Abstract
EMOS Master thesis competition 2023

‘Integration of land use vector data from administrative sources for agri-environmental analysis’

Author: Pietro Macedoni, University of Bologna

Keywords: Open IACS, land-use/land-cover, plant protection product

1. Introduction

In georeferenced studies, a common problem is having information layers where the spatial and thematic resolution is low (for instance, the Corine land cover database [1]). This leads to results that may not be accurate due to the little information available and provide not enough support to policymakers. Overcoming this matter is of big relevance, especially in the agricultural world, where new interest has been put into evaluating the impact of farming on the local environment. Developing new and more robust methodologies for estimating the effects of agriculture on natural resources is crucial to improve the farmers’ management practices.

The use of administrative data as auxiliary information can aid this process, especially in the European community, where, due to CAP requirements [2], the subsiding agencies have detailed datasets. The Open IACS project [3] stems from the necessity of having open, uniform, and comprehensive data, available for the researchers.

In this context, the thesis proposes a basic footprint in using administrative data for studying the relationship between land use and water quality measures.

2. Objective

This thesis proposes a methodology for generating a land use map with a very high geometric, thematic and temporal resolution, especially for agricultural land use. Then, it aims at understanding the relationships between specific chemicals traced in surface water and the type of land use located around the water monitoring stations.

I employed geospatial administrative data from European agricultural subsiding agencies to produce an improved land-use map compared to the ordinary land cover/use map available at the National/European level (e.g., Corine Land Cover [1]). The methodology is complemented by the integration of the land-use map with the georeferenced water monitoring stations.

Assessing the impact that different land uses have on water quality can support the definition of sustainable management practices for agricultural production as well as evaluate the performance of policies and directives.

3. Methods

The study was conducted on a large water basin located in the Foggia Province (Apulia region), South of Italy (*Figure 1*)

This work uses the geo-referenced data from the Italian Subsiding Agency (AGEA), the body managing the CAP agricultural subsidies, and from the Italian Institute for Environmental Protection and Research (ISPRA). Three land-use vector layers, from AGEA, with polygon geometry were acquired and processed: Land Parcel Identification System—LPIS (2016) [4]; Geo-spatial Aid Application—GSAA (2018); Gis Soil (2018).

The database containing the georeferenced surface water monitoring stations was acquired from ISPRA (2016) [6].

All the vector layers were converted to the common geographic reference system WSG 84/32N.

LPIS:

The layer is the land-use/cover map created through photointerpretation of very high resolution imagery (20 cm) carried out with a three-year cycle to cover the whole Italian territory [4]. The data is structured in polygons associated with information such as a numerical identifier and a generic type of land use/cover (e.g., arable land, permanent crops, forests, urban areas). In some cases, the polygons are classified with detailed land use codes (e.g., vine instead of permanent crop) through the integration of the photo interpreted information with ancillary data such as farms' data and field checks.

Geo-Spatial Aid Application (GSAA):

The GSAA vector layer includes only the agricultural areas digitized annually by the Italian farms during the administrative procedures for requesting the CAP agricultural subsidies. The thematic resolution of the layer is very high since it reports, for each cultivated parcel, the crops (wheat, vines, etc.), the intended use (forage, industry, etc.) and quality.

GIS Soil:

The layer can be described as an "intersection" of the layer of the cadastral parcels with the LPIS. GIS Soil is constantly updated by AGEA using its own administrative procedures, such as objective checks on land-use declarations or reviews provided by farms.

Water Monitoring Stations:

The dataset contains the location of the surface water monitoring stations for the period 2015–2016 and the relative tables with the typology and average amount of chemicals. In the study area, there are 26 survey stations unevenly distributed with some clustering in specific areas.

Among the chemical substances traced, we analyzed the presence of isoproturon (CAS 34123-59-6) [6], a plant protection product used as herbicide in agriculture.

Integration of the Three Land-Use Vector Layers: The Hybrid Layer:

We performed the integration of the three land-use layers to generate a very high-resolution map with an improved geometric, thematic, and temporal resolution compared to the original layers. Before performing the spatial intersection of the vector layers, harmonization of the spatial reference system and geometric and topological check were applied to the original datasets. This process was very challenging due to the nature of the single layers that were produced at different stages of the administrative process, by different actors and with different standards, procedures, and quality controls.

After the pre-processing phase, the three layers were intersected in this order: LPIS-Gis Soil-GSAA. It should be noted that the areas not covered by GIS Soil are mostly roads, city buildings and natural areas that are quite stable during the years. The last step in the generation process is a check on the combination of the two different land-use codes to resolve possible conflicts.
The result is a new hybrid layer with the highest possible thematic and spatial resolution due to the specificity of the GSAA code system.

Assigning Concentration Values to the Hybrid Layer:
The frequencies for the levels of concentration of isoproturon 0.05, 0.1, 0.15, 0.17, 0.2 in the 26 monitoring stations are 1, 1, 3, 1 and 20, respectively.
To extend the observed values to the surrounding area, a buffer of 5 km was created around each survey station [7]. The underlying assumptions are that the concentration levels are homogeneous in each circle and the observed values are due to the land use within the buffer. The polygons intersecting buffer areas were selected, rasterized, and assigned the respective concentration level.
The resulting land uses were aggregated from 735 classes to 630, and after adding a minimum of 50,000 observations per land use, a total of 13,584,287 points were kept (over the initial 15,772,668, 86.1%) for 25 land uses.
The concentration can assume six values that do not seem to be originated from a continuous space but are more likely to be rounded. For this reason, they are considered as an ordered categorical variable and not a continuous one.

4. Results

The results from the elaboration process were put in a 21 × 6 table analysed through a correspondence analysis [8,9] in R software [10] with the 'ca' package.
The inertia decomposition resulted in 5 dimensions of which the first two are considered since they explain 70.5% of the total inertia.
Table 1 is an extract of the summary for the Rows (land use) while *Table 2* is the summary for the columns (concentration of isoproturon).
Starting the analysis for the rows (*Table 1*), we observe that the coordinates for many land uses are clustered in the origin of the axes; therefore, they are close to the average profile. The quality parameter shows that large lakes and water basins, Beans EFA (Ecological Focus Area), polyphite pastures and unspecified tree crops are well represented. Considering now the EFA-type areas, we have three couples of land uses with this specification. While some have low quality levels, their pairs are spaced apart. This shows that the establishment of this type of area could lead to a change in the use of plant protection products (isoproturon).
An unusual behavior can be observed for large lakes and water basins, since it is the furthest from all points and the best represented. It is also associated with the highest inertia. This distance confirms the hypothesis that large water basins have a different behavior than other land uses regarding the concentration of isoproturon (and probably other chemicals).
As regards permanent trees, the behavior of specialized tree cultivations is similar to the behavior of vineyards. On the other hand, olive trees are located more distant and closer to the origin of the axes. Considering the columns in *Table 3*, we observe that the points are distant from each other and do not present a specific pattern. The quality levels are different; the higher concentrations are better represented (probably given the greater sample size) as opposed to the lower ones. It is noted that the highest contraction value is close to the origin of the axes, close to the average profile. In fact, studying the absolute contribution, the level 0.15 contributes the most to the first dimension, while 0.17 does the same for the second.

5. Contribution

The illustrated approach shows a clear path for employing geo-referenced administrative data in the research field. The drawback caused by their generating process (such as topological error, or slivering polygons) and the complexity of having large datasets can be fixed with time and computational power. Nevertheless, the improvement in the geometry of the polygon, and in the thematic/temporal resolution provides a better option to coarse land-use/cover maps available (e.g., Corine Land Cover with a minimum mapping unit of 25 hectares [1]).

The study uncovered methodological and data source limitations that are being addressed in current research. The hydrographic basins can be improved through the identification of the secondary basins; also, considering the terrain slope and the precipitation pattern as additional variables, it would be possible to improve the understanding of the chemical runoff and infiltration processes. The estimation process around the monitoring station can be refined, so that the buffer and assumption of homogeneity can be removed in favor of a more complex methodology (for example, kriging).

Despite the incurred limitations, it is important to refine the research process to provide better tools to foster the reduction of chemicals products in the agricultural process, to maintain the biodiversity and to reduce the environmental impacts.

References

[1] CORINE Land Cover. (s.d.). Available online: https://land.copernicus.eu/pan-european/corine-land-cover (accessed on 29 February 2020).

[2] European Commission. Available online: https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agriculturalpolicy/cap-glance_en (accessed on 29 February 2020).

[3] Open IACS. website: https://open-iacs.eu/

[4] SIAN. Aggiornamento Refresh 4 ciclo Specifiche Tecniche di Rilevazione. Available online: https://www.sian.it/downloadpub/zfadlx010?id=316975 (accessed on 29 February 2020).

[5] Rapporto Nazionale Pesticide Nelle Acque, Dati 2015–2016. Available online: https://collaboration.sina.isprambiente.it/s/BWidaPsR2JMbq26 (accessed on 29 February 2020).

[6] ISPRA. Available online: https://www.isprambiente.gov.it/contentfiles/00008000/8043-isoproturon.pdf (accessed on 29 February 2020).

[7] Juan Antonio Pascual Aguilar, V.A. Pesticide occurrence in the waters of Júcar River, Spain from different farming landscapes. Sci.Total Environ. 2017, 607–608, 752–760.

[8] Greenacre, M. Correspondence Analysis in Practice, 2nd ed.; Chapman & Hall: London, UK; CRC: Boca Raton, FL, USA, 2007.

[9] Beh, E.J. Simple Correspondence Analysis: A Bibliographic Review. Int. Stat. Rev. 2004, 72, 257–284.

[10] R Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: https://www.R-project.org (accessed on 29 February 2020).

Figure 1

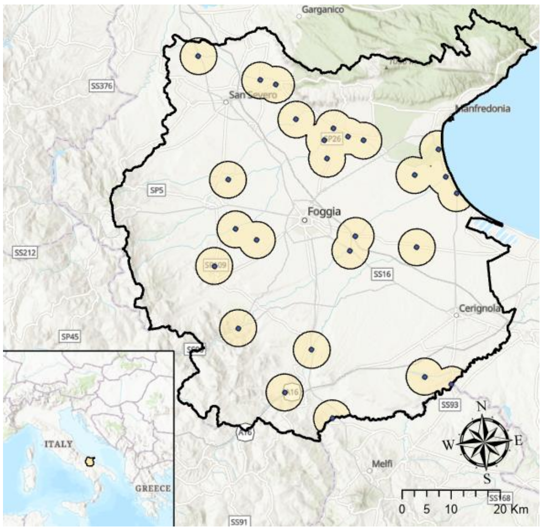
Manfredonia
Foggia
Cerignola
Melfi
ITALY
GREECE
SS376
SP5
SS212
SP45
SS16
SS93
A16
SS91
N
E
S
W
0 5 10 20 Km

Table 1

Land Use	Mass	Quality	Inertia	Dimension 1 Coordinates	Dimension 1 Squared Correlation	Dimension 1 Absolute contribution	Dimension 2 Coordinates	Dimension 2 Squared Correlation	Dimension 2 Absolute contribution
Tomato	35	416	26	-160	350	20	-70	66	7
Durum Wheat	411	144	35	9	11	1	-34	133	19
Agricultural Area Withdrawn from Production	23	41	11	41	34	1	-19	7	0
Agricultural Area Withdrawn from Production EFA	25	470	14	-163	468	15	-9	2	0
Deciduous Woods	5	373	7	-161	180	3	167	193	5
Residential Urban Fabric	4	774	8	-257	362	6	274	412	13
Isolated Buildings	5	352	2	95	209	1	79	143	1
Industrial and Commercial Areas	4	133	4	71	53	0	87	80	1
Transport Infrastructure	8	446	1	81	444	1	-5	2	0
Lakes and Water Basins of Significant Surface	5	996	363	2553	961	771	-482	34	49
Vineyard	29	504	85	176	109	21	335	395	133
Olive Trees	46	346	24	112	249	13	70	97	9
Oats	22	655	27	-156	203	12	-234	452	48
Chickpea	22	122	12	-31	19	0	73	103	5
Chickpea EFA	10	10	4	-17	7	0	-11	3	0
Beans	11	362	6	-52	48	1	-132	314	8
Beans EFA	12	830	23	-80	34	2	-391	796	73
Polifita Pasture	33	898	19	-200	714	29	101	184	14
Specialized Tree Cultivations Not Specified	7	876	8	188	337	6	238	539	16
Tree pasture - 50%	4	603	12	-331	401	10	234	202	9
Generic Building – Road	25	752	19	236	752	31	5	0	0
Seminative By Photointerpretation	151	528	115	3	0	0	198	528	240
Waters	25	67	20	20	5	0	69	62	5
Grassland	13	727	12	-254	683	19	-65	45	2
Barley	47	745	88	-118	77	15	-349	668	233
Asparagus	18	678	56	-227	173	21	-387	504	111

Table 2

				Dimension 1			Dimension2		
Concentration	Mass	Quality	Inertia	Coordinates	Squared Correlation	Absolute contribution	Coordinates	Squared Correlation	Absolute contribution
0,03	11	353	167	28	1	0	715	353	233
0,05	29	136	70	-88	33	5	-156	103	29
0,1	19	149	89	56	7	1	253	142	50
0,15	120	999	390	559	979	845	-80	20	31
0,17	47	829	221	-240	124	61	-570	704	617
0,2	774	798	62	-71	632	87	36	166	41