Proceeding Paper

Spatial Exploration of the Relationships between Agricultural Land Use and Water Quality Measures †

Pietro Macedoni [1], Elisabetta Carfagna [1] and Flavio Lupia [2,*]

1   Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Via Delle Belle Arti, 41, 40126 Bologna, Italy; pietro.macedoni@unibo.it (P.M.); elisabetta.carfagna@unibo.it (E.C.)
2   CREA—Research Centre for Agricultural Policies and Bioeconomy, Via Barberini 36, 00187 Rome, Italy
*   Correspondence: flavio.lupia@crea.gov.it
†   Presented at the 1st International Online Conference on Agriculture—Advances in Agricultural Science and Technology, 10–25 February 2022; Available online: https://iocag2022.sciforum.net/.

Abstract: Agricultural land-use effect on water quality (surface and groundwater) is a well-known issue, and actions are needed to reduce the impacts of farm inputs management. Direct and indirect links can also be found with some of the SDGs (i.e., SDG 6—Clean water and sanitation). Assessing these impacts can support the definition of sustainable management practices for agricultural production as well as evaluating the performance of policies and directives (e.g., European Common Agricultural Policy). In this work, we performed an analysis of the relationship between agricultural land use and pesticides found in surface waters located in South Italy (Foggia province—Apulia). Land-use data were produced with a complex data integration process using administrative geospatial data from the Italian agricultural paying agency. Land-use data were jointly analyzed with water quality measures on surface waters within a large watershed in the study area. A statistical analysis was carried out in order to assess the relationships between specific types of chemicals in water and land uses within a circular buffer of 5 km around each monitoring station.

Keywords: Land Parcel Identification System; LPIS; CAP; Apulia; water pollution

1. Introduction

The use of chemical agents in agriculture is a necessary practice to maintain high production levels. However, the role of policies (e.g., the European Common Agricultural Policy—CAP [1], pesticides, and water quality directives) is to foster the reduction in chemicals in the agricultural production process to maintain the biodiversity and to reduce the environmental impacts. At the same time, there is the need to collect suitable data on the agricultural inputs and to monitor the effects on the environment (soil and water). Monitoring is also a steppingstone to assess the performance and results of the policies in the agri-environmental domain.

The monitoring process should not stop at the surveying step, it needs to follow a more complex analysis to identify the origin and destination of the chemical substances in the environment. Similar works have been conducted in the Netherlands, where an atlas for surface water was created for analyzing the concentration level of pesticides, their evolution over time and how they can be linked to land uses [2]; in Spain, the dispersion behavior of pesticides was studied along the water basin of the Júcar river [3].

The aim of this work is to understand the relationships between specific chemicals traced in surface water and the typology of land use located around the water monitoring stations. We propose a methodology for the production of a land use map with a very high geometric, thematic, and temporal resolution, especially for the agricultural land use. We use geospatial administrative data from European agricultural paying agencies to produce an improved land-use map compared to the ordinary land cover/use map available at national/European level (e.g., Corine Land Cover [4]). The methodology is complemented

by the integration of the land-use map with the georeferenced water quality measures and a sound statistical analysis.

2. Materials and Methods

This work uses the geo-referenced data from the Italian Paying 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) [5]; Geo-spatial Aid Application—GSAA (2018); Gis Soil (2018).

One vector layer with point geometry containing the georeferenced surface water monitoring stations with the associated database was acquired from ISPRA (2016) [6]. All the vector layers were converted to the common geographic reference system WSG 84/32N. The area of interest is a large water basin located in the Foggia Province (Apulia region), South of Italy (Figure 1).
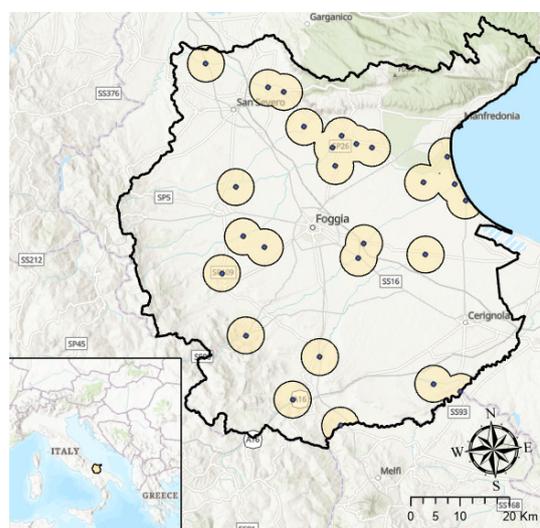
Figure 1. Study area (water basin in the Foggia province) with the surface water monitoring stations and the corresponding 5 km buffer.

2.1. 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 [5]. 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.

2.2. 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.

2.3. 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.

2.4. 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) [7], a plant protection product used as herbicide in agriculture.

2.5. 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.

2.6. 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 [3]. 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 12,960,454 points were kept (over the initial 15,772,668, 82.1%) for 21 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.

3. Results and Discussion

The results from the elaboration process were put in a 21 × 6 table analyzed through a correspondence analysis [8,9] in R software [10] with the 'ca' package. The value of the chi-square statistic is 12,77,406 with 100 degrees of freedom and *p*-value = 0; therefore, it is safe to assume some degree of association between these two categorical variables. The total inertia, computed as $\varphi^2 = \chi^2/N$, which describes the variation in the contingency table is: 0.098.

In Table 1, the level of inertia associated with each dimension is reported. The use of the first two dimensions explains 71.2% of inertia; adding the third and fourth dimensions would amount to 93.9%, but it would complicate the analysis too much. For this reason, only the first two dimensions were considered.

Table 1. Result summary for the inertia.

Dim	Principal Inertia	%	Cumulative%
1	0.04513	45.8	45.8
2	0.025015	25.4	71.2
3	0.014644	14.9	86
4	0.007776	7.9	93.9
5	0.005997	6.1	100
Total:	0.098562	100.0	

Table 2 is an extract of the summary for the Rows (land use) while Table 3 is the summary for the columns (concentration of isoproturon):

Table 2. Results summary for the first dimension.

Land Use	Mass	Quality	Inertia	Dimension 1			Dimension 2		
				Coordinates	Squared Correlation	Absolute Contribution	Coordinates	Squared Correlation	Absolute Contribution
Agricultural Area Withdrawn from Production	24	63	11	53	60	1	−11	3	0
Agricultural Area Withdrawn from Production EFA *	26	450	14	−154	450	13	−3	0	0
Large Lakes and Water Basins	5	994	385	2577	961	809	−476	33	50
Vineyard	31	536	87	188	126	24	338	410	140
Olive Trees	48	330	26	115	249	14	66	81	8
Oats	23	646	27	−149	192	11	−229	454	47
Chickpea	23	103	13	−31	18	1	67	85	4
Chickpea EFA	10	10	4	−16	6	0	−14	4	0
Beans	12	335	6	−42	33	0	−125	301	7
Beans EFA	12	809	23	−66	24	1	−379	785	71
Polifita Pasture	34	886	19	−197	705	29	100	181	14
Specialized Tree	7	887	8	194	354	6	238	533	17
Pasture with Tree 50%	4	598	12	−337	415	11	224	183	9
Seminative-Photointerp	158	565	117	10	1	0	202	564	259
Grassland	14	736	13	−253	688	19	−67	48	2
Barley	50	716	87	−104	63	12	−336	653	224
Asparagus	19	664	56	−218	166	20	−377	498	109

* EFA: Ecological Focus Area, the area of land upon which you carry out agricultural practices (e.g., fallow land, nitrogen-fixing crops) that are beneficial for the climate and the environment and to improve biodiversity.

Starting the analysis for the rows (Table 2), 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).

Table 3. Result summary for the second dimension.

Concentration	Mass	Quality	Inertia	Dimension 1			Dimension 2		
				Coordinates	Squared Correlation	Absolute Contribution	Coordinates	Squared Correlation	Absolute Contribution
0.03	12	390	168	53	2	1	748	389	258
0.05	30	126	68	−78	27	4	−150	99	27
0.1	20	175	88	81	15	3	265	160	55
0.15	118	999	396	570	980	848	−79	19	30
0.17	48	808	217	−226	115	54	−555	693	594
0.2	773	810	63	−73	661	90	35	149	37

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.

4. Conclusions

The illustrated approach shows how it is possible to use geo-referenced administrative data and integrate them with environmental datasets for understanding the impacts of the agricultural activities on the environment (water quality). Despite administrative vector layers not being designed for this kind of application, after a process of integration and data preparation, it is possible to obtain a significant improvement in the polygons geometry and in the thematic and temporal resolution. This process can be time consuming and demands a lot of computational power, but it provides a better option to coarse land-use/cover maps available (e.g., Corine Land Cover with a minimum mapping unit of 25 hectares [4]).

The conducted analysis shows how land uses differ from each other in relation to different concentrations. The used approach is a somewhat simplified approximation of reality, as it can be extended to the entire regional surface and not just to the buffer around the survey station.

The study uncovered methodological and data source limitations that can be addressed in future research. First, it would be advisable to improve the detail of the hydrographic basins. Through the identification of the secondary basins and by 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, also in relation with the farms' management practices.

Author Contributions: Conceptualization, P.M, F.L. and E.C.; methodology, P.M. and F.L.; software, P.M.; validation, P.M. and F.L.; data curation, P.M.; writing—original draft preparation, P.M. and F.L.; writing—review and editing, P.M. and E.C.; supervision, F.L.; project administration, F.L. All authors have read and agreed to the published version of the manuscript.

References

1. European Commission. Available online: https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/cap-glance_en (accessed on 29 February 2020).
2. Vijver, M.G.; Van'TZelfde, M.; Tamis, W.L.; Musters, K.J.; De Snoo, G.R. Spatial and temporal analysis of pesticides concentrations in surface water: Pesticides atlas. *J. Environ. Sci. Health Part B* **2008**, *43*, 665–674. [CrossRef] [PubMed]
3. 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. [CrossRef] [PubMed]
4. CORINE Land Cover. (s.d.). Available online: https://land.copernicus.eu/pan-european/corine-land-cover (accessed on 29 February 2020).
5. SIAN. Aggiornamento Refresh 4° ciclo Specifiche Tecniche di Rilevazione. Available online: https://www.sian.it/downloadpub/zfadlx010?id=316975 (accessed on 29 February 2020).
6. Rapporto Nazionale Pesticide Nelle Acque, Dati 2015–2016. Available online: https://collaboration.sina.isprambiente.it/s/BWidaPsR2JMbq26 (accessed on 29 February 2020).
7. ISPRA. Available online: https://www.isprambiente.gov.it/contentfiles/00008000/8043-isoproturon.pdf (accessed on 29 February 2020).
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. [CrossRef]
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).