A model for the Geographical Analysis and monitoring of agricultural areas example and tests in south Italian regions

Giuseppe Tradigo
DIMES Dept., Univ. of Calabria, Italy
Cosenza, Italy
tradigo@dimes.unical.it

Carmine Pagliaro
Surgical & Clin. Science Dept., Univ. of Catanzaro
Catanzaro, Italy
carminepagliaro@gmail.com

Giovanni Canino
Surgical & Clin. Science Dept., Univ. of Catanzaro
Catanzaro, Italy
Department of Computer Science and Engineering, State University of New York
Buffalo, USA
canino@unicz.it

Francesco Casalinuovo
Ist. Zooprofilattico
Catanzaro, Italy
casalinuovo@izsmportici.it

Caterina Graziani
Veterinary Epidemiology Unit
Department of Veterinary Public Health and Food Safety
Istituto Superiore di Sanitá
Rome, Italy
caterina.graziani@iss.it

Pierangelo Veltri
Surgical & Clin. Science Dept., Univ. of Catanzaro
Catanzaro, Italy
veltri@unicz.it

ABSTRACT

Water quality has an high impact on human health. Chemical mixtures deployed in the environment can lead to multi-level contamination, going from groundwater through surface water, crops, vegetables, livestocks and, finally, humans. Groundwater can be contaminated by the percolation of chemical compounds due to various management errors, which can be mainly re-conducted to vulnerability of the ground to pollution or wrong management of waste sites. The impact of groundwater contamination on human health is a serious problem and the risk of pathogen outbreaks must be predicted and, when possible, contained. Our study has the goal of automatically selecting high risk areas to reduce the risks of infection. We performed analyses on a dataset of farm companies containing analyses on their livestocks urine and faeces samples, water sourcing, crops samples and ground samples.

We searched for multi-pathogens in these samples and we built an epidemiological model inferring groundwater contamination from indirect evidence, considering a normalized severity score for the analyses results. With this information we selected the contamined waterways, which define potential areas of high impact on the health of people living nearby.

HealthGIS'14, November 04-07 2014 Dallas/Fort Worth, TX, USA

http://dx.doi.org/10.1145/2676629.2676639.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Data Analysys, Spatial Data

Keywords

Information Systems Applications, Geographical Data Analysis

1. INTRODUCTION

Water in agriculture is certainly a critical asset, due to its potential in transmitting contaminants and infections to crops, animals and, potentially, humans. Fresh fruits and vegetables are in contact with water during the various stages of the production process: from early stages with irrigation to the post-harvest washing operations. It has been shown how using contaminated water leads to potential health problems [1, 2, 3]. The situations where contaminated water sources affects crop production are, for example the use of contaminated water for the application of pesticides or herbicides, crop irrigation or washing in post-harvest [2]. Farm animals need to drink water of adequate quality, free from contamination. Pathogenic agents contained in water can spread rapidly where animals drink from the same pool or eat from the same tanks. In addition, animal manure is often used as fertilizer, taking into consideration only the nitrogen and a few other nutrients for crops [4]. Contaminants found in the manure produced by animals could cause an indirect contamination of waterways, soil and food

[5], hence the risks for human health increase. Bacteria contained in meat, dairy or vegetables, once ingested, colonize the intestine and adhere to the intestinal villi and microvilli causing the absorption of toxins which affect organs like colon or kidneys. Haemorrhagic colitis and haemolytic the uraemic syndrome are the main manifestations [6]. Our aim is to find a procedure and a model to correlate contaminants present in the flow of water used in agriculture with bacterial contamination found in plants and in animals. To achieve this goal we use GIS (Geographical Information System) technologies, providing an adequate set of tools for the management, storage, visualization and evaluation of geolocalized information [7, 8]. GIS tools allow the analysis of spatial data organized in layers of various types and dimensions (1d points, 2d lines and areas) able to contain information regarding complex environmental features. This offers an integrated view of environmental data which can be joined with social or clinical data, making them an ideal tool for the decision-making process. Data and information about analyses, stored and manipulated through GIS tools, can be used to develop models and study their spatial features for the evaluation of potential water contamination.

2. METHODS AND TECHNIQUES

The water analysis activity can be divided into two phases. In the first phase samples taken from the water, animals and plants are collected and stored. In the second phase we developed a model to estimate the potential contamination risk of waters.

2.1 Phase 1

We implemented a Java-based system for the storage, management and visualization of data working on cloud computational environments called SMAT (*Sistema per il Monitoraggio Ambientale e Territoriale, System for Environmental and Territorial Monitoring*). The system has a web-based interface and has been developed using the Grails framework. Data is persisted on a relational PostgreSQL database with PostGIS geographical extensions. Data storage, management and visualization can be performed by the user in two ways: the first one is through a web client, the second is through the QGis tool. As show in Figure 1, the user access data through SMAT, leveraging the resources and services offered by the Heroku platform. Data managed via the web interface can be extracted in various formats (e.g. excel, CSV).

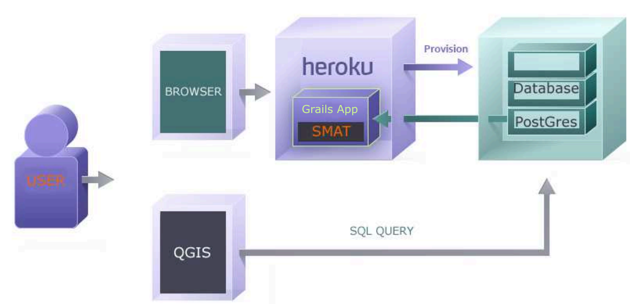

Figure 1: SMAT system architecture, two ways access data: web-client, qgis software

Finally, data is stored in a PostgreSQL[1] database instance. Data can also be accessed and manipulated via the QGIS software tool, as depicted in Figure 1, with geographical tools and manipulation primitives and functions. The core database entity/relation schema adopted is show in Figure 2. Each farm, represented by the *Company* entity, is associated with a number of *Assets*, which can be vegetables (representing coltures), crops and waters. Each Asset is in relationship with *Samples*, acquired during an *Analysis* procedure.

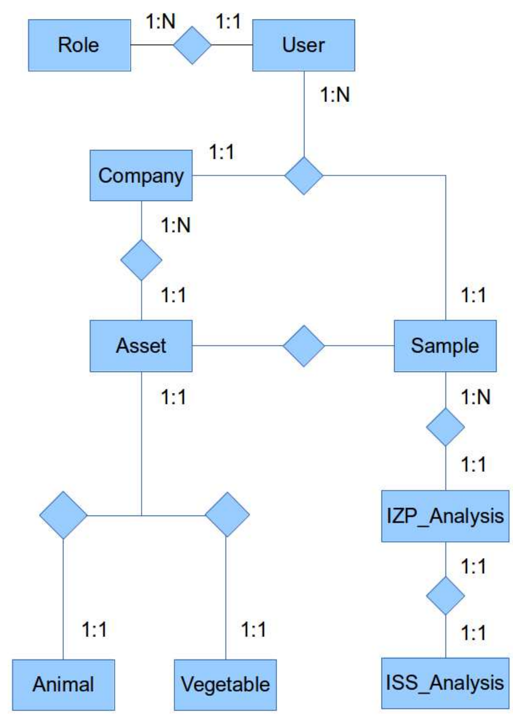

Figure 2: Database architecture

2.2 Phase 2

We have a dataset of twenty companies operating on the Calabria region (Italy), for each of which a number of analyses have been performed. Analyses have been performed on waters (for animal consumption and sewage), faeces (individual or groups) and vegetables. The goal of bacteriological analyses has been to search for pathogens like *Salmonella*, *Escherichia Coli* and others, particularly dangerous for human health. Each analysis has been associated with a geographical location and then inserted in our system.

3. EXPERIMENTAL RESULTS

A procedure written in the SQL language has been developed in order to generate a temporary table in which to store the scores and ids related to analyses of interest (i.e. containing pathogens). A scoring function has been create to measure the performance of a farm company by weighting its contamination level (i.e. analysis showing pathogenic contamination) according to the following scores: (*i*) high score (i.e. 5 points) for irrigation or drinking water, (*ii*)

[1]http://www.postgresql.org/

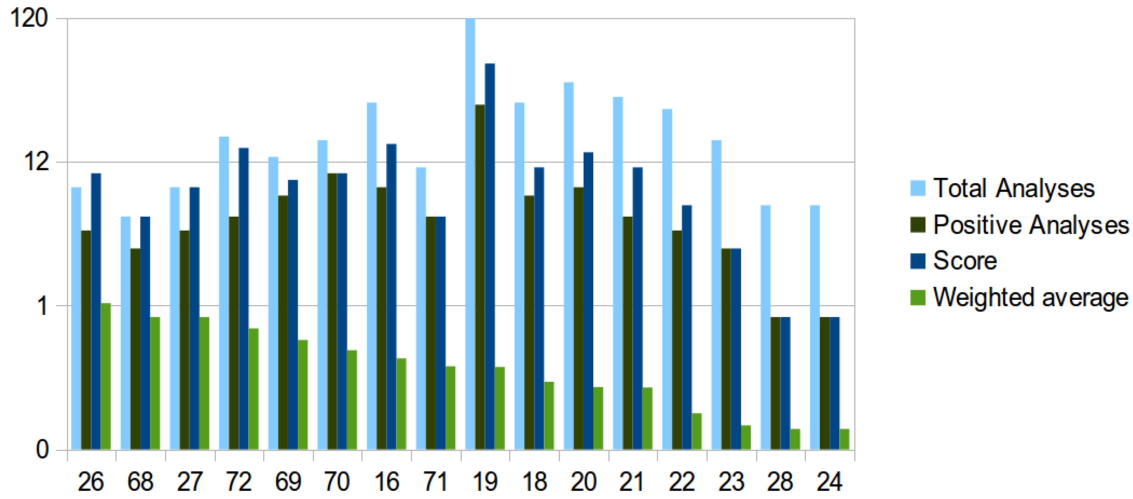

Figure 3: Weighted averaged score of the Farms involved in the analyses with logarithmic Y axis. Farm Company ids are depicted on the X axis.

medium score (i.e. 3 points) for individual faeces or vegetable analysis and (*iii*) normal score (i.e. 1 point) for other positive analyses (e.g. cumulative faeces, sewage waters). The rationale here is to penalize the most potentially dangerous results (contaminated drinking water, positive faeces analysis from single individual). The following shows the pseudo-code of the described procedures.

```
procedure scoring
begin
  int score, count, water, faeces, other;
  id company_id;
  create temporary table
    temp_score(company_id int, score int)
  Q1 = SELECT ad.id INTO water FROM company a
     JOIN asset ass ON a.id=ass.company_id
     JOIN animal an ON ass.id=an.asset_id
     JOIN sample p ON ass.id=p.asset_id
     JOIN izp_analysis aizs ON aizs.sample_id=p.id
     JOIN iss_analysis aiss ON aiss.analisiizs_id=aizs.id
     WHERE p.tipo_prelievo LIKE 'acqu%'
     AND aizs.result LIKE 'pos%'
  Q2 = SELECT ad.id INTO feces FROM company a
     JOIN asset ass ON a.id=ass.company_id
     JOIN animal an ON ass.id=an.asset_id
     JOIN sample p ON ass.id=p.asset_id
     JOIN izp_analysis aizs ON aizs.sample_id=p.id
     JOIN iss_analysis aiss ON aiss.analisiizs_id=aizs.id
     WHERE p.sample_type LIKE 'sing%'
     AND aizs.result LIKE 'pos%'
  Q3 = SELECT ad.id INTO other FROM company a
     JOIN asset ass ON a.id=ass.company_id
     JOIN animal an ON ass.id=an.asset_id
     JOIN sample p ON ass.id=p.asset_id
     JOIN izp_analysis aizs ON aizs.sample_id=p.id
     JOIN iss_analysis aiss ON aiss.analisiizs_id=aizs.id
     WHERE p.sample_type NOT LIKE 'singo%'
     AND p.sample_type NOT LIKE 'acqu%'
     AND aizs.result NOT LIKE 'pos%'
  while water > 0 INSERT INTO temp_score (water, 5) end
  while feces > 0 INSERT INTO temp_score (feces, 3) end
  while other > 0 INSERT INTO temp_score (other, 1) end
end procedure
```

The final ranking score is calculated as a weighted average of the score of each company by the total number of analyses performed, as shown in Table 1. We were able to identify farms having appreciable contamination values either in water or in the animals samples. Figure 4 depicts a visual rendering of the farms showing the highest contamination factors in our dataset. With other GIS functions we could plan a spatial prediction model in order to foresee possible spreads of contamination and areas interested by the potential epidemiological phenomenon. Furthermore we could interrogate the system and intersect or correlate with other geographic layers in order to discover new features.

Farm id	*Score*	*Weighted Average*	*Total Analyses*	*Positive Analyses*	*Crops*
26	10	1.25	8	4	100
68	5	1.00	5	3	10
27	8	1.00	8	4	50
72	15	0.83	18	5	36
69	9	0.69	13	7	200
70	10	0.59	17	10	0
16	16	0.52	31	8	750
71	5	0.45	11	5	0
19	58	0.45	129	30	346
18	11	0.35	31	7	32
20	14	0.33	43	8	750
21	11	0.32	34	5	108
22	6	0.21	28	4	63
23	3	0.18	17	3	27
28	1	0.17	6	1	30
24	1	0.17	6	1	25

Table 1: Scoring farms sorted by Weighted Average calculated from the score assigned to water, individual faeces, cumulative/sewage analyses. Zero crops means the Farm has an agricultural production only.

We have 17 total farm companies in our dataset, 16 of which (shown in Table 1) have been associated with at least one analysis showing a contaminating pathogen and 11 of which show contamination for all three subsets we identified (i.e. water, individual faeces and cumulative faeces/sewage. Only one out of the 17 companies was negative to all of the pathogenic analyses. Figure 3 shows how Farm Companies can be ranked considering the weighted average which weights the score obtained by analyzing the Companies assets, normalized by the total number of analyses. High risk

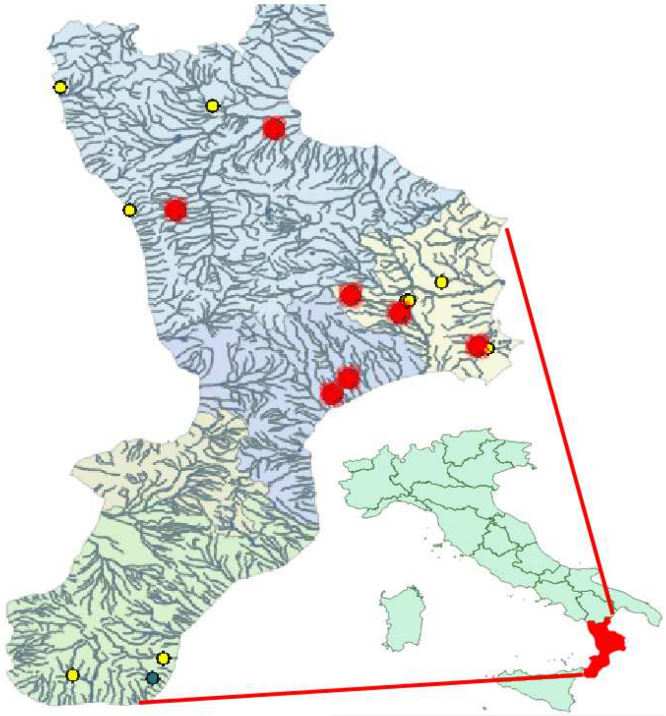

Figure 4: Farms with the highest risk of contamination are shown in red on the map.

companies having high contamination scores are depicted in Figure 4 as red markers. These were used to select the nearest waterways from the geographical layer containing waterways multi-line geometries. River basins selected can be used to build a buffer geometry to define areas with high probability of contamination risk. We then checked the official data from an agency of the Italian Ministry of Environment known as SINAnet[2] (*Rete del Sistema Informativo Nazionale Ambientale, Network of the National Environmental Informative System*). We used a geographical layer reporting data about water pollution and contamination measured by a network of fixed sensors. The areas defined by the waterways surroundings do correlate with great precision with official pollution data. According to their data, we have 6 farm companies marked by us as *high risk* matching the SINAnet areas marked as *high risk of contamination*.

4. CONCLUSIONS

This preliminary study shows how it is possible to correlate areas and risks with evidences of water contamination by analyzing crops and vegetables in farms. We had a dataset of 17 companies with crops, vegetables and water facilities over which a set of analyses have been performed. Since both vegetables and crops are tightly related to human health (e.g. though direct consumption), having a map of contamination risk is of high interest. We note that almost all of the companies we analyzed showed at least one analysis to be positive to contaminants, while the vast majority showed contamination at many levels. Data found by us are in accordance with the italian Ministry of Environment network for the automatic monitoring of contamination in waters.

[2] www.sinanet.isprambiente.it/

Acknoledgements

This work has been partially funded by the Project CCM ”Valutazione del rischio sanitario di infezioni da agenti zoonosici attraverso prodotti vegetali ottenuti in aree ad elevata pressione zootecnica” (CUP J65JI2000110001), funded by italian health ministry and Calabrian Region. Also, Carmine Pagliaro is funded by SH 2.0 PON Project funded by MIUR, while Giuseppe Tradigo is funded by PRIN GenData2020 funded by italian MIUR.

5. REFERENCES

[1] S.D. Pillai, C.H. Rambo, WATER QUALITY ASSESSMENT, Routine Techniques for Monitoring Bacterial and Viral Contaminants, *Encyclopedia of Food Microbiology (Second Edition)*, 766-772, 2014

[2] M.W. Diebel, J.T. Maxted, D.M. Robertson, S. Han, M.J.V. Zanden, Landscape Planning for Agricultural Nonpoint Source Pollution Reduction III: Assessing Phosphorus and Sediment Reduction Potential, *Environmental Management*, 43(1), 69-83, 2009

[3] R.M. Kirby, J. Bartram, R. Carr, Water in food production and processing: quantity and quality concerns , *Food Control*, 14(5), 2003

[4] J.R. Jones, J.A. Downing, Agriculture, *Reference Module in Earth Systems and Environmental Sciences Encyclopedia of Inland Waters*, 2009

[5] A. Azizullah, M.N.K. Khattak, P. Richter, Water pollution in Pakistan and its impact on public health - A review, *Environmental Modelling & Software*, 17(5), 413-425, 2002

[6] J.T. Srinivasan, V.R. Reddy, Impact of irrigation water quality on human health: A case study in India, *Ecological Economics*, 68(11), 2800-2807, 2009

[7] D.C. McKinney, X. Cai, Linking GIS and water resources management models: an object-oriented method, *Environmental Modelling & Software*, 17(5), 413-425, 2002

[8] D.Z. Sui, R.C. Maggio, Integrating GIS with hydrological modeling: practices, problems, and prospects, *Computers, Environment and Urban Systems* , 23(1), 413-425, 1999

[9] S.T.Y. Tong and W. Chen, Modeling the relationship between land use and surface water quality, *Journal of Environmental Management*, 66(4), 413-425, 2002

[10] A. Rebelo, I. Ferra, I. Goncalves, A.M. Marques, A Risk Assessment Model for Water Resources: Releases of dangerous and hazardous, *Journal of Environmental Management*, 140, 2014

[11] R.S.H. Yang, E.J. Rauckman, Toxicological studies of chemical mixtures of environmental concern at the National Toxicology Program: Health effects of groundwater contaminants, *Toxicology*, 47(1-2), 15-34, 1987

[12] D.J. Lapworth, N. Baran, M.E. Stuart, R.S. Ward, Emerging organic contaminants in groundwater: A review of sources, fate and occurrence, *Environmental Pollution*, 163, 2012