Exploring patterns and determinants of the agro carbon footprint in the Po Valley (Italy) using spatio-temporal small area models

Paolo Maranzano
University of Milano-Bicocca &
Fondazione Eni Enrico Mattei (Italy)

Riccardo Borgoni
University of Milano-Bicocca

Riccardo Pajno
University of Milano-Bicocca

Felicetta Carillo
Italian Council for Agricultural Research and Economics (CREA)

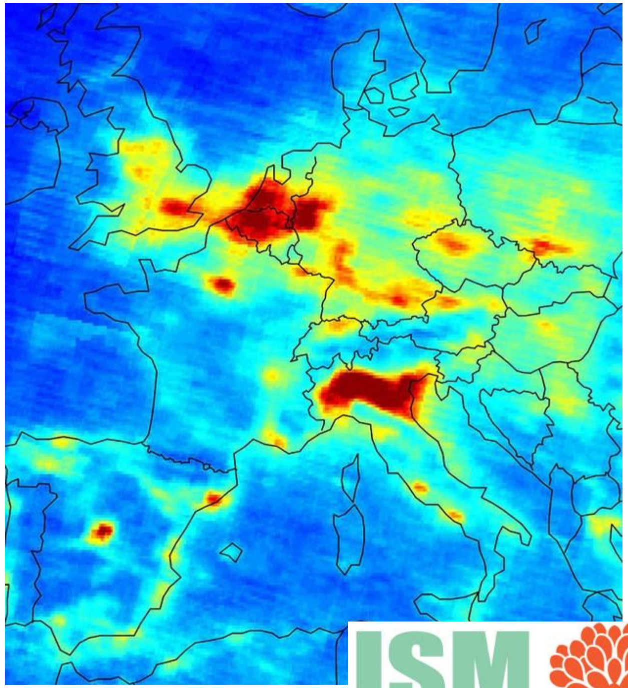
August 4th, 2024

Presentation line-up

1. Po Valley in Italy and agricultural carbon footprint
2. Available data on environmental impact of agriculture in Italy
    1. Survey data on agriculture in Europe: the RICA-FADN database
    2. Census data on agriculture: the Italian Agricultural Census 2020
3. The survey weights from RICA and the need to build a spatial-representative weighting scheme using census information
4. Small area models for the agricultural carbon footprint in the Po Valley: Fay-Herriot models
5. Conclusive remarks

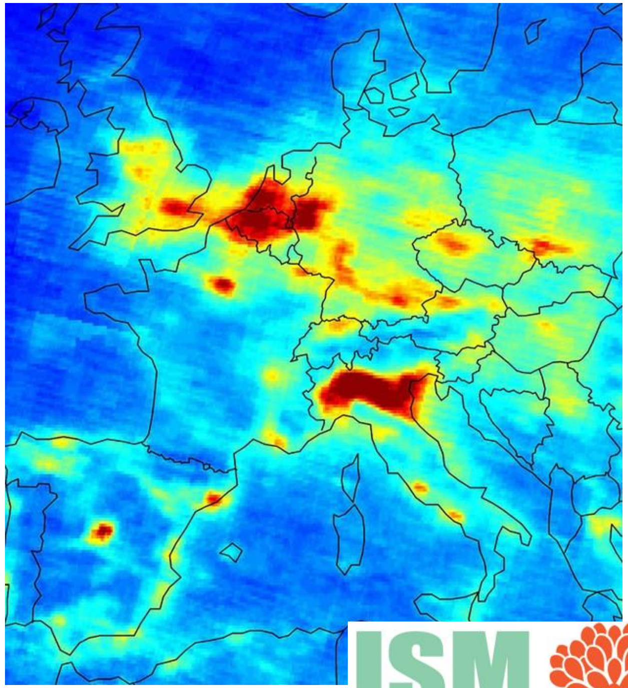
Context: the Po Valley

- The Po Valley appears to be **one of the most polluted areas on the World** due to:
    - Unfavorable geographical conditions (Raffaelli, Deserti et al. 2020)
    - High population density and industrialization (Regional Statistical Yearbook 2017)
    - <u>Intense farming activity </u>(Fassò, Rodeschini et al. 2023)



Context: the Po Valley

	Bovine density (heads/km2)	Swine density (heads/km2)	Number of livestock farms	Average n. of swines for farm	Average n. of bovines for farm
Brescia	109.3	289.7	4946	1086.6	87.0
Cremona	193.4	521.9	1598	2413.8	252.1
Lodi	160.4	489.2	664	2359.8	257.1
Mantova	137.3	451.4	2219	2396.3	171.4
Milano	53.5	43.3	806	912.1	107.5
Veneto	51.9	44.7	15388	331.9	55.6
Emilia-Romagna	31.1	57.9	7451	883.7	86.1
Piemonte	28.9	41.9	13326	575.9	44.6
Other italian regions	10.2	6.7	114751	101.1	22.4



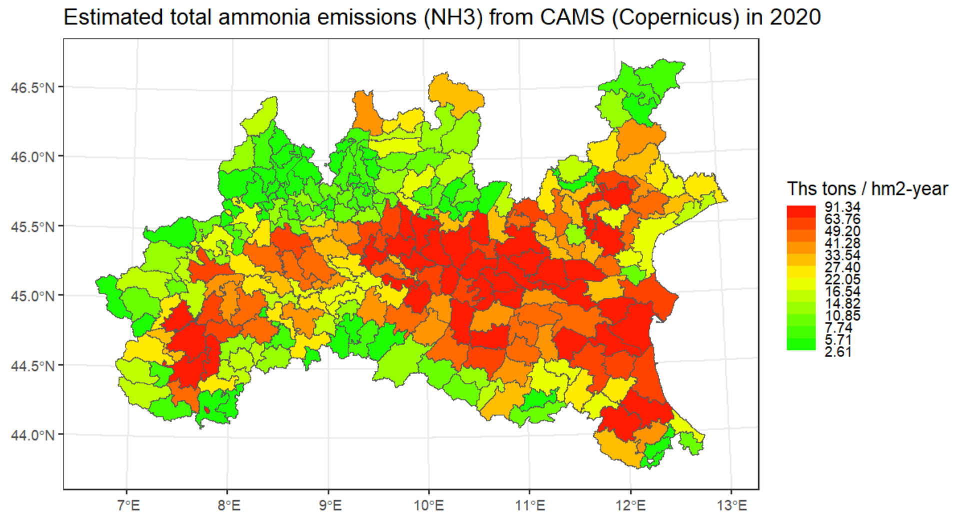
Agricultural Carbon Footprint

- International treaties about CO_2 reduction reveal the need to evaluate Carbon Footprint (CF) around the world
- Two main methods:
  - **IPCC** (*International Panel on Climate Change)* -> emissions per process
  - **LCA** (*Life Cycle Assessment)* -> emissions per product
- IPCC appears to be more convenient for sector studies
- This method assumes a linear relation between 6 different economic activities and 2 greenhouse gasses. Eventually, the emissions from this gasses are converted in equivalent CO_2 kg
- In Italy, **CF for agricultural sector is officially estimated by CREA** (Council for Agricultural Research and Economics) using IPCC methodology

JSM PORTLAND OREGON

Research task

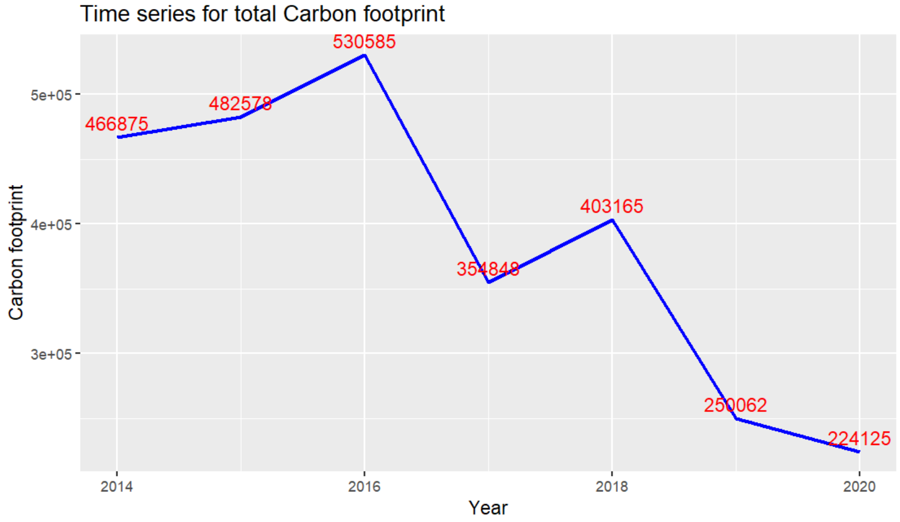
- We are interested in investigating the environmental impact of agricultural activities at the finest possible spatial scale: we need a measure of **farming carbon footprint**
- CREA quantifies **ACF at the farm level** → the **RICA database**
- In this work, using both sample and census data, we want to give an **accurate small area estimate** of the agricultural carbon footprint
- We rely on a official small area classification given by the **agrarian subregions** (i.e., internally homogeneous areas with similar agricultural characteristics): 256 spatial units across the Po Valley (Maranzano, McConville et al. 2023)

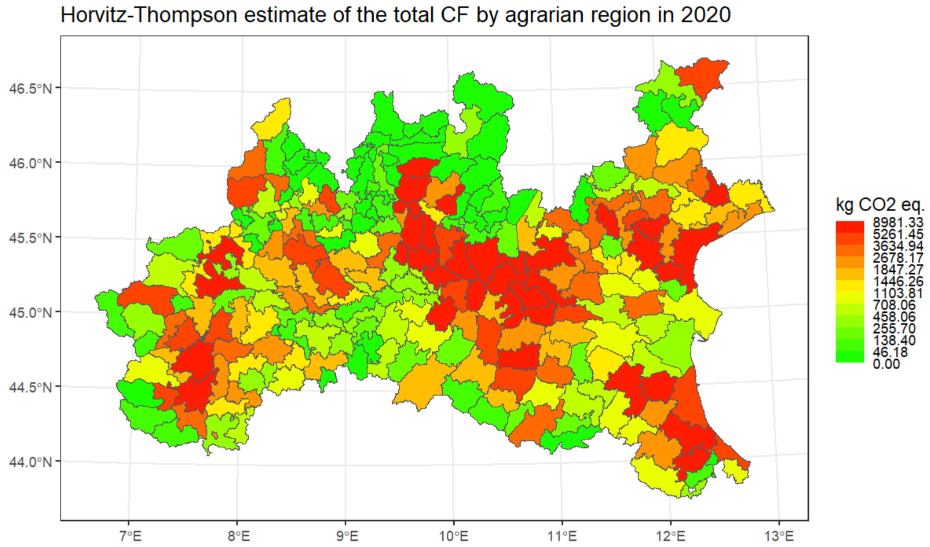


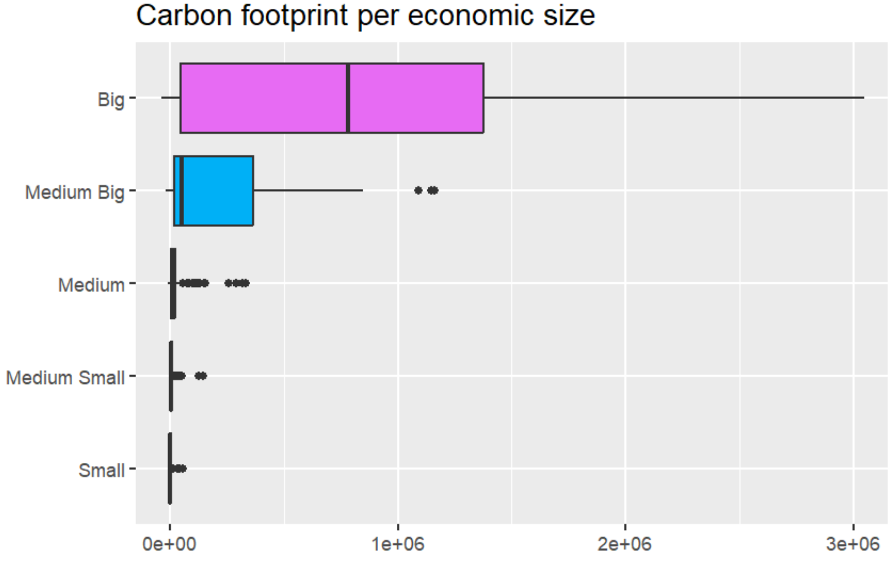
Sample data: the Italian FADN database

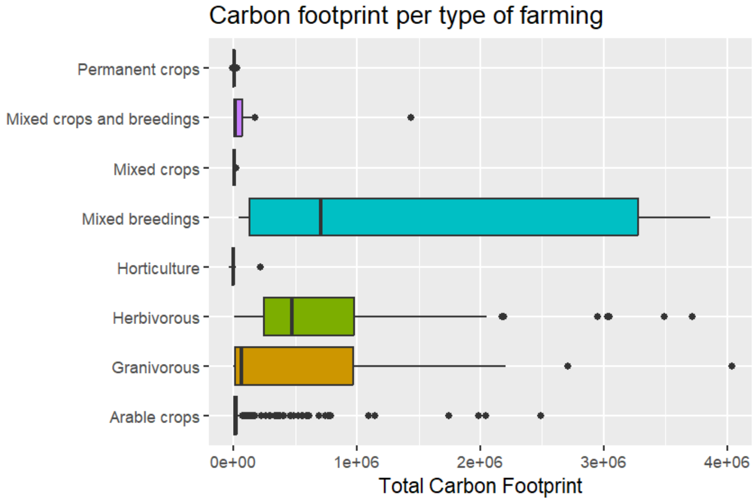
- The **RICA database** is made available by CREA, which collects information on agricultural activities in Italy on a yearly basis and harmonize data within the FADN (Farm accountancy data network) database of the EU;
- The database includes farms with a Standard Output bigger then 8.000€ and it is stratified according to:
  - **Regions** (Eurostat NUTS-2 level, e.g., Lombardia or Veneto)
  - Type of farming or technical specialization (e.g., crop specialized, breeding, mixed farming etc.)
  - Economic size (e.g., standard output classes)
- Our sample includes unit-level data from 4 Italian Regions (Lombardia, Piemonte, Veneto and Emilia-Romagna) in 2020
- In particular, we have considered the Carbon footprint (the model target variable) and the number of activities by **type of farming** and **economic size** from that sample.

Sample data: Carbon Footprint





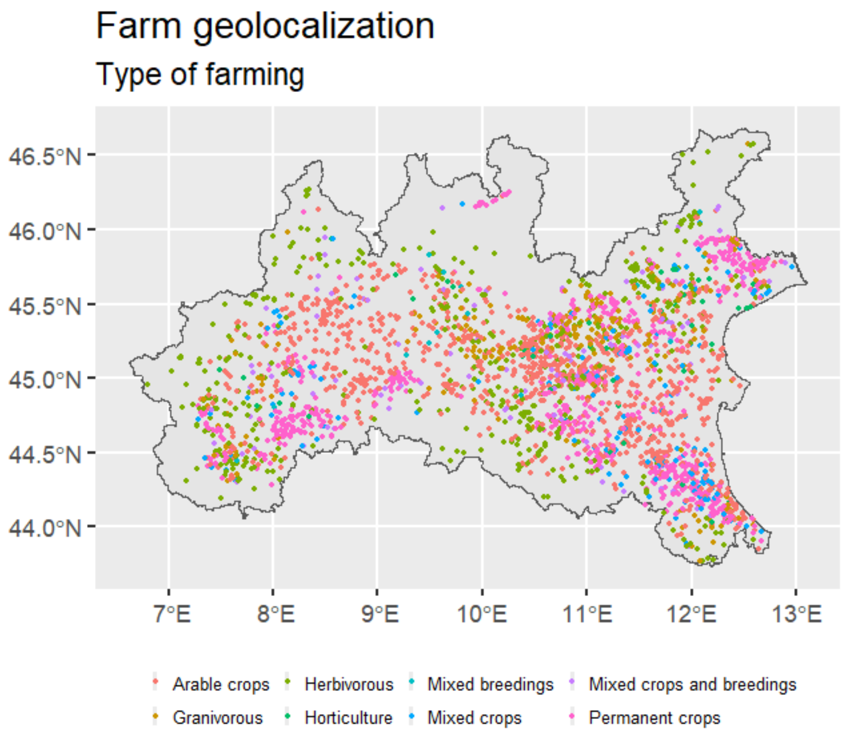


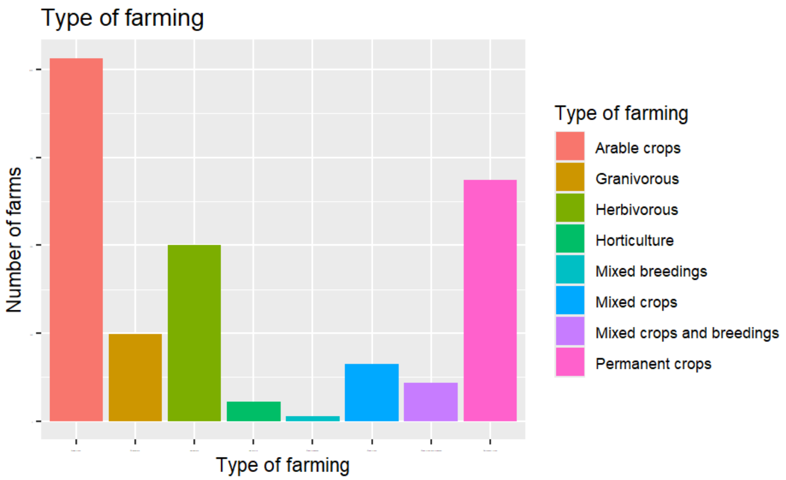


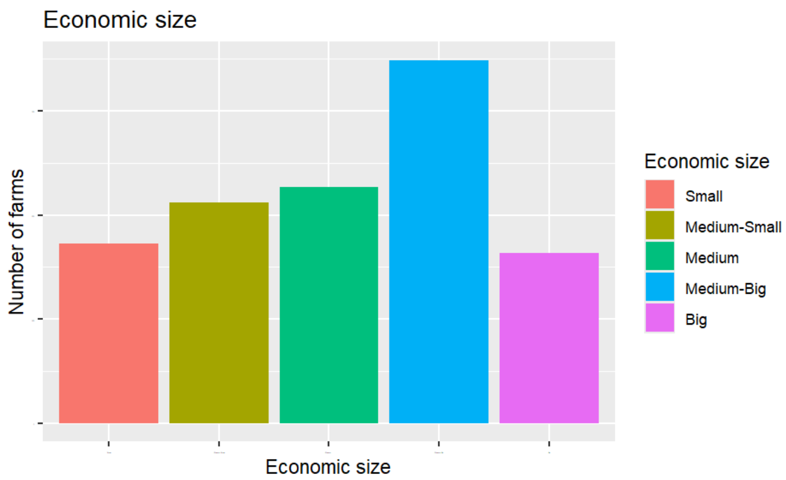
- Decreasing trend in the last five years
- Heterogeneous spatial distribution across agrarian regions
- Bigger farms tend to be more pollutant
- Breeding are more pollutant than crops

Sample data: number of farms

- 2806 farms spread across the four regions
- Major number of crops farms
- Prevalence of medium size farms
- Strong correlation between high farm density and most polluted areas







Census data: Italian Agricultural Census 2020

- National census survey on the Italian agricultural sector (data at municipal level)
- We took into consideration
  1. Total number of swines and bovines → To control for livestock farming
  2. Total utilized agricultural land → To control for the relevance (size) of farming
  3. Total standard working days (in agriculture) during the year → To control for the relevance (labor market) of farming
  4. The number of farms by type of farming, economic size, and agrarian region → **To build a spatial-representative weighting system**
- Total: about 226K farms → RICA sample ~ 1.23% of the total number of farms

A spatially-representative weighting scheme

- For each farm, RICA provides a sampling weight…
  - Which is representative of the regional techno-economic specialization of farm holdings
  - But it is not representative of the spatial distribution of farming activities within the administrative regions (e.g., Lombardia)
- As we are interested in estimating the agricultural carbon footprint at the finest spatial scale **we need a proper weighting scheme** to compute reliable small area estimates
- We rely on the following facts:
  - Each municipality belongs to one and only one agrarian subregion
  - Census data for 2020 contain information on the number of farms by economic size and technical specialization at the municipal level

A spatially-representative weighting scheme

We compute the joint frequencies of farm holdings by economic size (**small VS large farms**) and technical specialization (**crop specialized VS livestock specialized VS mixed farming**) for each agrarian subregion (i.e., N_{sjk}): each farm will be weighted as

$$w_{sjk} = \frac{N_{sjk}}{N_s}$$

- s=1,2,..,256 stands for agrarian subregion
- $j = 1,2,3$ for type of farming
- $k = 1,2$ for economic size

	Crop	Lives.	Mixed
Large	N_{25611}	N_{25612}	N_{25613}
Small	N_{25621}	N_{25622}	N_{25623}

s=256

…

	Crop	Lives.	Mixed
Large	N_{s11}	N_{s12}	N_{s13}
Small	N_{s21}	N_{s22}	N_{s23}

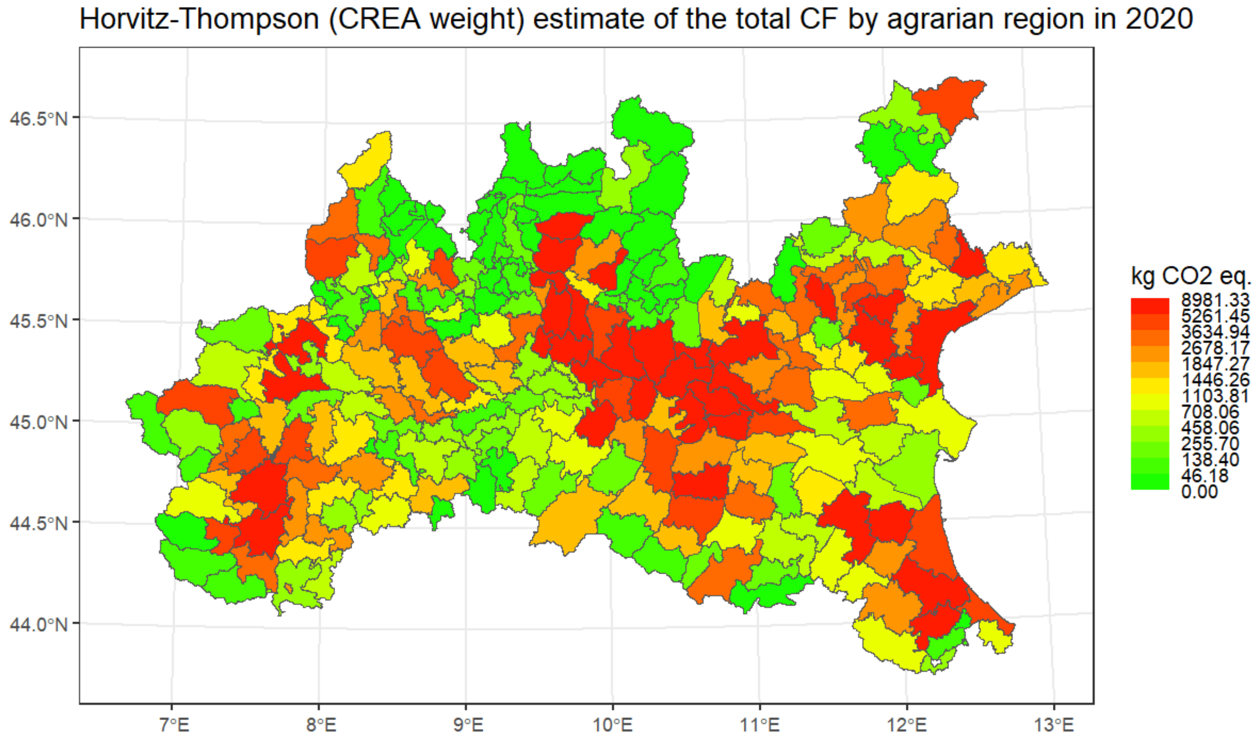
$$N_s = N_{s11} + N_{s12} + N_{s13} + N_{s21} + N_{s22} + N_{s23}$$

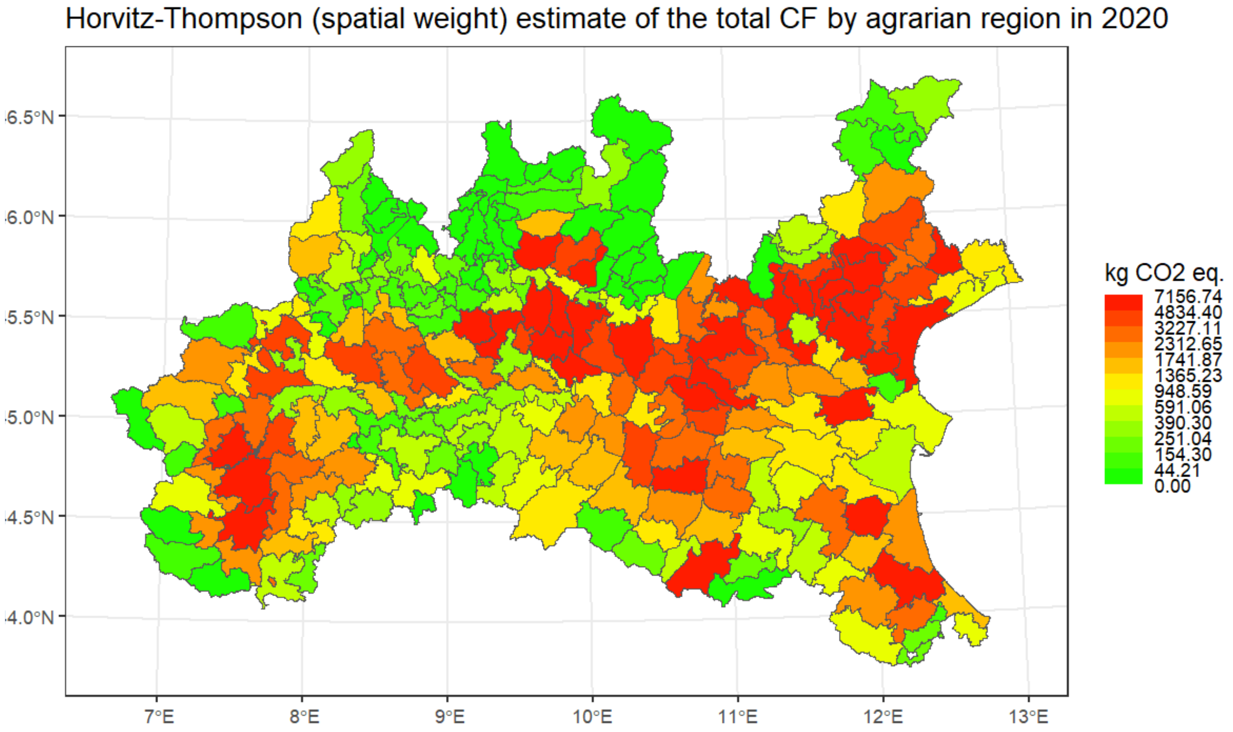
…

	Crop	Lives.	Mixed
Large	N_{111}	N_{112}	N_{113}
Small	N_{121}	N_{122}	N_{123}

i=1

Horvitz-Thompson direct estimate





Fay-Herriot model

Most common regression model for small area estimation, defined as following:

$$CF_s = X_s\beta + u_s + \varepsilon_s \quad s = 1, \dots, 256$$

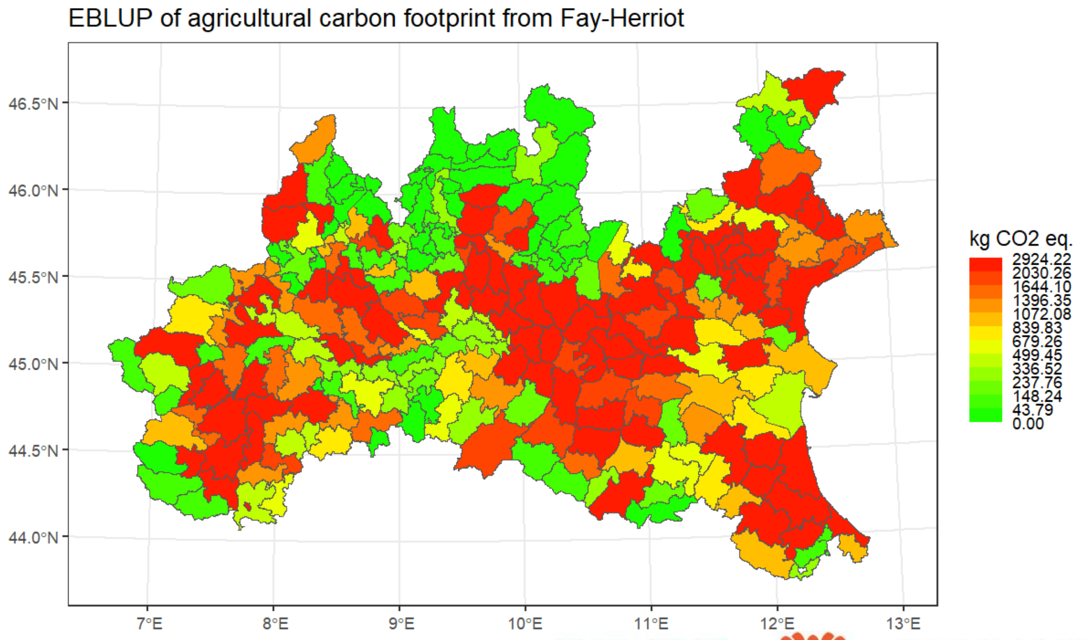
Where:

- CF_s is the HT direct estimate for agrarian region s
- X_d is a **census area-level 256×6 design matrix:** in addition to the agro-related variables, we included the average altitude of the agrarian region and the standard deviation (proxy of landscape wiggliness)
- u_s is a model-specific **random effect** (in this case we are defined as independently Normally distributed, i.e. $u_s \sim N(0, \sigma_{\mathrm{u}}^2)$)
- ε_s is the **model error term** with classical assumptions, i.e. $\varepsilon_s \sim N(0, \sigma_\varepsilon^2)$

Fay-Herriot model: results

	Estimate	SE	t-value	p-value
Intercept	-70.4157	76.5902	-0.9194	0.3579
Utilized agricultral surface	-0.0077	0.0044	-1.7336	0.0830
Standard working days in agriculture	0.0028	0.0003	8.7939	**0.0000**
Total bovine heads	0.0263	0.0065	4.0259	**0.0001**
Total swine heads	0.0024	0.0024	1.0216	0.3070
Average altitude	0.1293	0.1633	0.7917	0.4285
SD altitude	-0.4232	0.5097	-0.8303	0.4064

- **Labor** and **bovines** are strongly significant and their estimated coefficients are positive;
- The **EBLUP** estimates of the carbon footprint are smoother than the direct counterparts.



JSM PORTLAND OREGON

Spatial Fay-Herriot model

Most common regression model for small area estimation, defined as following:

$$CF_s = X_s\beta + u_s + \varepsilon_s \quad s = 1, \dots, 256$$

Where:

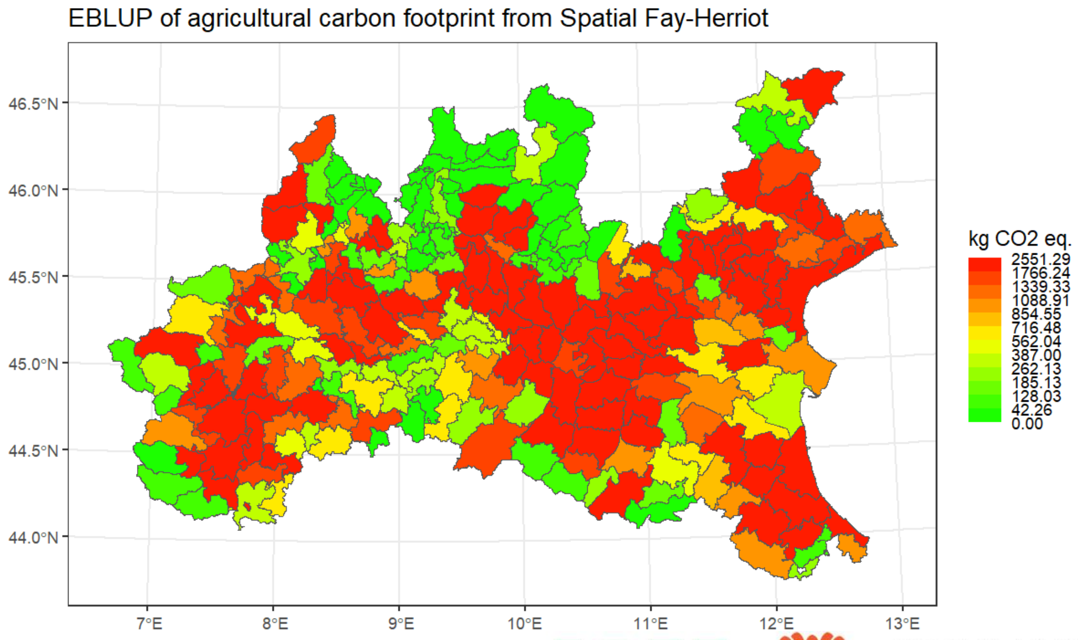
- CF_s is the HT direct estimate for agrarian region s
- X_d is a **census area-level 256×6 design matrix**
- u_s is a spatial **random effect following a SAR(1) process, i.e.,** $u_s = \rho W u_s + \varepsilon_s$ with ρ being the spatial autoregressive coefficient and W being the spatial first order adjacency matrix
- ε_s is the **model error term** with classical assumptions, i.e. $\varepsilon_s \sim N(0, \sigma_\varepsilon^2)$

JSM PORTLAND OREGON

Spatial Fay-Herriot model: results

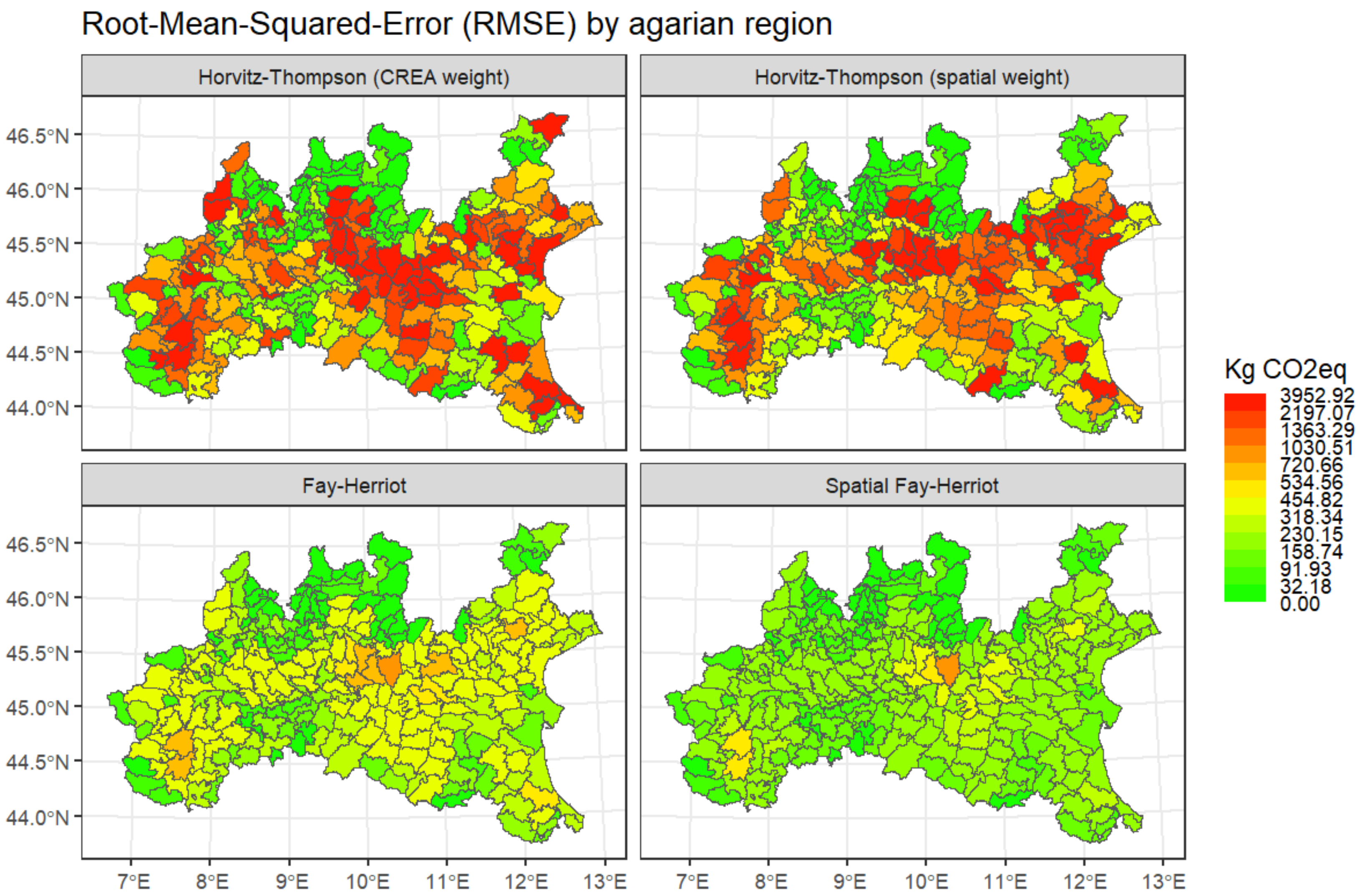
	Estimate	SE	t-value	p-value
Intercept	-39.9614	62.5573	-0.6388	0.5230
Utilized agricultral surface	-0.0058	0.0035	-1.6408	0.1008
Standard working days in agriculture	0.0024	0.0002	9.8119	**0.0000**
Total bovine heads	0.0102	0.0051	1.9991	**0.0456**
Total swine heads	0.0059	0.0019	3.1075	**0.0019**
Average altitude	0.0979	0.0906	1.0798	0.2803
SD altitude	-0.3434	0.2999	-1.1451	0.2522

- **Labor** and **animals** are strongly significant and positive;
- Signs and magnitudes are consistent with the FH model
- The **EBLUP** estimates of the carbon footprint are similar to the FH ones.



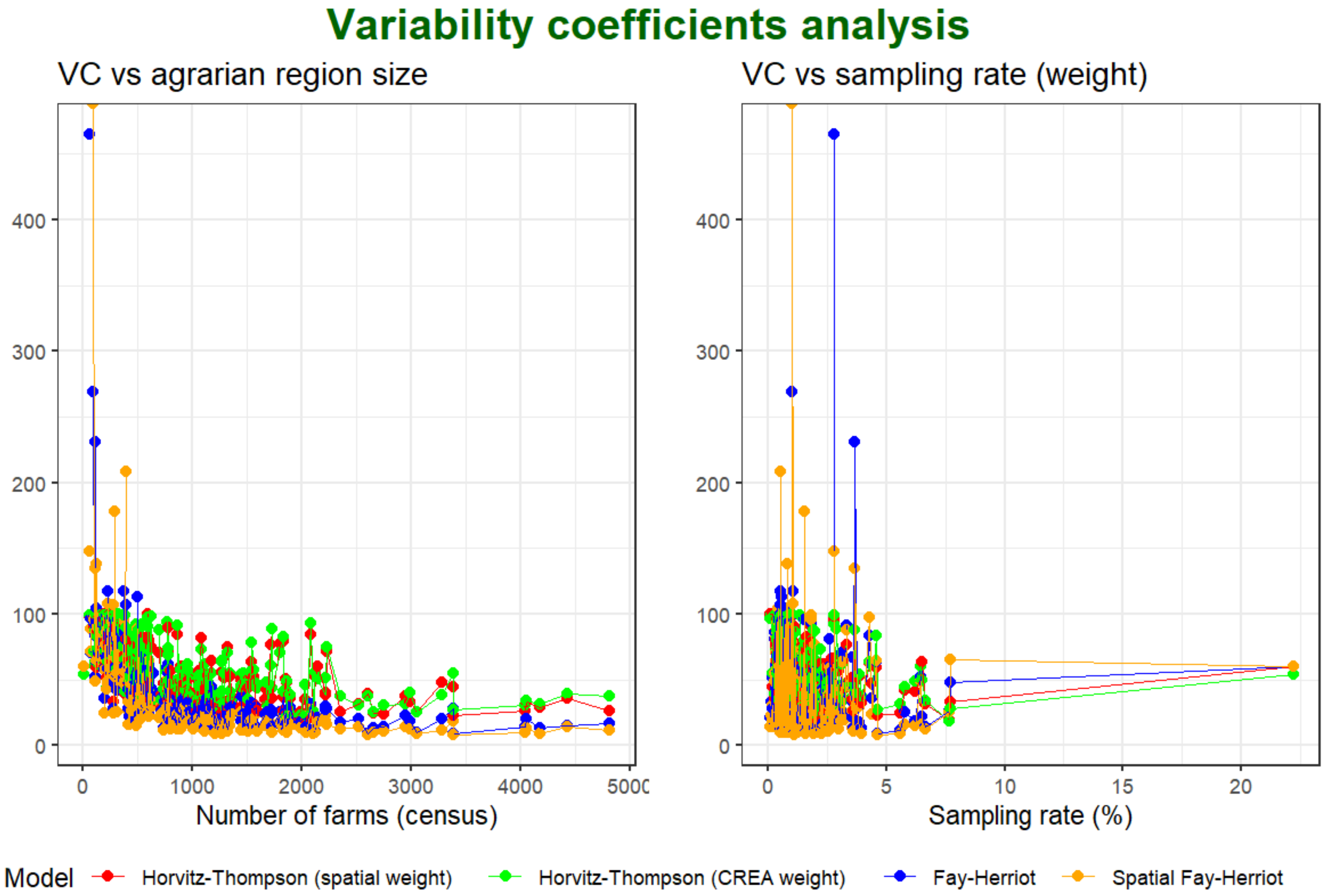
JSM PORTLAND OREGON

Model comparison: variability



- The root-mean-squared errors across space show large variability (consistently with the spatial distribution of farms and poor air quality)
- The linear models are able to remarkably improve the RMSE estimate. Also, the inclusion of spatial autoregressive terms further reduces the variability of the estimates.

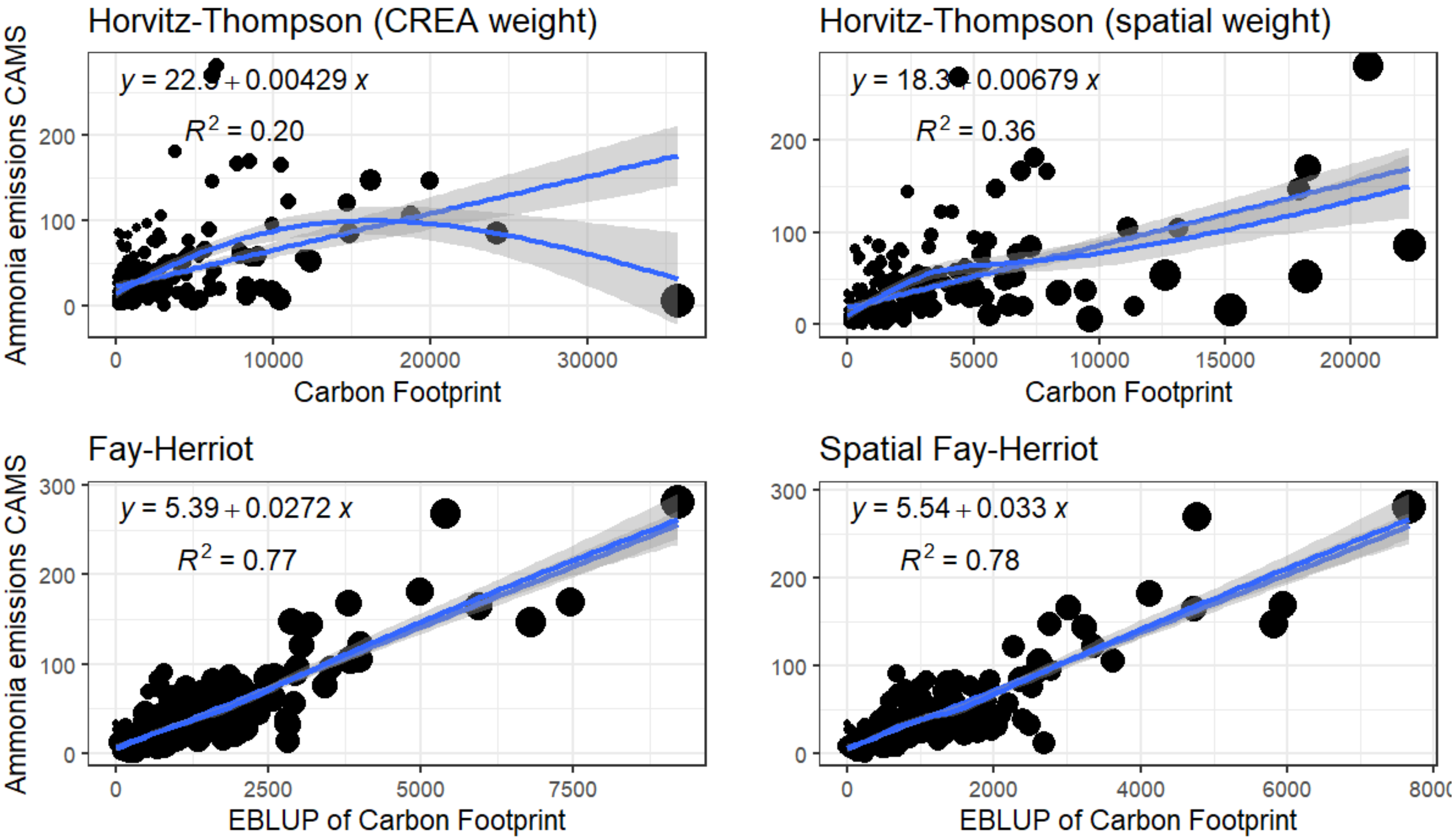
Model comparison: variability



- The coefficients of variability (VC) for modelled estimates (FH and SFH) are everywhere lower than the unmodelled estimates;
- When the number of farms increases, the variability reduces and the difference between modelled and unmodelled estimates shrinks to zero.

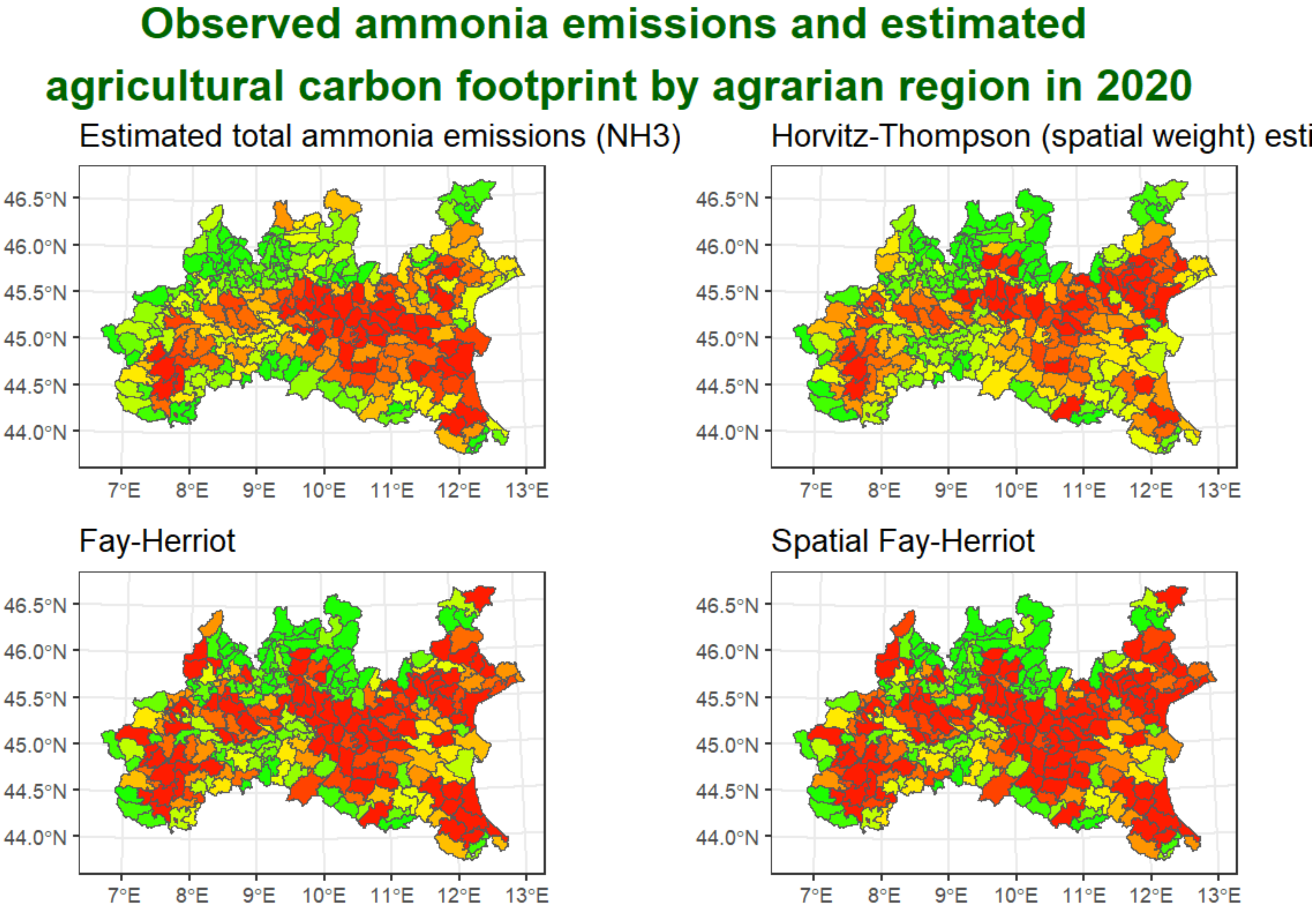
Model comparison: consistency

Empirical relationship among ammonia emissions (CAMS) and agricultural carbon footprint at agrarian regions



- We compare the carbon footprint estimates (direct and EBLUPs) with the estimated ammonia emissions from Copernicus-CAMS;
- Direct estimates seems poorly correlated with ammonia emissions, while modelled ACFs are strongly correlated;
- Modelled estimates are strongly consistent with actual values from institutions.

Model comparison: consistency



- We compare the carbon footprint estimates (direct and EBLUPs) with the estimated ammonia emissions from Copernicus-CAMS;
- Direct estimates seems poorly correlated with ammonia emissions, while modelled ACFs are strongly correlated;
- Modelled estimates are strongly consistent with actual values from institutions.

Conclusive remarks and future develpments

- We investigated the environmental impact of agricultural activities (**farming/agro carbon footprint**) in a critical region of Europe (the **Po Valley**), strongly affected by poor air quality and intensive farming;
- **We improved the original survey weights from CREA** by building area-specific weights (representative of the spatial distribution) based on the type of farming and economic size of farms;
- We **modelled** the direct estimates of agricultural CF using the Fay-Herriott approach with spatial structure:
    - Labor intensity and livestock farming seem to be the most relevant determinants of agro CF;
    - Modelled estimates are always lower compared to unmodelled (direct) estimates;
    - Modelled estimates are strongly consistent with actual values from institutions (CAMS emissions).
- Analysis could be easily extended to a panel structure considering the **temporal dimension** (CF estimates from CREA are available since 2014, while census information are only collected for 2020).


That's all Folks!


Q & A

Bibliographic references

- *Calderoni, S., Bonati, G.* (2013) “ICAII-Impronta Carbonica Aziende Agricole Italiane”, Working Paper, INEA, Rome
- *Morales, D., Esteban, M., Pérez, A., Hobza, T.* (2021) “A course on Small Area Estimation and Mixed Models”, Springer
- Carillo, F., et al. (2024). The spatio-temporal Fay-Herriot model using the state-space method: an application to Italian Lombard agrarian sub-regions. Book of Short Papers - 2nd Italian Conference on Economic Statistics (ICES 2024) - Statistical Analysis of Complex Economic Data: Recent Developments and Applications. Casa Editrice Bonechi, 66-69.
- Fassò, A., et al. (2023). "Agrimonia: a dataset on livestock, meteorology and air quality in the Lombardy region, Italy." Scientific Data 10(1): 143.
- Maranzano, P., et al. (2023). A geostatistical investigation of the ammonia-livestock relationship in the Po Valley, Italy. Book of the Short Papers SIS 2023. F. M. Chelli, M. Ciommi, S. Ingrassia, F. Mariani and M. C. Recchioni, Pearson: 200-205.
- Raffaelli, K., et al. (2020). "Improving Air Quality in the Po Valley, Italy: Some Results by the LIFE-IP-PREPAIR Project." Atmosphere 11(4): 429.
- Regional Statistical Yearbook, R. (2017). "Regional Statistical Yearbook of Lombardia in Europe 2017/2018."