Developing a Web Service Based Application for Demographic Information Modeling and Analyzing

Li Lin, Liping Di*, Chen Zhang, Lei Hu, Junmei Tang, Eugene Yu
Center for Spatial Information Science and Systems
George Mason University
Fairfax, USA
llin2,ldi,czhang11,lhu4,jtang8,gyu@gmu.edu

***Abstract*—Public safety has been discussed for many years, but how to use and understand crime data is still difficult. Limitations of previous research were mainly restricted by technology. The significant technological advance since last century provided a tool for researchers to compute large amount of data and complicate models while shrinking process time too. This research tried to combine results from previous work and current technology together in order filling following two gaps: 1) Predict fine spatial resolution crime information from demographic backgrounds for large coverage. 2) Implement web-based crime information access system for real time update and access.**

***Index Terms*—Web service, public safety, modeling, Hexagon Smart M.App**

I. Introduction

Public safety has been discussed for many years, and there were many studies tried to evaluate safety by using quantitative methods [1], [2], [3], [4], [5], [6], [7]. Crime data is a good representation of crime hotness for a city, but these figures were usually not accessible for public. In recent decades, general public started getting access to some crime data due to the effort from two sides: 1) the continuing release of historical crime record from government agencies; 2) the developing of open source crime data from Volunteered Geographic Information (VGI) [8], [9].

Although crime statistic is the quantitatively measurement of the safety level for a city, using crime information could be a difficult task for two reasons. First of all, there are many crime types while one could be significantly different from another. For example, [2] found that people with different level of education were usually connected with different crime types. Simply using the summation of all crime activities may not be able to precisely represent the safety information. Secondly, the reliability of crime data sources could be questionable [9], [10]. [10] argued that low report rate was one of the strongest points people argued against to the authority of official crime data in developing counties. Moreover, the distortion of location accuracy in crime data caused by privacy protection also contributed to the decreasing reliability of crime data [9].

Quantitative geography opened new windows for people to study crime from its spatial distribution. However, there are some limitations on these findings. First of all, previous works focused on providing understanding of statistical correlation between crime and demographic or economic parameters, but these simply regressions were not enough to completely explain how crime occurred. Secondly, most of previous research results are limited to small regions. For example, [11]s result only used data from few cities to illustrate crime. This small scale analysis may produce confusion for people who study crime at different places or large scale research such as state or country level. Lastly, previous researches were not able to concluded a model to represent safeness or predict crime rate. Simple understanding of crimes correlation was not able to effectively prevent crime. In other words, results from previous works were not easy to find their social value in real world.

Compared with crime researches from last century, more works have been done in recent years and scientists tried to improve their techniques from previous works. [1] proposed an idea of combining different demographic information together in order to represent the safeness of different communities. Using [1]s work as the foundation, a product called Human Security Index (HSI) was generated by [4] from United Nation. HSI provided safe indexes for each country at globe scale by using dozens of inputs. However, HSI was not able to explain the variation within a country since the product did not provide a fine spatial resolution information. In addition, the complicated concept and model of HSI did not allow frequently production. As a result, only two versions of HSI were produced in past few decades [12].

Limitations of previous research were mainly restricted by technology. The significant technological advance since last century provided a tool for researchers to compute large amount of data and complicate models while shrinking process time too. This research tried to combine results from previous work and current technology together in order filling following two gaps: 1) Predict fine spatial resolution crime information from demographic backgrounds for large coverage. 2) Implement web-based crime information access system for real time update and access.

II. Background

A. Crime and Demographic Information

Quantitative method was introduced in crime analysis in late twenty-century [3], [6], [13], [14]. Early studies were focused on discovering the spatial distribution of criminal activities [3], [6], [13], [14]. For example, [11] proposed a model to study

* Corresponding Author. ldi@gmu.edu

why some regions had higher crime and fear level than others. [11] generated three metrics to describe crime rate: defensible space feature, local social stability, and territorial facilities. Result indicated that these factors were significant related with crime at small social scale. Furthermore, research discovered a strong correlation between crime and social disorganization (such as household size) [13]. These research indicated a negative relationship between social stability and crime rate in different places.

After scientists realized that crime activities were not random happened across regions, people started to propose and analysis different factors in order to explain why crime happening more frequently at certain locations. Researchers were mainly focused on two directions to explain the variation of crime rate: socio economic variables, and demographic factors. For instance, [14] used regression models to test the relation between household burglary rate and poverty information. Result indicated that there was a strong correlation between crime rate and absolute poverty. To explain the finding between absolute poverty and crime rate, the author pointed out a theory: the economic shortage leaded to the poor development of the social infrastructures, and this underdevelopment eventually may lead to a higher crime rate. [6] found that crime rate and unemployment are statistically correlated among all age group though more young criminals were found in property crime than other types of crimes. [6] also concluded that the unemployment and crime relationship was statistically significant in whites while this relation was not so clear in African Americans.

Other researches paid more attention on demographic background for explaining crime. [5] tried to determine relationship between population change and crime rate in rural areas. [5] discovered that the changing in population increase rate will have positive influence on crime rate for a region. On the contrast, stable population will contribute to a low crime rate community [5]. Other research used other demographic information to evaluate safety as well. [13] studied on the relation between social stableness and household size, and found that the household size was correlated with potential crime. In decades, these works helped people to have better understanding of the relation between crime and other factors.

III. Framework

In this research, we try to combine results from previous work and state-of-the-art techniques together in order filling following two gaps: 1) predict fine spatial resolution crime information from demographic backgrounds for large coverage, 2) implement a web service based operational system for real time update and access.

To resolve the above problems, we established the Crime Index model to measure the crime rate of the specific area and develop a web service based application named Safe Neighborhood for users to access, visualize, analyze, and update Crime Index data. The framework of Safe Neighborhoods is shown as Figure 1, which can be divided into 5 parts: data provider, modeling, sever, client, and users. The basic workflow of the framework is: 1) querying and modelling raw data from different data sources, 2) feeding formatted data to server, 3) displaying the data on client, and 4) interoperating with users.

A. Data Provider

As described in the section 1, the safety condition of a specific area could be affected by a series of influential factors. To establish the Crime Index model, we apply multiple datasets from different data sources. Data source and datasets we applied in this research include:

- United States Census Bureau (http://www.census.gov/did/www/saipe/data/statecounty/data/index.html)
  Small Area Income and Poverty Estimates
- United States Department of Labor, Bureau of Label Statistic (http://www.bls.gov/lau/tables.htm)
  Labor force data by county
- United States Department of Agriculture, Economic Research Service (http://www.ers.usda.gov/data-products/county-level-data-sets/download-data.aspx)
  Poverty estimates for the U.S., States and counties
  Population estimates for the U.S., States and counties
  Unemployment and median household income for the U.S., States and counties
- ESRI (https://www.arcgis.com/home/item.html?id=b3802d8a309544b791c2304fece864dc)
  USA Crime Index

B. Modeling

Based on the work of HSI [4], [12], we define a new demographic information based index of crime with few enhancements. To establish the Crime Index model, we feed the raw data of demographic information queried from multiple sources and perform evaluation in a statistical model to detect variables which are statistically correlated with the given crime data. Correlated variables will be used to generated a linear regression model to represent crime index for each county in the United States. The model is defined as:

$$Y = \sum_{i=1}^{n} CiVi \tag{1}$$

where Y represents crime index, Ci represents the coefficient, Vi represents correlated variables.

Based on the previous study discussed in section 1, we feed different kinds of county level demographic data into Matlab to calculate the coefficients in the model. The linear regression model could be created by simply applying LinearModel function, and the modelling result shows the crime index is affected by four major factors: poverty rate, unemployment rate, average household size, and population increase rate. The crime index model could be described as the following equation:

$$Y = V_1 1.737 + V_2 1.2691 - V_3 0.62767 + V_4 0.50406 \tag{2}$$

where V1 represents the value of poverty rate, V2 represents the value of unemployment rate, V3 represents the value of average household size, V4 represents the value of population increase rate. We can see the crime index has positive correlation with poverty rate, unemployment rate, and population increase rate, negative correlation with average household size.

C. Server/Client

There are about 3,000 counties in the United States and more than 200,000 block groups. It means that the county crime layer will be relatively small; however, the size of block group crime data could be large. Displaying fine spatial resolution data could be challenging on desktop. For this reason, this research will design a cloud-based service for the final crime product. The first objective of this web-based platform is designed to handle large volume of crime data. Secondly, the web-based platform could be utilized to update crime information as soon as input data get updated. Lastly, the mobility of web-based application could benefit people who want the information immediately while they are on the street.

The server end and client end are two core component to develop the web service based operational system. In this research, we implemented Safe Neighborhoods on Hexagon Smart M.App Development Platform, the detail of the platform will be discussed in the next section. The server end is powered by Amazon Web Services which offers a suite of cloud-computing services. In the proposed framework, Crime Index data are managed in the Safe Neighborhoods server end. The client end is used to interact between server end and users, which allow users to perform a series of analytical operations and statistical analysis.

IV. Implementation

As described in the section 2, the major programming languages and tools used for developing Safe Neighborhoods include Matlab, ArcGIS, MapShaper, HTML5, JavaScript, and CSS. Figure 1 shows the architecture and its major implementation tools of Safe Neighborhoods.

[htbp]

In the development, Matlab is used as a core analyzing platform to calculate and establish the crime index model. To determine the correlation between safety level and its different influential factors, data from different sources could be feed and processed by Matlab build-in functions very easily.

As a powerful one-stop integrated platform for geospatial data analysis, ArcGIS plays an essential role for performing national scale data processing in county level based on the crime index model. Also, ArcGIS can efficiently visualize the outputted crime index data and save the result in Shapefile format. On the other hand, running speed is a significant measurement for the performance of web service. To improve the performance, we used the generalize tool in ArcToolbox to simplify the county level polyline features and compress data size.

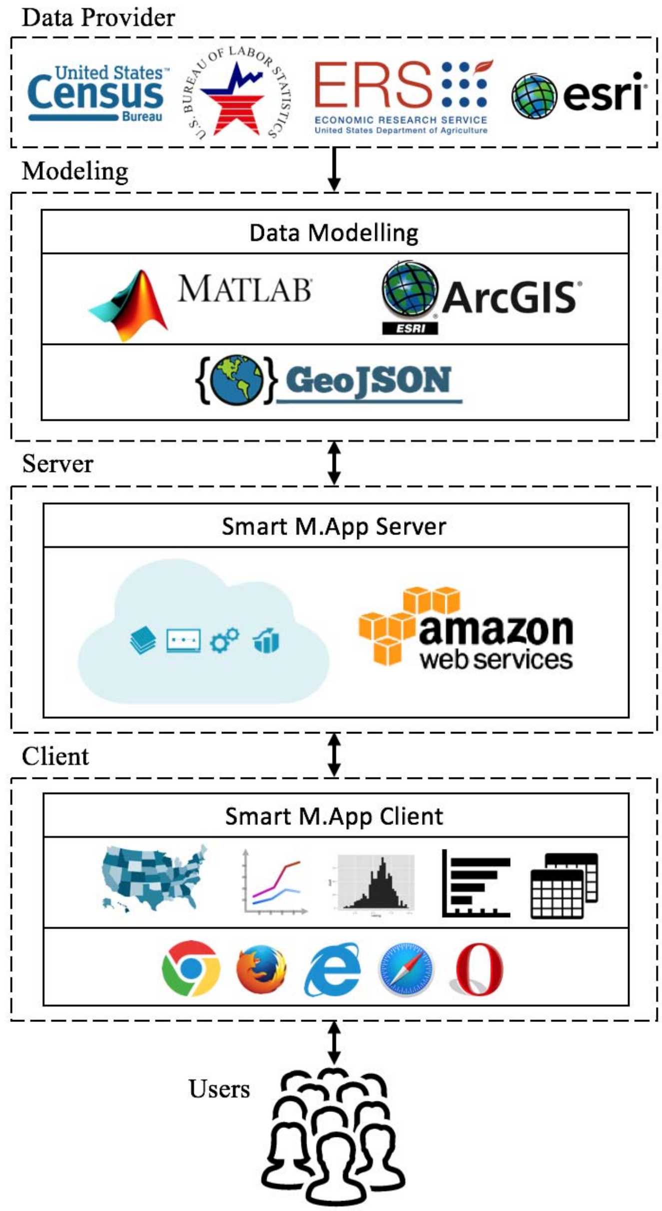

Fig. 1. The framework of Safe Neighborhoods

We implemented Safe Neighborhoods on Hexagon Smart M.App platform which provide an easy-to-go web based IDE integrated development environment (IDE) for developers to build lightweight and dynamic applications targeted to solve a specific problem [16]. As the supported geospatial data type of Hexagon Smart M.App platform, GeoJSON is a JSON (JavaScript Object Notation) based format for encoding a variety of geographic data structure [16], which supported by a series of GIS software such as OpenLayers, MapServer, GeoServer, and GDAL. We use MapShaper, a free-to-use generalization web service [17], to convert the type of county level crime index data from shapefile to GeoJSON.

As the most popular tool in the web development, HTML5, JavaScript, and CSS are three major programming languages

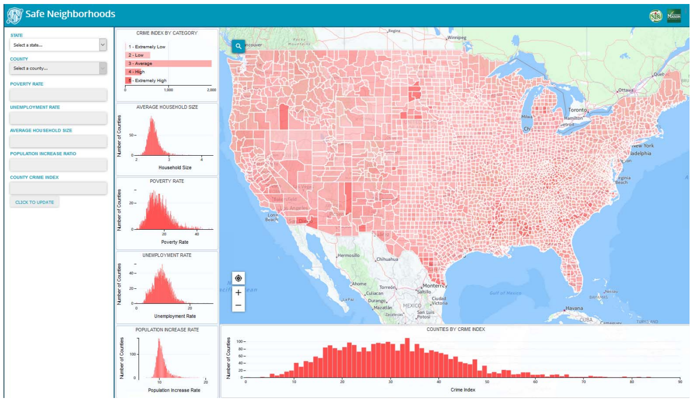

Fig. 2. The user interface of Safe Neighborhoods

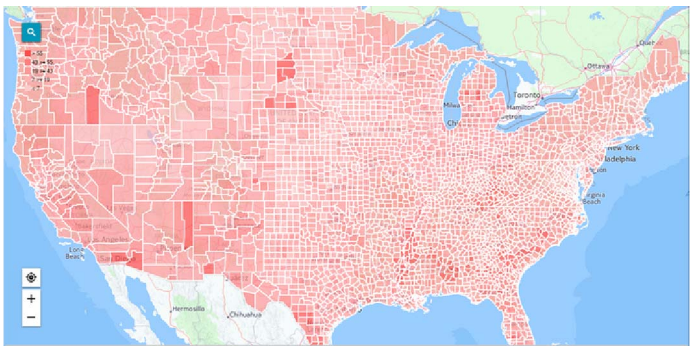

Fig. 3. evaluating the national scale safeness

used for developing the client-end of Safe Neighborhoods. Supported by the powerful extension of geospatial operations and statistics functionalities of Smart M.App Development Platform, tons of useful functions could be easily interoperated by Smart M.App API (application program interface).

V. Experiments and Results

Safe Neighborhoods is developed as an instance of Smart M.AppTM and published on Hexagon Geospatial Market-Place (https://store.hexagongeospatial.com/apps/138892). Figure 2 shows the user interface of Safe Neighborhoods.

As a Smart M.App application, Safe Neighorhoods provides a series of geospatial statistics operations and visualizations. Figure 3 - 6 show the major functions of Safe Neighborhoods which include: a) evaluate the safeness of each county in the contiguous United States using a crime model that correlates crime and demographic data, b) select a particular county to see detail statistics (such as New York county in New York state), c) provide fresh demographic data to generate a new Crime Index map (example shown is Monroe County, Florida), d) use the charts to define a range or select a category to display which counties are affected.

VI. Conclusion and Future Work

This paper proposes a framework to develop a web service based application for national scale crime index modeling and analyzing. The framework integrates data pre-processing,

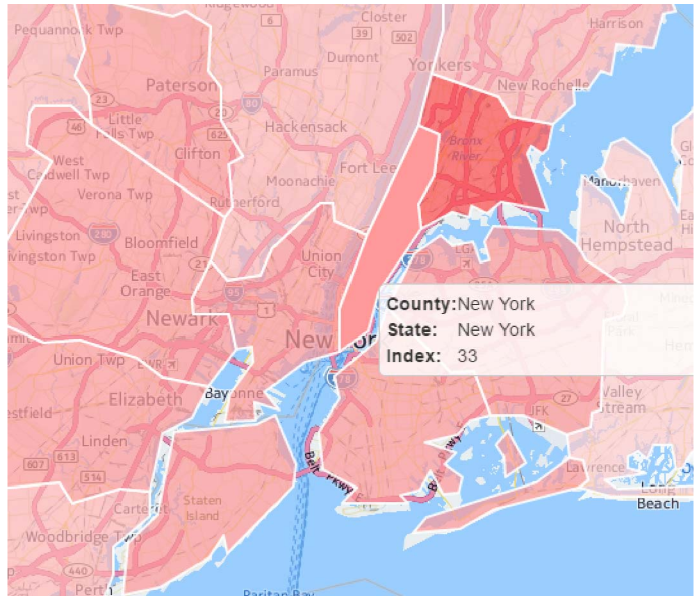

Fig. 4. detail statistics of each county

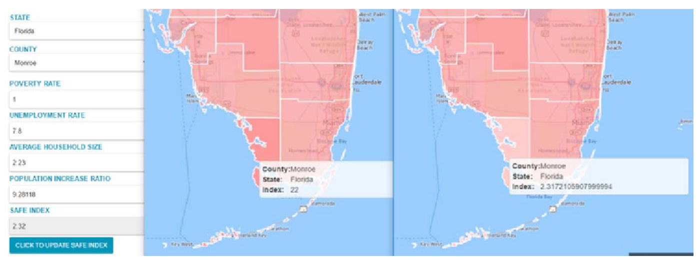

Fig. 5. updating and refreshing Crime Index

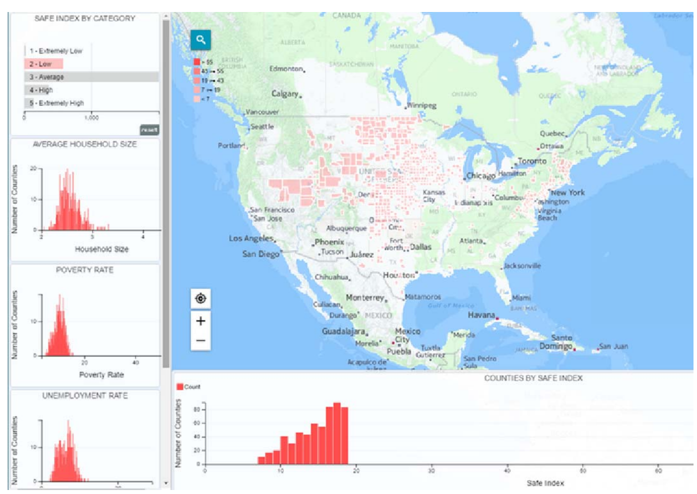

Fig. 6. highlighting with specific category

data modelling, server, client, and a series of state-of-the-art techniques. These techniques, especially for displaying, transferring, processing geospatial information, have been widely used in the GIS industry [18],[19],[20],[21],[22],[23],[24]. To measure the county level crime rate, we define the Crime Index which has positive correlation with poverty rate, unemployment rate, and population increase rate, negative correlation with average household size. Then we apply the model on all counties in the United States based on the real data from official data sources such as United States Census Bureau, United States Department of Labor, and United States Department of Agriculture. Crime index at blockgroup level could be calculated using downscaling method based on previous regression model. Although it is difficult to acquire blockgroup level demographic data for all counties, but result could be generated where data is available. Based on the Crime Index data, an web service based application named Safe Neighborhoods is developed as an instance of Smart M.AppTM and published on Hexagon Geospatial MarketPlace.

Current version of Safe Neighborhoods supports accessing, visualizing, and updating the latest Crime Index data. In the future, we will process more historical raw data from more data sources and improve the Crime Index Model. Furthermore, the historical Crime Index data would be added to server end and users are able to inquire and analyze the historical Crime Index data. With the extension of Crime Index data, more advanced functions, such as time series and predicting analysis, could be performed on Safe Neighborhoods client.

ACKNOWLEDGMENT

We thank Hexagon Geospatial IGNITE team for the technical support on Smart M.App platform. We would also like to show our gratitude to the Robert Gammon, Hexagon Geospatial Devision for sharing his pearls of wisdom with us during the development of "Safe Neighborhoods".

REFERENCES

[1] K. P. Bajpai, Human security: concept and measurement. Joan B. Kroc Institute for International Peace Studies, University of Notre Dame Notre Dame, 2000.

[2] F. G. Carneiro, P. R. Loureiro, and A. Sachsida, Crime and social interactions: a developing country case study, The Journal of Socio-Economics, vol. 34, no. 3, pp. 311318, 2005.

[3] T. G. Chiricos, Rates of crime and unemployment: An analysis of aggregate research evidence, Social problems, vol. 34, no. 2, pp. 187212, 1987.

[4] D. A. Hastings and others, Describing the human conditionfrom human development to human security: an environmental remote sensing and GIS approach, in Proceedings, 2008.

[5] M. R. Lee and S. A. Thomas, Civic community, population change, and violent crime in rural communities, Journal of Research in Crime and Delinquency, vol. 47, no. 1, pp. 118147, 2010.

[6] M. D. Smith, J. A. Devine, and J. F. Sheley, Crime and unemployment: Effects across age and race categories, Sociological Perspectives, vol. 35, no. 4, pp. 551572, 1992.

[7] J. Terpstra, Citizen involvement in local security networks, Security Journal, vol. 22, no. 2, pp. 156169, 2009.

[8] A. Keler and J. D. Mazimpaka, Safety-aware routing for motorised tourists based on open data and VGI, Journal of location Based services, vol. 10, no. 1, pp. 6477, 2016.

[9] L. Tompson, S. Johnson, M. Ashby, C. Perkins, and P. Edwards, UK open source crime data: accuracy and possibilities for research, Cartography and geographic information science, vol. 42, no. 2, pp. 97111, 2015.

[10] O. Marenin, Victimization surveys and the accuracy and reliability of official crime data in developing countriesPublications received, Journal of Criminal Justice, vol. 25, no. 6, pp. 463475, 1997.

[11] R. B. Taylor, S. D. Gottfredson, and S. Brower, Block crime and fear: Defensible space, local social ties, and territorial functioning, Journal of Research in crime and delinquency, vol. 21, no. 4, pp. 303331, 1984.

[12] D. Hastings, The Human Security Index: An Update and a New Release. Document Report Version 1.0 (March). 2011.

[13] T. D. Miethe, M. Hughes, and D. McDowall, Social change and crime rates: An evaluation of alternative theoretical approaches, Social forces, vol. 70, no. 1, pp. 165185, 1991.

[14] E. B. Patterson, Poverty, income inequality, and community crime rates, Criminology, vol. 29, no. 4, pp. 755776, 1991.

[15] G. Hammerer, Introducing the Map of the Future: The Hexagon Smart M. App, GeoInformatics, vol. 18, no. 8, p. 14, 2015.

[16] H. Butler, M. Daly, A. Doyle, S. Gillies, T. Schaub, and C. Schmidt, The GeoJSON format specification, Rapport technique, vol. 67, 2008.

[17] M. Harrower and M. Bloch, MapShaper. org: A map generalization web service, IEEE Computer Graphics and Applications, vol. 26, no. 4, pp. 2227, 2006.

[18] G. Y. Eugene et al., Online parameterization for WOFOST for United States using open geospatial standards, in Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on, 2016, pp. 16.

[19] Z. Sun et al., Embedding Pub/Sub mechanism into OGC web services to augment agricultural crop monitoring, in Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on, 2016, pp. 14.

[20] Z. Sun et al., CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models, Earth Science Informatics, pp. 117, 2017.

[21] Z. Sun et al., GeoFairy: Towards a one-stop and location based Service for Geospatial Information Retrieval, Computers, Environment and Urban Systems, vol. 62, pp. 156167, 2017.

[22] Z. Sun et al., Combining OGC WCS with SOAP to faciliate the retrieval of remote sensing imagery about agricultural fields, in Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on, 2016, pp. 14.

[23] C. Zhang, Z. Sun, G. Heo, L. Di, and L. Lin, A GeoPackage implementation of common map API on Google Maps and OpenLayers to manipulate agricultural data on mobile devices, in Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on, 2016, pp. 14.

[24] C. Zhang, Z. Sun, G. Heo, L. Di, and L. Lin, Developing a GeoPackage mobile app to support field operations in agriculture, in Agro-Geoinformatics (Agro-Geoinformatics), 2016 Fifth International Conference on, 2016, pp. 14.