Visualization of Black Sea Water Pollution using GIS Software and Mobile Technologies

<https://doi.org/10.3991/ijxx.vx.ix.xxxx>

Julian Vasilev1, Pavel Petrov()1, Jordan Jordanov1

1 University of Economics - Varna, Varna, Bulgaria

[petrov@ue-varna.bg](mailto:example@example.org)

**Abstract—**The purpose of this paper is to analyze public data for Black Sea water pollution on Varna’s beaches. Geographic Information System are used for visualization, and statistical software is used for correlation analysis. Open-source software products are used: QGIS and PSPP. One dataset is used, and after some ETL procedures, additional analysis is conducted. The results of the correlation analysis show interesting dependencies. Other statistically significant dependencies are merely ambiguous. Similar analyses may be conducted for other seas and other periods.

**Keywords—**software engineering, water pollution, mobile applications, data analysis, data visualization

1. Introduction

Geographic information systems (GISs) are designed to collect, store, retrieve, analyze, and visualize information related to spatial data. GIS-type systems provide different functionalities for creating dynamic maps, studying relationships between spatial objects, and analyzing territorial data for modeling, forecasting, and resource optimization. Primary data usually comes from various sources, and after their processing, new information is created, and knowledge is extracted to serve in making informed and justified decisions in the context of the relevant subject area of application.

GIS-type systems improve various processes and decision-making in urban planning, transportation, environment, agriculture, ecology, and health. Hence, their application is not limited to only a narrow circle of specialists in the field of geography or geology but is useful to architects, ecologists, farmers, and household users.

Historically, in first-generation GISs, spatial data was stored in simple "flat" files without a well-defined format. Second-generation GISs store part of the data in a relational database [1, 2, 3], while third-generation GIS data is stored and processed entirely in a relational database. Thus, we can conclude that the evolution of GISs is in the direction of fully integrating spatial data with a relational database (Fig. 1).

A picture containing text, diagram, screenshot, plan

Description automatically generated

1. Evolution of GIS Architecture [4]

However, a significant problem is that the main data types in traditional relational database management systems are limited to integers and real numbers, fixed and variable length character strings, date and time, binary large objects (BLOBs), and those similar that do not fully cover the needs of GISs. They usually work with spatial objects such as points, lines, and polygons, which include multiple components for representing the various spatial characteristics of the object (Fig. 2).

A diagram of a structure

Description automatically generated

1. Basic Types of Spatial Objects and Hierarchical Dependencies [4]

Different approaches have been used to solve this problem in various relational database management systems, often by adding an additional program library as an extension to the main system. The syntax used when working with these extensions is often incompatible. For example, the approach applied to the Oracle Spatial [5] and Graph extension to the Oracle Database cannot be applied directly to the SQLServerSpatialTools extension to Microsoft SQL Server [6].

Open-source projects have aimed to extend the capabilities of relational databases that are also open-source, for example, PostGIS. The PostGIS system builds on the capabilities of SQL for working with spatial objects as an extension of the PostgreSQL system [7]. The MySQL system natively supports most of the features of PostGIS, (i.e., they can be used directly) [8]. Unfortunately, the MariaDB system, based on MySQL, does not currently support PostGIS features related to geohash encoding/decoding, specifically the functions ST\_GeoHash, ST\_LatFromGeoHash, ST\_LongFromGeoHash and ST\_PointFromGeoHash [9].

PostGIS is a tool for processing spatial data in the popular relational database management systems, MySQL, and PostgreSQL. A key place in PostGIS is occupied by the program library liblwgeom [10], and, in connection with geohash coding - the file lwalgorithm.c. The library is open source and is therefore used in many other systems as well. For example, the SpatiaLite extension to SQLite uses the liblwgeom library [11].

1. Methodology

Public data from the Regional Health Inspection Varna (RHIV: https://www.rzi-varna.com/) was used with information about the water pollution of Varna’s beaches in May, June, July, and August 2023 [12].

The RHIV provides data for the water pollution of Varna’s beaches in tabular format, one table per beach. The columns are date (of the observation), indicator 1 (“Intestinal enterococci”), and indicator 2 (“Escherichia coli”). The following indicators, “Intestinal enterococci” and “Escherichia coli”, are monitored and measured in MPN/100 cm3. Levels below 15 considered normal. The RHIV does not give values below 15. They are given as “less than 15”. For data analysis, the value “7” was used. Six measurements were taken every two weeks (bi-monthly).

QGIS was used [13], and an OSM layer [14] was added to visualize Varna’s Bay. A new vector layer was also added: “Water\_pollution”. QGIS created relevant shape files with extensions: SHP (for the shape file), DBF (for the attributes tables), and PRJ (for the projections). The JASP open-source statistics program was used to analyze correlations among the variables [15].

1. Results and Discussion

Information from RHIV was entered in the data attributes table of the newly created layer “Water pollution”.

A screenshot of a computer

Description automatically generated

1. Data Attributes Table of the Newly Created Layer “Water pollution”

The variables en1, en2, …, en6 pointed to six measurements of indicator 1 (“Intestinal enterococci”), while the variables es1, es2, …, es6 pointed to six measurements of indicator 2 (“Escherichia coli”).

A map of a beach

Description automatically generated

1. Visualization of time series data for indicator 1 “Intestinal enterococci”; Varna’s beaches May-August 2023

QGIS allowed the display of single labels. However, several values were visualized with the help of the concatenation function.

A screenshot of a computer

Description automatically generated

1. Showing several values (time series values) for “Intestinal enterococci”

Similarly, the values for single labels were changed *(“es1 || ', ' || es2 || ', ' || es3 || ', ' || es4 || ', ' || es5 || ', ' || es6 ||', ' || location”)*. Time series data for the second indicator (“Escherichia coli”) were visualized.

A map of a beach

Description automatically generated

1. Visualization of time series data for “Escherichia coli” ; Varna’s beaches May-August 2023
2. Correlation analysis for water pollution in Varna’s Bay; May-August 2023 (**dataset 1**)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable 1** | **Variable 2** | **Formal description** | **Explanation** |
| en3 | en5 | r=0.825, p<0.05, n=7 | The values of “Intestinal enterococci” in the fifth period are positively correlated with the values in the third period. Possible lag effect in pollution. |
| es3 | en4 | r=0.820, p<0.05, n=7 | Ambiguous dependency; no meaningful explanation |
| en3 | es4 | r=0.885, p<0.05, n=7 | Ambiguous dependency; no meaningful explanation |
| es4 | en6 | r=0.891, p<0.01, n=7 | Ambiguous dependency; no meaningful explanation |
| es5 | en5 | r=0.784, p<0.05, n=7 | Same time period; Possible period of pollution |
| es6 | en6 | r=0.989, p<0.01, n=7 | Same time period; Possible period of pollution |
| es4 | es6 | r=0.925, p<0.01, n=7 | Possible lag effect after pollution with “Escherichia coli” in time period 4, visible in time period 6. |

In all rows in Table 1 “n=7”, because there are 7 beaches which were included in the analysis.

The initial dataset used in QGIS (Fig. 1) was transposed to create time-series variables. Each variable was a certain beach and a specific indicator. A new dataset was created (dataset 2) with 14 variables (seven beaches multiplied by two indicators), and rows (six cases; six measurements in equal periods). The variables in dataset 2 were constructed as time-series variables.

1. Creating **dataset 2** using transpose.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Location** | **3rd Buna Beach** | **Officers' Beach** | **Central Beach** | **South Beach** | **Aspar-uhovo Beach** | **Veteran Beach** | **Cherno-morets Beach** |
|  | **b1\_en** | **b2\_en** | **b3\_en** | **b4\_en** | **b5\_en** | **b6\_en** | **b7\_en** |
| en1 | 15 | 15 | 15 | 7 | 7 | 7 | 7 |
| en2 | 7 | 7 | 7 | 15 | 15 | 7 | 15 |
| en3 | 7 | 77 | 77 | 46 | 7 | 7 | 7 |
| en4 | 93 | 15 | 15 | 77 | 7 | 15 | 7 |
| en5 | 30 | 93 | 125 | 126 | 7 | 61 | 46 |
| en6 | 7 | 126 | 7 | 30 | 7 | 7 | 15 |
|  | **b1\_es** | **b2\_es** | **b3\_es** | **b4\_es** | **b5\_es** | **b6\_es** | **b7\_es** |
| es1 | 15 | 15 | 7 | 7 | 30 | 7 | 7 |
| es2 | 30 | 7 | 15 | 7 | 7 | 7 | 30 |
| es3 | 195 | 94 | 127 | 144 | 7 | 7 | 7 |
| es4 | 15 | 209 | 93 | 94 | 30 | 7 | 7 |
| es5 | 15 | 127 | 307 | 327 | 127 | 30 | 7 |
| es6 | 7 | 287 | 15 | 94 | 15 | 7 | 7 |

Dataset 2 was imported into PSPP, and correlation analysis was conducted.

1. Correlation analysis for water pollution in Varna’s Bay; May-August 2023 (**Dataset 2**)

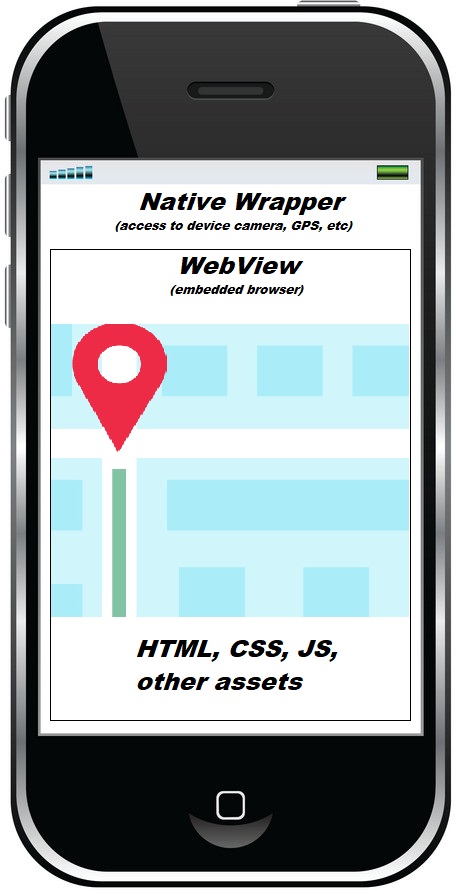
|  |  |  |  |
| --- | --- | --- | --- |
| **Variable 1** | **Variable 2** | **Formal description** | **Explanation** |
| b3\_en | b6\_en | r = 0.816, p < 0.05, n = 6 | Beach 6 is situated to the south of beach 3. Water flows in the Black Sea move water from north to south, an expected correlation. |
| b4\_en | b6\_en | r = 0.891, p < 0.05, n = 6 | Beach 6 is situated to the south of beach 4. Water flows in the Black Sea move water from north to south, an expected correlation. |
| b4\_en | b4\_es | r = 0.918, p < 0.05, n = 6 | Possible pollution of beach 4 (South beach) with sewage water or water flows are moving contamination from north to south. |
| b6\_en | b7\_es | r = 0.940, p < 0.01, n = 6 | Continuous contamination on neighbor beaches. |
| b6\_en | b6\_es | r = 0.989, p < 0.01, n = 6 | Possible pollution of beach 6 (Veteran beach) water flows are moving contamination from north to south. The wind moves sea water to the beach. |
| b3\_es | b4\_es | r = 0.964, p < 0.01, n = 6 | Water flows are moving contamination from north to south. |
| b5\_es | r = 0.868, p < 0.05, n = 6 |
| b6\_es | r = 0.905, p < 0.05, n = 6 |
| b5\_es | b6\_es | r = 0.974, p < 0.01, n = 6 |

In all rows in Table 3, “n=6” because there are six measurements (six periods). Some statistically significant correlations were found in dataset 2, but they were not meaningful, so they are not included in the analysis.

1. Mobile and hybrid applications for data gathering and visualization

Water pollution requires rapid and precise monitoring. Using mobile applications to capture data in real time eliminates the traditional delay associated with data acquisition and processing. By providing mechanisms for immediate capture and feedback, mobile applications ensure that the information accessed is the most recent and pertinent, allowing decision-makers to respond quickly and effectively.

Hybrid mobile technologies, which combine the benefits of native and web applications, offer promising avenues for the collection, visualization, and analysis of water pollution data in a timely manner. Hybrid mobile applications are built with web technologies such as HTML, CSS, JavaScript and frameworks like Ionic, and then packaged in a native container for deployment (fig. 7).



1. Application of the WebView component for hybrid applications strategy

They offer several benefits [16, 17, 18]:

* Cross-platform deployment permits the use of a single codebase for both Android and iOS.
* Avoid the need for distinct codebases and teams to save money.
* Access to Device Features: Despite being web-based, the application provides access to device-specific features, such as GPS, camera, etc.
* Updates are available without requiring users to obtain new versions from app stores.

However, hybrid applications encounter obstacles such as:

* Reliance on web page views sometimes causes performance problems.
* The hybrid nature of app stores results in extended approval times.
* Delays in accessing the most recent native features.

The application integrates with a GIS backend that provides visualization and analysis capabilities for geospatial data. On a dynamic map, water quality information, including pollutant concentrations, is plotted. Real-time insights increased public awareness through the democratization of data, and scalability to include additional data sources and aquatic bodies are among the benefits. Concerns regarding data security and dependable integration with IoT devices are among the obstacles.

QField and Mergin Maps are instruments for collecting and visualizing geospatial data [19, 20]. Utilizing QGIS projects, QField is predominantly a mobile utility for visualizing and capturing geospatial data in the field. Mergin is a cloud-based infrastructure designed for geospatial data and project storage, synchronization, and collaboration. For a GIS workflow involving field data collection, one could design their project in QGIS, deploy it to the field using QField, and then transfer and share the data using Mergin for purposes of collaboration or backup. The table 4 provides a comparative description to distinguish between the two.

1. Comparition between QField and Mergin Maps

|  |  |  |
| --- | --- | --- |
| **Mobile Geospatial Instrument** | **Indicator** | **Description** |
| **QField** | Nature | Mobile GIS application for Android devices |
| Main Use | Designed for data collection and visualization in the field. Allows users to collect geospatial data in the field using their QGIS initiatives. |
| Key Features | Exploration and presentation of spatial data. Editing and incorporating geospatial data. Integration of GPS for positioning and monitoring. Support for offline mapping for fieldwork without Internet connectivity. |
| Integration | Compatible with QGIS projects. The map or data collection project is designed in QGIS and then transferred to QField for fieldwork. |
| **Mergin Maps** | Nature | Cloud-based platform for storing, sharing, and versioning of geospatial data. |
| Main Use | Users are able to synchronize their geospatial projects and data, which facilitates collaboration and ensures that data is stored up and version-controlled. |
| Key Features | Control of versions for geospatial data. Capabilities for collaboration in real time. Integration with QGIS by means of a plugin. Provides a record of all modifications made to a dataset. |
| Integration | Mergin is not a direct substitute for QField, but it complements it. QField allows users to capture data in the field, which is then synchronized with the Mergin platform for archiving, sharing, and collaboration.  Mergin also integrates with QGIS, allowing users to synchronize projects and data between the desktop and cloud platforms. |

1. Conclusions

The scientific study examined water contamination on Varna's beaches in the Black Sea region, using GISs for visualization and statistical tools for correlation analysis. The use of open-source software tools such as QGIS and PSPP aided in investigating water quality indicators and their potential interrelationships. The findings highlight the complicated mechanisms of water contamination and its regional variations.

Several significant correlations were discovered via correlation analysis of water pollution data, offering insight into potential dependencies between different pollution indicators. Notably, correlations between the indicators of "Intestinal enterococci" and "Escherichia coli" were discovered, implying potential contamination sources and temporal trends. However, certain dependencies remained ambiguous, highlighting the complexities of water pollution and the need for further research.

The study approach highlighted the growth of GIS architecture, from early flat file storage to spatial data integration with relational databases. The difficulties of expressing spatial items within ordinary relational databases were recognized, spurring the creation of extensions such as PostGIS to successfully overcome these restrictions.

Furthermore, the article addressed the use of mobile GIS technology in the context of water contamination analysis. Integrating hybrid mobile applications with GIS backends improves real-time visualization and data collection and enables more efficient data collection, greater public awareness, and scalability for additional data sources.

The paper also discussed using field data collecting and collaboration technologies such as QField and Mergin Maps, which integrate seamlessly with QGIS projects and cloud-based storage, simplifying data synchronization and boosting collaboration among researchers and professionals.

In conclusion, this study offers insightful information on the challenges of employing GIS technologies to analyze water pollution. The relationships between pollution indicators show the importance of a comprehensive approach comprehending pollution trends and probable sources. Researchers and stakeholders are further empowered by integrating mobile applications and cloud-based platforms to handle water pollution concerns successfully, enabling collaboration and reasoned environmental management decision-making. After the study's conclusion, similar assessments in other geographical areas and periods can be built upon the methodology and insights offered here.

1. Acknowledgment

This research is financially supported by NPI-67/2023 from University of Economics - Varna Science Fund.

1. References
2. M. F. Goodchild, “Reimagining the history of GIS,” Annals of GIS, vol. 24, no. 1, pp. 1–8, Jan. 2018, doi: 10.1080/19475683.2018.1424737.
3. O. Conrad et al., “System for Automated Geoscientific Analyses (SAGA) v. 2.1.4*”, Geoscientific Model Development*, vol. 8, no. 7, pp. 1991–2007, Jul. 2015, doi: 10.5194/gmd-8-1991-2015.
4. T. Chafiq, M. Ouadoud, H. J. Oulidi, and A. Fekri, “Application of Data Integrity Algorithm for Geotechnical data quality Management,” *International Journal of Interactive Mobile Technologies*, Dec. 2018, doi: 10.3991/ijim.v12i8.9569.
5. M. Leslie et al. *Introduction to PostGIS*, 2023 https://postgis.net/workshops/en/postgis-intro/PostGISIntro.pdf
6. “Spatial database,” Oracle.com. [Online]. Available: https://www.oracle.com/database/technologies/spatialandgraph.html. [Accessed: 25-Aug-2023].
7. S. Tarmur and C. Özturan, "Parallel Classification of Spatial Points Into Geographical Regions," *18th International Symposium on Parallel and Distributed Computing (ISPDC)*, Amsterdam, Netherlands, 2019, pp. 9-15, doi: 10.1109/ISPDC.2019.000-3.
8. O. Regina and L. S. Hsu, *PostGIS in action*, 2021. Manning Publications
9. E. Zhuk, “Improvements of the Black Sea oceanographic GIS,” in *Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019)*, 2019.
10. “MySQL/MariaDB Spatial Support Matrix,” MariaDB KnowledgeBase. [Online]. Available: https://mariadb.com/kb/en/mysqlmariadb-spatial-support-matrix/. [Accessed: 25-Aug-2023].
11. Postgis, “postgis/liblwgeom at master · postgis/postgis,” GitHub. [Online]. Available: https://github.com/postgis/postgis/tree/master/liblwgeom [Accessed: 25-Aug-2023].
12. “SpatiaLite SQL functions reference list.” [Online]. Available: http://www.gaia-gis.it/gaia-sins/spatialite-sql-4.2.0.html [Accessed: 25-Aug-2023].
13. Rzi-varna.com. [Online]. Available: https://www.rzi-varna.com/health/mvodi-2023.pdf. [Accessed: 25-Aug-2023].
14. A. Nielsen, K. Bolding, F. Hu, and D. Trolle, “An open source QGIS-based workflow for model application and experimentation with aquatic ecosystems,” *Environmental Modelling and Software*, vol. 95, pp. 358–364, Sep. 2017, doi: 10.1016/j.envsoft.2017.06.032.
15. S. S. Sehra, J. Singh, and H. Singh, “Assessing OpenStreetMap data using Intrinsic quality Indicators: an extension to the QGIS Processing toolbox,” *Future Internet*, vol. 9, no. 2, p. 15, Apr. 2017, doi: 10.3390/fi9020015.
16. J. Love et al., “JASP: Graphical Statistical Software for Common Statistical Designs,” *Journal of Statistical Software*, vol. 88, no. 2, Jan. 2019, doi: 10.18637/jss.v088.i02.
17. G. A. Farsi et al., “A Review of Virtual Reality Applications in an Educational Domain,” *International Journal of Interactive Mobile Technologies*, vol. 15, no. 22, p. 99, Nov. 2021, doi: 10.3991/ijim.v15i22.25003.
18. C. Vitolo, Y. Elkhatib, D. E. Reusser, C. Macleod, and W. Buytaert, “Web technologies for environmental Big Data,” *Environmental Modelling and Software*, vol. 63, pp. 185–198, Jan. 2015, doi: 10.1016/j.envsoft.2014.10.007.
19. H. A. Manaf and M. N. S. Man, “Mobile Application and Web 2.0 as an E-Participation Mechanism: A Literature Analysis”, *International Journal of Interactive Mobile Technologies (iJIM)*, 15(06), pp. 185–192. doi: 10.3991/ijim.v15i06.20673.
20. R. Montagnetti and G. Guarino, “From Qgis to Qfield and Vice Versa: How the New Android Application Is Facilitating the Work of the Archaeologist in the Field,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, Oct. 2021, doi: 10.3390/environsciproc2021010006.
21. “Mergin maps: Collect, store and analyze your Geo-data easily,” Merginmaps.com. [Online]. Available: https://merginmaps.com/. [Accessed: 25-Aug-2023].
22. Authors

**Julian Vasilev** is a [German](https://en.wikipedia.org/wiki/Germans) computer scientist, engineering educator, and Vice Chancellor at Carinthia University of Applied Sciences (CUAS), Europastrasse 4, in Villach, Austria. He works also as a visiting professor at the Universities of [Amman](https://en.wikipedia.org/wiki/University_of_Jordan) ([Jordan](https://en.wikipedia.org/wiki/Jordan)), [Braşov](https://en.wikipedia.org/wiki/Braşov), ([Romania](https://en.wikipedia.org/wiki/Romania)) and [Patras](https://en.wikipedia.org/wiki/Patras), ([Greece](https://en.wikipedia.org/wiki/Greece)). He is the Editor-in-Chief of the journals iJOE, iJET, and iJIM at online-journals.org (email: example@example.org).

**Pavel Petrov** is Associate Professor in the Department of Informatics at the University of Economics - Varna, Bulgaria. His research interests include distributed web systems, big data, and mobile computing. (email: petrov@ue-varna.bg).

**Jordan Jordanov** is an assistant professor and a doctoral candidate in the Department of Informatics at the University of Economics in Varna, Bulgaria. His research interests include modern cloud services and mobile technologies. (email: jordanov.jordan@ue-varna.bg).

Article submitted 2020-10-19. Resubmitted 2021-11-18. Final acceptance 2020-11-27. Final version published as submitted by the authors.