A Geospatial Farm Typology for the EU based on FADN data

Marc Müller, Hugo Scherer, Michael Wögerer
Walter Rossi-Cervi, Tamás Krisztin, Felicity Addo

Pacioli workshop, Montegrotto, 06.-09.10.2024

Funded by the European Union

BrightSpace Horizon Europe Grant Agreement No 101060075 | LAMASUS Horizon Europe Grant Agreement No 101060423

Views are of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Project background

- BrightSpace:
  - Duration 2022-2027, 14 Partners
  - Definition and operationalization of a Safe and Just Operating Space for EU agriculture
  - Identification of quantifiable indicators
  - Projections until 2050

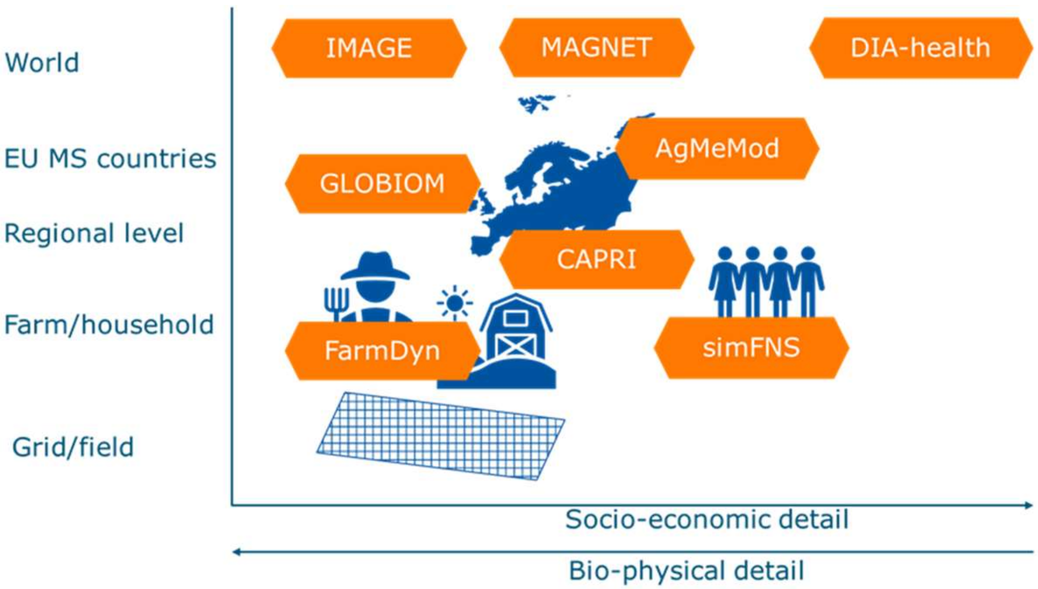
- LAMASUS:
  - Duration 2022-2026, 17 Partners
  - Ex post and ex ante impact assessment of European policies on land use and land management
  - Projections until 2050

Research background

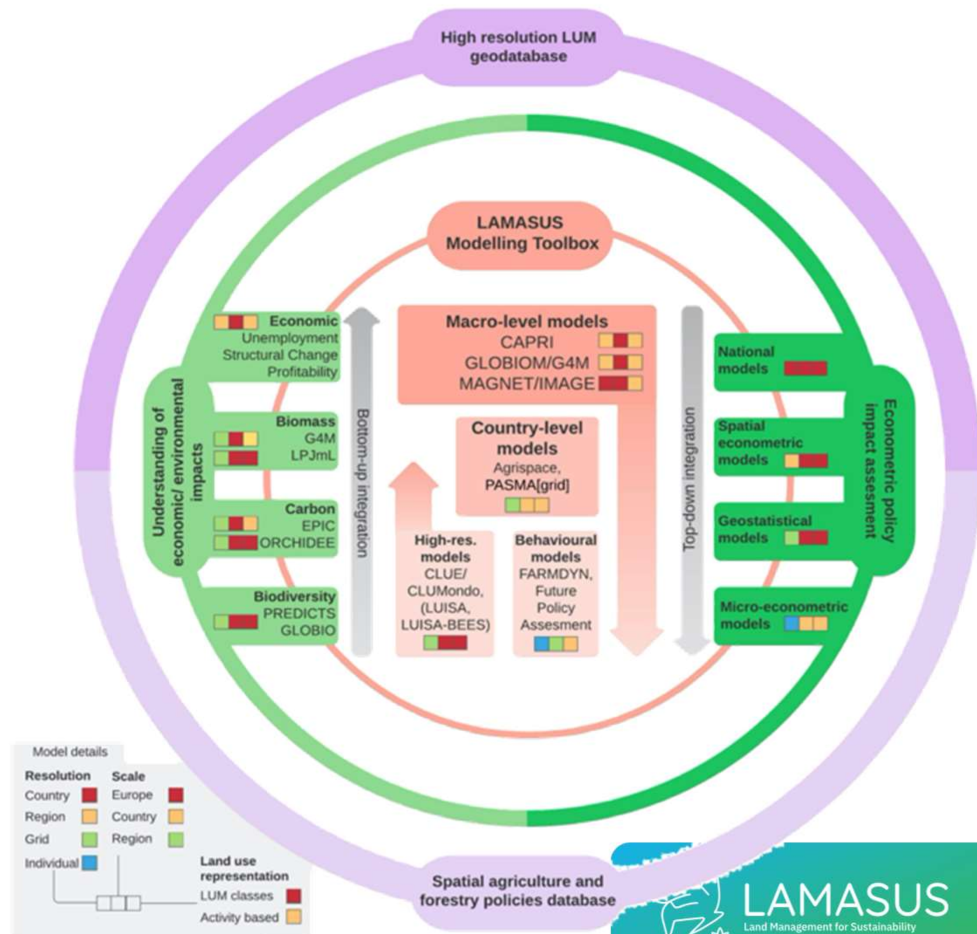
- Ex-ante policy analysis
  - Identification of innovative policy and governance measures in agriculture
  - Which farms are likely to get involved in agro-environmental measures?

- Ex-ante technology analysis
  - Identification of promising agricultural technologies
  - Which farms are likely to invest in GHG saving technologies?

BrightSpace and LAMASUS model toolboxes



- Use of FADN data in:
  - **FarmDyn – Farm level**
  - **GLOBIOM – Sector level**
  - CAPRI– Sector level
  - MAGNET– Economy-wide



BRIGHT SPACE

LAMASUS
Land Management for Sustainability

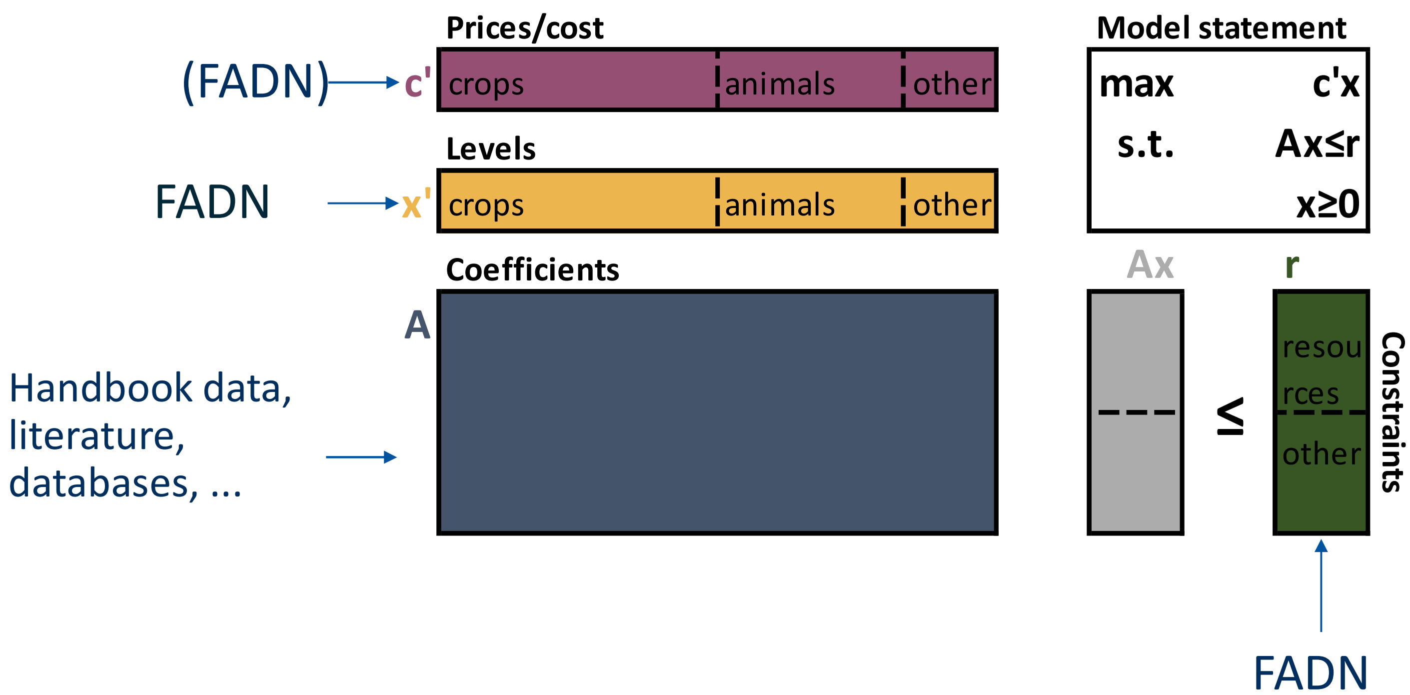
How FADN is used in FarmDyn

- Optimization of farm production plan
- Very detailed representation of technology (inputs per activity, e.g. field operations, feed nutrient contents, manure application, yield responses to fertilization,...)

- Data from FADN:
  - Outputs (yields and levels of cropping and animal activities)
  - Endowments (areas, herds, labour)

- Additional data:
  - Input prices
  - Activity specific input levels (variable cost, machinery requirements)
  - Sources: Handbook data, dedicated databases (KTBL in DE, KWIN in NL)

Illustrative farm level programming tableau & data sources



BRIGHT SPACE

LAMASUS Land Management for Sustainability

Typical vs individual farms

- Obstacles for modelling individual farms:
  - Technology coefficients
  - Environmental characteristics
  - Computation time (e.g. FarmDyn may take 90 sec for a single dairy farm)

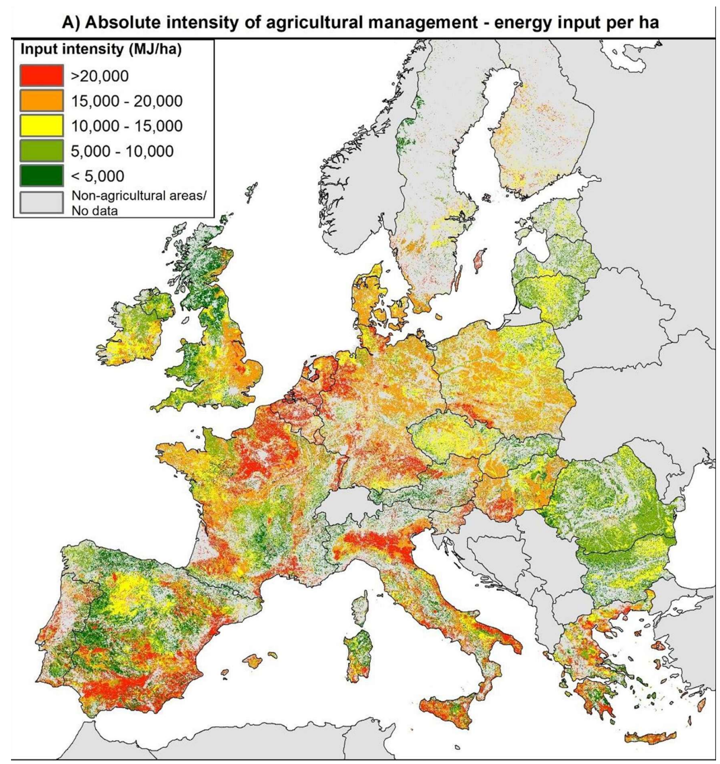
- Two solutions:
  - Summarize non-observable costs in a farm-specific cost term (PMP approach)
    - JRC's IFM-CAP model does this!
  - Focus on few typical farms!
    - Farm types in FADN (TF14) are great, but require further split as e.g. dairy farms are one single group.

How FADN is used in GLOBIOM

- Estimation of activity-specific variable cost
- Cost structure of technologies
- Upscaling of case studies

Energy intensity as criterion for a farm typology

- Article:
    - Rega, C., Short, C., Pérez-Soba, M., & Paracchini, M. L. (2020). A classification of European agricultural land using an energy-based intensity indicator and detailed crop description. Landscape and urban planning, 198, 103793.
- Main argument: anthropogenic energy intensity is relevant for design of policies aiming at GHG and nutrient emission reduction



Selected FADN variables

- Standard output (SE005)
- Expenditures for:
  - Fertilizer (SE295)
  - Plant protection (SE300)
  - Seeds (SE285)
  - Energy (SE345)

- Share of grassland (SE028) in UAA (SE025)
- Livestock density per UAA

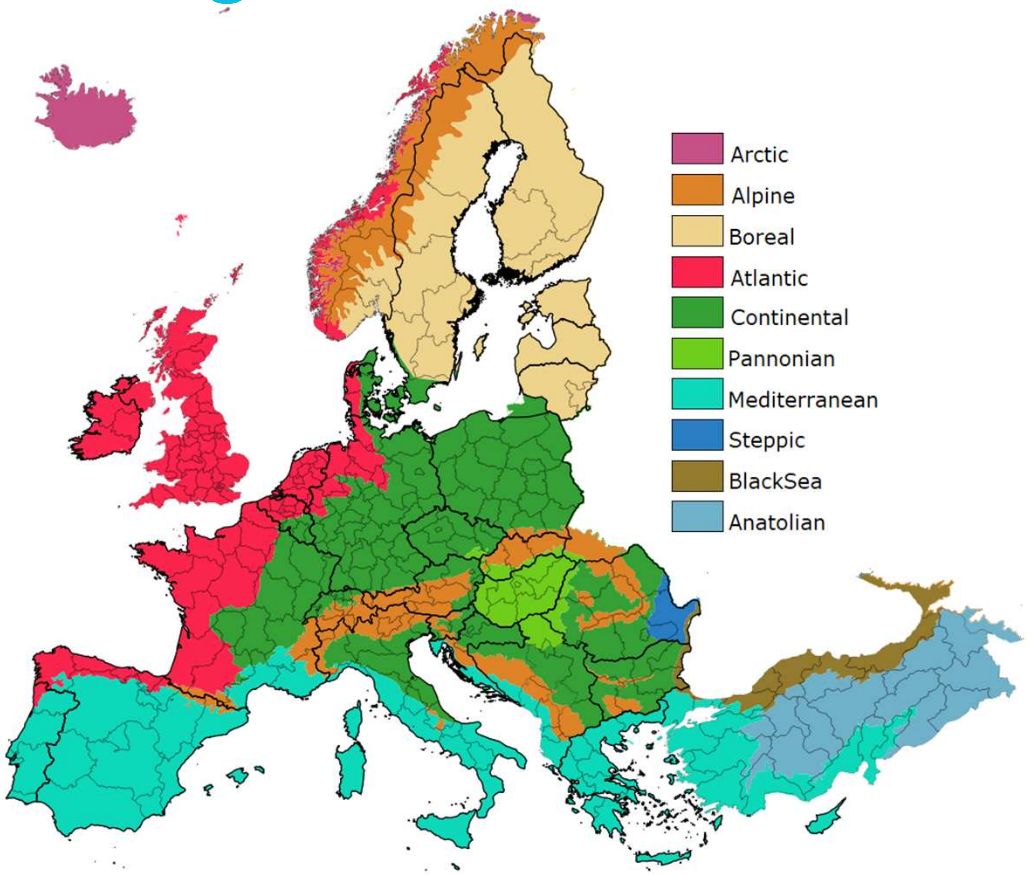
Adding spatial data to FADN

- We want to bring spatial data to FADN, not the other way around!

- Bio-geographic regions
- Soil texture

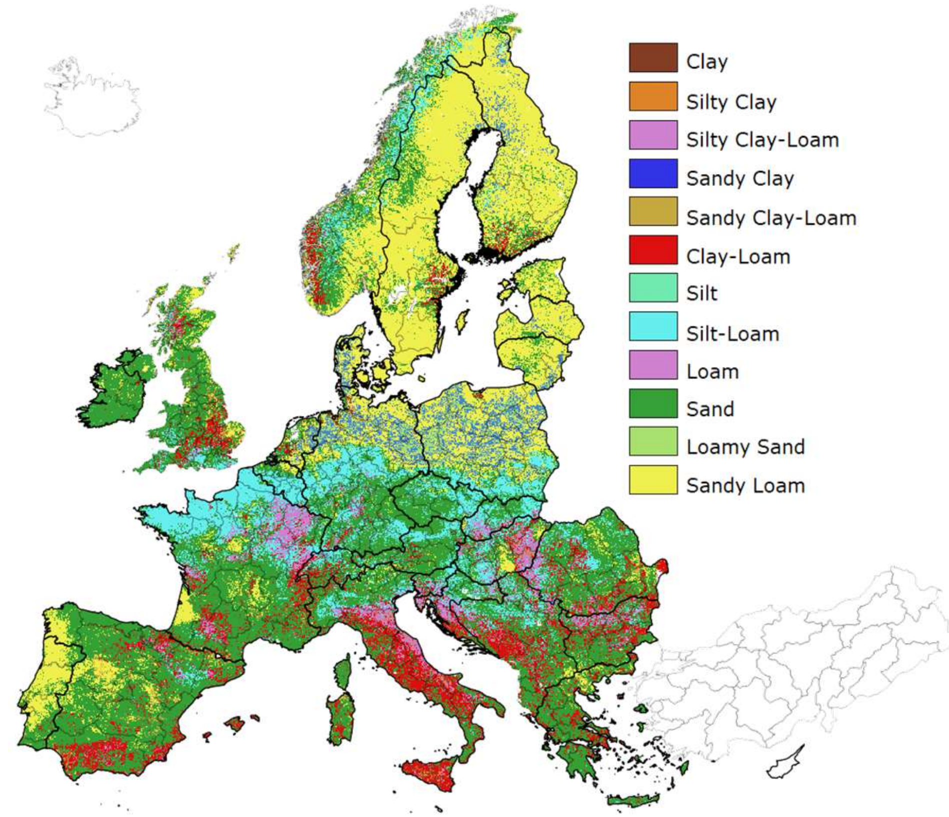
- Challenge: FADN provides only NUTS2 for location of farms
- Approach: Mask out non-agricultural areas to increase precision

Bioregions and soil texture at NUTS2 level



EEA Bio-geographic regions

(https://www.eea.europa.eu/en/analysis/maps-and-charts/biogeographical-regions-in-europe-2)



JRC LUCAS Data (Texture classes based on USDA)

(https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data)

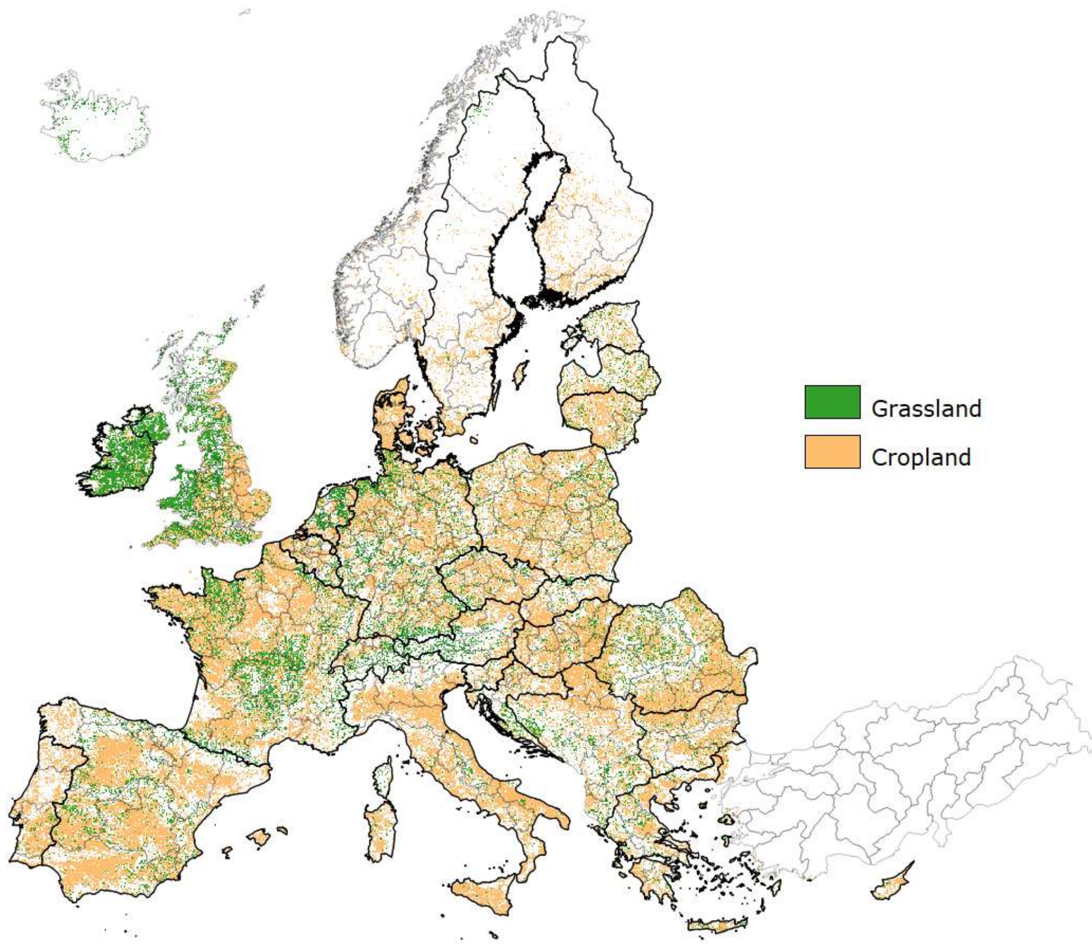
BRIGHT SPACE

LAMASUS Land Management for Sustainability

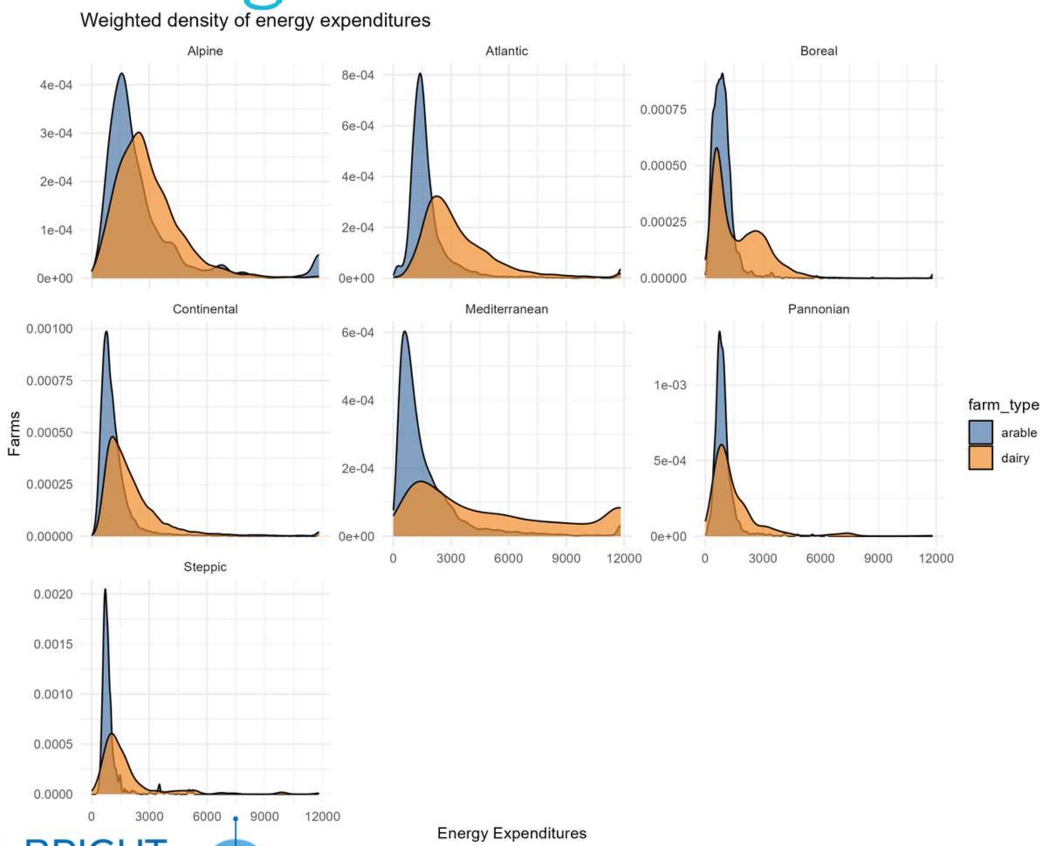
Input data: Land use maps

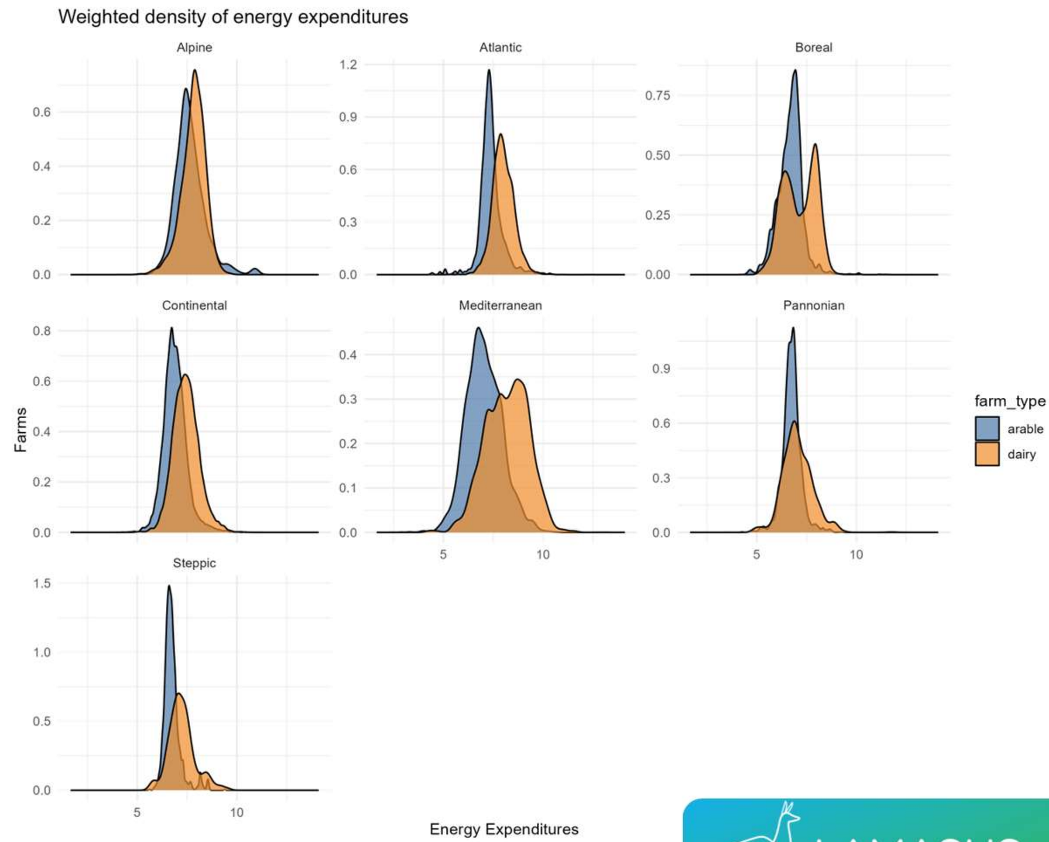
Corine Land Cover Map

https://land.copernicus.eu/en/products/corine-land-cover



Explorative analysis – farm type and biogeo regions

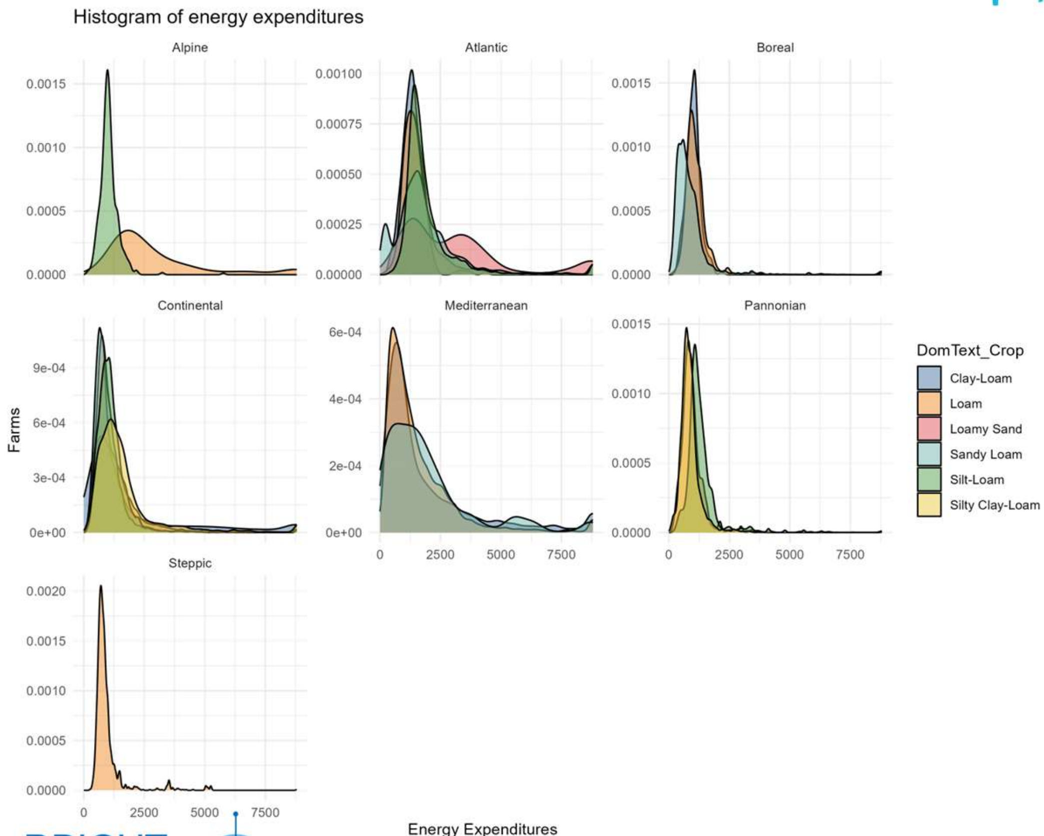


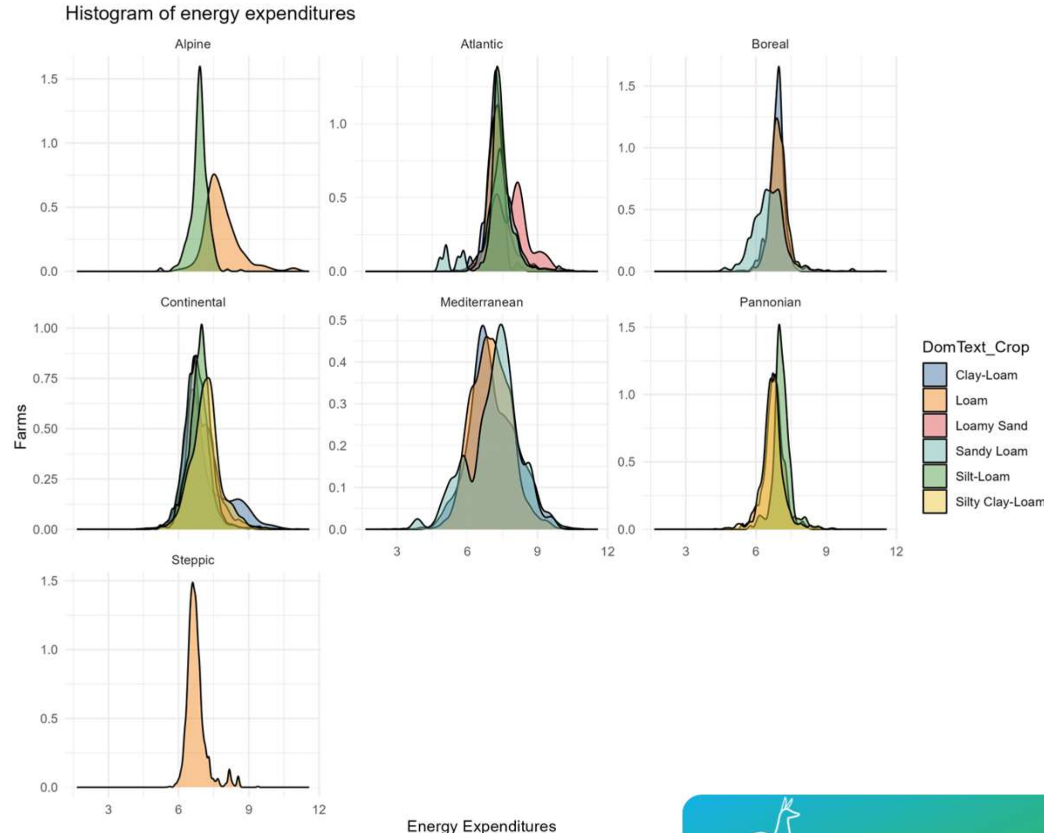


BRIGHT SPACE

LAMASUS Land Management for Sustainability

Explorative analyisis – biogeo regions, dominant texture crop, arable farms





BRIGHT SPACE

LAMASUS Land Management for Sustainability

Explorative data analysis - Summary

- No clear pattern visible
- Variables appear log-normal distributed
- Multiple modes only in few cases

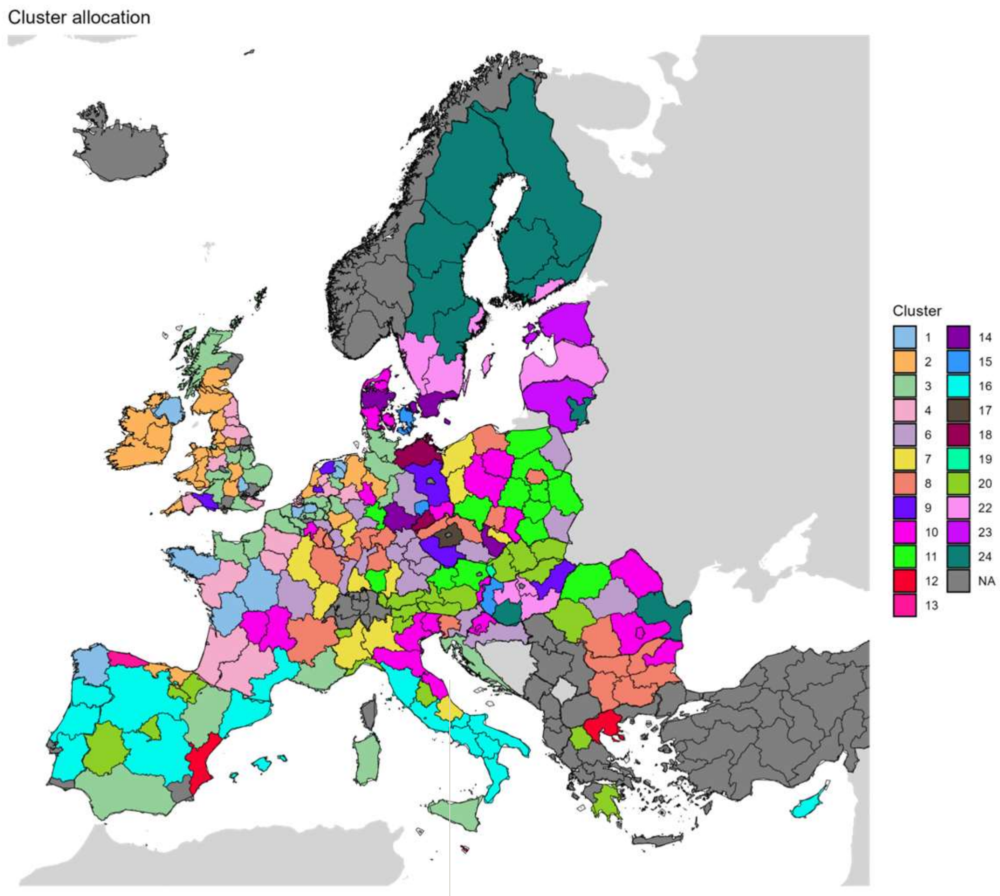
- Solution: Cluster analysis

Clustering techniques

- Permit the identification of similar groups within large number of variables
- Inclusion of categorical variables (e.g. bio-geo regions) using dummy variables (1-0)
- Based on minimization of distances to
  - a central point (e.g. k-means)
  - each other (e.g. hierarchical clustering)
- Number of clusters set by researcher (no specific rule)
- Stepwise reduction of cluster number

Example clustering result

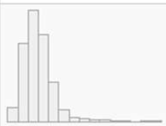
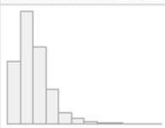
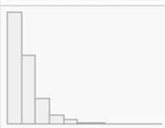
- Hierarchical clustering
- 27 possible clusters
- 22 dominant per NUTS2 region



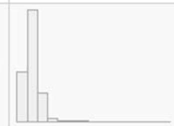
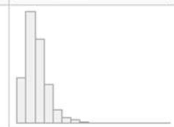
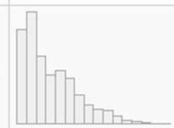
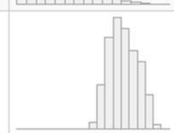
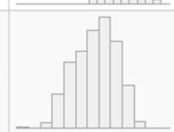
18

Example clusters 7 & 8

Cluster 7: Continental, all grassland

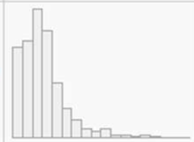
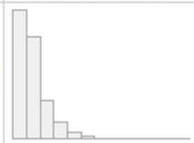
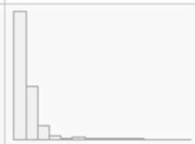
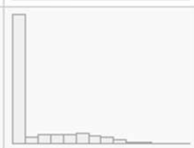
Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
BioReg_crop [character]	1. Continental	3281 (100.0%)		3281 (100.0%)	0 (0.0%)
ls_density [numeric]	Mean (sd) : 1.5 (0.9) min ≤ med ≤ max: 0.1 ≤ 1.4 ≤ 7.5 IQR (CV) : 0.9 (0.6)	3267 distinct values		3281 (100.0%)	0 (0.0%)
energy_expenditures [numeric]	Mean (sd) : 2046.9 (1292.9) min ≤ med ≤ max: 56 ≤ 1814 ≤ 11789.4 IQR (CV) : 1509.4 (0.6)	3281 distinct values		3281 (100.0%)	0 (0.0%)
SE005 [numeric]	Mean (sd) : 128.6 (108.1) min ≤ med ≤ max: 6.7 ≤ 99.1 ≤ 1085.8 IQR (CV) : 122 (0.8)	3245 distinct values		3281 (100.0%)	0 (0.0%)
grass [numeric]	Mean (sd) : 1 (0.1) min ≤ med ≤ max: 0.8 ≤ 1 ≤ 1 IQR (CV) : 0.1 (0.1)	1096 distinct values		3281 (100.0%)	0 (0.0%)
arable [numeric]	Mean (sd) : 0 (0.1) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.2 IQR (CV) : 0.1 (1.5)	1079 distinct values		3281 (100.0%)	0 (0.0%)

Cluster 8: Continental, mixed grass and cropland

Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
BioReg_crop [character]	1. Continental	5450 (100.0%)		5450 (100.0%)	0 (0.0%)
ls_density [numeric]	Mean (sd) : 1.5 (0.9) min ≤ med ≤ max: 0.2 ≤ 1.4 ≤ 14.3 IQR (CV) : 0.8 (0.6)	5443 distinct values		5450 (100.0%)	0 (0.0%)
energy_expenditures [numeric]	Mean (sd) : 2211.8 (1385.2) min ≤ med ≤ max: 103.8 ≤ 1958.9 ≤ 16163.8 IQR (CV) : 1540 (0.6)	5450 distinct values		5450 (100.0%)	0 (0.0%)
SE005 [numeric]	Mean (sd) : 175.8 (138.2) min ≤ med ≤ max: 7.6 ≤ 135.9 ≤ 769.7 IQR (CV) : 194.6 (0.8)	5422 distinct values		5450 (100.0%)	0 (0.0%)
grass [numeric]	Mean (sd) : 0.7 (0.1) min ≤ med ≤ max: 0 ≤ 0.7 ≤ 0.9 IQR (CV) : 0.1 (0.1)	4427 distinct values		5450 (100.0%)	0 (0.0%)
arable [numeric]	Mean (sd) : 0.3 (0.1) min ≤ med ≤ max: 0 ≤ 0.3 ≤ 0.6 IQR (CV) : 0.1 (0.3)	4434 distinct values		5450 (100.0%)	0 (0.0%)

Example cluster 20

Cluster 20: Several biogeo regions, dominantly alpine

Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
BioReg_crop [character]	1. Alpine 2. Mediterranean 3. Pannonian 4. Steppic	2199 (76.6%) 487 (17.0%) 115 (4.0%) 68 (2.4%)		2869 (100.0%)	0 (0.0%)
ls_density [numeric]	Mean (sd) : 1.6 (1.2) min ≤ med ≤ max: 0.1 ≤ 1.4 ≤ 8.7 IQR (CV) : 1.3 (0.8)	2864 distinct values		2869 (100.0%)	0 (0.0%)
energy_expenditures [numeric]	Mean (sd) : 2805.8 (2287.1) min ≤ med ≤ max: 52 ≤ 2321.9 ≤ 24375.8 IQR (CV) : 2594.6 (0.8)	2869 distinct values		2869 (100.0%)	0 (0.0%)
SE005 [numeric]	Mean (sd) : 128 (168.5) min ≤ med ≤ max: 7.8 ≤ 77.4 ≤ 1420.3 IQR (CV) : 90.2 (1.3)	2848 distinct values		2869 (100.0%)	0 (0.0%)
grass [numeric]	Mean (sd) : 0.9 (0.1) min ≤ med ≤ max: 0.3 ≤ 1 ≤ 1 IQR (CV) : 0.2 (0.2)	862 distinct values		2869 (100.0%)	0 (0.0%)
arable [numeric]	Mean (sd) : 0.1 (0.1) min ≤ med ≤ max: 0 ≤ 0 ≤ 0.7 IQR (CV) : 0.2 (1.6)	791 distinct values		2869 (100.0%)	0 (0.0%)

Summary

- Need for a farm typology for EU beyond TF14 for modelling purposes
  - Analysis of policy and technology/management scenarios

- Spatial data were combined with FADN to derive a farm typology that takes farm location into account
  - Combination was done at NUTS2 level
  - Non-agricultural areas were excluded

- Hierarchical clustering with step-wise reduction of cluster numbers
- 22 Clusters for dairy farms in the EU were identified

Next steps

- Refine clustering
- Apply method to arable farms
- Ensure that cluster farms can be recognizable by EU member states farming experts
- Add additional data (spatial, FADN variables)

Remaining questions

- Are we selecting the right variables from FADN?
- Which additional spatial data should be included?
  - Rainfall? Average temperature?
  - Share of area equipped for irrigation?
- Why else would apparently similar farmers find different optimal cropping plans?

BRIGHT SPACE

LAMASUS
Land Management for Sustainability

www.brightspace-project.eu | www.brightspace-horizon.eu

www.lamasus.eu

Funded by the European Union

BrightSpace Horizon Europe Grant Agreement No 101060075 | LAMASUS Horizon Europe Grant Agreement No 101060423

Views are of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.