Geo-Visual Analytics for Pest Population Dynamics

Gustavo A. García-Chapeton
Faculty of Geo-information
Science and Earth Observation
ITC, University of Twente
Enschede, The Netherlands
g.a.garciachapeton@utwente.nl

Frank O. Ostermann
Faculty of Geo-information
Science and Earth Observation
ITC, University of Twente
Enschede, The Netherlands
f.o.ostermann@utwente.nl

Abstract

Our research focuses on advancing the support for collaborative work and long-term analysis in Geo-Visual Analytics. At the same time, it brings the analytical power offered by Geo-Visual Analytics to the study of pest population dynamics, which addresses important societal, ecological and economic challenges. In this paper, we describe the first steps towards an integrated Collaborative Geo-Visual Analytics system: First, an analysis of requirements to support the monitoring and control of pests in general, and specific requirements for our application case concerning the Olive Fruit Fly in Málaga, Spain. Second, we designed a system architecture that supports novel and established collaboration techniques, and developed an initial prototype focused in data integration, data enrichment, and geo-visual exploration. We conclude after initial testing that the system architecture properly accommodates all the requirements, and provides a reference model for a Geo-Visual Analytics system for pest population dynamics. Once completed, the prototype will allow analysts to explore the outputs of a statistical model based on monitoring data, as well as topographic, environmental, and climatic predictors. Further, a novel technique called "Spatio-Temporal Discussion Board" will allow users to define discussion boards bounded in space and time, helping to focus the contributors' attention on features of interest, and elicit actionable information.
Keywords: Geo-visual Analytics, System architecture, Data exploration, Pest population dynamics, Olive Fruit Fly.

1 Introduction

Our research aims to link the opportunities offered by Internet-based technologies, with the societal, ecological and economic challenges posed by pests. In particular, it identifies Geo-Visual Analytics (GVA) and collaborative analysis as a feasible approach to improve the understanding of pest population dynamics, and to support effective and ecologically viable monitoring and control strategies.

Many researchers agree that collaboration is of key importance and offers unexploited opportunities for GVA (Thomas and Cook, 2005, Andrienko et al., 2007, Isenberg et al., 2011). Several collaborative methods have been implemented in GVA systems, such as annotations (Hopfer and Maceachren, 2007), chat (Hardisty, 2009), snapshot (Ho, 2013) and storytelling (Lundblad, 2013). However, despite this significant research, collaboration is still not a common feature in off-the-shelf GVA systems.

The analysis of pest populations is important due to the significant negative effects of pests on environment, economy, and human health. The outbreaks of a pest can threaten local flora and fauna (Paritsis, 2012), and some pests can significantly damage crop fields, producing serious economic impacts for farmers and the food supply chain, threatening food security (Fao, 2005). Finally, some pests represent a threat for human health by providing infection vectors for diseases (Williams et al., 2013). Despite those negative impacts, these species are part of the natural balance of ecosystems. Therefore, to minimize their negative effects without affecting the natural balance, it is necessary to understand their population dynamics (Gilioli et al., 2016).

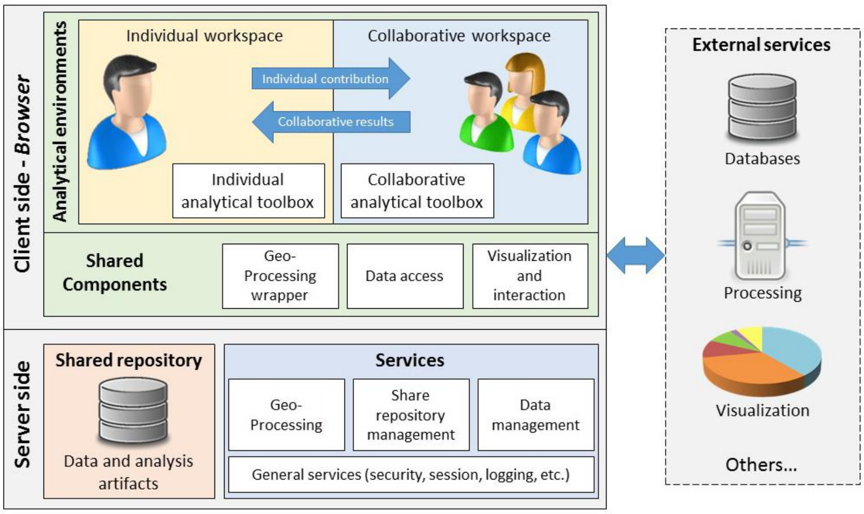
The main outputs of our research described in this paper are an architecture for Collaborative GVA (CGVA) systems that support novel and established collaboration techniques, and a prototype aimed at studying pest population dynamics. The selected application case is the monitoring and control of the Olive Fruit Fly (OFF) in Andalucía, Spain. The specific application case requirements and our objective to support a wide range of collaboration techniques led to the decision of developing a custom GVA system, instead of adopting and modifying an existing one.

The initial prototype focusses on data integration, data enrichment, and geo-visual exploration. This prototype represents the first step towards our GVA system for pest population dynamics, and is currently being extended to support the analysis of a statistical model outputs, and a flexible individual-collaborative analysis workflow.

2 Background

GVA aims to produce a synergy between human analytical skills and computer storage and processing power, and to support effective understanding, reasoning and decision making on the basis of complex spatio-temporal datasets (Andrienko et al., 2007). Current examples include the analysis of criminal activity (Roth et al., 2015), self-organizing mobile networks (Van Quan et al., 2009), planning ship routes based on weather conditions (Lundblad et al., 2009), ocean

Figure 1 System architecture.



Source: the authors.

temperature and salinity volume data (Ho and Jern, 2008), and disaster and crisis management (Tomaszewski et al., 2007).

Despite recent advances in research on GVA, support for collaborative work and long-term analysis remains challenging. Collaboration is one of the grand challenges for GVA, because analysts increasingly face large, complex, ill-defined, and broadly scoped problems (Thomas and Cook, 2005, Andrienko et al., 2007, Isenberg et al., 2011). Additionally, analysis is often a complex, multi-staged and dynamic task that resembles a long-term process (Keim et al., 2010, Isenberg et al., 2011).

To the best of our knowledge, GVA has not been applied in the analysis of pest population dynamics, which could benefit significantly from the analytical capabilities it offers. Additionally, this application domain requires the support of collaborative analysis because it involves practitioners and experts from different domains, agencies, and geographical and administrative regions.

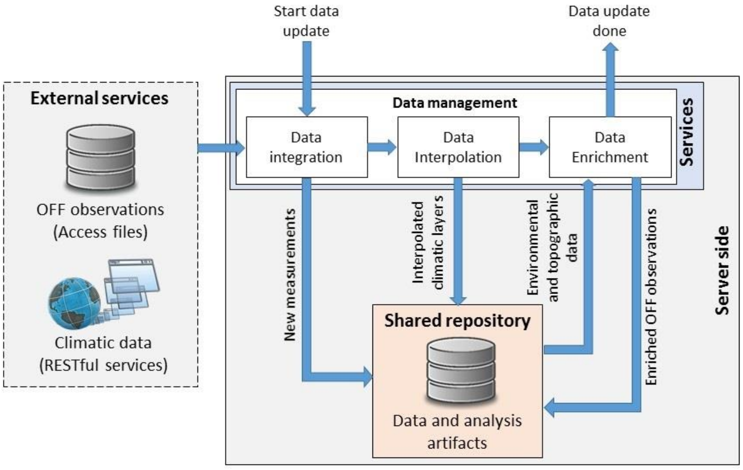
3 System design

The first step in designing a GVA system for pest population dynamics was to identify common features of projects on monitoring and control of pests. Interviews and discussions with experts (i.e., researchers and field technicians), and literature review provided information on the following four common features:

- **Continuous data collection:** the monitoring and control of pests require continuous collection of field data to measure their abundance in a given moment, to determine when and where control measurements are needed, and to assess how well these control measurements worked. Therefore, the system should be capable to automatically integrate the field data as soon as it is available in an electronic format. Additionally, it is of key importance for analysts that the data is enriched with information about factors that affects the pest population (e.g., topographic, environmental and climatic information).
- **Long-term analysis:** the monitoring and control efforts usually last for several years. Data is collected in regular basis and analyzed to identify features of interest such as rapid increasing of the population (i.e., a potential outbreak), rapid decreasing of the population (i.e., a potential population collapse), and changes in population dynamics compared with other observation periods, among others. For this reason, the system should be capable to keep all the data and analysis results available at any time, so analysts can build upon them.
- **Use of geo-statistical models:** the capacity to collect data is limited, and therefore, analysts make use of geo-statistical models to interpolate the information in space and time. In this regards, the system should offer access to geo-processing capabilities.
- **Collaboration:** the analysis of pest population dynamics is of interest for several stakeholders, such as farmers, environmental authorities, and environmental scientists, among others. They can provide variety of ideas and perspectives, diversity of expertise and, domain and local knowledge, which can lead to better understanding and decision making. Therefore, the system should offer tools to allow collaborative work.

Based on these features, we designed the system architecture shown in Figure 1, which is based on the Client/Server model. These architecture fulfills the identified requirements by including services keeping the database up-to-date and enriching the observations with relevant information ("Data management" component); offering services to keep data and analysis artifacts (e.g., hypothesis, evidence and conclusions) available for long-term analysis ("Shared repository" component); providing geo-processing services, which includes access to application-specific algorithms and general purpose processing libraries ("Geo-processing" component); finally, by offering two analytical environments, "Individual workspace" which offers a private working space, and

Figure 2 Workflow to integrate and enrich OFF observations.



Source: the authors.

"Collaborative workspace" which is shared by several analysts working in a specific feature of interest, they include functionality to exchange analysis artifacts seamlessly.

Additionally, we identified specific requirements for our application case:

- **Hierarchical access to the data:** the data should be accessible following the hierarchy of cycle (i.e., year), observation period (i.e., week of the year) and monitoring location. Depending on the species under study, this hierarchy may change, but the concept of cycle and observation period are likely to be universal to all pest population analyses, although potentially with different time boundaries and granularity.
- **Temporal evolution of observation sites:** it is relevant for analysts to be able to visualize the evolution over the year of a monitoring site, and to be able to compare it with other sites. This feature not only allows them to compare different location, but to analyze how different topographic, environmental and climatic conditions affect the dynamics of the pest.
- **Comparison over cycles:** it is important for analysts to be able to compare specific observation periods (i.e. week of the year) over different cycles (years), which allows them to identify variations in the population dynamic due to climatic factors or human intervention.

These case-specific requirements do not affect the system architecture, but they do affect the database structure and the implementation of the prototype.

4 Prototype implementation

Based on the previously described system architecture, we developed an initial prototype with the following functionality: data integration and enrichment ("Data management" component), interactive visualizations ("Visualization and interaction" component), and data exploration ("Individual workspace"). These functions act over the "Share repository" by creating, editing and querying data. The aim of developing this prototype was to implement the basic functionality for our GVA system for pest population dynamics, and to implement the case-specific requirements.

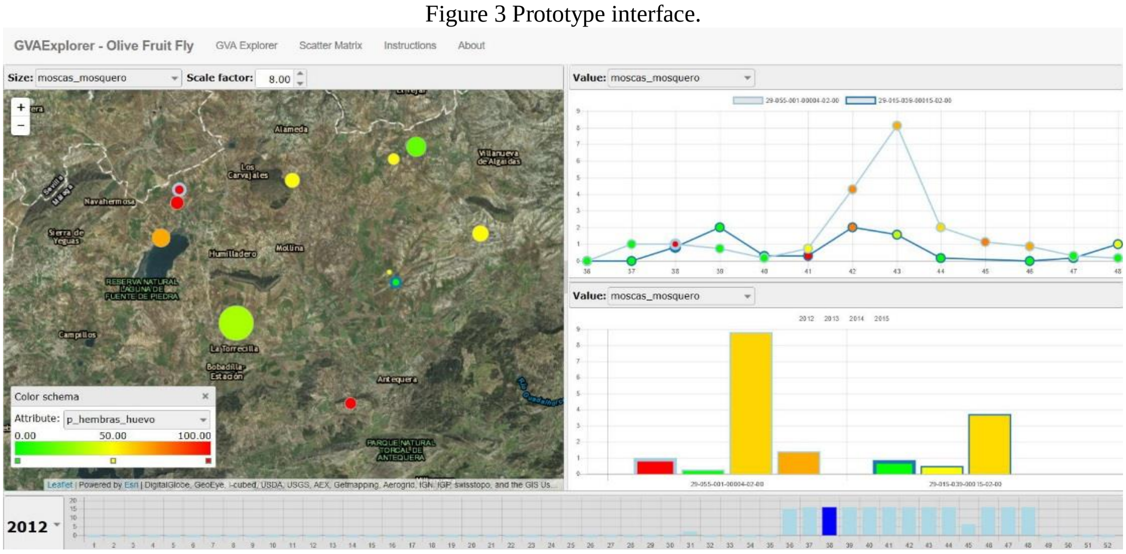
The first step was the design of the database schema and the methods to integrate and enrich the OFF observations. This step involves the "Shared repository" and the "Data management" components. The "Shared repository" was implemented as a PostgreSQL[1] database and a plain files system. The "Data management" component was implemented as Python[2] scripts and includes three processes: data integration, data interpolation and data enrichment (See Figure 2). The data integration reads the OFF observations (provided by field technicians as MS Access files) and climatic data from official web services[3], and inserts them into the database. The data interpolation process uses the climatic observations (point data) to create climatic data layers, and stores them in the plain files system. Finally, the enrichment process loads climatic, environmental and topographic data from the "Shared repository", and based on the locations and timestamps of the observations, it updates the observations with the respective information.

The second step was to develop a web-based interface to explore the data (See Figure 3). The interface was developed using Bootstrap, JQuery, jQWidgets and Chart.js[4]. The interface fulfils the case-specific requirements; therefore, it allows to access data in a hierarchical manner, visualize the evolution of observation stations over the cycle, and compare observation periods amongst cycles. In the following paragraphs we explain the different components in the interface and their purpose.

The time line component is located in the bottom part of the interface (See Figure 3), it allows to access the data on a two-

[1] https://www.postgresql.org/
[2] https://www.python.org/
[3] http://www.juntadeandalucia.es/agriculturaypesca/ifapa/ria/
[4] http://getbootstrap.com/, https://jquery.com/, http://www.jqwidgets.com/ and http://www.chartjs.org/

Figure 3 Prototype interface.



Source: the authors.

level hierarchy (and temporal scales), where the levels are cycle (i.e., year) and observation period (i.e., week of the year). On the left side of the component, the analysts can select the cycle, and as response the system creates a bar chart showing the temporal distribution of the observations, where the x-axis represents the observation periods and the y-axis the number of observations on each period. The analysts can click on this chart to display the observations of a given period in the map component.

The Map component is located in the left side of the interface (See Figure 3), it allows to display the observations of a specific observation period (e.g., week 38 of 2012). The data representation is based on the observation locations and an attribute that is used to set the size of the symbols (i.e., graduated symbol map). Additionally, another attribute is represented as the colour of the symbol, this is set in a shared colour schema as discussed later. Finally, the analysts can select a station by clicking on the map symbols (or Ctrl + clicking for multiple selection) to display detailed data.

On the right side of the interface, the analysts have access to two time-based charts (See Figure 3), these charts get activated when one or more observation stations are selected. The chart on the top displays the evolution of a given attribute for the selected stations over the cycle (e.g., values of the stations 1 and 2 for the attribute "flies captured per week" over the year 2012). The chart on the bottom displays the comparison of values for a given attribute in the current observation period over the cycles for the selected stations (e.g., "flies captured per week" on week 38 for the years 2012, 2013 and 2014, for stations 1 and 2).

The prototype implements the established mechanism of multiple linked views based on selection and shared colour scheme. In the case of the selection-based linking, it commonly works by highlighting in all the views the objects that are selected in one of them. However, due to the requirement of hierarchical access to data, in our system when an object is selected in one view, the other views show detailed data related to it. Regarding the shared colour schema, the analysts are allowed to select an attribute (e.g., "flies captured per week") and a colour map that is applied to all the views.

5 Conclusions and future work

This paper reports the results of a requirement analysis for a GVA system to study pest population dynamics. The main requirements are to support continuous data collection, long-term analysis, use of geo-statistical models, and collaborative analysis. Additional requirements specific to our case study are to access data in a hierarchical manner, visualize the evolution of observation stations over the cycle, and compare observation periods amongst cycles. After initial testing, we conclude that all these requirements can be properly accommodated in a system architecture as shown in Figure 1 and discussed in Section 3 (System design).

Additionally, we developed a prototype as proof-of-concept. The prototype focuses in the processes of data integration, data enrichment, and geo-visual exploration. While this prototype has not yet implemented all the architecture's components, it constitutes an important and significant step towards the full realization of the system.

Following this reported work, we are extending the prototype to allow analysts to explore the results of a statistical model based on the monitoring data and, topographic, environmental and climatic predictors. Additionally, we are working in the design and implementation of a method for collaborative long-term analysis called "Spatio-Temporal Discussion Board" (García-Chapeton and Ostermann, 2016). This method allows analysts to define discussion boards bounded in space and time. Our hypothesis in the design of this method is that having a clear defined context and boundaries for discussions will help in focusing analysts' attention, and elicit sensible information. Finally, once completed, we will conduct functional and usability tests; and then, use the system in the context of the application case, aiming to validate the design of the system architecture and collaborative methods, and their implementation. Specifically, it will be used during the months from July to December (i.e., critical monitoring and control period) of 2017.

Acknowledgement

The authors thank the valuable knowledge and data provided by Emilio García (Departamento de Sanidad Vegetal, Málaga), Jesús Olivero (Departamento de Biología Animal, Universidad de Málaga), Bert Toxopeus (Faculty ITC, University of Twente), Antonio Sánchez ("Antequera" producers association, Málaga), and Gonzalo Reina ("La Camorra" producers association, Málaga).

References

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. J., Maceachren, A. and Wrobel, S. (2007) Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science,* 21(8)**,** 839-857.

Fao (2005) Impact of climate change, pests and diseases on food security and poverty reduction. FAO.

García-Chapeton, G. A. and Ostermann, F. O. (2016) Spatio-Temporal Discussion Board. *Presented at the 19th AGILE Conference on Geographic Information Science.* Helsinki, Finland.

Gilioli, G., Pasquali, S. and Marchesini, E. (2016) A modelling framework for pest population dynamics and management: An application to the grape berry moth. *Ecological Modelling,* 320**,** 348-357.

Hardisty, F. (2009) GeoJabber: enabling geo-collaborative visual analysis. *Cartography and Geographic Information Science,* 36(3)**,** 267-280.

Ho, Q. (2013) *Architecture and Applications of a Geovisual Analytics Framework.* PhD, Linköping University.

Ho, Q. and Jern, M. (2008) Exploratory 3D geovisual analytics. *In:* Research, Innovation and Vision for the Future, 2008. RIVF 2008. IEEE International Conference on, 13-17 July 2008. IEEE, 276-283.

Hopfer, S. and Maceachren, A. M. (2007) Leveraging the potential of geospatial annotations for collaboration: a communication theory perspective. *International Journal of Geographical Information Science,* 21(8)**,** 921-934.

Isenberg, P., Elmqvist, N., Scholtz, J., Cernea, D., Ma, K.-L. and Hagen, H. (2011) Collaborative visualization: Definition, challenges, and research agenda. *Information Visualization,* 10(4)**,** 310-326.

Keim, D., Kohlhammer, J., Ellis, G. and Mansmann, F. (2010) *Mastering the Information Age: Solving Problems with Visual Analytics.* Florian Mansmann.

Lundblad, P. (2013) *Applied Geovisual Analytics and Storytelling.* PhD, Linkoping University.

Lundblad, P., Eurenius, O. and Heldring, T. (2009) Interactive Visualization of Weather and Ship Data. *In:* Information Visualisation, 2009 13th International Conference, 15-17 July 2009. 379-386.

Paritsis, J. a. V., T.T. (2012) Outbreak species. *In:* Craig, R. K., Nagle, J. C., Pardy, B., Schmitz, O. and Smith, W. (eds.) *The Berkshire Encyclopedia of Sustainability.* Great Barrington, MA.: Berkshire Publishing.

Roth, R., Ross, K. and Maceachren, A. (2015) User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. *ISPRS International Journal of Geo-Information,* 4(1)**,** 262.

Thomas, J. and Cook, K. (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* National Visualization and Analytics Ctr.

Tomaszewski, B. M., Robinson, A. C., Weaver, C., Stryker, M. and Maceachren, A. M. (2007) Geovisual analytics and crisis management. *In:* Proceedings of the 4th International ISCRAM Conference, 2007. Delft, the Netherlands, 173-179.

Van Quan, H., Astrom, T. and Jern, M. (2009) Geovisual analytics for self-organizing network data. *In:* Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on, 2009. IEEE, 43-50.

Williams, D. T., Straw, N., Townsend, M., Wilkinson, A. S. and Mullins, A. (2013) Monitoring oak processionary moth Thaumetopoea processionea L. using pheromone traps: the influence of pheromone lure source, trap design and height above the ground on capture rates. *Agricultural and Forest Entomology,* 15(2)**,** 126-134.