turkey: a case study,” *Energy*, vol. 28, no. 9, pp. 929–939, Jul. 2003.
- [20] World Bank Group. [Online]. Available: <http://www.worldbank.org/>. [Accessed: 09-Jun-2015].

On the uncertainty in modeling urban air quality under imprecise emission data

Piotr Holnicki¹ and Zbigniew Nahorski¹

¹Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland
holnicki@ibspan.waw.pl, nahorski@ibspan.waw.pl

Abstract

Air pollution dispersion models have recently been used for supporting decisions concerning air quality management and emission control. Emission inventory is the basic input dataset in air quality evaluation. To select the best strategy of emission reduction and to assess the possible environmental effects, there is a need to estimate the contribution of the respective emission sources to the resulting air pollution. This paper addresses the problem of uncertainty of air pollution models, related imprecision and uncertainty of the emission data. The problem is discussed in a case study for Warsaw agglomeration, where the main urban scale air pollutants: SO₂, NO_x, PPM₁₀, PPM_{2.5} are considered. CALMET/CALPUFF modeling system is used as the main forecasting tool which links the emission data with the resulting concentrations. Uncertainty analysis, based on a Monte Carlo algorithm, shows the main factors which decide on the resulting uncertainty of the model forecast.

Keywords: air pollution, emission inventory, computer model, uncertainty analysis

1. Air quality models in decision support systems

A direct application of air quality models is in forecasting dispersion of pollutants, analysis of ecological results of some specific meteorological episodes, or evaluation of the final environmental impact of emission sources. Recently developed Integrated Assessment Models (IAM) [2, 3, 9, 16] are used for supporting decisions concerning air quality management and emission control policy. The air quality model is a key module of such a system which enables to assess environmental, economic or health benefits of emission abatement, and to select the best strategy of emission reduction. In such applications, there is a need to estimate the contribution of emission sources to ambient concentrations with required accuracy. However, due to a very complex structure of such systems, there exist many sources of environmental effects of atmospheric pollution as well as in the resulting regulatory decisions.

The quantitative assessment of uncertainty brings the modeling prediction closer to reality. It increases decision maker's confidence in the modeling results and improves the quality of the final decisions. To assess the accuracy of modelling results and a connected decision support process, inaccuracy and uncertainty of the model should be evaluated. The main sources of results' variability (temporal or spatial) and uncertainty (imprecise information or lack of information about unknown quantity) should be identified and assessed [11, 15, 17].

It is a common view in the literature that emission field inventory is one of the main sources of uncertainty in modeling of air pollution dispersion. The problem is particularly significant in urban agglomerations [1, 4, 10, 12, 14]. Emission field in such cases comprises a variety of sources, point-, area- and line-, with different technological parameters, emission intensities, composition of polluting species, and also – with different uncertainty which is introduced to the system. This uncertainty

must be taken into account in complex analysis, when the results are to be used in supporting regulatory decisions.

2. Urban scale uncertainty analysis

The computations performed in the framework of the study relate to the forecasts and analysis of air pollution dispersion in Warsaw agglomeration. The aim was to evaluate the environmental impact of the main categories of emission sources as well as to estimate the uncertainty of this forecast, which is related to the uncertainty of emission field inventory. The analysis covers a rectangular domain, approx. 30 x 40 km² of Warsaw Metropolitan Area shown in Fig. 1. The regional scale, Gaussian puff dispersion model CALPUFF [18, 19, 20] was used to simulate the air pollution transport and transformations within the domain.

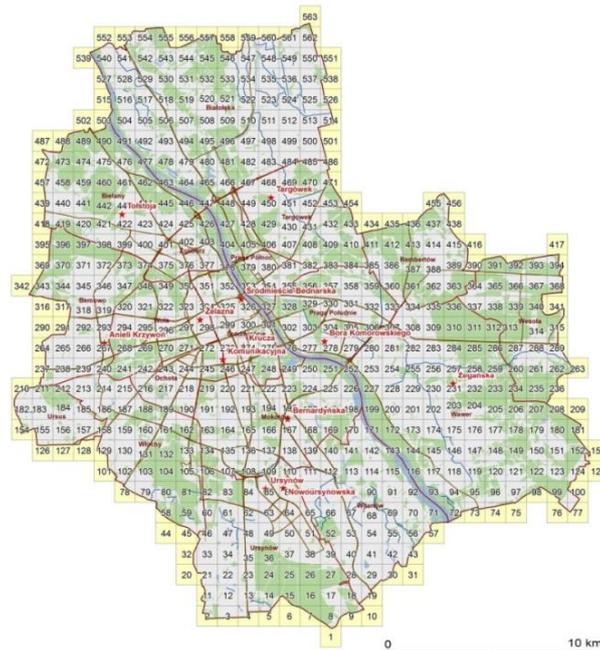


Figure 1. The study domain and the receptor points ([8], due to CCA License)

In case of the discussed Warsaw study the total emission field was decomposed into four basic categories, mainly according to the emission parameters and the intrinsic uncertainty [7]. According to the previous remarks, assumed emission field was categorized into following four classes:

- 16 high point sources (power/heating plants – low uncertainty),
- 1002 low point sources (industry – medium uncertainty),
- 872 area sources (residential sector – high uncertainty),
- 1157 linear sources (transportation – high uncertainty).

For computational purposes, the domain is discretized with a regular square grid with the step size $h = 1$ km. The point sources are located according to the geographical coordinates, the area and linear sources are represented as one grid element, 1 km² (compare Fig. 1). Computations take into account temporal variability of the meteorological and emission input data with the assumed step-size of time resolution,

$\tau = 1$ h. The annual mean concentrations of the main polluting species are recorded at 563 fictitious receptor points, located in the center of grid elements shown in Fig. 1. The list of the main primary and secondary pollutants considered in this study encompasses sulfur oxides (SO₂), nitrogen oxides (NO_x), sulfate and nitrate aerosols, particulate matter (PM₁₀ and PM_{2.5}) and Pb.

3. Uncertainty analysis

The uncertainty of the modeling results has been assessed using a Monte Carlo algorithm [6, 13] and the input uncertainties of the emission data. Applied to all the sources and pollutants, 2000 random sets of emission data were preprocessed within the assumed ranges of uncertainty. Each random set of the emission data encompasses a one-year time interval. As stated in [8], to avoid generating unrealistic emission episodes, a correlation between emission intensities of key individual pollutants from each emission source was established.

Table 1. The input uncertainty range depending on emission category (95 CI) ([8], due to CCA License).

Pollutant	Emission sources			
	High point	Low point	Area	Linear
SO ₂	± 15%	± 20%	± 30%	± 30%
NO _x	± 20%	± 30%	± 40%	± 40%
PPM ₁₀	± 25%	± 40%	± 40%	± 40%
PPM _{2.5}	± 25%	± 40%	± 40%	± 40%
PPM _{10_R}	–	–	–	± 40%
PPM _{2.5_R}	–	–	–	± 40%
Pb	± 30%	± 40%	± 50%	± 50%

Table 1 presents these ranges, assumed for 4 categories of emission sources, at the 95% confidence interval. The applied input uncertainties have been mainly based on expert opinions as presented in [8]. The normal distribution of the input emission data was assumed. The relative uncertainty range of the resulting pollution concentrations at a receptor point is calculated as a ratio: $(c_{97.5} - c_{2.5})/c_M$, where $c_{2.5}$ is the 2.5 and $c_{97.5}$ is the 97.5 percentile concentration value, and c_M is the mean value.

In the previous papers [7, 8] violation of air quality limits (EU 2008) by NO_x, PM₁₀, PM_{2.5} concentrations, mainly in central districts, was indicated. The accuracy of the forecasts was confirmed by comparison of the results with measurements (FA2 index). Below the main factors which decide on the final uncertainty are discussed.

The first factor which determines the resulting uncertainty is the category of emission sources with the dominating share in polluting the receptor (see Table 1). For example, it is known that NO_x and Pb are typical traffic-related compounds for which the concentration maps are correlated with the topology of the arterial streets. The correlation is also seen on uncertainty maps. The mobile sources also contribute to PM₁₀ pollution, but in the case of PM_{2.5} and SO₂ there is a considerable share of the area

sources (residential sector) and also of some point sources. The above correlations are reflected in the uncertainty maps.

An important factor influencing uncertainty is the relative share of emission categories which contribute to the selected receptor point. Exemplary maps shown below compare distributions of a typical traffic-related NO_x pollution and the pollution of PM_{2.5}, which strongly depends on other emission categories, e.g. the area emission from the residential sector. Figure 2 contains two related pairs of maps, pollutant concentration and uncertainty, for NO_x (top) and PM_{2.5} (bottom), respectively.

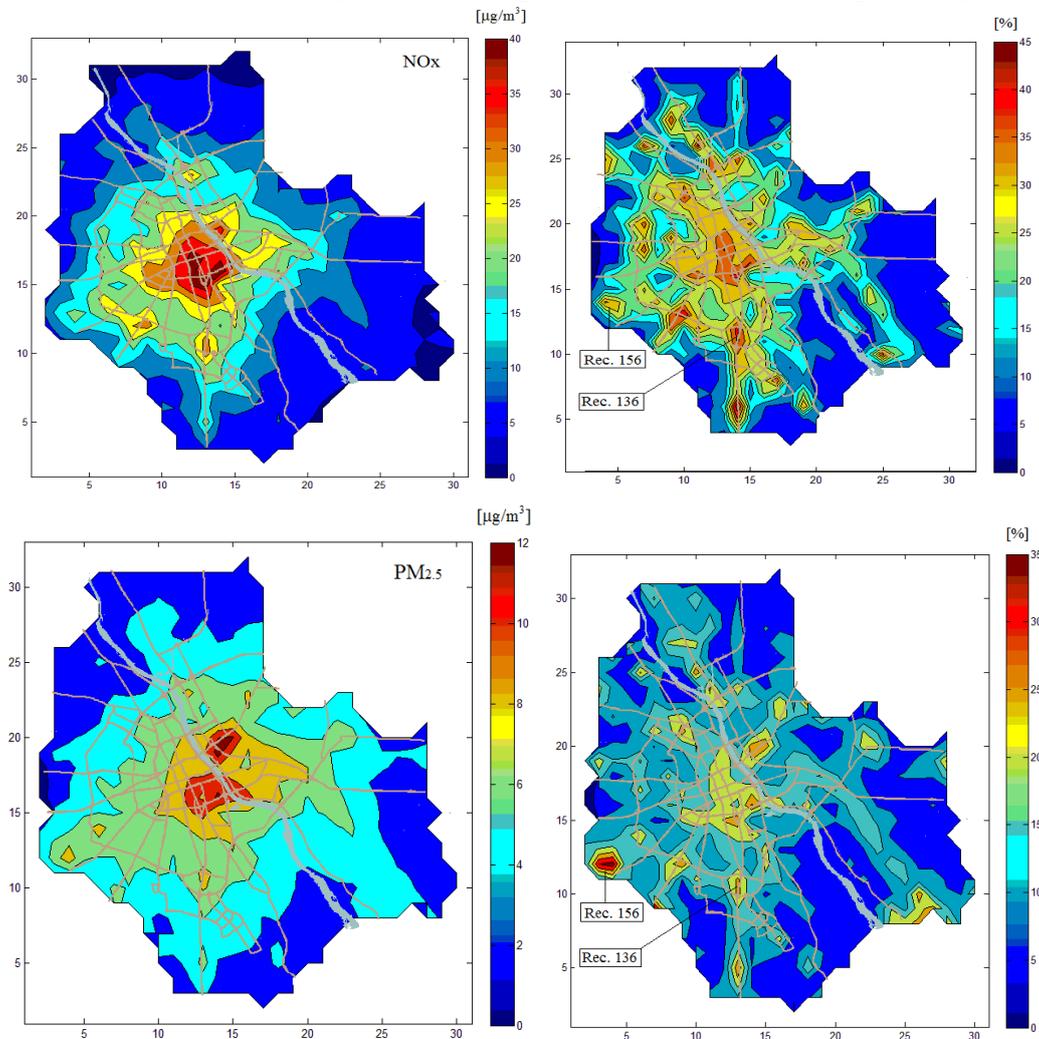


Figure 2. The spatial maps of concentration (left) and uncertainty (right): NO_x – top and PM_{2.5} – bottom ([8], due to CCA License)

It can be observed that the maximum concentrations (left maps) occur in the central districts of the city, but the spatial variability of uncertainty (right map) is much more evident and is not strictly correlated with the concentration map. For NO_x the maximum uncertainties are obtained near the main crossroads (similar properties represent the other traffic dependent pollutants, such as PM₁₀ or Pb), while the global uncertainty maximum for PM_{2.5} represents the residential sector (individual housing area).

In these cases, high uncertainties correlate to some extent with the concentration values, but in fact they depend also on the location of the receptors. The location determines the relative share of the contributing emission categories and the quantity

of the individual emission sources which affect a given receptor point. A specific coincidence of these factors leads to extreme values of the overall uncertainty at some locations. The quantity of individual sources that substantially contribute to a spot strongly influence the uncertainty. Due to the averaging effect, the rising number of such emission sources leads to the lower aggregate level of the relative uncertainty, while a low number of the unbalanced sources mean high uncertainty. This fact is illustrated by the data for pollutions NO_x and $\text{PM}_{2.5}$ recorded at receptors #136 (crossroad) and #156 (housing), presented in Table 2 and in Table 3, respectively.

Table 2. Uncertainty of NO_x concentration depending on the receptor location

Sources	Receptor #136 uncertainty range 45%			Receptor #156 uncertainty range 28%		
	Concentration [$\mu\text{g}/\text{m}^3$]	Share [%]	Dominating sources	Concentration [$\mu\text{g}/\text{m}^3$]	Share [%]	Dominating sources
LINEAR	29,8	94,3	4	15,94	80,1	9
AREA	1,0	3,1		2,86	14,4	2
LOW	0,7	2,2		0,4	3,5	
HIGH	0,1	0,3		0,7	2	
Total	31,6		4	19,9		11

Table 3. Uncertainty of $\text{PM}_{2.5}$ concentration depending on the receptor location

Sources	Receptor #136 – uncertainty range 23%			Receptor #156 – uncertainty range 33%		
	Concentration [$\mu\text{g}/\text{m}^3$]	Share [%]	Dominating sources	Concentration [$\mu\text{g}/\text{m}^3$]	Share [%]	Dominating sources
LINE	6,4	74,5	5	3,1	33,7	1
AREA	1,9	22	1	5,8	60	3
LOW	0,2	2,0		0,2	2,2	
HIGH	0,08	0,9		0,1	1,1	
Total	8,58		6	5,97		4

So, the fewer sources contribute to the pollution level, the higher level of the relative uncertainty may be expected. At the same time, an unbalanced contribution of the individual generally increases the aggregate level uncertainty for the forecasted pollution. This general conclusion is illustrated in the two selected receptor points, namely receptor #136 (crossroad) where the high level of uncertainty occurs for traffic-related pollutants (NO_x), and receptor #156 (residential area) in a peripheral district where fine particulates $\text{PM}_{2.5}$ predominate in the emission field and induce a high level of uncertainty. On the other hand, in such cases, the impact of the input emission uncertainty assumed in Table 3 becomes less important.

4. Summary

The paper addresses the problem of uncertainty of urban scale air pollution models under uncertainty of emission data. The case study discussed deals with Warsaw agglomeration where Monte Carlo algorithm is used. and sources contribute to the pollution level, the higher level of the relative uncertainty may be expected. Exemplary results illustrate the spatial distribution of uncertainty in the domain. The main factors are indicated which decide on resulting uncertainty of the forecast. It relates to the receptor's location, but also depends on the share of emission classes that affect the receptor site and on the number of the individual emission sources contributing to the overall concentration.

References

- [1] Buchholz, S., Krein, A., Junk, J., Heinemann, G., Hoffmann, L. (2013). Simulation of urban-scale air pollution patterns in Luxembourg: Contributing sources and emission scenarios. *Environmental Modeling and Assessment*, 18, 271– 283.
- [2] Carnevale, C., Finzi, G., Pisoni, E., Volta, M., Guariso, G., Gianfreda, R., Maffei, G., Thunis, P., White, L., Triacchini, G. (2012). An integrated assessment tool to define effective air quality policies at regional scale. *Environmental Modelling & Software*, 38, 306 – 315.
- [3] Calori, G., Clemente, M., De Maria, R., Finardi, S., Lollobrigida, F., Tinarelli, G. (2006). Air quality integrated modelling in Turin urban area. *Environmental Modelling & Software*, 21, 468 – 476.
- [4] Colville, R.N., Woodfield, N.K., Carruthers, D.J., Fisher, B.E.A., Rickard, A., Neville, S., Hughes, A. (2002). Uncertainty in dispersion modeling and urban air quality mapping. *Environmental Science & Policy*, 5, 207 – 220.
- [5] EU (2008) Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe .
- [6] Holnicki P., Nahorski, Z., Tainio, M. (2010). Uncertainty in air quality forecasts caused by emission uncertainty. In: *HARMO 13th Conference on Harmonisation within Atmospheric Dispersion Modelling*, 119 – 123.
- [7] Holnicki, P., Nahorski Z. (2013). Air quality modeling in Warsaw Metropolitan Area. *Journal of Theoretical and Applied Computer Science*, 7, 56 – 69.
- [8] Holnicki, P., Nahorski Z. (2015). Emission Data Uncertainty in Urban Air Quality Modeling – Case Study. *Environmental Modeling & Assessment*. DOI: 10.1007/s10666-015-9445-7.
- [9] Lim, L.L., Hughes, S.J., Hellowell, E.E. (2005). Integrated decision support system for urban air quality assessment. *Environmental Modelling & Software*, 20, 947– 954.
- [10] Malherbe, L., Ung, A., Colette, A., Derby, E. (2011). Formulation and quantification of uncertainties in air quality mapping. *ETC/ACM Technical Paper 2011/9*.
- [11] Maxim, L., van der Sluijs J. (2011). Quality in environmental science for policy: Assessing uncertainty as a component of policy analysis. *Environmental Science & Policy* 14, 482– 492.

- [12] Mediavilla-Sahagún, A., ApSimon, H.M. (2006). Urban scale integrated assessment for London: Which emission reduction strategies are more effective in attaining prescribed PM10 air quality standards by 2005? *Environmental Modelling & Software*, 21, 501–513.
- [13] Moore G.E., Londergan R.J. (2001) Sampled Monte Carlo uncertainty analysis for photochemical grid models. *Atmospheric Environment*, 35, 4863–4876.
- [14] Oxley, T., Valiantis, M., Elshkaki, A., ApSimon, H.M. (2009). Background, road and urban transport modeling of air quality limit values (The BRUTAL model). *Environmental Modelling & Software*, 24, 1036–1050.
- [15] Park, S.-K., Cobb, C.E., Wade, K., Mulholland, J., Hu Y., Russel, A.G. (2006). Uncertainty in air quality model evaluation for particulate matter due to spatial variations in pollutant concentrations. *Atmospheric Environment*, 40, S563 – S573.
- [16] Pisoni, E., Carnevale, C., Volta, M. (2010). Sensitivity to spatial resolution of modeling systems designing air quality control policies. *Environmental Modelling & Software*, 25, 66–73.
- [17] Sax, T., Isakov, V., 2003. A case study for assessing uncertainty in local-scale regulatory air quality modeling applications. *Atmospheric Environment*, 37, 3481 – 3489.
- [18] Scire, J.S., Strimaitis, D.G., Yamartino, R.J. (2000). A User's Guide for the CALPUFF Dispersion Model. Earth Technology Inc.
- [19] Tartakovsky D., Broday D. M., Stern E. (2013) Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. *Environmental Pollution*, 179, 138-145.
- [20] Trapp, W. (2010). The application of CALMET/CALPUFF models in air quality assessment system in Poland. *Archives of Environmental Protection*, 36, 63 – 79.

On the possibility of selenium air emission inventory from small domestic sources

Damian Zasina¹

Jarosław Zawadzki²

¹ Institute of Environmental Protection - NRI,
Warsaw, Poland
damian.zasina@kobize.pl

² Warsaw University of Technology,
Warsaw, Poland
j.j.zawadzki@gmail.com

Abstract

The main purpose of this paper will be examining the possibility of effective selenium (Se) air emission inventory from small domestic combustion sources. Authors would like to raise issues connected with data collection, its consistency and also stress the context of uncertainty analysis of air emission from small geographically scattered sources.

Although emission inventory of selenium is covered by UN ECE LRTAP Convention, the selenium emission inventory has not been performed so far nor in Poland, neither in its part. Data derived from working paper [1] suggested underestimation of selenium air emission from domestic stoves. Moreover official emission inventory guidelines [2] reported significant difference between Tier 1 Se emission factors for energy industries (NFR: 1A1a) and residential combustion (1A4bi). This data hinder emission inventory and may significantly affect results of uncertainty analysis.

Keywords: selenium, air emission, uncertainty analysis

1. Introduction

This paper presents using scarce, incomplete and uncertain data in air emission inventory. Due to many lacks in estimation of selenium air emission such as: using data derived from simplified official guidelines (Tier 1) [2], lacks in measurements, no data on economical or any other national dependencies, lacks of data on uncertainties the official data for international submission is still under preparation and analysis [1].

The main task of presented analysis is to merge pieces of information from various sources and determine range of uncertainty of Se air emission from domestic (small) combustion sources (included in NFR 1A4bi category).

As initial analysis authors would like to present quantitative assessment of uncertainty according to previous works [3,4] in particular using *mixture model with two components* [3], expressed as:

$$f(x) = wf_1(x) + (1 - w)f_2(x), \quad (1)$$

where:

$f(x)$, probability density function (PDF) of emitted air pollutant;

$0 \leq w \leq 1$, weight of component PDF;

$f_1(x), f_2(x)$, PDFs of components.

Methodology presented in this paper considers results of selenium content analysis in Polish coals, due to their importance in national economy. Also introduced selected methods of data integration due to small amount of direct measurements available. Due to its volatility in combustion process Se content in fuel could be treated as an emission factor. Moreover, the lack of measurements of selenium content in coal makes impossible straightforward analysis of selenium air emission.

According to formula (1), functions $f1(x)$ and $f2(x)$ could be treated as PDFs of emission source activity and also emission factor respectively. For determining of total uncertainty of emission, assumptions of partial uncertainties are needed.

Case of incomplete data forced use of bootstrapping or maximum likelihood estimation. Basing on assumptions from [3], considered distributions were: lognormal, Weibull and Gamma. Apart from adjusting distributions technical aspect of independent distributions combination was considered.

In second step, authors would like to analyze spatial aspects of estimated emission considering disaggregation of *top-down* emission estimation by surrogates.

References

- [1] B. Dębski, 2013. First step Se air emission estimation from Poland. IEP-NRI, NECD (in Polish; unpublished).
- [2] EEA, 2013. EMEP/EEA Air Pollutant Emission Inventory Guidebook, ISBN 978-92-9213-403-7, <http://dx.doi.org/10.2800/92722>
- [3] J. Zheng, H.C. Frey, 2004, Quantification of Variability and Uncertainty Using Mixture Distributions: Evaluation of Sample Size, Mixing Weights and Separation Between Components, Risk Analysis, Vol. 24, No. 3, 553-571, <http://dx.doi.org/10.1111/j.0272-4332.2004.00459.x>
- [4] Y. Zhao, H.C. Frey, 2006, Uncertainty for Data with Non-Detects: Air Toxic Emissions from Combustion, Human and Ecological Risk Assessment, Vol. 12, No. 6, 1171-1191, <http://dx.doi.org/10.1080/10807030600977178>

Author Index

Amann, Markus	111	Marland, Eric	117
Andres, Robert J.	117	Marland, Gregg	117
Bartyzel, Jakub	177	Nahorski, Zbigniew	25, 41, 97, 137, 159, 201
Bun, Rostyslav	16, 25, 41, 49, 57, 64, 124, 193	Necki, Jarosław	177
Charkovska, Nadiia	16, 25, 41, 57, 193	Nocera, Silvio	103
Chmura, Lukasz	177	Oda, Tomohiro	49, 124
Danylo, Olha	16, 25, 41, 49, 124, 193	Ott, Lesley	124
Forsell, Nicklas	152	Paasonen, Pauli	111
Fritz, Steffen	1, 9	Pałka, Piotr	159
Galkowski, Michal	177	Paunu, Ville-Veikko	110
Gao, Junlian	71	Radziszewska, Weronika	159
Gon, Hugo Denier van der	111	Robison, Charles	117
Gorczyca, Zbigniew	177	Rosicki, Marek	187
Gusti, Mykola	152	Rovenskaya, Elena	78, 90
Halushchak, Mariia	41, 49, 57, 64, 124	Rozanski, Kazimierz	177
Heldstab, Jürg	32	Schäppi, Bettina	32
Hogue, Susannah	117	Schepaschenko, Dmitry	1, 9
Holnicki, Piotr	201	See, Linda	9, 193
Horabik-Pyzel, Joanna	16, 25, 41, 124, 137	Shvidenko, Anatoly	1, 9
Jasek, Alina	177	Stańczak, Jarosław	159
Jarnicka, Jolanta	97	Striamets, Oleksandr	41
Jonas, Matthias	16, 64, 57, 78, 90	Tonin, Stefania	103
Khabarov, Nikolay	152	Topylko, Petro	41, 49, 64, 124, 193
Klimont, Zbigniew	111	Tymków, Przemysław	193
Kraxner, Florian	1	Verstraete, Jörg	130
Kuc, Tadeusz	177	Visschedijk, Antoon	111
Kupiainen, Kaarle	110, 111	Woodard, Dawn	117
Kupper, Thomas	32	Xu, Xiangyang	71, 193
La Notte, Alessandra	103	Załupka, Magdalena	187
Lesiv, Myroslava	9, 41, 49, 124	Zasina, Damian	208
		Zawadzki, Jarosław	208
		Zimnoch, Mirosław	177
		Żebrowski, Piotr	78, 90

