

A novel architecture of Web-GIS for mapping and analysis of echinococcosis in Poland

Marcin Kulawiak^{1*}, Natalia Kulawiak², Małgorzata Sulima², Katarzyna Sikorska²

¹ Department of Geoinformatics, Faculty of Electronics, Telecommunications and Informatics Gdansk University of Technology, Gdansk

² Department of Tropical and Parasitic Diseases, University Center of Maritime and Tropical Medicine, Gdynia, Poland; Division of Tropical and Parasitic Diseases, Institute of Maritime and Tropical Medicine, Medical University of Gdansk, Gdansk, Poland

* Correspondence: Marcin.Kulawiak@eti.pg.edu.pl

Abstract

Echinococcosis is an infectious disease transferred through ingestion of food or water which have been contaminated with eggs of the Echinococcus tapeworm, which are spread by intermediate parasite hosts. Because the latter are primarily territorial, research related to diagnosis and prevention of echinococcosis requires investigation of environmental factors, which can be supported with the use of a Geographical Information System (GIS). In particular, since this type of research is usually performed by a team of medical researchers working individually with many patients, it can greatly benefit from the use of a Web-GIS, which provides remote access to data integration, storage and analysis tools. However, most existing Web-GIS solutions dedicated for disease mapping either exhibit limited functionality and interactivity (often due to the use of a thin-client approach), or are tied to a particular technological platform through a thick client based on a browser plugin. Moreover, few of the existing solutions provide user-friendly tools for data input, and none are directly integrated into the workflow of clinicians. This paper proposes a novel system architecture for Web-GIS dedicated to disease mapping, which attempts to optimize the use of client and server hardware resources while providing the user with a variety of analytic functionalities. The proposed architecture has been implemented in the form of a system for remote mapping and analysis of echinococcosis cases in Poland, which has been designed in cooperation with a major Polish hospital centre.

Keywords: Web-GIS, echinococcosis, disease mapping, geoprocessing, architecture

Introduction

Echinococcosis is an infectious, zoonotic disease which develops after ingestion of food or water contaminated with eggs of the Echinococcus tapeworm. As such, its spread is largely determined by environmental factors which facilitate the tapeworm's growth. In consequence, research related to diagnosis and prevention of echinococcosis requires investigation of those environmental factors (Lass et al, 2017). Issues involving the environment have an innate geospatial component, and thus they are

best analysed with the help of a Geographic Information System (GIS). Geographic Information Systems provide dedicated tools for creation, integration, editing, storage and visualization of geographic data (Longley et al., 2005). Over the years they have been successfully applied to investigation of diverse issues such as monitoring and analysis of marine areas (Dawidowicz & Kulawiak, 2017), management of education data (Agrawal & Gupta, 2020), digital reconstruction of real-world objects (Kulawiak & Kulawiak, 2017) and agricultural land use planning (Wojcik-Len et al., 2018). Moreover, geospatial problems which require cooperation of several researchers are best investigated with the use of a Web-GIS, which provides remote access to a set of standard GIS functionalities such as data integration, mapping and analysis. This being said, integration of Web-GIS to disease mapping is a complex topic, which requires careful consideration of many factors. Research in this field has been carried out for many years, and employed different solutions and technologies. Initially, Web-GIS solutions dedicated to disease mapping focused on using thin clients with limited interactivity, often based on commercial server applications such as ArcIMS and ArcSDE and focused on data presentation (Blanton et al., 2006). More advanced functionalities, such as map animation, were provided by thick-client solutions, which usually required the installation of a web browser plugin such as Adobe Flash. Such systems still used server-side data processing, however it was often provided with the use of Open Source solutions such as MapServer and PostGIS (Reinhardt et al., 2008). In time, even thin-client services for mapping and sharing disease information started to allow for more complex data analysis methods, such as displaying the state of selected variables at different times on choropleth maps, by utilizing server-side processing through the open Web Map Service (WMS) protocol's Styled Layer Descriptor (SLD) component (Gao et al., 2008). This method was later expanded into a Web Processing Service (WPS) for server-side generation of thematic disease maps using XML input data (Gao et al., 2009). A few years later, the functionality of Web-GIS for disease mapping has been further extended by the inclusion of geocoding services, as well as server-side execution of analysis methods including buffering, centroids and density estimation (Dominkovics et al., 2011). For this purpose the system used commercial technologies such as PostGIS and ArcGIS Server in tandem with the Adobe Flex browser plugin, and relied on a geocoding service provided by Yahoo. Soon after, a similarly advanced Web-GIS for analysis of health datasets was developed using Open Source technologies (Sopan et al., 2012). On the server side the system used GeoServer along with PostGIS, and the javascript OpenLayers library was used to build the thin client. The system allowed for mapping the results of database queries (eg. for the projected life expectancy in different counties), however it did not integrate data collection tools. In the same year, GeoServer and OpenLayers have also been used for building a Web-GIS for server-side processing and generation of disease risk and vulnerability maps. The system allowed analysis and mapping of user-uploaded spreadsheets with the use of remote sensing Land Surface Temperature (LST) data derived from the MODIS sensor (Porcasi et al., 2012). Not long after, a new element of Web-GIS disease mapping has been introduced in the form of server-side infection spread simulation (Ramirez-Ramirez et al., 2013). The proposed system for simulation and analysis of influenza infection spread has been implemented with the use of ArcGIS Server and R framework, with the client component realized in Silverlight. In



consequence the system required a browser plugin despite the fact that most functionalities were implemented on the server side. At the same time, Open Source technologies have become widely adopted in construction of Web-GIS tools for disease mapping. PostGIS, GeoServer and OpenLayers together with Ext.js constituted a geospatial framework for online input and mapping of Dengue cases in India (Tiwari and Jain, 2013). The combination of PostGIS and GeoServer has also been used for server-side data storage and processing in a prototype Web-based cancer atlas (MacEachren et al., 2013), and PostGIS has also been paired with GeoDjango, PySal and GeoExt to dynamically generate maps on the server-side and present them to remote users by means of WMS in a framework for analytical mapping of health data (Moncrieff et al., 2014). Different Open Source technologies in the form of MySQL, Python, PHP and OpenLayers have been used to implement a Web-based GIS for analysis of dengue fever outbreaks in developing countries. The system used Kernel Density Estimation (KDE) to perform server-side spatial analysis of remotely uploaded data via density estimation (Delmelle et al., 2014). This being said, the simultaneous progress of commercial solutions has allowed for the introduction of new Web-based disease mapping tools, exemplified by a concept Web-GIS for pollen allergy mapping (Lupa et al. 2017). The system used a two-tier architecture in which the user interacted with a Web application implemented in a Software as a Service (SaaS) model on the ArcGIS Online platform for remote collection and mapping of allergen data, while the actual maps were created and updated by the system administrator using ArcGIS Desktop. In parallel, ArcGIS Server has been used for server-side geoprocessing of diverse spatial datasets for the purpose of identifying areas prone to spread of tropical diseases in East Java (Vitianingsih et al. 2017). While that system required the Silverlight browser plugin to function, more recent disease mapping systems based on commercial solutions tend to use a more modern approach. For instance, the Web-GIS for collection, storage and distribution of information on West Nile disease spread in the Mediterranean basin (Savini et al., 2018) used ArcGIS Server and ArcSDE for data storage and dissemination, with data visualization provided by the ArcGIS Javascript API along with the PhyloCanvas library. While it has been designed for the purpose of data integration and dissemination, and only used Javascript API for building the client, the system allowed for a certain level of interactivity, including highlighting of regions affected by a selected virus strain as well as launching an animation of available temporal data such as disease distribution history.

Recently, geospatial investigation of infectious diseases has intensified due to the worldwide spread of COVID-19. Various spatial analysis methods have been applied during research of the publicly available spatio-temporal infection datasets. Choropleth mapping has been used to depict the spatial distribution of infections in different countries (Bođová & Kollár, 2021)(Rahman et al., 2021), while hot spot mapping has enabled the detection of disease clusters on the global as well as national scale (Charandabi & Gholami, 2021). Spatial autocorrelation has been applied for investigation of disease spread characteristics in China (Li et al., 2020), analysis of the distance from testing sites for different groups of citizens in Florida (Tao et al., 2020) as well as the association between citizen mortality and proximity to available hospital beds in São Paulo (Alcântara et al., 2020). In conjunction with remote sensing, the available datasets have allowed for the investigation of the correlation of weather and



environmental conditions such as temperature, precipitation and air pollution with infection rates, suggesting their association with the number of new COVID-19 cases in several countries (Ma et al., 2020)(Zhu et al., 2020)(Tehrani et al., 2021). At the same time, the lack of reliable testing methods in the early days of the pandemic has stimulated the research of new data sources for modelling the spread of the virus, including analysis of citizen mobility data (Jia et al., 2020), detecting anomalous trends in ambulance calls (Gianquintieri et al., 2020) as well as analysis of social network posts via machine learning (Mackey et al., 2020). Moreover, the global nature of the pandemic has increased the variety of performed research. New topics include methods of contact tracking for early detection of possible infections (Benreguia et al., 2020)(Boulos & Geraghty 2020) and analysis of the availability and adequacy of open datasets for studying the spatio-temporal spread of the virus (Brovelli & Coetzee, 2021)(Mooney et al., 2021). Additionally, the introduction of lockdowns has given the scientific community a unique opportunity to study the impact of travel restrictions on both mental and physical health of people (Adams-Prassl et al., 2020)(Basu et al., 2020)(Marino et al., 2021) as well as the state of environment pollution (Arora et al., 2020)(Oxoli et al., 2020)(Aman et al., 2020).

The ongoing effort for online monitoring of the worldwide spread of COVID-19 has been supported by both commercial as well as Open Source technologies (Boulos & Geraghty 2020). The Center for Systems Science and Engineering (CSSE) at Johns Hopkins University produced the Novel Coronavirus Visual Dashboard (Dong et al.,2020), which aggregates COVID-19 infection data obtained from national and international health institutions around the world. The online tool provides information about the number of infections, deaths and administered vaccines in the form of map markers as well as time series plots. The data may be displayed in the global context as well as individually for selected countries and U.S. states. A similar service has been established by the WHO in the form of the Coronavirus Dashboard. It provides a similar set of information (COVID-19 infections, deaths and vaccinations) through similar means (map and time series plots). One of the key differences lies with data mapping: in contrast to CSSE's markers, WHO opted for a choropleth map, which assigns different shades of colour to countries depending on their total number of infections, deaths or vaccinations (WHO, 2020). Several smaller-scale COVID-19 mapping initiatives exist, however they appear to be using a selection of the aforementioned data presentation techniques (scale-dependant marker clusters, time series plots and choropleth maps showing the number of new cases/deaths/vaccinations as well as their cumulative count) (Allen et al.,2020)(Arab-Mazar et al.,2020) (Mooney & Juhász, 2020). This quantitative approach to data presentation and analysis is understandable, as COVID-19 has shown to be transmitted via direct contact with an infected individual, and other potential aspects of its spread (such as environmental conditions) are yet to be explored (Al Huraimel et al., 2020). The system architecture employed by these solutions focuses on providing a thin and lightweight Javascript-based client, while the server side is primarily used for data storage and serving of background map tiles.

As it can be seen, during the last decade Web-GIS tools dedicated to disease analysis and mapping have made considerable progress from simple data dissemination websites to sometimes quite advanced

tools for data processing and analysis. In terms of system architecture, however, the common characteristic of the presented systems is that they either provide a thin web client and rely on server-side processing, or require the installation of a browser plugin (Flex, Silverlight) in order to provide a higher level of interactivity on the client. Moreover, from the design point of view, many of the presented systems require the user to upload health data in a particular format, which requires technical experience and thus may reduce the number of potential users. From the user perspective, it is important that a Web-GIS provides interactive data analysis tools through a responsive interface (Sakairi et al., 2008), (Kulawiak & Lubniewski, 2014), (Go & Gretzel, 2016), (Chakraborty & McMillan, 2018). In this context, the responsiveness of a thin Web-GIS client will always be limited by several factors, including network bandwidth and server load (Kulawiak et al., 2019). On the other hand, the limited performance of both web browsers as well as Personal Computers (PC's) in general has for many years limited the applicability of plugin-free thick-client solutions. This being said, recent years have seen a substantial improvement in the performance of both Web browsers as well as PC's, to the extent where processing of smaller geographic datasets may actually be faster when performed on the client (Kulawiak et al., 2019). In this context, the article presents a novel thick-client architecture of a Web-GIS dedicated to disease mapping, based purely on Open Source technologies. The architecture has been implemented in the form of a system applied to remote mapping and analysis of echinococcosis cases in Poland.

Materials and Methods

As it has been shown in the preceding section, existing Web-GIS tools for disease analysis and mapping either use a thin-client approach, which puts a high strain on the server and limits user interactivity, or employ a thick client which requires installation of a browser plugin. Moreover, few of the existing solutions provide user-friendly tools for data input, and none are directly integrated into the workflow of clinicians. In this context, the presented work aims to propose a novel system architecture for Web-GIS dedicated to disease mapping, which reduces the strain on the server while providing the user with a variety of analytic functionalities. The proposed architecture has been applied to a system for remote mapping and analysis of echinococcosis cases in Poland, which has been designed in cooperation with a major Polish hospital centre.

The characteristics of echinococcosis

Echinococcosis (EC) is a parasitic disease caused by the larval forms of tapeworm from the family of Taeniidae. There are several species known to cause infections, however only *Echinococcus granulosus* sensu lato and *Echinococcus multilocularis* are significant from the epidemiological and clinical standpoint. The parasite infects carnivores, mainly foxes and dogs, which excrete the parasite eggs into the environment. Man is an intermediate host, which becomes infected through close contact with dogs (*E. granulosus* sensu lato) or by eating unwashed fruits or vegetables covered in parasite eggs of *E. multilocularis*. Clinically the disease is characterized by a long-lasting asymptomatic phase. Symptoms depend on the location and the size of the parasitic lesion, which means that for many years the disease



may remain hidden (McManus et al., 2003). In most cases the primary lesions are localized in the liver in the form of a single-chamber cyst in the case of *E. granulosus sensu lato* (also known as Cystic Echinococcosis - CE) and a tumour-like mass in the case of *E. multilocularis* (also known as Alveolar Echinococcosis - AE). In some cases the cysts reach very large sizes, and extrahepatic lesions including remote locations like lungs, spleen or brain may be detected at the time of diagnosis. Diagnosis is based on typical imaging findings, positive results of antibody detection tests, or the histopathological and molecular examinations of intraoperative material. Treatment consists of surgical interventions in combination with pharmacological therapy or is limited to pharmacological therapy with benzimidazoles (albendazole or mebendazole). Liver transplantation is regarded as a salvage therapy in advanced hepatic EC. Studies have shown that the risk of infection is greatly increased by the vicinity of forests, while deforestation causes quick recession of new infection cases (Giraudoux et. al, 2006). Echinococcosis prevention also involves systematic de-worming of domestic dogs and reduction of the population of foxes in endemic areas (Kern et al., 2003). In Poland around a dozen new cases of echinococcosis are discovered on a yearly basis, primarily in the Voievodships of Warmia-Mazury, Podkarpackie and Malopolskie. The highest risk of infection applies to humans working in forestry or accustomed to collection of undergrowth (Gawor, 2016). New cases of echinococcosis are diagnosed and treated in a few dedicated medical facilities, one of the most prominent being the University Center of Maritime and Tropical Medicine located in the city of Gdynia in northern Poland (UCMTM).

The University Center of Maritime and Tropical Medicine

UCMTM is one of the oldest medical scientific and research institutions in Poland. It was established in 1938 under the name of the Institute of Maritime Medicine, and since 1939 it has been operating as an independent institution. After the second World War, it was reactivated in 1947 and incorporated into the structures of the Medical Academy in Gdansk. In 2003, after the establishment of the Academic Center of Maritime and Tropical Medicine, the scientific part of UCMTM was incorporated into the structures of the Medical Academy in Gdansk (now Medical University of Gdansk) as the Interdepartmental Institute of Maritime and Tropical Medicine (IIMTM). UCMTM consists of four wards (Tropical and Parasitic Diseases, Cardiology and Internal Diseases, Occupational Diseases, as well as Hyperbaric Medicine and Marine Rescue) and their respective outpatient clinics. It currently employs 28 specialists who represent inter alia internal diseases, infectious diseases, marine and tropical medicine, cardiology and anesthesiology. In addition, over a dozen residents are currently undergoing specialization training. Annually, UCMTM treats more than 3,500 patients, and a similar number is provided with medical advice in the outpatient clinics.

During the last decade, the ward of Tropical and Parasitic Diseases hospitalized nearly 100 patients with newly diagnosed alveolar echinococcosis. The majority of patients came from rural areas, 55% of them were women. Half of these patients required radical treatment, about 10% of patients qualified for liver transplantation, and another 10% died (Sulima et al., 2019). The collection and processing of data on cystic echinococcosis is still ongoing.

In order to provide the UCMTM clinicians with the best possible tools, an analysis of their workflow has been performed, beginning with patient interviews and data collection, including data storage and processing, to the methods of its analysis and presentation. Moreover, a literature review on the applied methods of GIS-based echinococcosis analysis has also been performed.

State-of-the-art in GIS-based analysis of echinococcosis

Since echinococcosis infections are transferred indirectly, they depend primarily on environmental factors, which in turn makes the disease well suited to analysis with tools provided by GIS. In literature, diseases determined by environmental factors are analysed using hot spot mapping and spatial autocorrelation (Louis et. al, 2014) as well as simple dot-based prevalence depiction (Kelly-Hope et. al, 2004) (Kistemann et. al, 2002) (Dangendorf et. al, 2002). Looking at echinococcosis specifically, the most commonly applied spatial analysis algorithms are choropleth mapping (Brundu, et. al, 2014) (Berke, 2001) (Berke, 2005) as well as hot spot and dot-based prevalence mapping (Larrieu et. al, 2013). These analysis methods have been used to establish the environmental factors which may influence the disease spread, as well as visualize the relative rates of infection in different regions. Recently, choropleth mapping has been used to depict the relative infection rates in different countries in order to map the global distribution of alveolar and cystic echinococcosis (Deplazes et. al, 2017). The research has been conducted based on reported results of testing dogs and foxes, and thus it cannot be directly related to infections of humans. A somewhat more advanced analysis (however vastly more limited in scope) has been performed in the Xiji County of China's Ningxia Hui Autonomous Region (NHAR), where landscape profiles have been analysed to select suitable habitats for conducting a study of echinococcus occurrence in domestic dogs (Liu C. et. al, 2018). The study has shown that echinococcosis was prevalent in mountainous regions and relatively scarce on the plateau, which was attributed to the occurrence of small mammal carriers. The risk of transmission to humans has not been discussed in much detail, however it was deemed to be highest in rural regions, where domestic dogs are malnourished and allowed to roam freely. Cases of human cystic echinococcosis have also been spatially analysed in the China Sichuan Province (Liu L. et al., 2018), where choropleth mapping was used to depict spatial autocorrelation and cluster analysis of echinococcus occurrence in different towns of the region. The results have shown five spatial clusters with significantly higher prevalence rates. In comparison to other places in the region, these areas have been characterized with a prevalence of grasslands, as well as with a higher number of owned and stray dogs. These findings were corroborated by an independent GIS-based study of the prevalence of human cystic echinococcosis in Western China (Huang et al. 2018). A regression analysis of human CE cases obtained from literature has shown that prevalence of echinococcosis shows positive spatial correlation with grassland area and population illiteracy rate, and negative correlation with population density, Gross Domestic Product (GDP) and Land Surface Temperature (LST). Recently, a similar analysis has been performed for the southwestern Iran province of Fars (Ghatee et. al, 2020). The research, performed in a different climate and for partially dissimilar list of potential disease spread factors, has revealed positive spatial correlation of human CE with density of livestock, location of irrigated

farmlands and low terrain slope. In contrast to research performed in China, the study has shown no relation between LST values and human EC occurrences.

Summing up, the presented study of the application of GIS and spatial analysis to research of echinococcosis cases has shown that a system dedicated for such a task should provide a certain set of analytic functionalities, including dot-based prevalence mapping, hot spot mapping, choropleth mapping and spatial autocorrelation as well as tools for comparison with other types of spatial data such as land use and terrain height. Moreover, the system should be well integrated into the workflow of clinicians, providing support for collection and management of patient data, including information on initial interviews and diagnosis as well as treatment history and results.

Results

This section presents the software architecture of the proposed solution, describes the used datasets and processing methods as well as illustrates the application of the developed system to mapping and analysis of echinococcosis cases in Poland.

System architecture

The proposed architecture of the Web-GIS for mapping and analysis of echinococcosis is shown in Figure 1.

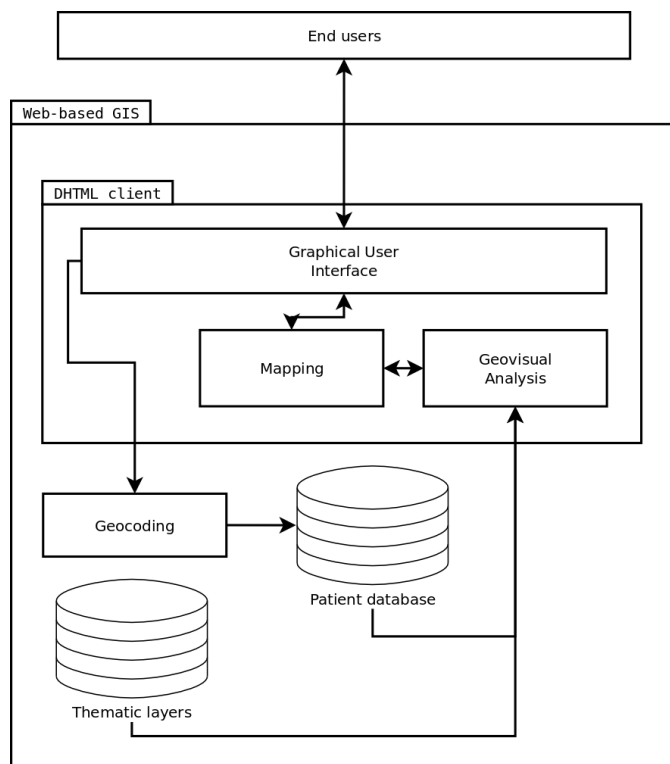


Fig. 1: Architecture of the proposed system for mapping and analysis of echinococcosis occurrences

The proposed Web-GIS is based on the client-server architecture and uses the thick client approach.

Access to system functionalities is provided through the Graphical User Interface (GUI) module, which uses PHP, HTML and Javascript to provide a set of easy to use data management tools. These include forms for input of information such as patient data and disease history, applied medicaments, performed tests and their results, treatment types and their duration, as well as tools for database query which include searching and filtering the results. Aside from tabular data provided by the user, the system database can also hold raster data formats such as scans of test results. Because the database contains potentially sensitive patient data, such as name, address, weight, height or type of infected organ, access to the system is restricted through an account hierarchy which recognizes three types of users. The administrator account provides full access to the entire system, and enables management of other user accounts. Regular user accounts are set up for all doctors working with EC cases. Each clinician can view all the patient data and perform analyses, however they may enter and edit data regarding their patients only. Limited accounts are reserved for hospital interns, who are allowed to browse collected data, but have no means of modifying it.

The server also hosts a collection of Thematic layers, which serve as background data and may be used by the Geovisual Analysis module for research purposes. The server is primarily used for data storage, with the exception of the Geocoding module, which performs a look-up of the patient's address and converts it into geographic coordinates using technology of the Open Source Nominatim project (Hoffmann, 2018). The module takes up about 10 GB of disk space and requires less than 2 GB of RAM to process. Considering that the Patient and Thematic layer databases take up less than 100 MB of disk space and require considerably less RAM, the system's hardware requirements for the server are very low by modern standards.

The geospatial information collected in the system databases may be displayed by means of the Mapping module, which is based on the Open Source OpenLayers library (<https://openlayers.org>). OpenLayers is a modern web mapping Javascript library which uses HTML5 and CSS3 features such as canvas and WebGL data rendering to provide fast and efficient visualization of geospatial data. Applied optimizations such as buffering vector features in raster format and caching batches of drawing instructions allow the library to maintain good performance even when displaying relatively large vector datasets (Kulawiak et al., 2019). In addition, OpenLayers can import and display geospatial data from a variety of standard sources, including open data formats such as KML and GeoJSON as well as web-based dissemination protocols such as OGC WMS and Web Feature Service (WFS). Imported features may be displayed using custom styles created using HTML and CSS, and their spatial characteristics may be investigated using built-in data visualization methods such as feature clustering and hot-spot mapping. Moreover, OpenLayers is written in pure Javascript, which not only minimizes application loading time, but also enables its easy integration with various types of GUI and simplifies the process of enhancing and expanding its functionalities with the use of external libraries. The latter is particularly important as the Mapping module performs dynamic data exchange with the Geovisual Analysis module.

Processing methods

The system's Geovisual Analysis module implements advanced methods of spatial analysis in addition to providing a standard set of geoprocessing tools. The latter are realized by means of the Open Source Javascript Topology Suite (JSTS) library, which is an ECMAScript 2015 port of the well established Java Topology Suite (JTS) (Harrtell, 2020). The library provides a set of spatial predicates and functions for processing geometry conforming to the Simple Features Specification for SQL published by the Open Geospatial Consortium. In consequence, it enables convenient computation of overlay functions such as intersection, difference or union as well as operations of buffering or distance measurement. The operations may be performed on standard geometric features such as point, polyline or polygon, as well as their collections (MultiPoint, MultiLineString, MultiPolygon). As part of the system's Geovisual analysis module, JSTS is used to solve vector feature problems such as point-in-polygon, which are substantial in implementation of the spatial analysis methods. The spatial analysis methods implemented by the module include choropleth mapping and spatial autocorrelation of patient data. The former is a method of mapping spatial statistics data, where the occurrences of a given phenomenon (such as wildfires or individuals earning above-average wages) are aggregated within given spatial boundaries such as districts, cities, states or countries. Every boundary is then assigned a colour which represents the number of such occurrences within this particular boundary. The colours are taken from a pre-computed scale, which is delimited by the smallest concentration of occurrences from one end, and the largest one from the other. This means that if a greyscale bar is used for representation, the smallest number of occurrences would be represented by white colour, the largest one would be represented by black colour, and all other areas would be assigned different shades of grey (Wright, 1938). Spatial autocorrelation, on the other hand, is an extension of the mathematical measure of temporal autocorrelation, in which consecutive values of the same variable are similar to each other. In the geographical context, spatial autocorrelation measures the extent to which neighbouring values of a given variable are similar to one another. While temporal autocorrelation is measured in the function of time, spatial autocorrelation is measured in the function of distance, and is usually computed for values measured within given boundaries (such as districts, cities, states or countries). One of the most prominent measures of spatial autocorrelation is Moran's I (Anselin, 1995), which produces values in the range (-1; 1). Lower values indicate negative spatial autocorrelation, higher values indicate positive spatial autocorrelation, and values near zero suggest no spatial autocorrelation (meaning that values in neighbouring boundaries are distributed randomly). The spatial autocorrelation algorithm used by the system's Geovisual Analysis module is based on the Javascript implementation of Moran's I provided by Huang (2018). The final data analysis algorithm provided by the Geovisual Analysis module is heat mapping (Wilkinson & Friendly, 2009). A heat map represents the magnitude of a spatial phenomenon (such as population density or radio signal coverage) by means of colour intensity. In a heat map, a single point of data is represented by a structure resembling a two-dimensional Gaussian distribution, where its amplitude is determined by the value of the analysed variable at this point. The depictions of neighbouring points are aggregated within a given distance (eg. a matrix of screen pixels), producing more visually significant spots. Due to this characteristic, heat mapping is often used for representing the spatial density of a given phenomenon. The implementation



of heat mapping used in the system's Geovisual Analysis module is provided by the OpenLayers library.

Datasets used by the system

The available datasets describe the area of Poland and include administrative boundaries, terrain type and height. The former have been obtained from the Polish Central Office of Geodesic and Cartographic Documentation (GUGIK, 2020). The dataset contains 376 vector features consisting of 68207 vertices and measuring 3164949 bytes bytes in GeoJSON format. Terrain type data has been obtained from the CORINE Land Cover database which is a Pan-European collection of land cover and land use information made available within the Copernicus programme. Since its introduction in 1985, the database has been periodically updated for the last 30 years by visual interpretation of latest high resolution satellite images. The dataset, divided into 44 classes, exhibits a geometric accuracy of less than 25 m and a thematic accuracy of over 85% (Copernicus, 2018). For the purpose of this work, the dataset has been divided into classes describing land types known to have positive or negative associations with echinococcosis: municipal areas, forests, farmlands and grasslands as well as water bodies. The resulting dataset has been generalized to 1 km accuracy, resulting in 10363 vector features consisting of 191033 vertices and measuring 9597317 bytes in GeoJSON format. Terrain height information has been obtained from the GTOPO30 DEM (USGS, 1996). GTOPO30 is a raster dataset containing terrain height information for the entire planet with the spatial resolution of approximately 1 kilometre. The DEM has been created in 1996 by the Center for Earth Resources Observation and Science (EROS) of the U.S. Geological Survey (USGS), and is freely available for download and use. Because the state-of-the-art analysis did not reveal a direct relationship between certain values of terrain height and EC occurrence, and instead suggested a possible link to the type of terrain (such as plains or mountains), the GTOPO30 DEM has been classified into 100 m intervals representing 11 types of terrain height. The result of the classification has been vectorized, producing a dataset containing 1362 vector features consisting of 17408 vertices and measuring 810926 bytes bytes bytes in GeoJSON format. The database of addresses in Poland used by the Geocoding module is based on data obtained from OpenStreetMap through Geofabrik (2021).

The presented system has been applied to collection and analysis of echinococcosis cases in Poland by the University Center of Maritime and Tropical Medicine in Gdynia. Insights revealed by this practical application of the system are presented in the following section.

Application of the system to analysis of EC cases in Poland

Analysis of patients with echinococcosis in a spatial context is a complex and time-consuming task. It requires the doctor not only to maintain a detailed disease histories of many patients, but also to manually analyse them and apply selected cases to the map. Moreover, every change to the dataset equals a requirement for manual reanalysis. This approach is not efficient and results in a high probability of making a mistake. The presented system answers the above challenges by combining in one tool a comprehensive module for creation, edition and display of disease history with a module for

dynamic mapping of selected aspects of the disease in a geographical context. The patient database not only allows to store detailed disease histories, but also enables their display and filtering depending on characteristics such as patient age, severity of the disease or time of hospitalization. When creating a patient card, the patient's geographical location is automatically assigned to the record, which allows the user to immediately observe the patient's position on the map. This, combined with the filter capabilities of the patient database, enables spatial visualization of thematic data such as the location of patients in various stages of the disease, or the number of cases in a given city. With a single click on the top bar of the GUI, the user may switch from the Patient database to the Mapping module. The latter presents users with a multi-layer map of Poland. The available layers include a base map, obtained from the OpenStreetMap service, with overlaid administrative boundaries representing areas belonging to different cities. By default, the map also displays the locations and types of all recorded echinococcosis cases. Patients are represented by markers which vary in shape, style and colour depending on personal details. Patient's sex is represented by marker shape (circles for women, triangles for men), type of echinococcosis is displayed by marker style (alveolar echinococcosis is represented by a uniform colour, while cystic echinococcosis is represented with a line pattern), while colour indicates treatment status (patients undergoing treatment are coloured red, while patients considered fully treated are shown in green). Figure 2 presents a sample view of the Mapping module presenting cases of echinococcosis discovered in the Pomerania region of northern Poland.

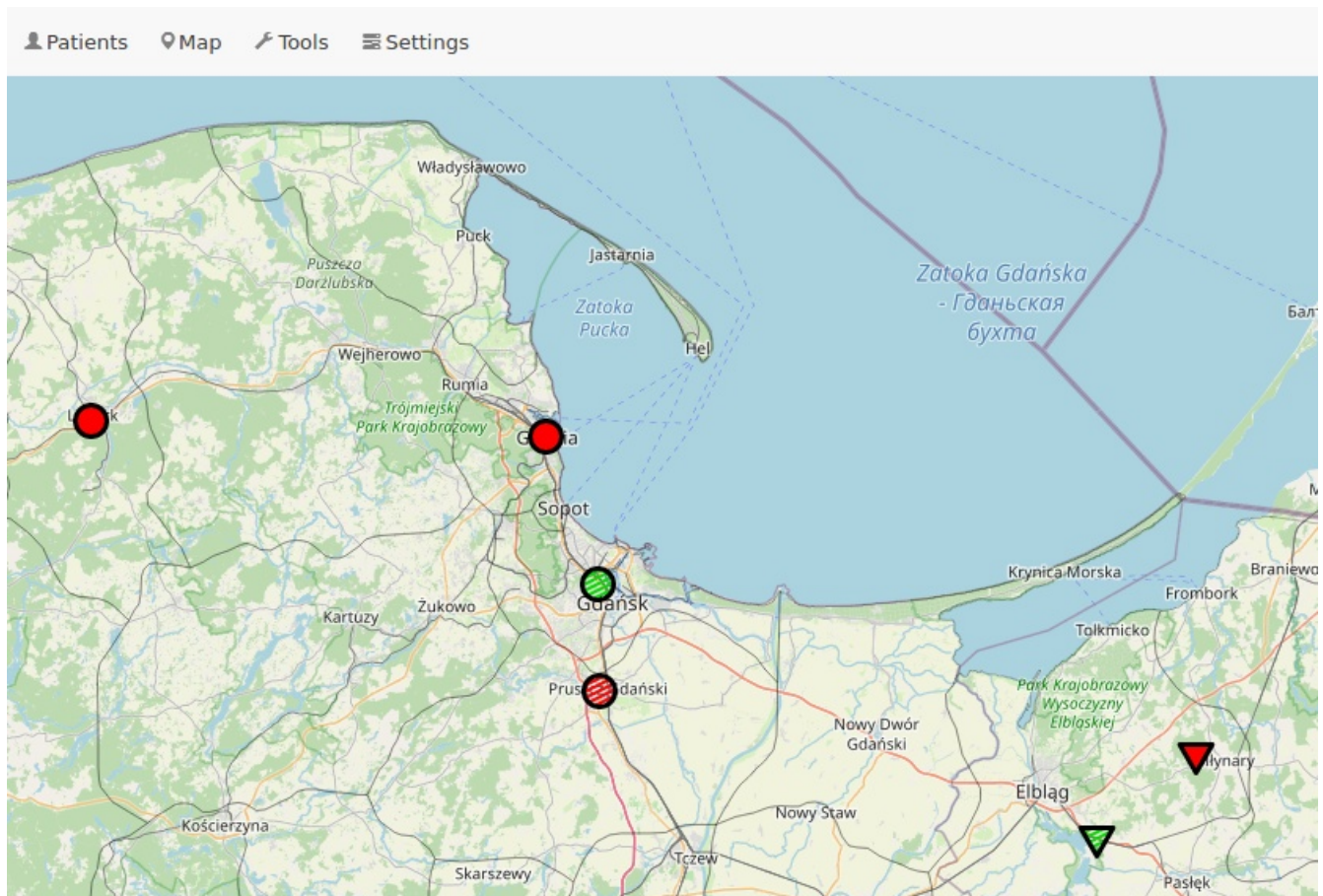


Fig. 2: Cases of echinococcosis discovered in the Pomerania region of northern Poland, displayed by the system's Mapping module. Marker shape represents patient's sex (men are marked with triangles), colour denotes whether treatment has been completed (green marks fully treated patients), while style denotes type of echinococcosis (cystic echinococcosis is represented with a line pattern)

The remaining map layers, which depict land use and terrain height, are initially hidden and need to be enabled by the user. This makes the default view much easier to interpret, while also ensuring a high level of system responsiveness by allowing the layers to asynchronously load in the background when the Mapping module is first started. These datasets become particularly useful when combined with the research tools provided by the client's Geovisual Analysis module, which include heat mapping, choropleth mapping and spatial autocorrelation.

Because Poland is located on relatively flat terrain, the provided terrain height classification layer makes it fairly simple to establish whether the collected EC occurrences are in any way related either to a given terrain type (such as highlands, plains or mountains) or a given range of terrain heights. Figure 3 presents a heat map of EC cases overlaid on different terrain height classes. As expected due to the topology of Poland's terrain, the majority of reported EC infections took place on plains located no more than 200 m above the sea level.

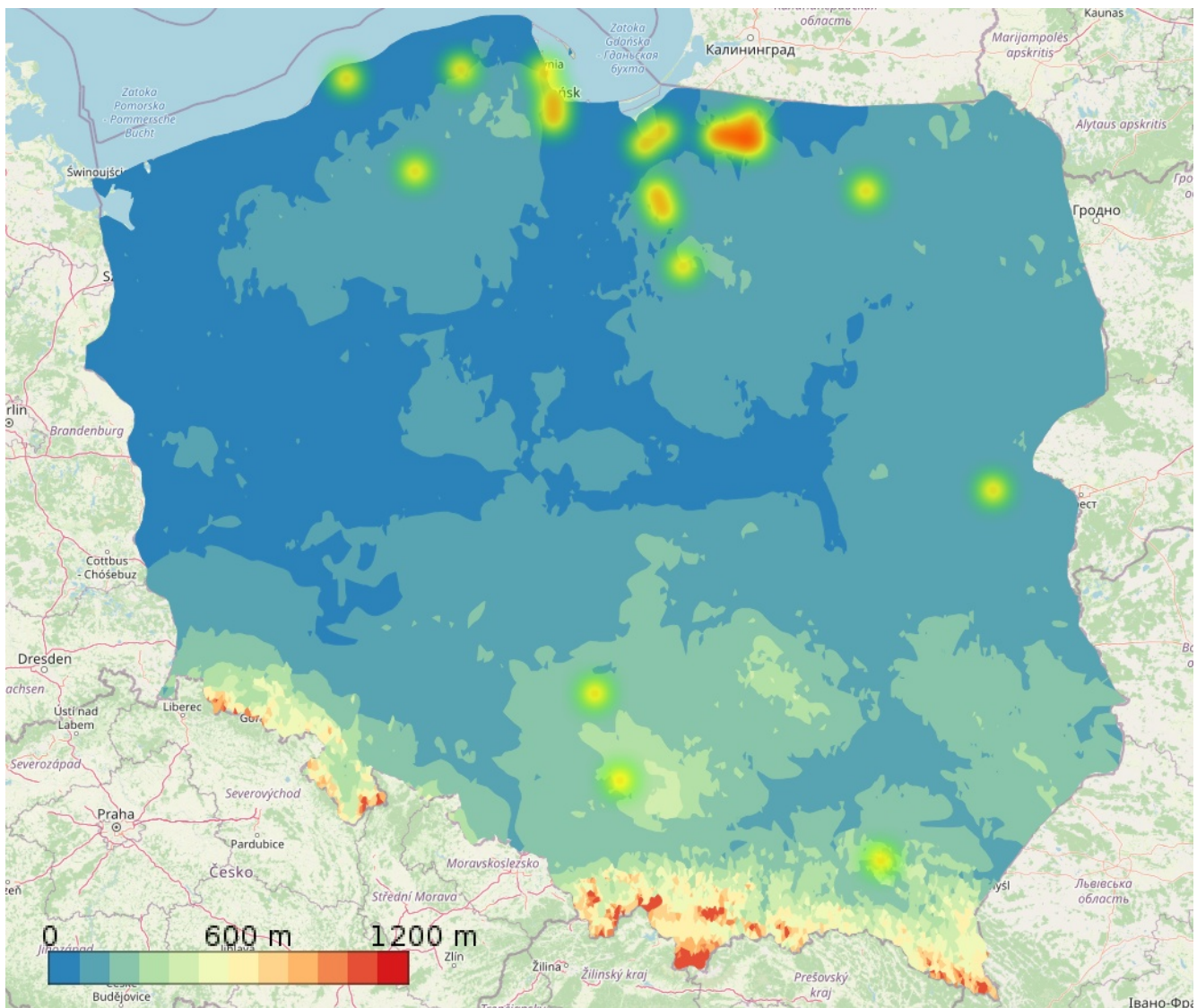


Fig. 3: A heat map of EC cases overlaid on different terrain height classes. As it can be seen, the majority of reported EC infections took place on plains located no more than 200 m above the sea level

While analysis of EC cases in relationship to terrain characteristics is important from the scientific viewpoint, the practical medical approach requires quick identification of new infections and limiting their number. For this purpose it is necessary to identify municipal areas particularly prone to infections and implement preventive measures, eg. in the form of informative campaigns. For this purpose, the system provides several tools, including choropleth mapping and spatial autocorrelation. Figure 4 presents a sample choropleth map of EC infections in different administrative areas of the country. The number of recorded EC cases is represented by a colour scale spanning from green (no cases) through yellow (moderate number of cases) to red (the largest concentration of EC cases).

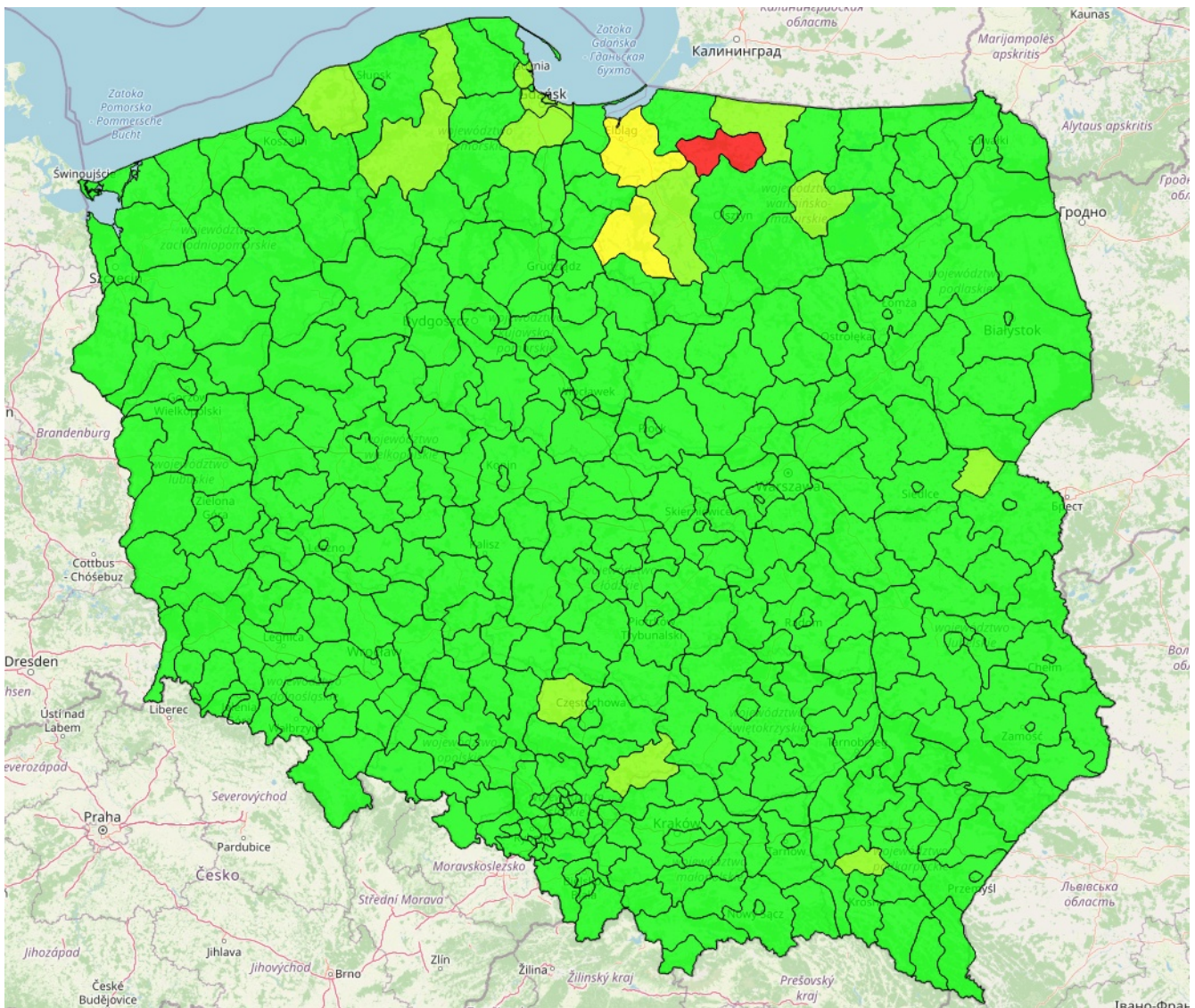


Fig. 4: The system enables spatial analysis of collected echinococcosis cases by means of a choropleth map. The largest number of EC cases has been recorded in the city of Lidzbark (marked in red), while significantly large concentrations may also be found in the cities of Elbląg and Ilawa (shown in yellow)

As it can be seen in Figure 4, the country-wide spread of EC cases looks substantially different when analysed in the context of administrative areas. The EC concentrations which may be found in Figure 3 (heat map) in the northernmost part of the country are divided into several city areas, which reduces their significance. On the other hand, the concentrations in the north-eastern area (cities of Elbląg, Ilawa and particularly Lidzbark) remain meaningful. Their significance may be further analysed with the help of spatial autocorrelation. The system's Geovisual Analysis module provides a tool which uses points-in-polygon data (which also constitutes the choropleth layer) to calculate local Moran's I for every administrative area and returns identified clusters of significantly high values, which are depicted

on a dedicated thematic layer.

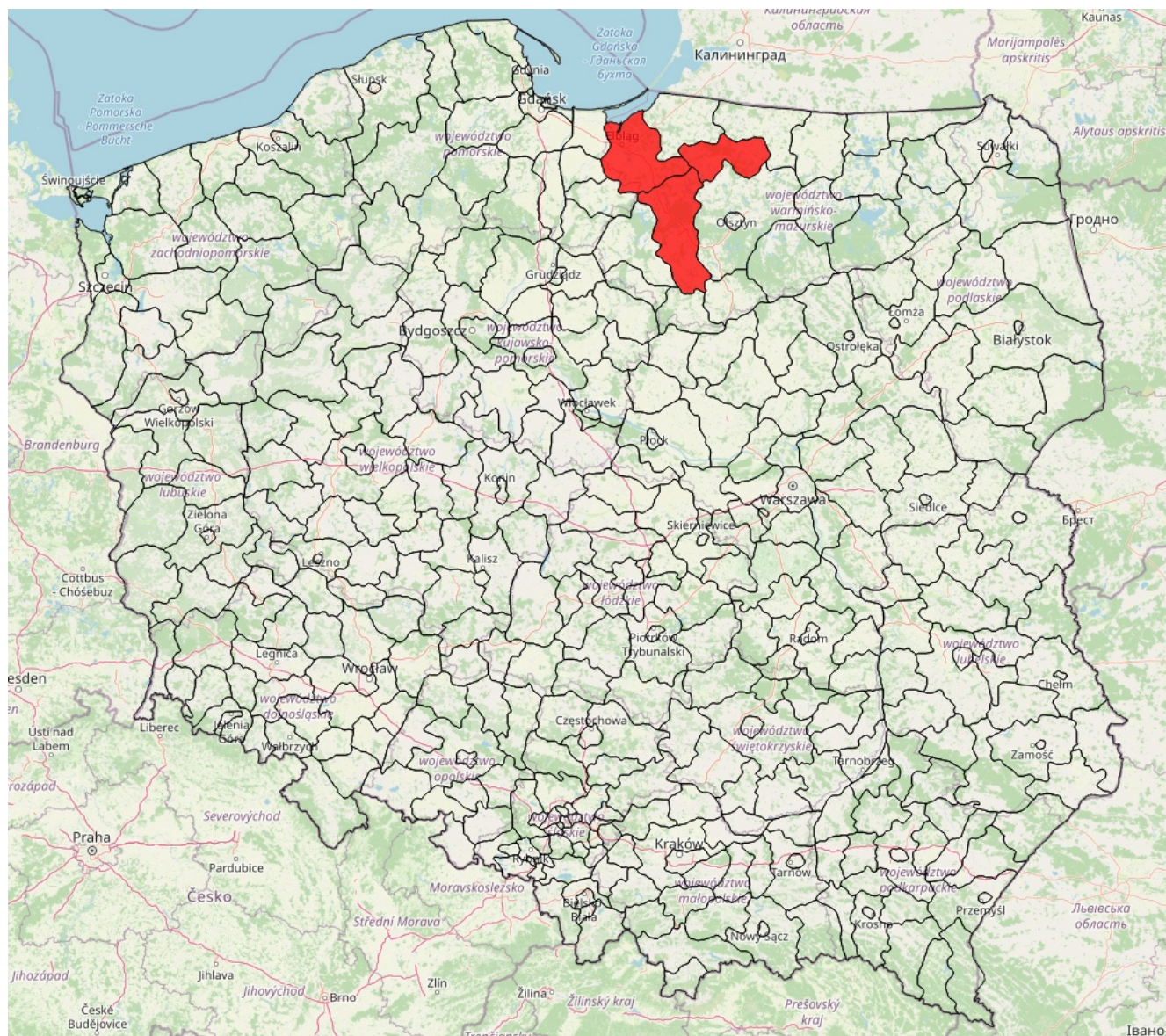


Fig. 5: Cluster of significantly high neighbouring values of local Moran's I calculated for the number of EC cases reported within the boundaries of every city

Since reported EC infections are not prevalent on the entire area of Poland, and instead are distributed rather sparsely, there is little point in analysing the spatial clustering of low infection rates. For this purpose the system's spatial autocorrelation tool focuses on presenting spatial clustering of cities with high infection rates, which may provide the end user with new insights. Figure 5 presents a cluster map of high values of local Moran's I calculated for the number of EC cases reported within the boundaries of every city. The identified cluster consists of the neighbouring cities of Lidzbarsk, Elbląg and Ostroda. While the latter contains a smaller number of reported EC infections, it is surrounded by Lidzbarsk, Elbląg and Ilawa which all have significant numbers of infections. In consequence, the provided tool

has identified a potentially interesting and not immediately obvious spatial autocorrelation cluster, which warrants further research. Such an investigation may be performed with the use of other tools provided by the system. A detailed analysis of this area, conducted with the use of the system's land use layer, reveals that it consists of several relatively small urban centres, surrounded with swaths of rural and forest areas. A close-up view of this area as seen on the land use layer is shown in Figure 6.

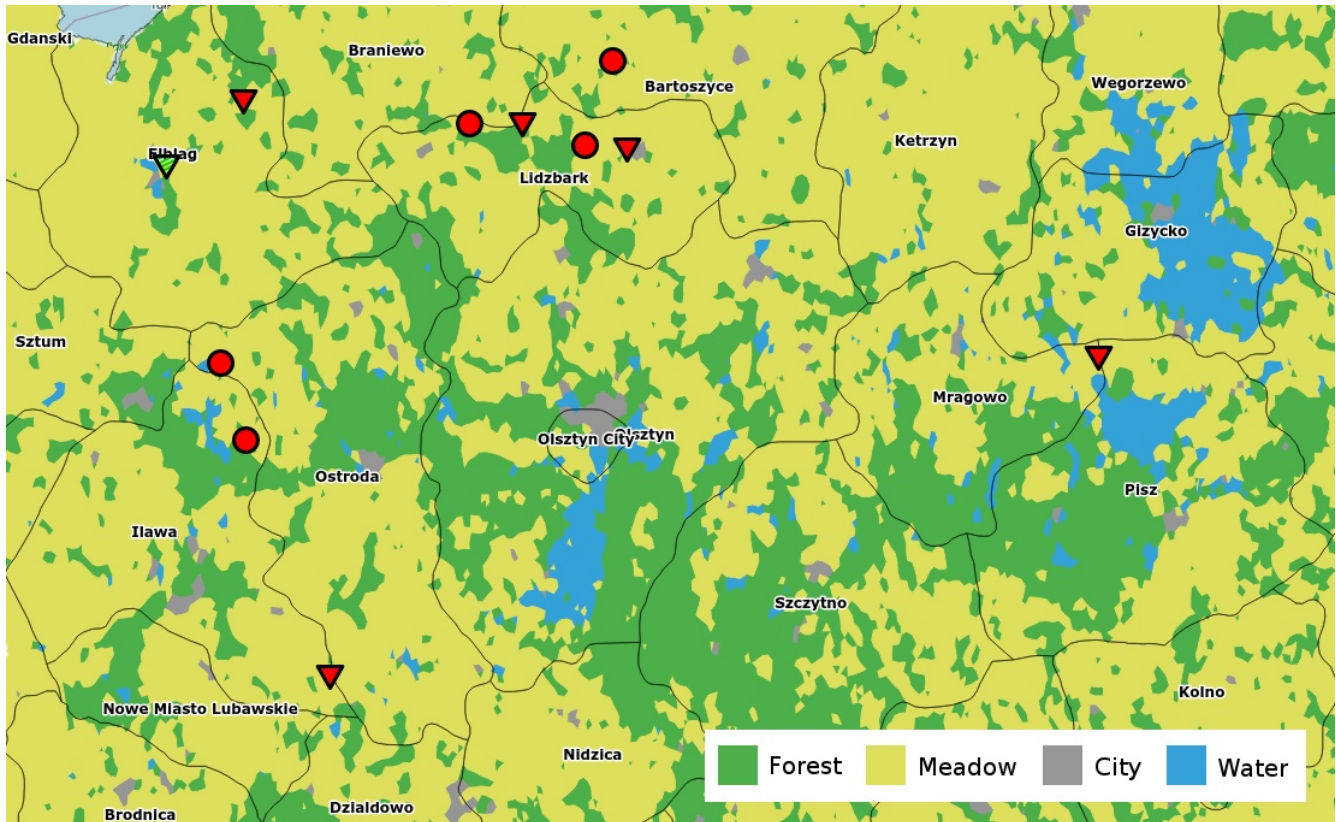


Fig. 6: Dot-based prevalence of EC cases in north-eastern Poland overlaid on different land use classes. As it can be seen, the majority of reported EC infections took place in rural areas

As it can be seen in Figure 6, virtually all EC infections reported in Lidzbark, Elblag, Iława, Ostroda, Bartoszyce and Mragowo took place in rural areas, far away from local city centres. When analysing the entirety of the dataset collected thus far in the system, it becomes clear that the vast majority of reported EC cases take place in such environments. Moreover, most of the remaining infections, which have been registered in large agglomerations, turn out to be located near forests. This is generally consistent with previous reports (Golab and Czarkowski, 2014; Waloch, 2012) and may point to a deficit in public awareness of EC and/or methods of its prevention. Analysis of EC data collected in the system thus far suggests that an EC prevention campaign targeted at the cities of Lidzbark, Elblag and Iława could significantly help in reducing the overall number of infections in the country. This being said, it should be noted that the relatively low number of EC observations currently present in the system's database does not allow to make general assumptions regarding the nature of echinococcosis infections in Poland.

Discussion

This section discusses the obtained results, including system architecture, functionality and employed data modelling techniques, in the context of the current state-of-the-art.

Discussion of system functionality

As it has been discussed previously, studies involving spatial mapping of echinococcosis have thus far been performed manually with the help of standard GIS applications. The majority of those studies have involved choropleth mapping (Deplazes et. al, 2017)(Brundu, et. al, 2014) (Berke, 2005)(Berke, 2001), while some also used hot spot (Ghatee et. al, 2020)(Huang et al., 2018) and dot-based prevalence mapping (Larrieu et. al, 2013). Aside from administrative boundaries, some of the studies also analysed EC infection locations in the context of land use (Huang et al. 2018) and terrain height (Ghatee et. al, 2020)(Liu C. et. al, 2018). In all aforementioned cases, the medical data needed to be manually collected, geocoded and pre-processed into a compatible format for the purposes of mapping. In comparison, the presented system constitutes a dedicated solution, which has been closely integrated within the workflow of clinicians. It supports the entire process of EC analysis, from collecting patient interview and medical tests data, through monitoring the treatment process, to mapping and analysis of results. Moreover, since the mapping process is automated, the system makes it possible to dynamically filter patient data depending on the values of attributes such as age, severity of the disease or length of hospitalization, with the results being immediately reflected on the map. The presented system relieves medical researchers of time-consuming tasks by providing automatic geocoding of collected information as well as delivering all relevant datasets and analysis algorithms in one user-friendly package. Additionally, the system also provides a novel method of mapping patient data, which exploits the shape, style and colour of a marker to convey information regarding patient's sex, infection type and treatment status, respectively. Thus, in the context of EC mapping, the system provides all state-of-the-art tools and methods, integrated alongside automated geocoding and novel mapping techniques into a single package designed to fit the needs of clinicians. In fact, interviews with UCMTM staff have revealed that the developed Web-GIS has become a useful tool for streamlining the process of diagnosis and treatment of echinococcosis cases. Its introduction has helped in systematizing patients, including identification of risk groups which required more careful observation. Due to it being closely integrated into the existing medical workflow, the system has made it possible to create and maintain detailed medical records in an organized manner, which is particularly important due to the EC treatment process often taking many years. Moreover, the provided analytical tools have shown to be useful during medical research. This was to be expected, as the analytical functionality of the presented system has been designed in the general context of GIS-based disease mapping applications. Recent applications of GIS tools to disease mapping employed heat mapping (Rasam & Shariff, 2019), land cover analysis (Dlamini, 2020) choropleth mapping (Zahrieh et al., 2020) as well as choropleth combined with heat mapping via kernel density estimation (Murad & Khashoggi, 2020). Some researchers have also employed data interpolation by means of Inverse Distance Weighting (IDW) (Belief, 2018) and Kriging (Paireau et al., 2018) as well as time series plots (Dong et al., 2020). In this

context, the system provides all state-of-the-art data analysis and processing methods aside from data interpolation and time series plots. This limitation is by design, however, as these tools were not deemed necessary in the current version of the system due to the characteristics of the analysed disease. In the case of time series plots, they can be very useful in the analysis of fast-spreading diseases, however EC infections are generally counted on a yearly basis. This means that such plots are constructed very rarely, and although it would be technically feasible, there was no need for them to be automatically constructed and updated by the system. However, it is possible that after several more years of expanding the system database plotting a time series of infection rates may reveal new insights, at which point this functionality may be introduced. As far as data interpolation is concerned, such techniques work best with a large number of evenly spaced input datasets, whereas EC infections are relatively sparse (particularly in developed countries) and have a tendency for clustering around areas inhabited by intermediate hosts. Thus, the accuracy of EC interpolations would be limited, at least until a sufficiently large and spatially varied infection dataset is collected. It should be noted, however, that there are no technical obstacles in the way of implementing this functionality in the future, as it has been shown that data interpolation in DHTML can work well even on a relatively weak client hardware (Kulawiak & Wycinka, 2017).

Discussion of system architecture

As far as dedicated Web-GIS for disease mapping are concerned, the latest works primarily feature a client-server architecture with a thin DHTML client providing data mapping and visualization, combined with a server which hosts the patient database and data processing functionalities. In all cases, maps are prepared server-side and sent to the client in raster format. Interactivity is thus limited to browsing and zooming the map and (occasionally) clicking on interface elements in order to display additional information (such as regional infection statistics in the case of the COVID-19 mapping services (Boulos & Geraghty 2020)). Moreover, all analytical functionalities (such as spatial autocorrelation (He, 2021) temporal data animation (Savini et al., 2018), density estimation (Delmelle et al., 2014) and choropleth mapping (Lupa et al. 2017)) are also realized on the server. This approach allows the systems to run on very weak client machines, at the cost of requiring a relatively powerful server infrastructure. Moreover, although existing systems attempt to collect some data automatically from online sources, large parts of their databases still require manual data collection and update (He, 2021)(Dong et al., 2020)(Savini et al., 2018)(Lupa et al. 2017). In comparison, the presented system employs a thick-client architecture which places very little strain on the server, and instead runs all of its data presentation and analysis algorithms on the client machines. Moreover, the thick-client is implemented strictly with the use of DHTML, which makes it independent from the software and hardware platform of the client machine (provided that the client is equipped with a modern web browser). While this could potentially have a negative impact on user experience depending on the performance of the client machines, in practice even computers which are considered low-end by modern standards have enough processing power to sufficiently support such an architecture, if it has been properly optimized (Kulawiak et al., 2019). This is why by default the presented system's

Mapping module only displays the base map and the patients layer, which are both lightweight and do not impact startup time or display performance in any noticeable way. Since OpenLayers supports asynchronous loading of resources, the remaining layers are downloaded in the background and displayed on user demand. Moreover, since OpenLayers also supports buffering vector features in raster format, the system does not need to re-render the vector layers on every update of the map viewport. As a result, the only noticeable slowdown in data display occurs during the first render of the largest dataset in the system, which is the terrain type layer. On average this operation takes about 1.5 s on a relatively old Core i7-3630QM, a notebook-class processor produced in 2012. Most of the system's spatial analysis operations are also instantaneous, with the exception of calculating Moran's I which on average requires over 6 s to complete. With all rendering and analysis operations running on the client, the strain on the server is limited to serving and receiving relatively small amounts of data (including datasets, the entire DHTML client is less than 15 MB in size, and database update requests usually measure between 2 and 6 KB) and performing occasional geocoding operations (which usually complete within several seconds).

Aside from performance concerns, a potential disadvantage of the employed thick-client architecture could lie in the dissemination of the results of performed analyses. Since those are performed on the client, they cannot be easily saved on the server for convenient presentation to other researchers. While this could potentially be an issue in larger and spatially dispersed research teams, it did not pose a problem for the small and geographically focused team at UCMTM. On the whole, the system architecture, implemented and delivered in the SaaS model, has proven to be easy to maintain and update, while at the same time providing considerable data processing and analysis capabilities. Being written purely in DHTML and not dependant on plugins, the system has shown to operate without issues on all leading Web browsers. Moreover, with no reliance on third-party services, the system has remained fully functional throughout all encountered internet access outages (with the only issue being the unavailability of the OSM background layer). This includes the automatic geocoding functionality, which is realized by an integrated address geodatabase.

Discussion of employed modelling approaches

In order to model the spatial characteristics of EC cases in Poland, the system uses several techniques, including styled markers, choropleth mapping, heat mapping and spatial autocorrelation. All of them have advantages and disadvantages, which are discussed in this section.

Styled markers

The principles of geovisual analytics advocate the use of shape, style and colour for the purpose of conveying tabular information in graphical form (Andrienko et al., 2007). Such an approach facilitates the use of a person's visual cognition capabilities for the purpose of broadening the stream of data simultaneously delivered to the brain. In practical terms, it allows the end user to obtain a certain number of informations without specifically looking for them (which would involve interacting with the GUI to search the system database). The system uses this methodology to represent the patient's

sex, infection type and treatment status by means of marker shape, style and colour, respectively. While a lot of thought has gone into making those associations possibly intuitive, interviews held with clinicians have shown that the scheme can be initially confusing and generally requires a period of adjustment, during which users need to get acquainted with the meaning of at least one of the aforementioned characteristics. Moreover, long-term observations indicate that a lengthy delay between consecutive interactions with the system may result in the need to re-adjust oneself to the said depiction scheme. Despite this, in general users reacted favourably to this method of data visualization, and none of the respondents indicated a preference for a more traditional uniform way of visualizing EC cases.

Choropleth mapping

The choropleth map aggregates occurrences of the investigated phenomenon within given spatial boundaries. The system uses it to depict the number of EC infections in every city in Poland. An innate limitation of this approach is that it firmly frames the depicted situation within the chosen administrative boundaries. Therefore it is not possible to analyse the conditions within any given city, nor is it possible to investigate and compare the situation in different Voievodships. It also produces abrupt changes at the boundaries of neighbouring regions, which may give a false impression regarding the spatial distribution of patients. Moreover, this method of data analysis does not take into account variables such as population density. In consequence a small but densely populated region could be given a greater visual emphasis than a larger but sparsely populated area, potentially leading to inaccurate conclusions regarding the percentage of infected population in both areas. Finally, although choropleth maps are a good way of depicting the characteristics of different regions, they are not recommended for conveying actual data values. This is particularly true of similar values, which are often displayed with undistinguishable shades of colour. In the above context, It is important to use the choropleth map for the appropriate purpose, which (in the case of the presented system) is the comparison of EC infection rates in different cities. Moreover, the conclusions of such an analysis should be interpreted in the context of other available data, such as terrain use. As shown in the Results section, these differences may be caused by the effectiveness of local education or health services as well as by the proximity of natural habitats of intermediate hosts. This is also why the system allows for analysis of the choropleth map in conjunction with overlaid heat map or patient markers, which will provide information regarding the spatial spread and density of individual EC cases (as this information cannot be derived from the choropleth map alone).

Heat mapping

Heat maps are widely used for depicting the clustering of spatially distributed values of a given variable, eg. the occurrences of a given phenomenon. They convey the differences in the magnitude of the phenomenon at different clusters by graduating the intensity of their colour representation. Although heat maps are a good way of depicting value clustering, they generally fail to properly represent both the values themselves as well as the differences between them. This is particularly evident in the case of similar values, which are often displayed with undistinguishable shades of colour.

Moreover, using a single linear colour scale for the heat map means that a large group of values in a small area may easily overshadow even a large number of smaller and more dispersed groups. While this issue could be somewhat alleviated by using a scale consisting of several consecutive colours, this is complicated by the fact that humans demonstrate variable sensitivity to different colours, which makes construction of a continuous and coherent colour scale rather difficult (Cavonius & Estevez, 1975). As a result, heat maps should be primarily used for detecting of spatial feature clusters, and particular consideration should be given to the applied colour scales. This is why the heat map used by the presented system employs a colour scale consisting of green, yellow and red, which are among the most easily distinguishable colours for the majority of humans (Green, 1968). Moreover, the heat map available in the presented system functions as an overlay, which allows for its use in conjunction with other datasets. This way the spatial density of EC cases may be investigated eg. in relation to administrative boundaries, land use or terrain height, providing additional context to the detected spatial clusters.

Spatial autocorrelation

Spatial autocorrelation is a method of measuring the similarity of values of a spatial variable recorded at neighbouring locations. The existence of autocorrelation (be it positive or negative) may indicate the existence of an external factor or phenomenon influencing the spatial spread of values. As a statistical tool, spatial autocorrelation is most effective when operating on datasets which exhibit a consistent spatial pattern across the entire study area (Legendre & Fortin, 1989), which is virtually impossible to achieve in a constantly evolving database of a discontinuous spatial phenomena such as infectious disease occurrences. Moreover, Moran's I specifically is known to disregard local spatial dependence in favour of global autocorrelation, which means that certain small-scale relations may remain undetected (Getis & Ord, 2010). While these are valid points from the statistical point of view, it is important to note that in the presented system spatial autocorrelation is used not for the purpose of statistical analysis, but as a tool for supporting clinicians in their investigation of a disease. In this context, spatial autocorrelation is used to detect possible spatial dependencies between groups of infections, and not to study the nature of these dependencies in relation to the entire area of interest. Thus, the fact that some relations may remain undetected does not discredit the dependencies that the tool has managed to identify. As it has been shown the Results section, such a high-level approach can have its merits, given that the results are critically analysed with the use of other available data and methods.

Conclusions

The goal of this work was to implement a novel architecture of a Web-GIS dedicated to disease mapping and analysis. The architecture has been realized in the form of a system applied to remote mapping and analysis of echinococcosis, and has been tested in a real-world environment by the staff of the University Center of Maritime and Tropical Medicine located in the city of Gdynia in northern Poland. Because the system has been designed in order to fit into the existing medical workflow, it could be seamlessly deployed and did not require a lot of user training. According to expectations, the

introduction of the system has helped to better organize patient records and streamlined the process of data management.

The system has been equipped with state-of-the-art tools for EC analysis, including heat mapping, choropleth mapping and spatial autocorrelation. Moreover, it also integrates the datasets needed for investigating the local characteristics of EC infections, including land use and terrain height data. Currently, none of the existing EC analysis systems are known to provide such a comprehensive analysis toolset. In addition, the system provides an innovative way of presenting patient data, including sex, infection type and treatment status, via the style of their map marker, which has been generally well received by the users. This being said, while the provided analysis tools and reference datasets enable discovery of infection trends and investigation of their causes, the results of their application need to be critically assessed by experienced researchers, and therefore they shouldn't be used in a fully automated manner.

The system has been built around a novel thick-client architecture in which data analysis is performed on a platform-independent DHTML client, minimizing the strain on the server and thus ensuring a balanced exploitation of hospital resources. In practice, both the client and the server components have been successfully deployed and tested on relatively old hardware, showing that the analysis algorithms performed on the client provide acceptable performance even on a laptop manufactured in 2012. Furthermore, the proposed architecture has been implemented in the SaaS model, allowing the system to be centrally maintained and updated by a single administrator. At the time of writing the proposed architecture is still unique in the realm of disease mapping systems. However, the current trends of increasingly more interactive web-based mapping clients and constantly improving computer hardware, combined with mentioned benefits of thick DHTML clients would indicate that such architectures will likely become more and more common in the upcoming years.

Initial results of applying the system to analysis and mapping of EC cases in Poland have confirmed the identification of Warmia-Mazury and Podlasie as endemic regions by prior works, and expanded upon the current knowledge by specifically identifying a significant concentration of EC cases in the cities of Lidzbark, Elblag and Ilawa. Narrowing down the endemic areas from Voievodship to city level may be of vital importance, as screening endemic areas is a crucial element of EC prevention and early diagnosis and treatment has a significant impact on the prognosis.

Future work could involve the introduction of time series plots of infection cases integrated with the patient search mechanism, which would allow for temporal analysis of infection rates eg. for different types of patients, locations, or disease types (CE vs AE). Moreover, integration of a data interpolation method (eg. IDW) and development of a convenient way of sharing the results of performed analyses between research team members would constitute a universal template for a comprehensive and cost-effective tool for analysis and mapping of various types of diseases.

Declarations

Funding

This work received no additional funding.

Competing interests

The authors declare that they have no conflict of interest.

Availability of data and materials

The datasets generated and/or analysed during the study are not publicly available because they contain sensitive patient data.

Code availability

Not applicable.

Author Contributions: Conceptualization: M.K. and N.K.; methodology: M.K. and N.K.; software: M.K.; validation: M.S and K.S; resources: M.S.; writing—original draft preparation M.K.; writing—review and editing N.K. and K.S.; supervision: K.S. All authors have read and agreed to the published version of the manuscript.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

No details, images or videos of an individual person are available in this manuscript.

References

- Adams-Prassl, A., Boneva, T., Golin, M., Rauh, C. 2020. The impact of the coronavirus lockdown on mental health: evidence from the US.
- Agrawal, S., Gupta, R. D. 2020. Development of SOA-based WebGIS framework for education sector. *Arabian Journal of Geosciences*, 13(13), 1-20.
- Alcântara, E., Mantovani, J., Rotta, L., Park, E., Rodrigues, T., Carvalho, F. C., Souza Filho, C. R. 2020. Investigating spatiotemporal patterns of the COVID-19 in São Paulo State, Brazil. *Geospatial health*, 15(2).
- Allen, W. E., Altae-Tran, H., Briggs, J., Jin, X., McGee, G., Shi, A., ... & Lin, X. 2020. Population-



scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nature human behaviour*, 4(9), 972-982. DOI: <https://doi.org/10.1038/s41562-020-00944-2>

Al Huraimel, K., Alhosani, M., Kunhabdulla, S., & Stietiya, M. H. 2020. SARS-CoV-2 in the environment: Modes of transmission, early detection and potential role of pollutions. *Science of The Total Environment*, 140946.

Aman, M. A., Salman, M. S., Yunus, A. P. 2020. COVID-19 and its impact on environment: improved pollution levels during the lockdown period—a case from Ahmedabad, India. *Remote Sensing Applications: Society and Environment*, 20, 100382.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M. J., MacEachren, A., & Wrobel, S. 2007. Geovisual analytics for spatial decision support: Setting the research agenda. *International journal of geographical information science*, 21(8), 839-857.

Anselin, Luc. 1995. Local indicators of spatial association—LISA. *Geographical analysis* 27(2), pp. 93-115.

Arab-Mazar, Z., Sah, R., Rabaan, A. A., Dhama, K., & Rodriguez-Morales, A. J. 2020. Mapping the incidence of the COVID-19 hotspot in Iran—Implications for Travellers. *Travel Medicine and Infectious Disease*, 34, 101630.

Arora, S., Bhaukhandi, K. D., Mishra, P. K. 2020. Coronavirus lockdown helped the environment to bounce back. *Science of the Total Environment*, 140573.

Basu, S., Karmakar, A., Bidhan, V., Kumar, H., Brar, K., Pandit, M., Latha, N. 2020. Impact of lockdown due to COVID-19 outbreak: lifestyle changes and public health concerns in India. *Int J Ind Psychol*, 8(2).

Belief, E. 2018. GIS based spatial modeling to mapping and estimation relative risk of different diseases using inverse distance weighting (IDW) interpolation algorithm and evidential belief function (EBF)(Case study: Minor Part of Kirkuk City, Iraq). *Int J Eng Technol*, 7(4.37), 185-91.

Benreguia, B., Moumen, H. Merzoug, M.A., 2020. Tracking covid-19 by tracking infectious trajectories. *IEEE Access*, 8, pp.145242-145255.

Berke, O., 2001. Choropleth mapping of regional count data of *Echinococcus multilocularis* among red foxes in Lower Saxony, Germany. *Preventive Veterinary Medicine*, 52(2), pp.119-131.

Berke, O., 2005. Exploratory spatial relative risk mapping. *Preventive veterinary medicine*, 71(3), pp.173-182.

Blanton, J.D., Manangan, A., Manangan, J., Hanlon, C.A., Slate, D. and Rupprecht, C.E. 2006. Development of a GIS-based, real-time Internet mapping tool for rabies surveillance. *International Journal of Health Geographics*, 5(1), p.1.

Bod'ová, K., Kollár, R. 2021. Spatial scales, patterns, and positivity trends of SARS-CoV-2 pandemics in mass rapid antigen testing in Slovakia. *PloS one*, 16(8), e0256669.

Boulos, M. N. K., Geraghty, E. M. 2020. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *Int J Health Geogr* 19, 8.

Brovelli, M.A. Coetzee, S., 2021. Open Geospatial Data for Responding to the COVID-19 Challenge. *COVID-19 Pandemic, Geospatial Information, and Community Resilience: Global Applications and Lessons*, p.31.

Brundu, D., Piseddu, T., Stegel, G., Masu, G., Ledda, S. and Masala, G., 2014. Retrospective study of human cystic echinococcosis in Italy based on the analysis of hospital discharge records between 2001 and 2012. *Acta tropica*, 140, pp.91-96.

Cavonius, C. R., & Estevez, O. 1975. Contrast sensitivity of individual colour mechanisms of human vision. *The Journal of physiology*, 248(3), 649-662.

Chakraborty, A., McMillan, A. (2018). GIS and Scenario Analysis: Tools for Better Urban Planning. In *Comprehensive Geographic Information Systems* (pp. 371-380). Elsevier.

Charandabi, N. K., & Gholami, A. 2021. COVID-19 Spatiotemporal Hotspots and Prediction Based on Wavelet and Neural Network. In *COVID-19 Pandemic, Geospatial Information, and Community Resilience* (pp. 211-226). CRC Press.

Copernicus. 2018. The CORINE Land Cover database. Available at: <http://land.copernicus.eu/pan-european/corine-land-cover> [Accessed on: 01.04.2021]

Dangendorf, F., Herbst, S., Reintjes, R., Kistemann, T. 2002. Spatial patterns of diarrhoeal illnesses with regard to water supply structures—a GIS analysis. *International journal of hygiene and environmental health*, 205(3), pp.183-191.

Dawidowicz, A., Kulawiak, M. 2018. The potential of Web-GIS and geovisual analytics in the context of marine cadastre. *Survey Review*, 50(363), pp.501-512. DOI: 10.1080/00396265.2017.1328331

Delmelle, E. M., Zhu, H., Tang, W., Casas, I. (2014). A web-based geospatial toolkit for the monitoring of dengue fever. *Applied Geography*, 52, 144-152.

Deplazes, P., Rinaldi, L., Rojas, C. A., Torgerson, P. R., Harandi, M. F., Romig, T., Antolova D., Schurer J.D., Lahmar, S., Cringoli G., Magambo, J., Thompson, R.C.A., Jenkins, E. J. (2017). Global distribution of alveolar and cystic echinococcosis. *Advances in parasitology*, 95, 315-493.

Dlamini, S. N. 2020. Remote Sensing Applications in Disease Mapping. In *Remote Sensing*. IntechOpen.



Dominkovics, P., Granell, C., Pérez-Navarro, A., Casals, M., Orcau, À. and Caylà, J.A., 2011. Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in Barcelona, Spain. *International journal of health geographics*, 10(1), p.1.

Dong, E., Du, H., & Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, 20(5), 533-534.

Gao, S., Mioc, D., Anton, F., Yi, X. and Coleman, D.J., 2008. Online GIS services for mapping and sharing disease information. *International Journal of Health Geographics*, 7(1), p.1.

Gao, S., Mioc, D., Yi, X., Anton, F., Oldfield, E. and Coleman, D.J., 2009. Towards Web-based representation and processing of health information. *International Journal of Health Geographics*, 8(1), p.1.

Gawor, J. 2016. Alveolar echinococcosis in Europe and Poland. Threats to humans. *Przegląd epidemiologiczny*, 70(2), 281-288.

Geofabrik GMBH. 2021. Download OpenStreetMap data for this region: Poland. Available at: <http://download.geofabrik.de/europe/poland.html> [Accessed on 01.04.2021]

Getis, A., & Ord, J. K. 2010. The analysis of spatial association by use of distance statistics. In *Perspectives on spatial data analysis* (pp. 127-145). Springer, Berlin, Heidelberg.

Ghatee, M. A., Nikaein, K., Taylor, W. R., Karamian, M., Alidadi, H., Kanannejad, Z., ... & Pouladfar, G. 2020. Environmental, climatic and host population risk factors of human cystic echinococcosis in southwest of Iran. *BMC Public Health*, 20(1), 1-13.

Gianquintieri, L., Brovelli, M.A., Pagliosa, A., Dassi, G., Brambilla, P.M., Bonora, R., Sechi, G.M. and Caiani, E.G., 2020. Mapping Spatiotemporal Diffusion of COVID-19 in Lombardy (Italy) on the Base of Emergency Medical Services Activities. *ISPRS International Journal of Geo-Information*, 9(11), p.639.

Giraudoux, P., Pleydell, D., Raoul, F., Quéré, J.P., Wang, Q., Yang, Y., Vuitton, D.A., Qiu, J., Yang, W. and Craig, P.S., 2006. Transmission ecology of *Echinococcus multilocularis*: What are the ranges of parasite stability among various host communities in China?. *Parasitology international*, 55, pp.S237-S246.

Go, H., Gretzel, U. (2016). The role of interactive maps and spatial ability in creating virtual tourism experiences: A measurement framework.

Golab, E., Czarkowski, M.P., 2014. Echinococcosis and cysticercosis in Poland in 2012. *Przegl. Epidemiol.* 68 (2), pp. 279-282.

Green, D. G. 1968. The contrast sensitivity of the colour mechanisms of the human eye. *The Journal of physiology*, 196(2), 415-429.



GUGIK. 2020. Data from the state register of borders and areas of country territorial division units. Available at: <http://www.gugik.gov.pl/pzgif/dane-bez-oplat/dane-z-panstwowego-rejestru-granic-i-powierzchni-jednostek-podzialow-terytorialnych-kraju-prg> [Accessed on 01.04.2021].

Harrtell, B. 2020. JavaScript Topology Suite. Available at: <https://github.com/bjornharrtell/jsts> [Accessed on 10.02.2021]

He, M. 2021. Mapping of Schistosomiasis Haematobia in Zanzibar. In Sino-African Cooperation for Schistosomiasis Control in Zanzibar (pp. 123-145). Springer, Cham.

Hoffmann, S. 2018. Open-source geocoding with OpenStreetMap data. Available at: <https://nominatim.org> [Accessed on 01.04.2021]

Huang, D., Li, R., Qiu, J., Sun, X., Yuan, R., Shi, Y., Qu, Y., Niu, Y. 2018. Geographical environment factors and risk mapping of human cystic echinococcosis in Western China. *International journal of environmental research and public health*, 15(8), 1729.

Huang, Z. 2018. Local Moran's I using JS. Available at <https://github.com/zhaosongH/Local-Moran-s-I-by-JS> [Accessed on 01.04.2021]

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582(7812), 389-394.

Kelly-Hope, L., Diggle, P., Rowlingson, B.S., Gypapong, J.O., Kyelem, D., Coleman, M., Thomson, M.C., Obsomer, V., Lindsay, S.W., Hemingway, J. and Molyneux, D.H., 2004. Negative Spatial Association Between Lymphatic Filariasis and Malaria in Africa (No. 1049). Berkeley Electronic Press.

Kern, P., Bardonnnet, K., Renner, E., Auer, H., Pawlowski, Z., Ammann, R.W., Vuitton, D.A., Kern, P. 2003. European echinococcosis registry: human alveolar echinococcosis, Europe, 1982-2000. *Emerging infectious diseases* 9(3), pp.343-349.

Kistemann, T., Dangendorf, F. and Schweikart, J., 2002. New perspectives on the use of Geographical Information Systems (GIS) in environmental health sciences. *International Journal of Hygiene and Environmental Health*, 205(3), pp.169-181

Kulawiak, M., Dawidowicz, A., Pacholczyk, M. E. 2019. Analysis of server-side and client-side Web-GIS data processing methods on the example of JTS and JSTS using open data from OSM and geoportal. *Computers & Geosciences*, 129, 26-37. DOI: 10.1016/j.cageo.2019.04.011

Kulawiak, M., Kulawiak, M. 2017. Application of Web-GIS for dissemination and 3D visualization of large-volume LIDAR data. In *The rise of big spatial data* (pp. 1-12). Springer, Cham. DOI:10.1007/978-3-319-45123-7_1

Kulawiak, M, Wycinka, W. 2017. Dynamic signal strength mapping and analysis by means of mobile Geographic Information System. *Metrology and Measurement Systems Volume 24, Issue 4*, pp. 596-606. doi: 10.1515/mms-2017-0057

Kulawiak, M., Lubniewski, Z. 2014. SafeCity—A GIS-based tool profiled for supporting decision making in urban development and infrastructure protection. *Technological Forecasting and Social Change*, 89, 174-187. DOI: 10.1016/j.techfore.2013.08.031

Larrieu, E., Herrero, E., Mujica, G., Labanchi, J.L., Araya, D., Grizmodo, C., Calabro, A., Talmon, G., Ruesta, G., Perez, A. and Gatti, A., 2013. Pilot field trial of the EG95 vaccine against ovine cystic echinococcosis in Rio Negro, Argentina: Early impact and preliminary data. *Acta tropica*, 127(2), pp.143-151.

Lass A, Szostakowska B, Myjak P, Korzeniewski K. 2017. Detection of *Echinococcus multilocularis* DNA in fruit, vegetable, and mushroom samples collected in the non-endemic territory of the Pomerania province and comparison of the results with data from rural areas of the neighbouring highly endemic Warmia-Masuria province, Poland. *Acta Parasitol.* 62(2), 459-465. DOI: 10.1515/ap-2017-0053.

Legendre, P., Fortin, M. J. 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80(2), 107-138.

Li, H., Li, H., Ding, Z., Hu, Z., Chen, F., Wang, K., Peng, Z. Shen, H., 2020. Spatial statistical analysis of coronavirus disease 2019 (covid-19) in China. *Geospatial Health*, 15(1).

Liu, C. N., Xu, Y. Y., Cadavid-Restrepo, A. M., Lou, Z. Z., Yan, H. B., Li, L., Fu, B-Q., Gray, D.J., Clements, A.A., Barnes, T.S., Williams, G.M., Jia, W-Z., McManus D.P., Yang, Y. R. (2018). Estimating the prevalence of *Echinococcus* in domestic dogs in highly endemic for echinococcosis. *Infectious diseases of poverty*, 7(1), 1-9.

Liu, L., Guo, B., Li, W., Zhong, B., Yang, W., Li, S. C., Wang, Q., Zhao, X., Xu, K.-J., Qin, S.-C., Huang, Y., Yu, W.-J., He, W., Liao, S., Wang, Q. 2018. Geographic distribution of echinococcosis in Tibetan region of Sichuan Province, China. *Infectious diseases of poverty*, 7(1), 1-9.

Longley, P. A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. 2005. *Geographic information systems and science*. John Wiley & Sons.

Louis, V.R., Phalkey, R., Horstick, O., Ratanawong, P., Wilder-Smith, A., Tozan, Y. and Dambach, P., 2014. Modeling tools for dengue risk mapping—a systematic review. *International journal of health geographics*, 13(1), p.1.

Lupa, M., Samulowska, M., Chmielewski, S., Myszkowska, D., Czarnobilska, E. 2017. A concept of webgis pollen allergy mapping. *Proceedings of the 17th International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management*, 29, 1141-1148.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B. 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the total environment*, 724, 138226.

MacEachren, A. M., Crawford, S., Akella, M., & Lengerich, G. (2008). Design and implementation of a

model, web-based, GIS-enabled cancer atlas. *The Cartographic Journal*, 45(4), 246-260.

Mackey, T., Purushothaman, V., Li, J., Shah, N., Nali, M., Bardier, C., Liang, B., Cai, M., Cuomo, R. 2020. Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data infoveillance study. *JMIR public health and surveillance*, 6(2), e19509.

Marino, R., Vargas, E., Flores, M. 2021. Impacts of COVID-19 lockdown restrictions on housing and public space use and adaptation: Urban proximity, public health, and vulnerability in three Latin American cities. Taylor & Francis, 363.

McManus, D.P., Zhang, W., Li, J., Bartley, P.B. 2003. Echinococcosis. *The Lancet* 362(9392), pp.1295-1304.

Mooney, P., Juhász, L. 2020. Mapping COVID-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*, 10(2), 265-270.

Mooney, P., Grinberger, A. Y., Minghini, M., Coetzee, S., Juhasz, L., Yeboah, G. 2021. OpenStreetMap data use cases during the early months of the COVID-19 pandemic. In *COVID-19 Pandemic, Geospatial Information, and Community Resilience* (pp. 171-186). CRC Press.

Moncrieff, S., West, G., Cosford, J., Mullan, N. and Jardine, A., 2014. An open source, server-side framework for analytical web mapping and its application to health. *International Journal of Digital Earth*, 7(4), pp.294-315.

Murad, A., & Khashoggi, B. F. 2020. Using GIS for disease mapping and clustering in Jeddah, Saudi Arabia. *ISPRS International Journal of Geo-Information*, 9(5), 328.

Oxoli, D., Cedeno Jimenez, J.R. Brovelli, M.A., 2020. Assessment of SENTINEL-5P Performance for Ground-Level Air Quality Monitoring: Preparatory Experiments Over the COVID-19 Lockdown Period. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4431, pp.111-116.

Paireau, J., Pelat, C., Caserio-Schönemann, C., Pontais, I., Le Strat, Y., Lévy-Bruhl, D., & Cauchemez, S. 2018. Mapping influenza activity in emergency departments in France using Bayesian model-based geostatistics. *Influenza and other respiratory viruses*, 12(6), 772-779.

Porcasi, X., Rotela, C.H., Introini, M.V., Frutos, N., Lanfri, S., Peralta, G., De Elia, E.A., Lanfri, M.A. and Scavuzzo, C.M., 2012. An operative dengue risk stratification system in Argentina based on geospatial technology. *Geospatial health*, 6(3), pp.31-42.

Rahman, M., Islam, A.H.M. Islam, M., 2021. Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling earth systems and environment*, 7(3), pp.2059-2087.

Ramírez-Ramírez, L.L., Gel, Y.R., Thompson, M., de Villa, E. and McPherson, M., 2013. A new

surveillance and spatio-temporal visualization tool SIMID: SIMulation of Infectious Diseases using random networks and GIS. *Computer methods and programs in biomedicine*, 110(3), pp.455-470.

Rasam, A. R. A., & Shariff, N. M. 2019. Disease Mapping and Spatial Landscape Characterization of Tuberculosis Ecology in Malaysia. *Abstracts of the ICA*, 1, NA-NA.

Reinhardt, M., Elias, J., Albert, J., Frosch, M., Harmsen, D. and Vogel, U., 2008. EpiScanGIS: an online geographic surveillance system for meningococcal disease. *International journal of health geographics*, 7(1), p.1.

Sakairi, T., Tamada, T., & Nakata, H. (2008, August). GIS Crisis-management systems using Ajax technology. In *2008 SICE Annual Conference* (pp. 3043-3046). IEEE.

Savini, L., Tora, S., Di Lorenzo, A., Cioci, D., Monaco, F., Polci, A., Orsini, M., Calistri, P., Conte, A. (2018). A Web Geographic Information System to share data and explorative analysis tools: the application to West Nile disease in the Mediterranean basin. *PloS one*, 13(6), e0196429.

Sopan, A., Noh, A.S.I., Karol, S., Rosenfeld, P., Lee, G. and Shneiderman, B., 2012. Community Health Map: A geospatial and multivariate data visualization tool for public health datasets. *Government Information Quarterly*, 29(2), pp.223-234.

Sulima M, Szostakowska B, Nahorski W, Sikorska K, Wołyniec W, Wąż P. 2019. The usefulness of commercially available serological tests in the diagnosis and monitoring of treatment in patients with alveolar echinococcosis. *Clin Exp Hepatol*. 5(4), pp. 327-333. doi: 10.5114/ceh.2019.89480.

Tao, R., Downs, J., Beckie, T. M., Chen, Y., McNelley, W. 2020. Examining spatial accessibility to COVID-19 testing sites in Florida. *Annals of GIS*, 26(4), 319-327.

Tehrani, N. A., Mollalo, A., Farhanj, F., Pahlevanzadeh, N., Janalipour, M. 2021. Time-Series Analysis of COVID-19 in Iran: A Remote Sensing Perspective. In *COVID-19 Pandemic, Geospatial Information, and Community Resilience* (pp. 277-290). CRC Press.

Tiwari, A. Jain, D.K. 2013, November. Geospatial Framework For Dengue using Open Source Web GIS Technology. In *Joint International Workshop of ISPRS WG VIII/1 and WG IV/4 on Geospatial Data for Disaster and Risk Reduction November* (pp. 21-22).

USGS. 1996. USGS EROS Archive - Digital Elevation - Global 30 Arc-Second Elevation (GTOPO30). Available at: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30> [Accessed on 01.04.2021].

Waloch, M., 2012. Cystic echinococcosis in Poland in 2010. *Przegl. Epidemiol.* 66 (2), pp. 311-313.

WHO. 2020. World Health Organization Coronavirus (COVID-19) Dashboard. Available at: <https://covid19.who.int> Accessed on: 22.09.2021.

Wilkinson, L., & Friendly, M. 2009. The history of the cluster heat map. *The American Statistician*,

63(2), 179-184.

Wojcik-Len, J., Sobolewska-Mikulska, K., Sajnog, N., Len, P. 2018. The idea of rational management of problematic agricultural areas in the course of land consolidation. *Land use policy*, 78, 36-45.

Wright, J. K. 1938. *Notes on Statistical Mapping: With Special Reference to the Mapping of Population Phenomena*. American Geographical Society.

Vitianingsih, A. V., Cahyono, D., Choiron, A. (2017, October). Analysis and design of web-geographic information system for tropical diseases-prone areas: A case study of East Java Province, Indonesia. In *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 255-260). IEEE.

Zahrieh, D., Golafshar, M. A., Patel, S. H., DeWees, T. A. 2020. Quantification of inequities in breast cancer incidence in New Mexico through Bayesian disease mapping.

Zhu, Y., Xie, J., Huang, F., Cao, L. 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *Science of the total environment*, 727, 138704.