A Reproducible Multiscale Workflow for Socioecological Indicator Calculation in European Agriculture

Alan Bernardo Palacio[1][0000-0002-7868-3601], Antonio Espinosa[1][0000-0002-6460-3789], and Joan Marull[2][0000-0002-7793-4165]

[1] Universitat Autònoma de Barcelona, Barcelona, Spain
alan.palacio@autonoma.cat, antoniomiguel.espinosa@uab.cat
[2] Minuartia, Barcelona, Spain
joan.marull@minuartia.com

Abstract. We present a reproducible computational workflow that transforms official European statistics (Eurostat), CORINE Land Cover, and Copernicus High Resolution Layers into spatially explicit sustainability indicators for any European NUTS2 region. The workflow performs three computational steps: (i) mass-balance-preserving dasymetric downscaling from NUTS2 regions to H3 hexagons ($\approx 0.74\,\text{km}^2$), enabling re-aggregation to arbitrary administrative boundaries; (ii) computation of socio-metabolic indicators—Energy Return on Investment (EROI), Energy-Landscape Integration Assessment (ELIA), and greenhouse gas (GHG) emissions—with explicit, configurable coefficients; and (iii) Shapley-value decomposition attributing observed indicator changes to land-area, intensity, and composition factors. Applied to three NUTS2 regions (NL23 Flevoland, ES53 Illes Balears, ITF2 Molise) over 2010–2020, composition shifts consistently dominate observed sustainability transitions (57–90%), while land-area and intensity changes together account for less than half. The workflow is implemented in Python and all coefficients are versioned, enabling transparent cross-regional comparison and independent replication.

Keywords: Multiscale modelling · Shapley decomposition · Dasymetric downscaling · Sustainability indicators · European agriculture

1 Introduction

Agricultural sustainability assessment requires methods that compare environmental performance across diverse regions using consistent indicators. In practice, however, assessments rely on partial metrics (e.g., yield per hectare, emissions intensity) or region-specific narratives that make cross-regional comparison difficult [1]. Decision-makers need transparent, reproducible tools that quantify sustainability consequences using well-accepted indicators across different agricultural systems.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

This paper presents a computational workflow that makes sustainability indicators *comparable, auditable, and reproducible* across European regions. The workflow addresses a specific research question: *How do socioecological sustainability indicators evolve across regions, and what factors drive the observed transitions?*

Our contributions are:

1. A *multiscale downscaling pipeline* that allocates NUTS2-level statistics to H3 hexagons ($\approx 0.74\,\text{km}^2$) using dasymetric weights, preserving mass balance by construction and enabling re-aggregation to arbitrary boundaries (municipalities, watersheds, custom zones).
2. An *integrated indicator computation module* producing socio-metabolic indicators (EROI, ELIA), GHG emissions, and nutrient balances, with all coefficients explicit and configurable.
3. A *Shapley-based attribution pipeline* that decomposes observed indicator changes into land-area, intensity, and composition contributions with axiomatic fairness guarantees.

2 Related Work

Existing frameworks each cover part of this space: agri-economic models (CAPRI [2], FSSIM [3]) capture policy responses at coarse resolution; land-use platforms (LUISA [4], CLUE-S [5]) model transitions without metabolic accounting; ecosystem-service tools (InVEST [6]) evaluate endpoints post-scenario without attribution; socio-metabolic frameworks (MuSIASEM [7], HANPP [8]) offer fund–flow theory but lack spatialisation and decomposition pipelines. No existing workflow combines mass-balance-preserving multiscale downscaling, socio-metabolic indicators, and Shapley attribution in a single auditable pipeline.

3 Data and Study Areas

We evaluate three NUTS2 regions—the Nomenclature of Territorial Units for Statistics level 2, a standardised European classification typically covering 3,000–15,000 km^2 [9]. NUTS2 is the finest level at which Eurostat systematically reports crop areas, livestock numbers, and production volumes. These three regions were specifically chosen to test the workflow's robustness across a vast gradient of European geographic and agricultural contexts: *NL23 Flevoland* (central Netherlands) represents flat, highly intensive arable and dairy farming on reclaimed coastal polders; *ES53 Illes Balears* (western Mediterranean, Spain) provides a closed island ecosystem under severe water stress and tourism-driven land-use pressure; and *ITF2 Molise* (southern Italy) serves as a case study for rugged, mountainous topography with lower-intensity traditional cereal cultivation and sheep grazing.

We analyse 2010–2020 (11 years). Land cover uses CORINE Land Cover (CLC) [10] epochs with linear interpolation between available years. Eurostat [9]

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

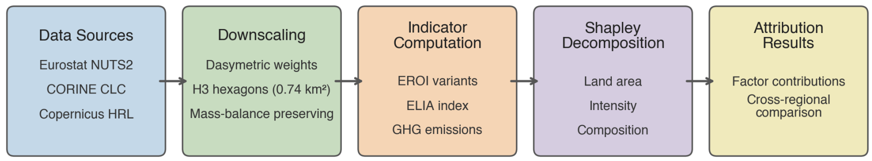

Fig. 1. Computational workflow. Official statistics (Eurostat), land cover (CORINE CLC), and high-resolution layers (Copernicus HRL) are harmonised, downscaled to H3 hexagons preserving mass balance, used to compute socio-metabolic and emissions indicators, and decomposed via Shapley values into land, intensity, and composition contributions.

provides annual agricultural statistics via datasets `agr_r_animal`, `apro_cpsh1`, and `aact_eaa01`. For 2017 onwards, we incorporate Copernicus High Resolution Layers (HRL) [11, 12] at 10 m resolution as supplementary allocation weights for downscaling.

4 Methods

Figure 1 shows the workflow. Data from Eurostat, CLC, and HRL are harmonised, downscaled to hexagonal cells, used to compute indicators, and finally decomposed via Shapley values to attribute observed changes.

4.1 Multiscale dasymetric downscaling

Official statistics are reported at NUTS2 scale, too coarse to identify sub-regional patterns. To enable finer analysis while preserving official totals, we use *dasymetric mapping* [14]—a spatial disaggregation technique that distributes aggregate quantities to finer zones using ancillary data as allocation weights. Each NUTS2 total X_r for quantity type c (e.g., wheat area, cattle count) is allocated to H3 hexagons [13] (resolution 8, $\approx 0.74\,\mathrm{km}^2/\mathrm{cell}$):

$$w_{h,c} = \frac{\mathrm{suit}_{h,c}\,(1 - \mathrm{builtup}_h)}{\sum_{h' \in r} \mathrm{suit}_{h',c}\,(1 - \mathrm{builtup}_{h'})}, \qquad \hat{x}_h = w_{h,c} \cdot X_r \tag{1}$$

where $\mathrm{suit}_{h,c} \in [0, 1]$ is the suitability of hexagon h for quantity c, derived from CLC land-cover classes (e.g., arable land is suitable for crops, pasture for livestock), and $\mathrm{builtup}_h \in [0, 1]$ is the built-up fraction from the Global Human Settlement Layer [15], which excludes urban areas from agricultural allocation. By construction $\sum_{h \in r} \hat{x}_h = X_r$, preserving official totals exactly.

This step enables re-aggregation to any spatial boundary (municipalities, watersheds, custom policy zones) while maintaining consistency with NUTS2 statistics. It also reveals intra-regional variation: within NL23, hexagon-level EROI ranges from 0.32 to 0.61, identifying hotspots invisible in the aggregate value of 0.45.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

4.2 Indicator computation

Following IPCC tiering principles [16], we distinguish Tier 1 indicators (accounting-based, directly derived from data) from Tier 2 indicators (coefficient-based proxies). All GHG values are CO_2-equivalents using AR5 Global Warming Potentials (CH_4=28, N_2O=265). LSU (Livestock Standard Units) follows Eurostat coefficients [9].

EROI. Energy Return on Investment measures energy yield relative to energy invested [17]. We compute: $\mathrm{EROI} = E_{\mathrm{out}}/(E_{\mathrm{ext}} + E_{\mathrm{int}})$, where E_{out} is metabolisable energy in products (MJ/yr), E_{ext} represents external fossil/industrial inputs, and E_{int} represents internally cycled biomass. Values below 1 are typical in fossil-subsidised agriculture (Tier 1).

ELIA. The Energy-Landscape Integrated Analysis index [18, 1] integrates energy storage E, metabolic organisation I (efficiency of energy flows), and landscape structure L (spatial heterogeneity of land cover):

$$\mathrm{ELIA} = \left(\frac{E \cdot I \cdot L}{\kappa}\right)^{1/3} \tag{2}$$

where κ is a normalisation constant calibrated for cross-regional comparability (Tier 1).

GHG. Total emissions combine livestock sources (enteric fermentation, manure management) and land-based sources (fertiliser, soil N_2O, machinery):

$$\mathrm{GHG}_{\mathrm{total}} = \sum_i N_i \cdot g_i + \sum_j A_j \cdot g_j \tag{3}$$

where N_i is the head count of livestock species i, g_i is the emission factor per head (kg CO_2e/head/yr, IPCC Tier 1 defaults), A_j is the area (ha) of land-use class j, and g_j is the emission factor per hectare (kg CO_2e/ha/yr, Tier 2).

4.3 Shapley decomposition of indicator changes

To attribute observed indicator changes to underlying drivers, we apply Shapley-value decomposition [19]. Any aggregate indicator I (GHG, energy) is expressed as a function of three factors, chosen as an *accounting factorisation* of the indicator identity: land area (A, total agricultural ha), intensity (f, yield or stocking density per ha in LSU), and composition (c, species mix). This is a deliberate modelling choice for attribution, not a claim that these are the only causal forces or that they are statistically independent; latent drivers act through changes in A, f, c, or remain outside the model scope.

For an observed change $\Delta I = I_{t_1} - I_{t_0}$, the Shapley value of factor k is:

$$C_k = \frac{1}{3!} \sum_{\pi \in \Pi} \left[I(S_\pi^k \cup \{k\}) - I(S_\pi^k) \right] \tag{4}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

where Π is the set of all orderings, S_π^k the set of factors applied before k in ordering π, and $k \in \{\text{land}, \text{intensity}, \text{composition}\}$. Each of the $2^3 = 8$ counterfactual states (e.g., $A_{2020}, f_{2010}, c_{2010}$) is evaluated by recomputing the full indicator model. By construction: $C_{\text{land}} + C_{\text{intensity}} + C_{\text{composition}} = \Delta I$, providing symmetric, axiomatically fair attribution [19].

4.4 Uncertainty and implementation

CORINE achieves ≥85% thematic accuracy. Downscaled hexagon values are estimates; aggregate errors approach zero by construction at NUTS2 scale. Tier 2 emission coefficients carry ±30% uncertainty [16]; Tier 1 indicators (EROI, ELIA) support reliable cross-regional comparisons, while Tier 2 outputs (GHG) should be interpreted directionally. Under ±30% coefficient perturbation, composition remains the largest Shapley contributor in all regions. The workflow is implemented in Python 3.11 (GeoPandas, H3-py). All coefficients are stored in versioned YAML files.

5 Results

5.1 Spatial discretisation

Figure 2 shows the downscaling result. NL23's intensive arable system produces high EROI values (green), while ITF2 and ES53 show greater heterogeneity reflecting their mixed agricultural systems.

5.2 Shapley decomposition across three regions

Table 1 presents the three-level Shapley decomposition for all regions. In every case, composition is the dominant contributor, explaining 57–90% of observed GHG changes and 57–84% of energy changes. Land-area contributions are small (1–11%), reflecting the stability of European agricultural land over this decade. Intensity contributes a moderate share (8–28% for GHG, 13–41% for energy), largest in ITF2 where yield changes accompanied the composition shift.

The pattern is consistent but not uniform. ES53 shows the strongest composition dominance (≥84%), ITF2 the weakest for energy (57%) where yield changes also contributed. In NL23, dairy cattle (58% in 2010) shifted toward sheep/goats (17% by 2020); this compositional change—not land loss or intensity reduction—explains most GHG decline. ES53 crop data are partially missing in Eurostat; its results primarily reflect livestock dynamics.

6 Discussion and Conclusions

We presented a reproducible multiscale workflow—dasymetric downscaling, indicator computation, and Shapley decomposition—for computing and comparing socioecological sustainability indicators across European NUTS2 regions.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

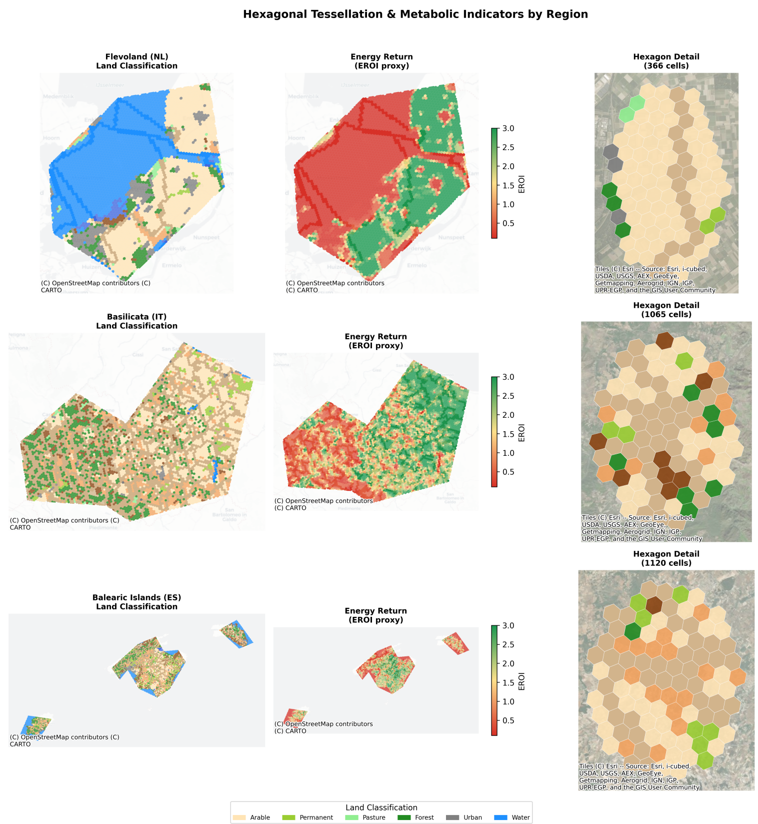

Fig. 2. Hexagonal tessellation for three NUTS2 regions. Left: dominant CLC land class. Centre: EROI proxy (green = higher). Right: hexagon close-ups over satellite imagery.

Table 1. Three-level Shapley decomposition of observed 2010–2020 changes for all three NUTS2 regions. Values are percentage contributions to the total observed change in each indicator; rows sum to 100%.

Region	Indicator	C_{land} (%)	$C_{\text{int.}}$ (%)	$C_{\text{comp.}}$ (%)	Dominant
NL23 Flevoland	GHG	11.1	13.2	75.7	composition
	Energy	15.0	20.3	64.7	composition
ES53 Illes Balears	GHG	1.5	8.2	90.3	composition
	Energy	2.1	13.8	84.1	composition
ITF2 Molise	GHG	2.3	27.8	69.8	composition
	Energy	3.0	40.6	56.5	composition

Composition shifts explain 57–90% of observed GHG transitions across three contrasting regions. Current CAP eco-schemes primarily target land manage-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

ment and area payments, which account for a minority of observed variation. Effective decarbonisation may require support for *species transition*—helping farmers shift production mix. The framework does not model market demand, price responses, or trade adjustments.

The multiscale step enables sub-regional analysis (municipalities, watersheds) while preserving consistency with official statistics. Hexagon-level resolution reveals intra-regional patterns invisible in aggregate data.

GHG accounting is territorial; trade-displaced emissions are not captured. Tier 2 coefficients carry ±30% uncertainty [16]; composition dominance may not generalise to regions with rapid land conversion. Future work will focus on verification against independent data sources, Monte Carlo sensitivity analysis, and extending the framework to prospective scenarios.

Acknowledgments. This work has been granted by the Ministerio de Ciencia e Innovación MCIN AEI/10.13039/501100011033 under contract PID2023-146193OB-100.

Disclosure of Interests. The authors declare no competing interests.

References

1. Marull, J., Font, C., Padró, R., Tello, E., Panazzolo, A.: Assessing landscapes socio-ecological functionality. *Land Use Policy* **131**, 106558 (2023)
2. Britz, W., Witzke, P. (eds.): CAPRI Model Documentation. https://www.capri-model.org/
3. Louhichi, K., et al.: FSSIM, a bio-economic farm model for EU farming systems. *Agricultural Systems* **108**, 1–21 (2012)
4. Lavalle, C., et al.: A high resolution land use/cover modelling framework for Europe. LNCS **6782**, 60–75. Springer (2011)
5. Verburg, P.H., et al.: Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management* **30**, 391–405 (2002)
6. Sharp, R., et al.: InVEST User Guide / The Natural Capital Project. https://naturalcapitalproject.stanford.edu/software/invest
7. Giampietro, M., Mayumi, K., Sorman, A.H.: *The Metabolic Pattern of Societies*. Routledge (2012)
8. Haberl, H., et al.: Quantifying and mapping the human appropriation of net primary production. *PNAS* **104**(31), 12942–12947 (2007)
9. Eurostat: Eurostat database (agriculture and regional statistics). https://ec.europa.eu/eurostat
10. Copernicus Land Monitoring Service: CORINE Land Cover (CLC). https://land.copernicus.eu/pan-european/corine-land-cover
11. Copernicus Land Monitoring Service: High Resolution Layer – Grassland. https://land.copernicus.eu/en/products/high-resolution-layer-grassland
12. Copernicus Land Monitoring Service: High Resolution Layer – Imperviousness. https://land.copernicus.eu/en/products/high-resolution-layer-imperviousness
13. Uber Technologies: H3: Hexagonal Hierarchical Geospatial Indexing System. https://h3geo.org/

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6

14. Eicher, C.L., Brewer, C.A.: Dasymetric mapping and areal interpolation. *Cart. Geogr. Inf. Sci.* **28**(2), 125–138 (2001)
15. European Commission, JRC: Global Human Settlement Layer (GHSL). https://ghsl.jrc.ec.europa.eu/
16. IPCC: 2019 Refinement to the 2006 IPCC Guidelines for National GHG Inventories (2019). https://www.ipcc-nggip.iges.or.jp/public/2019rf/
17. Hall, C.A.S., Balogh, S., Murphy, D.J.R.: What is the minimum EROI that a sustainable society must have? *Energies* **2**(1), 25–47 (2009)
18. Padró, R., Marull, J., Tello, E., et al.: Towards an energy-landscape integrated analysis? *Landscape Urban Plan.* **203**, 103905 (2020)
19. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games, Vol. II, pp. 307–317. Princeton UP (1953)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_6