Integration of multi-source information in disaggregation of spatial emission data

Joanna Horabik-Pyzel, Zbigniew Nahorski

4th International Workshop on Uncertainties in Atmospheric Emissions
Kraków, 7-9 October 2015

Systems Research Institute
Polish Academy of Sciences

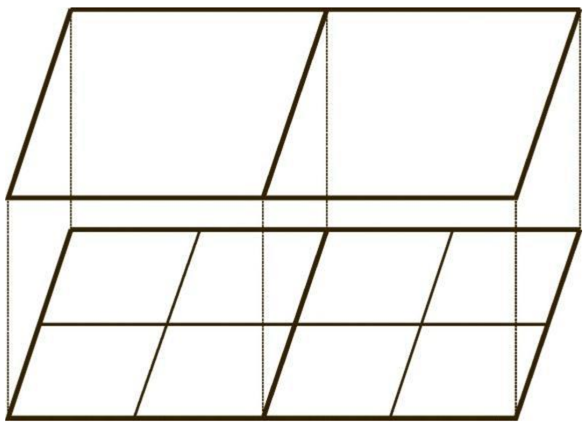
Outline of presentation

- Why to disaggregate?
- Disaggregation based on proxy data
- Some sources of additional information
- Methods of combining multi-model estimates
- Some questions

Why to disaggregate emissions?

coarse grid

- Quantification of GHG emissions at fine spatial scales is advantageous for many environmental, physical, and socio-economic analyzes.
- It can be easily integrated with other data in a gridded format.

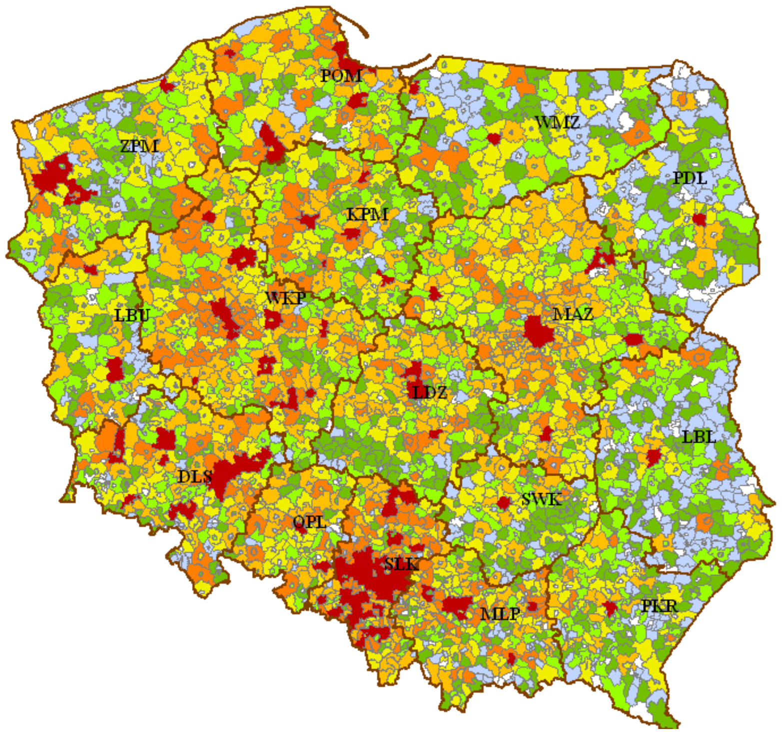


fine grid

- To better understand the transport of different pollutants atmospheric dispersion models are used; these models usually need emission fields in the gridded format.
- To improve assessment of carbon and other chemical component cycles, in order to better understand the climate change processes.

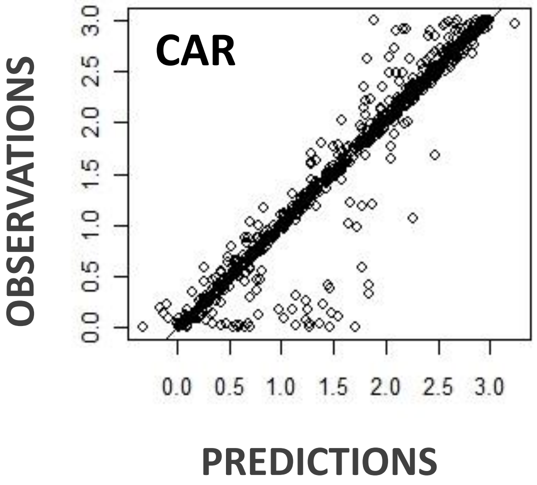
Disaggregation based on proxy data

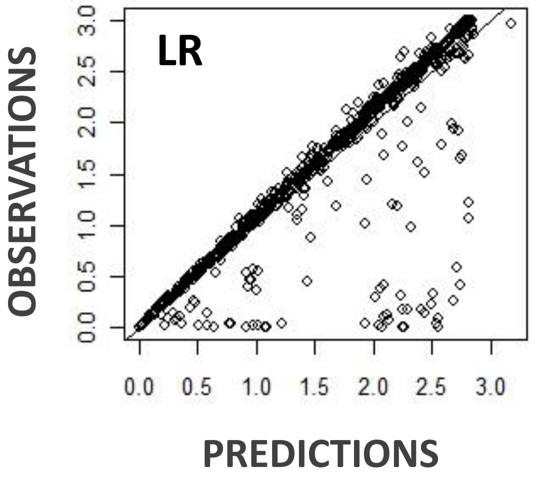
- Main tool – disaggregation of the coarse grid emissions proportionally to a proxy variable given in fine grid, like population, area of the cel, satellite observations of nighttime lights, etc., or modeling.
- Disaggregation should be done possibly individually for different categories of activities – GESAPU project (Bun et al.).
- Uncertainty can be estimated using IPCC methodology (preferably Tier 2: Monte Carlo method).



An example of 2km×2km disaggregated emissions from the GESAPU project.

Extensions



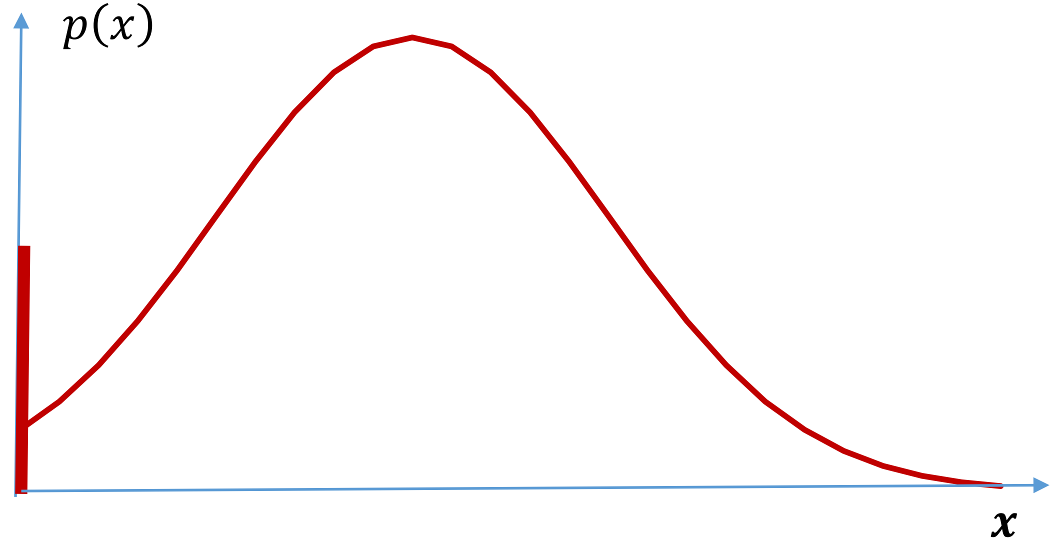


- Using more than one proxy data for disaggregation – linear regression (LR).

- Problem is to estimate its parameters, as no data for the fine resolution grid usually exist.
- Using data from other regions is questionable and rather prone to high errors (Ghosh et al., 2010).
- A method to use regression function – in a more general context of spatially autocorrelated data (CAR) – was proposed by Horabik & Nahorski: estimate the parameters from the coarse grid data and use them in the regression function for the fine grid. Uncertainty distribution are estimated by statistical inference.

Problems with values at zero

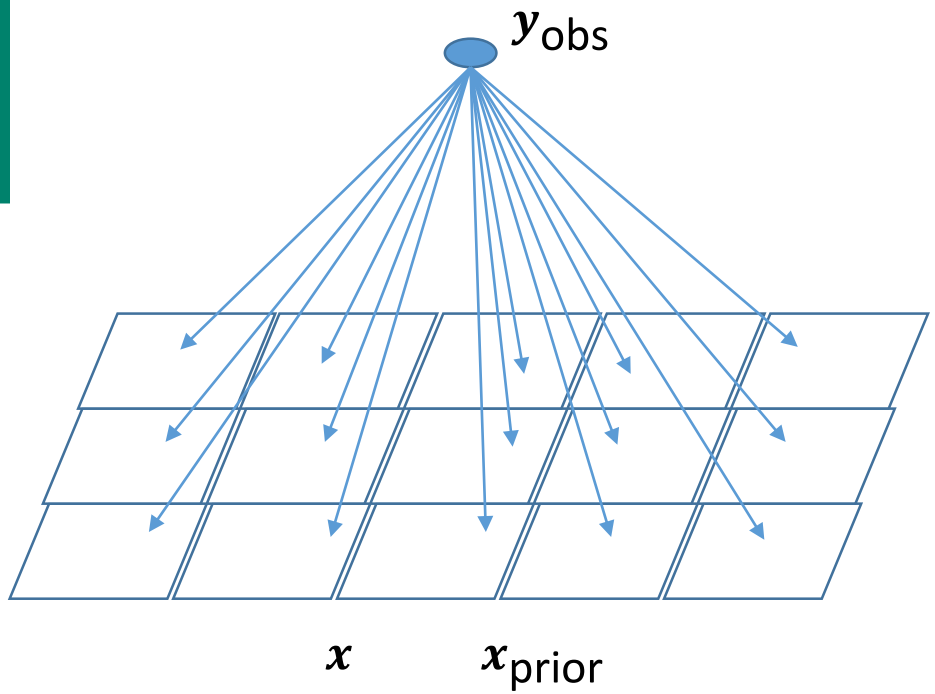
- Zero emissions in part of cells – small negative values from a model.
- Continuous probability distribution for positive arguments, probability mass at zero.
- Semicontinuous variables, Clumped-at-zero data, Zero-inflated data.



Other sources of information

- Atmospheric observations
- Measurements of tracers in the atmosphere
- Direct local measurements of fluxes
- Remote sensing
- Other information (values for some cells, independent emission assessments in subregions, common sense knowledge, etc.) that are difficult to use directly as proxies

Atmospheric observations



- The problem is to partition the estimated atmospheric load to emission sources.
- Typically done using the inversion methods to estimate CO_2 fluxes, the Bayes estimator is generally applied.

$$p(\boldsymbol{x}|\boldsymbol{y}_{\text{obs}}) = \frac{p(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y}_{\text{obs}})}$$

Function that relates emission with concentration $\boldsymbol{y}_{\text{obs}} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\psi}$

Gaussian distributions!

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_{\text{prior}} + \left(\mathbf{H}^T\mathbf{C}_y^{-1}\mathbf{H}+\mathbf{C}_x^{-1}\right)^{-1}\mathbf{H}^T\mathbf{C}_y^{-1}\left(\boldsymbol{y}_{\text{obs}} - \mathbf{H}\boldsymbol{x}_{\text{prior}}\right)$$

$$\widehat{\mathbf{C}}_x = \left(\mathbf{H}^T\mathbf{C}_y^{-1}\mathbf{H} + \mathbf{C}_x^{-1}\right)^{-1} = \mathbf{C}_x - \mathbf{C}_x\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_x\mathbf{H}^T + \mathbf{C}_y\right)^{-1}\mathbf{H}\mathbf{C}_x$$

$\mathbf{C}_x$, $\mathbf{C}_y$ - covariance matrices

Measurements of tracers in the atmosphere

- The ^{14}C isotope – half-life approximately 5700 years.
- The fossil fuels come from organisms which lived milion (10^6) to hundred milion (10^9) years ago.
- Intensive burning of the fossil fuels dilutes the atmospheric concentration of the ^{14}C isotope – may be used as a tracer of fossil fuel originated CO_2 emissions.
- Other tracers – CO, NO_x, $^{13}CO_2$, SF_6, other related with emissions from a source.
- Source apportionment – Lopez et al. (2013) for Paris.
- Scarce measuring observation stations (10 in Europe for ^{14}C in 2008) – low resolution of fluxes; corn leaves, rice, grape wine ethanol, grass, tree leaves, tree rings.

Direct local measurements of fluxes

- Flux tower observations – eddy covariance method.

  Scarce – USA and Canada, only 36 flux towers (2013).

- Chamber system flask measurements, e.g. measurement of emission from the soil.

  High spatial variability – 3 times difference in the distance of few metres (N_2O) – Gałkowski (2015) in vicinity of Kraków.

- Interpolation is needed to get areal data.

  Geostatistical methods (Kriging) – homogeneous fields.

  Bayesian melding – can cope with inhomogeneity of pollution fields.

Combining multi-model estimates: Averaging

Weighted ensemble average of model results $\bar{x} = \frac{\sum_i w_i x_i}{\sum_i w_i}$

Usually weights determined from historical records of performance.

Methods developed by climate projection community

Disaggregation: no historical data available (nothing is known on models accuracies).

Reliability ensemble method (REA) – weights calculated iteratively:

1. Simple average – deviations from averaged value.
2. Weights inversely proportional to the absolute deviations in the previous iteration.

Combining multi-model estimates: Bayesian approach

Adaptation of atmospheric inversion rule: $H = I$ (identity matrix)

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_{\text{prior}} + \left(\mathbf{C}_y^{-1} + \mathbf{C}_x^{-1}\right)^{-1} \mathbf{C}_y^{-1}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{x}_{\text{prior}}\right)$$
$$= \boldsymbol{x}_{\text{prior}} + \mathbf{C}_x\left(\mathbf{C}_y + \mathbf{C}_x\right)^{-1}\left(\boldsymbol{y}_{\text{obs}} - \boldsymbol{x}_{\text{prior}}\right)$$

$$\hat{\mathbf{C}}_x = \left(\mathbf{C}_y^{-1} + \mathbf{C}_x^{-1}\right)^{-1} = \mathbf{C}_x - \mathbf{C}_x\left(\mathbf{C}_x + \mathbf{C}_y\right)^{-1}\mathbf{C}_x$$

For diagonal covariance matrices $\mathbf{C}_x$ and $\mathbf{C}_y$

$$\hat{x}_i = x_{i,\text{prior}} + \frac{c_{ii,x}}{c_{ii,x} + c_{ii,y}}\left(y_{i,\text{obs}} - x_{i,\text{prior}}\right), \qquad i = 1, \dots, n$$

$$\hat{c}_{ii,x} = \frac{1}{\frac{1}{c_{ii,x}} + \frac{1}{c_{ii,y}}} = \frac{c_{ii,x}c_{ii,y}}{c_{ii,x} + c_{ii,y}}, \qquad i = 1, \dots, n$$

Combining multi-model estimates: Joint probability distribution approach

Proposed by Arkady Kryazhimsky (2013).

Independent uncertainty distributions $p(x)$ and $p(y)$. Joint distribution

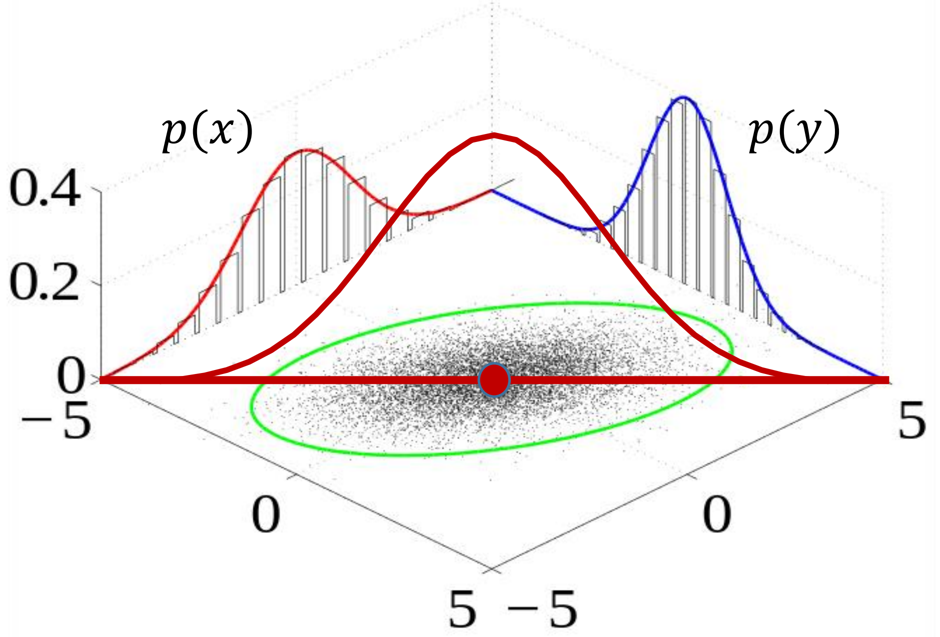
$$p(x, y) = p(x)p(y)$$

Find

$$\arg\max_{x=y} p(x, y) = \arg\max p(x, x)$$

Similarly for the multivariable case.

Kryazhimsky, Rovenskaya, Shvidenko, Gusti et al. (2015)



Some questions

- How much all additional knowledge can improve the estimates obtained by using proxy variables?

  Not much can be probably expected in the case of, say, CO_2 industrial emissions. For N_2O emission from the biosphere, very poorly estimated, , perhaps can give a considerable improvement.

- Introduction of additional knowledge will change the officially reported emission estimates. Is it a challenge to official national inventory reports (NIRs)?
- A lot of development is still needed to implement the ideas presented in the talk. Is the problem worth of taking it up?

Thank you for your attention!