EXPLORING VOLUNTEERED GEOGRAPHIC INFORMATION WITH DATA QUALITY CONTROL FOR INTEGRATED PEST MANAGEMENT

YAN YINGWEI

NATIONAL UNIVERSITY OF SINGAPORE

2016

EXPLORING VOLUNTEERED GEOGRAPHIC INFORMATION WITH DATA QUALITY CONTROL FOR INTEGRATED PEST MANAGEMENT

YAN YINGWEI

(M.Sc., The Chinese University of Hong Kong)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOGRAPHY

NATIONAL UNIVERSITY OF SINGAPORE

2016

Supervisor:
Associate Professor Chen-Chieh Feng

Examiners:
Assistant Professor Winston Chow
Assistant Professor Mi Diao
Professor Daniel Sui, The Ohio State University

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Yingwei Yan

Yan Yingwei

8th August 2016

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Associate Professor Chen-Chieh Feng, for his guidance, intellectual support, and mentorship throughout my doctoral research. My supervisor has been a superb academic role model. Without his inspirational and constant support, I would never have been able to finish my doctoral research. I am also honored that Associate Professor Yi-Chen Wang and Prof. Xixi Lu have acted as my thesis committee members. I appreciate the time they have spent on guiding my work and enjoyed all of our discussions over the years of my Ph.D candidature.

This thesis has been supported by National University of Singapore (NUS) research scholarship; and Singapore National Research Foundation under its Inter-national Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office through the Centre of Social Media Innovations for Communities (COSMIC). Various necessary computational resources were provided by the Department of Geography of NUS and COSMIC, with special thanks to Mr. Raymond Huang Jinliang for his technical supports. I would also like to thank Prof. Roger Zimmermann, Associate Professor Klarissa Ting-Ting Chang, Dr. Maffee Peng-Hui Wan, and their research assistants Mr. Adrian Tan, Mr. Avijit Sengupta, Miss Jia Rubio, Mr. Rax Chun Lung Suen, Ms. Wen Yong Chua, and Mr. Yeow Chuan Ng, at the Department of Information Systems of NUS, for their technical assistances in conducting pilot studies, collecting and processing relevant data of my research.

This thesis could not have been conducted without the assistances from the Agricultural Department of Xiajiang prefecture, Jiangxi province of China, and their dedicated pest management staffs and experts, with special thanks to Mr. Xiaojun Liao and Mr. Zhifeng Liu for their generous and invaluable intellectual, technical, and material assistances throughout my fieldworks.

Furthermore, I would like to thank my fellow friends, including Lishan, Xiankun, Shaoda, Jinghan, Siqi, Guanie, Seonyoung, Suraj, Piseth, Jiaqi, Jingfu, Ningning, and Valerie for the friendships over the past few years. I also thank Ms. Pauline Lee and other staffs from the Department of Geography and the Faculty of Arts and Social Sciences, for their kind administrative guidance and helps. Lastly, I would like to express my deep appreciation to my family for their continuous support during these years of my Ph.D candidature.

Table of Contents

Summary

Volunteered geographic information (VGI), the information sourced from the platforms for interactive geospatial data production and consumption fostered by Web 2.0 technologies, have enabled spatial decision-making at a much larger spatial scale and faster temporal scale. In the field of integrated pest management (IPM), utilizing VGI has been suggested as a means to improve the efficiency and effectiveness of IPM due to its inherent advantages in information collection and dissemination. However, current VGI-based IPM has been limited to a primitive conceptual framework with three components–information collection, sense making (knowledge discovery), and dissemination–and its software implementation for data collection and dissemination. Several questions related to VGI-based IPM remain unanswered: (1) How can the primitive VGI-based IPM framework be enhanced? (2) Can VGI sense making indeed generate meaningful outcomes for enhancing IPM? (3) How can the quality of VGI be assured? (4) Lastly, what are the roles of volunteer participants in VGI-based IPM, and by extension, how can the volunteer participation be motivated and sustained? Therefore, answering these questions become the research objectives of this thesis.

To answer the first question, an enhanced conceptual framework of VGI-based IPM based on the epistemological foundation of VGI was proposed. The framework incorporates quantitative, qualitative, or combined quantitative-qualitative methods that are suited for the transformative paradigm afforded by VGI. It serves as a framework of reference for enhancing IPM based on VGI as an alternative to the positivist paradigm adopted in traditional IPM, and for the development of more comprehensive VGI-based IPM. Regarding the second question, VGI were collected pertaining to both short-term (pest outbreaks) and long-term (pest invasions) pest risks. The study explored a range of sense making methods that are well suited to the characteristics of VGI, including hot spot analysis, phenological analysis, and ecological niche modelling. The sense making indeed revealed important directional, clustering, and temporal characteristics of crop pest outbreaks; and provided insights into the possible distributional changes of invasive crop pests. To answer the third question, an expert system based on fuzzy set theory was developed and tested in this study for VGI quality assurance. Regarding the last research question, three clusters of participant roles were identified and conceptualized based on the lessons learned from the explorations for research question two and three. These roles are: (1) basic geospatial data contributions for knowledge discovery and decision-making; (2) metadata creation; and (3) higher level data contributions for community building (cognitive engagement in IPM). Additionally, the

results of a participation incentive analysis through questionnaire surveys showed that farmers tend to concern temporal patterns (i.e., specific timings of pest risk management) more than spatial patterns of pest outbreaks (i.e., specific areas of pest risk management); an important factor contributing to a continued and more engaged user participation was the usefulness of information disseminated to individual participants.

By answering the research questions, this study has its important significances as follows. It moves beyond a primitive conceptual framework by offering an operational VGI-based IPM that enables sustainable VGI collection, effective VGI quality assurance, and insightful knowledge discovery. Practically, this research benefits relevant experts by alleviating them from collecting geospatial information to focusing on data analysis. It also has significances to pesticide reduction, pest managerial investments, agricultural productivity enhancement, and to the design, development, testing, and deployment of VGI-based IPM tools and systems.

In conclusion, VGI offers a promising way to improve traditional IPM. This study sheds light on how IPM can be enhanced through VGI efforts. Nevertheless, future work is still needed to explore VGI-based IPM further, particularly in terms of (1) big data collection, storage, analysis, and dissemination; (2) cognitive abilities of participants in IPM; and (3) user privacy in data production and consumption.

List of Tables

List of Figures

1 Introduction

1.1 Volunteered geographic information

VGI is the most frequently used term in geographic information science (GIScience) to describe the information sourced from the platforms for interactive geospatial data production and consumption fostered by Web 2.0 technologies (Goodchild, 2007a). In such platforms, how geospatial information is created, maintained, and used have been revolutionized. Traditional geospatial data production is restricted to experts who produce professional GIScience outputs to be consumed by non-experts. This is a top-down approach based on an expert-centric paradigm, which enforces a clear-cut boundary between information providers and consumers. However, due to Web 2.0 technologies, the top-down approach has given way to a bottom-up approach in which non-experts not just consume but also can contribute geospatial information. Such a bottom-up approach blurs the boundary between data providers and consumers as those traditionally seen as data consumers can now play the role of producers, which allows for the acquiring and consumption of ubiquitous, real-time, and near zero-cost spatially referenced information. One famous such example is OpenStreetMap (https://www.openstreetmap.org/). Users of OpenStreetMap can contribute their data to the online community anytime and anywhere, and are allowed to freely use, distribute, transmit, and adapt the data generated by other users.

Therefore, VGI phenomenon creates more opportunities for the general public to involve in GIScience. Hitherto, the value of VGI has been explored in many application domains, including, but not limited to, those for disaster, emergency, and crisis management (Pultar *et al.*, 2009; De Longueville *et al.*, 2010); surveillance or monitoring programs (Fowler *et al.*, 2013; Langley *et al.*, 2013); urban or environmental management and planning (Seeger, 2008; Song and Sun, 2010); new generation of gazetteer (Keßler *et al.*, 2009a; Keßler *et al.*, 2009b); and land use/cover mapping (Foody and Boyd, 2012; Vaz and Jokar Arsanjani, 2015). These explorations represent the belief in GIScience community that non-experts can be engaged in and benefit from geospatial data collection and analysis.

1.2 VGI-based integrated pest management

Exploring the value of VGI must be grounded in a specific application domain because without a domain, it is impossible to derive meaningful geographic knowledge from geographic data. In this thesis, VGI-based integrated pest management (IPM) was explored. IPM seeks multidisciplinary approaches for agricultural pest management in order to reduce pesticide use, which has become one of the most important ways for ensuring high agricultural productivity (Morse and Buhler, 1997; Peshin *et al.*, 2009a; Peshin and Dhawan, 2009a).

IPM comes in two branches, the tactical and strategic IPM (Barfield and Swisher, 1994; Morse and Buhler, 1997; Zalucki *et al.*, 2009). It has been suggested that IPM should be enhanced in both branches (Beddow *et al.*, 2010). Tactical IPM is a short-term and timely prevention of pest infestations for relatively small regions, e.g., predicting local pest outbreaks to alert the corresponding stakeholders to prepare for reducing pest damages. It is therefore conducted to react to problems at present or in the near-term. In contrast, strategic IPM focuses on long-term and large-area agricultural planning. It is implemented to react to potential problems that will be encountered in the future, and concerns less about current problems. Strategic IPM are therefore conducted to answer abstract or hypothetical questions, which may impact fundamental aspects about the future, leading to substantial consequences for pest management problems in the years ahead. For example, the siting of cropland, the optimal selection of crop variety in an area to avoid potential infestations caused by pest invasions associated with climatic change.

Each type of IPM is conducted possibly based on different types of information and faces its particular challenges in gathering the required information. On the one hand, tactical IPM is conducted using the information on what is actually occurring or is expected to occur soon. The challenge is, collecting such real-time information based on traditional data collection means is resource demanding. On the other hand, achieving strategic IPM requires pest information to be collected over large spatiotemporal extents. Studies are needed to use these pieces of information to counter the potential environmental and economic damages of pests. However, collecting such information across large areas and in long-term scales is even more challenging, especially given the fact that there exists a large number of pest species. Without efficient and sufficient data collections, it is also hardly possible to achieve effective IPM information disseminations to exert influences (Peshin and Dhawan, 2009a). Especially, traditional top-down IPM dissemination (e.g., radio programmes) is often aimed for a large heterogeneous audience without adjusting its information to individual needs. The information are packed and thrown to the audience without any personalization and further interaction (Peshin and Dhawan, 2009a). As a result, the over and indiscriminate use of pesticides (pesticide use reduction is the central management objectives of IPM (Lewis *et al.,* 1997; Morse and Buhler, 1997)) are still common in most countries despite the deployment of various IPM approaches. This problem is particularly serious in underserved farming communities in developing countries where access to effective information for solving or avoiding pest infestation is problematic (Van den Berg and Jiggins, 2007; Deng *et al.*, 2012; Deng and Chang, 2012).

The emergence of Web 2.0 technologies seems promising in overcoming the challenges mentioned above due to its inherent advantages in information collection and dissemination. Web 2.0 has the potential to foster interactive communities in which not only experts but also non-experts can create spatially referenced information (i.e., VGI). The information created by them can even be ubiquitous, real-time, and near zero-cost. In turn, the general public can benefit from the information collectively contributed by themselves, which are closely bound up with their own interests. Indeed, a VGI-based IPM approach has already been proposed with its conceptual framework (Figure 1.1) (Deng and Chang, 2012). This framework is composed of a VGI retrieval component that facilitates VGI collection (e.g., volunteered crop pest surveillance information), a VGI analysis component that performs sense making (knowledge discovery), and a dissemination component that disseminates pest management information. It resembles the two interactive and dynamic conceptual workflows pertaining to VGI-based crisis management proposed by Ostermann and Spinsanti (2011) and Craglia *et al.* (2012), but has its own specificities on crop pest risk management. Nevertheless, current VGI-based IPM has been limited to this primitive conceptual framework (Deng and Chang, 2012) and its software implementation for data collection and dissemination

(Suen *et al.*, 2014). Several issues related to VGI-based IPM remain unaddressed: (1) How can the primitive VGI-based IPM framework be enhanced conceptually and practically for it operation? (2) Can VGI sense making indeed generate meaningful outcomes for enhancing tactical and strategic IPM? (3) How can the quality of VGI be better assured despite the suggestion of ensuring the user friendliness of VGI collection tools for reducing mistakes in VGI (Deng and Chang, 2012)? (4) Lastly, what are the roles of volunteer participants in VGI-based IPM, and by extension, how can the volunteer participation be motivated and sustained?

Figure 1.1 VGI-based IPM framework.

1.3 Research objectives

To address the research questions raised above, this research aimed to achieve the following four objectives pertinent to VGI-based IPM:

(i) Propose and develop an enhanced conceptual framework of VGI-based IPM;

(ii) Explore VGI sense making to enhance IPM;

(iii) Develop an approach to assure the quality of VGI;

(iv) Explore the roles of volunteer participants in VGI-based IPM, and by extension the appropriate approaches to motivate and sustain the volunteer participation.

The first objective is concerned with the conceptual framework of IPM. As mentioned in Section 1.2, the conceptual framework proposed by Deng and Chang (2012) is primitive. An enhanced conceptual framework of VGI-based IPM is needed with more contributing components added, which could serve as a framework of reference for guiding the research on VGI-based IPM, and thus the development of more comprehensive VGI-based IPM.

The second objective is concerned with whether VGI sense making can indeed generate meaningful outcomes for solving both tactical and strategic IPM problems. As VGI defies conventional data collection, storage, and analysis processes, the data can be contributed anytime and anywhere, which are somewhat unstructured (Miller and Goodchild, 2015). As a result, raw VGI tends to be heterogeneous and diverse, which may hinder the discovery of valuable knowledge for IPM. It is therefore necessary to investigate whether VGI can indeed be made sense of to enhance IPM; in other words, to investigate whether it is feasible to utilize VGI as a source of data for IPM knowledge discovery.

The third objective is concerned with the quality of VGI. The extent to which a user can trust VGI has always been called into question because VGI is inherently heterogeneous and diverse, and because its creation tends not to adhere to standards required in the creation of conventional authoritative data (e.g., government generated data) and it lacks data quality descriptions (e.g., standard metadata) for determining its fitness for a particular purpose (Yanenko and Schlieder, 2012). Thus, it is imperative to develop an approach to assure the quality of VGI for IPM.

The fourth objective is concerned with the practical and potential roles of volunteer participants in VGI-based IPM, and by extension the ways they are incentivized. It is important to learn what a volunteer participant can do in VGI-based IPM, as maximizing the value of volunteer participants by understanding their roles in a concerned management issue contributes to the realization of the corresponding management goals (Seeger, 2008; Song and Sun, 2010; Li and Goodchild, 2012). Volunteer participation must also be motivated and sustained, in order to ensure the sustainability of effective management (Li and Goodchild, 2012).

1.4 Research significance

By achieving the research objectives indicated above, the significance of this research is as follows. It moves beyond a primitive conceptual framework by offering an operational VGI-based IPM that enables sustainable VGI collection, effective VGI quality assurance, and insightful knowledge discovery. Practically, this research benefits relevant experts by alleviating them from laborious and expensive geospatial information collection and thus enabling them to focus on data analysis. It also has significances to pesticide reduction, pest managerial investments, agricultural productivity enhancement, and to the design, development, testing, and deployment of VGI-based IPM tools and systems.

1.5 Research scope

This thesis focuses on the VGI related to quantitative spatiotemporal pest surveillance, based on which the research objectives mentioned in Section 1.3 are set. VGI related to the cognitive ability of volunteer participants (e.g., farmers' knowledge about pests' ecology, farmers' ability in defining pest management problems, farmers' ability in analyzing and interpreting spatial data) in pest management is beyond the scope of this thesis. The provision of information pertinent to pest surveillance is the most basic ability of volunteer participants, it therefore should be fully explored prior to extending the research to the cognitive ability of volunteer participants in IPM.

1.6 Thesis structure

The structure of the following parts of this thesis and the main contents are briefly described as follows.

Chapter 2 will provide literature reviews pertinent to the proposed research objectives. Contributions by the previous studies will be introduced. The related research progresses, gaps, and challenges will also be discussed.

Chapter 3 will propose an enhanced framework of VGI-based IPM.

Chapter 4 will cover the topic pertinent to VGI sense making for enhancing tactical IPM, participation incentive issues, and the roles of volunteer participants in tactical VGI-based IPM.

Chapter 5 will cover the topic pertinent to VGI sense making for enhancing strategic IPM, and the roles of volunteer participants in strategic VGI-based IPM.

Chapter 6 will discuss the development of a VGI quality assurance approach and the roles of volunteer participants in the quality assurance.

Lastly, **Chapter 7**, the major findings and the implications of this study will be summarized. The limitations and suggested future work will also appear in this chapter.

2 Literature review

This chapter provides a detailed literature review. In Section 2.1, the value of VGI will be identified and a typology of the diverse forms of VGI creation will be described, followed by a review in Section 2.2 about the shortcomings of IPM and the potential improvements through VGI. Section 2.3 will identify the research gaps pertaining to the development VGI-based IPM.

2.1 Volunteered geographic information

2.1.1 The value of VGI

VGI has attracted considerable attention from researchers as it can be an important source of understanding of the surface of the Earth (Goodchild, 2007a). In VGI settings, the general public are encouraged to share their spatially-referenced information, through ways that are easy for non-experts to master, including mainly (1) uploading, marking, and annotating geographic features using geospatial Web (or GeoWeb, e.g., Wikimapia: http://wikimapia.org); (2) contributing information collected through personal location-aware devices to VGI databases (e.g., OpenStreetMap: http://www.openstreetmap.org/); and (3) adding information to online shared media, such as location-based photographs, and texts (e.g., Flickr: http://www.flickr.com/). Through them, the creators of VGI establish virtual networks to work on a common task (or subtask) in either a synchronous or an asynchronous manner. They share their understanding of a common situation, shape contexts, and convey cognition through contextual knowledge of a place. VGI phenomenon thereby defies the traditional asymmetric power structure of geospatial information production and consumption, i.e., a minority of authorized data producers versus a majority of passive data consumers (illustrated in Figure 2.1a). On VGI platforms, geospatial data consumers are enabled to produce data and vice versa. The traditional division between data consumers and producers blurs. Dialogue between official and non-official voices on an equal footing is advocated (Figure 2.1b). A neologism "produser" has therefore emerged to describe this balancing effect that enables people to play the dual roles of geospatial data producers and consumers (Coleman *et al.*, 2009). Some have argued that the "produsers" may have knowledge that is unknown to experts; local people in a sense may themselves count as experts in their own local or indigenous knowledge (Cinderby and Forrester, 2005). As such, the creation of previously unrecorded spatial data for the discovery of previously unknown knowledge, may be on the list of the most exciting value of VGI (Cinnamon and Schuurman, 2013).

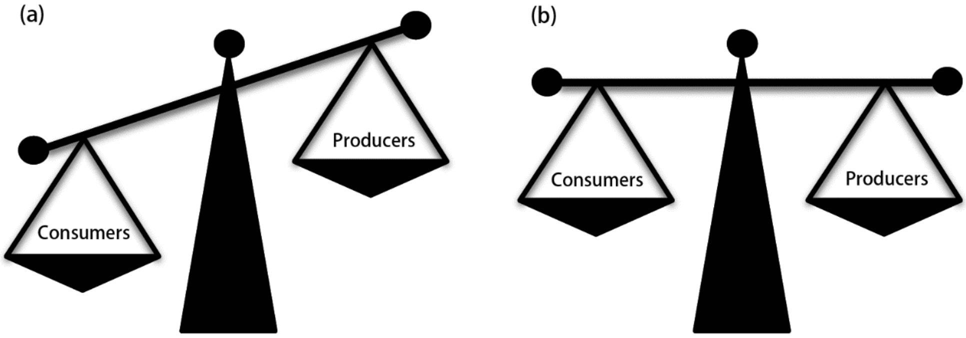


Figure 2.1 Change in the power structure of geospatial information production and consumption, from (a) the asymmetric power structure to (b) the symmetric power structure.

From the technical perspective, the rapid development of VGI is attributed to Web 2.0 technologies which favor participation and collaboration in the creation of common goods over the Internet. The Internet in the Web 2.0 era functions as a cyberspace of radical inclusion, bringing together indirectly-related physical communities into directly-connected virtual communities. It creates platforms for free and ubiquitous collaborations of intelligence and thus promotes a digital democracy, where techno-libertarian and egalitarian are the norms (Haklay *et al.*, 2008; Han, 2012). Among the key principles of Web 2.0, harnessing collective intelligence is at the core for sustaining VGI platform building (O'reilly, 2007). This principle embraces cyber-collectivism for the formation of Web 2.0 cyberspaces where great opportunities for achieving explosions of productivity and innovation are offered. In Web 2.0, information technologies are increasingly socially embedded; new forms of social interaction within information networks are formed, and netizens are no longer exogenous (Castells, 2000). Therefore, Web 2.0 has driven the general public to contribute information on an unprecedented scale and made real-time information interactions possible, leading to diverse initiatives using information by citizens (Elwood *et al.*, 2012). By contributing their collective intelligences, the general public is in their true sense involved in GIS democracy. Therefore, thanks to Web 2.0 technologies, the creation of sheer amounts of ubiquitous, cost-effective, and real-time geospatial information by the general public is also one of the most important value of VGI (Goodchild, 2007a).

2.1.2 The forms of VGI creation

This section provides a typology of VGI creations forms identified from the literature, which can be taken into consideration for scientific inquiries.

(1) Active VGI creation (Goodchild, 2007b): for active VGI creation, participants contribute personal knowledge or local/indigenous observations for specified purposes on their own initiative. The contribution is non-compensated and non-coerced. Most VGI projects fall in this category, e.g., Wikimapia, OpenStreetMap, and Flickr (https://www.flickr.com/). Underlying this type of VGI is the notion of participant emancipation and activism (Jones and Weber, 2012). There are two major concerns when using active VGI for scientific inquiries. The first concern is about the uncertain expertise of VGI creators (Keßler *et al.*, 2009a). The second concern is about digital vandalism, the act

that some “damagers” degrade the fruits of VGI platforms through ways such as making nonsense, spam, or false information (Coleman *et al.*, 2009).

(2) Passive VGI creation (Craglia *et al.*, 2012): for passive VGI creation, data are contributed implicitly in the sense that making contributions is not a contributor’s primary intention. Passive VGI creation has also been referred to as ambient geographic information (AGI). It is argued that active VGI creation is crowdsourcing (i.e., outsourcing specific tasks to the general public), while AGI is crowd harvesting (i.e., harvesting information created by the general public in a meaningful manner) (Stefanidis *et al.*, 2012). Social media applications, e.g., Facebook (https://www.facebook.com/) and Twitter (https://twitter.com/), are two such examples. Facebook and Twitter users communicate with one another or present themselves by posting articles, photos, and blogs. But in most cases, they are not conscious of their shared information being used for other purposes. Another example is the retrieval of geographic footprints from location-aware smartphones, such as the case in gamification of VGI (Matyas *et al.*, 2011). Passive VGI creation is less problematic with digital vandalism since this type of VGI creators are not directly involved in data collection. It does not mean, however, that data quality issue can be neglected. In fact, sometimes, Twitter and Facebook users publish misleading or biased information for better self-presentation (Bakshy *et al.*, 2011).

(3) Facilitated VGI creation (Seeger, 2008; Cinnamon and Schuurman, 2013): Cinnamon and Schuurman (2013) mentioned that many of the best-known VGI projects (e.g., OpenStreetMap) operate on a distributed model in which direct interactions between VGI creators and VGI users are not necessary. Facilitated VGI, however, is created through an assisted data contribution model in which a targeted group of participants (with necessary abilities) is requested to contribute geospatial data according to predefined questions or criteria in order to achieve a pre-established objective within an established geographic extent (Seeger, 2008; Cinnamon and Schuurman, 2013). Interactions between the facilitators and data creators can be through face-to-face communications, digital communications, or the combination of both, due to which data tend to be consciously contributed. For example, Cinnamon and Schuurman (2013) asked the emergency medical service paramedics from Cape Town to identify injury hotspots in their service areas through both face-to-face and digital communications. Girres and Touya (2010) stated that a balance must be sought between data contribution freedom and the need for contributors to comply with specifications, so as to improve data quality. Indeed, in the assisted data contribution model with knowledgeable participants, participants can be directly facilitated, encouraged, or even trained to provide accurate information for achieving objectives. Note that the core value of commonly referred conceptions of VGI (i.e., voluntariness) is retained in facilitated VGI creation.

However, compared to active and passive VGI creations, facilitated VGI creation approach may be limited in fostering the empowerment of marginalized communities due to its inherent constrains on data contribution freedom. Lastly, projects that are based on facilitated VGI creation may not produce information that are accessible to the general public (Cinnamon and Schuurman, 2013).

The following chapters of this thesis will explore active and facilitated VGI creations for enhancing IPM. Passive VGI is not considered in this thesis because crop pest observation so far is still a topic in the minority on passive VGI platforms (e.g., Twitter, based on searches using the keywords related to crop pests on these platforms via their application programming interface).

2.2 Integrated Pest Management

In the following sub-sections, a review about the shortcomings of traditional IPM will be reviewed first. The possibility of taking advantage of VGI to enhance IPM and the related challenges will be given next.

2.2.1 From traditional IPM to VGI-based IPM

Since the 1960s, due to the rising pest resistances to pesticides, IPM has been applied to improve the control of pests in the field. Traditional IPM strategies since the 1960s rely on linear, top-down, and research-driven strategies (Table 2.1). However, the ineffectiveness of such strategies, specifically the inappropriateness of the recommended pest control strategies and the lack of sense of farmer participation and thus the ownership of the programs (van den Berg *et al.*, 2002; Litsinger *et al.*, 2009), quickly surfaced as major obstacles toward effective implementations of IPM. Researchers could achieve only limited success in alleviating pest problems without being more inclusive of non-official voices. Despite working closely with professionally trained IPM extension workers, the IPM would make limited progress without involving farmers participating directly in official decision-making processes (Norton *et al.*, 2005).

To address the aforementioned problems, participatory IPM, a non-expert-led, non-closed-systems, and non-research-driven strategy, was proposed (Table 2.1). It advocates farmer involvement as a means to enhance traditional IPM, by leveraging on farmers' own experiences in their own crop pest managements (Pretty, 1995). In addition, it takes advantage of the complementarity of farmer and scientific knowledge to improve the effectiveness in managing pests. Perhaps one of the most widely adopted participatory IPM approaches is farmer-field-school, through which IPM has moved from training towards education, exploration, and empowerment (Peshin and Dhawan, 2009b; Peshin and Dhawan, 2009a). Despite of these benefits, participatory IPM is costly, measured in terms of per farmer reached, which severely limited its outreach capacity to a relatively small proportion of farming communities (Luther *et al.*, 2005). The questions remain as to how to enable participatory actions of millions of farmers to reveal IPM knowledge; and how personalized pest management information can be diffused to them cost-effectively.

Recently, scholars have started promoting information and communication technology (ICT) as a way to enhance participatory IPM (Peshin *et al.*, 2009a). Incorporating geo-aware and community-based technology into IPM, thus a VGI-based IPM, has been proposed as one such solution (Deng and Chang, 2012). A VGI-based IPM approach is envisioned to enable interactions amongst all pest management stakeholders (e.g., farmers, scientists, extension workers, and policy makers) beyond geographic boundaries, taking care of their daily observations, perceptions, resource constraints, and objectives in pest management. Such a bottom-up approach has the potential to drive IPM towards a new paradigm of greater participation, communication, collaboration, and transparency that necessitate a timely, ubiquitous, and constant flow of diverse pest management information.

The next sections will further review the advantages of VGI pertaining to the enhancement of IPM in detail, the challenges to achieving VGI-based IPM will then be idenfited.

Table 2.1 Development of IPM and its supporting techniques. Note that different IPM strategies (participatory and non-participatory strategies) coexist from 1970s onwards.

Period	IPM strategy	Main supporting technique	Characteristics	Reference
1960s onwards	Linear, top-down, and research-driven.	Training and visit extension, integration of biological control and chemical control, habitat management, genetic engineering (pest-resistant crop varieties), semio-chemicals, selective pesticides and botanicals, cultural control, ecological niche.	Expert-led operations, closed-systems, lack of farmer participations.	(Peshin and Dhawan, 2009a), (Peshin and Dhawan, 2009b)
1970s onwards	Participatory.	Farmer field school, farmer-first, rapid rural appraisal, participatory rural appraisal, focus groups, structured workshops, farmer congress.	Criticism on top-down IPM techniques, an emerging role as an active participatory approach for community engagement.	(Dlott *et al.*, 1994), (Norton *et al.*, 2005), (Peshin and Dhawan, 2009a), (Peshin and Dhawan, 2009b)
2000s onwards	Location- and community-based, participatory, and ubiquitous collaboration.	VGI, participatory GIS, user-generated content, GPS, Web 2.0 services, keyhole markup language (KML), application programming interface (API), webGIS.	Interactive geospatial ICT for multilateral social interactions and collaborations among all stakeholders, such as farmers, scientists, extension workers, and policy makers.	(Goodchild, 2007a), (McCall, 2008), (Deng and Chang, 2012)

2.2.2 The potential value of VGI in enhancing tactical and strategic IPM

Barfield and Swisher (1994), amongst others, have defined what they call two "schools of thought" under the umbrella of IPM, i.e., tactical IPM and strategic IPM (see Section 1.2 for their definitions). Essentially, Tactical IPM (the dominant form of current IPM) is seen as a more responsible approach to the use of pesticide (e.g., forecasting pest outbreaks to determine the best timings for pesticide uses) based on a "sample, spray, and pray" cycle; while strategic IPM emphasizes the needs for a thorough understanding of the pest physiology and ecology across the globe (Zalucki *et al.*, 2009; Zalucki *et al.*, 2015). Morse and Buhler (1997) proposed a tactical-strategic IPM axis model, of which the strategic end of the axis is continuously extending as the understanding of agro-ecosystem evolves while the tactical end of the axis is more stable as it is less ambitious.

Section 2.1.1 has described the value of VGI in general, the following sections will describe the value of VGI in improving IPM (both tactical and strategic) in particular.

2.2.2.1 Value in IPM information collection for sense making

In terms of tactical IPM, the sample phase of tactical IPM sends experts to fields (McMaugh, 2005) or deploy pest monitoring traps (Augustin *et al.*, 2012) to collect pest surveillance data. These data collection methods have the well-known issues pertaining to high human resources cost, experts' lack of indigenous knowledge, inaccessibility to remote rural areas, coarse temporal resolution to reflect changes on the ground, and inaccurate "geo-registration" of the data (e.g., pest traps can attract pests from outside the targeted area). However, the change towards real-time and ubiquitous pest surveillance data collection through the general public (i.e., VGI) has the potential to remediate these issues (Goodchild, 2007a).

In terms of strategic IPM, managing pests strategically is conducive to avoid potential pest risks in the years ahead. Studies are needed to counter the environmental and economic impacts of such risks. Large-area plus long-term pest surveillance is therefore needed to understand the ecology of pests pertaining to various external factors (e.g., environmental factors, habitat factors, and phenological factors) and the related risks. Such a task demands even higher human resources and expenses. The problem is further exacerbated given the fact that certain pest species are rare, elusive, or difficult to detect (Thompson, 2013) . Therefore, the strategic IPM has often been seen as only an inspirational goal, and the strategic IPM researchers and practitioners are in a minority compared to those of tactical IPM (Barfield and Swisher, 1994; Zalucki *et al.*, 2015). However, VGI gives hope to accelerate the development of strategic IPM, simply because VGI are ubiquitous and cost-effective (Goodchild, 2007a).

Due to these inherent advantages of VGI in information collection for both tactical and strategic IPM, it has the potential to reveal more meaningful knowledge that are unknown through traditional IPM approaches. The next section will further describe the value of VGI in IPM information dissemination.

2.2.2.2 Value in IPM information dissemination

In contrast to traditional participatory IPM which can only accommodate limited numbers of participants (e.g., Farmer field school and structured workshops, see Table 2.1), all pest management stakeholders can freely interact with each other to exchange ubiquitous information in VGI-based IPM platforms. VGI-based IPM therefore allows for the sense making of the ubiquitous data of individual participants and the dissemination of personalized IPM information of close interests of the participants; it also allows for feedback from information receivers in a cost-effective manner. This is a great advantage of VGI-based IPM dissemination compared to traditional ones (e.g., radio programmes) which are often aimed for a large heterogeneous audience without adjusting its information to individual needs and allowing for any further interaction (Peshin and Dhawan, 2009a). In addition, data collection in VGI platforms can be real-time across large areas. This can enable instantaneous and ubiquitous information disseminations for time-critical scenarios (e.g., pest outbreaks). Pest management stakeholders enabled with Internet or cell phone connections can receive and view the information almost anytime and anywhere.

Enhancing both types of IPM with VGI, however, comes with challenges. The following sections will identify several research gaps pertinent to achieving such a goal. This thesis specifically aims to fill these research gaps.

2.3 Research gaps in VGI-based IPM

2.3.1 Potential enhancement of VGI-based IPM framework

As mentioned in the Section 1.2, the current VGI-based IPM has been limited to the primitive conceptual framework consisting of three general components, i.e., information collection, sense making, and dissemination (Deng and Chang (2012). Computational implementation for VGI-based IPM can be readily built using existing off-the-shelf computing technologies such as the implementation presented by Suen *et al.* (2014). It is, however, necessary to enhance the framework by incorporating components regarding the methods of information collection, quality assurance, and sense making into it. In addition, a good conceptual framework of VGI-based IPM must emphasize the unique advantages or features of VGI in enhancing IPM. For example, it appears that public participation GIS (PPGIS), which is closely related to VGI (Cinnamon and Schuurman, 2013), is equally effective in enhancing IPM because it also engages public involvement and interactive data collection and dissemination. But distinctions between VGI and PPGIS do exist and should be identified (e.g., whether the focus is on information or outcomes) (Tulloch, 2008; Cinnamon and Schuurman, 2013; Brown *et al.*, 2014). Therefore, stressing the uniqueness of VGI in enhancing IPM in the framework is necessary.

An enhanced conceptual framework of VGI-based IPM can serves as a better framework of reference for guiding the research on VGI-based IPM, and thus the development of more comprehensive VGI-based IPM. This become the first issue this thesis aimed to address.

2.3.2 Differences between VGI and conventional data as obstacles to VGI sense making

In contrast to conventional data, VGI are generated, accessed, and adapted by the general public. The general public can be either authoritative or non-authoritative. They can be either skilled professionals or enthusiastic but unskilled amateurs. The various levels of expertise of VGI contributors lead to the diversity or heterogeneity of VGI (Foody *et al.*, 2013). This is further complicated by the fact that VGI can be explicitly generated (with a specific purpose in mind) or implicitly generated (without a specific purpose in mind) and comes with non-uniform formats and semantic descriptions (Coleman *et al.*, 2009; Craglia *et al.*, 2012). Coote and Rackham (2008) introduced eight characteristics of conventional datasets, based on which a detailed comparison between conventional datasets and VGI datasets can be provided (Table 2.2). As can be seen from Table 2.2, the differences between VGI datasets and the conventional datasets are manifold. Such differences may hinder the mining of data for discovering pest management knowledge, and thus poses a question as to whether VGI sense making can indeed generate meaningful outcomes for enhancing IPM. In other words, is it feasible to utilize VGI as a source of data for IPM knowledge discovery?

Table 2.2 A comparison between conventional datasets (adapted from Coote and Rackham (2008)) and VGI datasets.

Characteristic	Conventional datasets	VGI datasets
Purpose	Created for a specific and defined set of requirements whether for legal, administrative, or commercial purposes.	Created by VGI contributors for various personal purposes with different motivations (Coleman *et al.*, 2009; Craglia *et al.*, 2012).
Cost	Depending on the context, data may or may not be freely available, but usually there is at least some dissemination charge and, most likely, restrictions on access and use.	Depending on different VGI platforms, VGI may be freely accessible (e.g., OpenStreetMap) or partially accessible (e.g., Twitter). However, VGI platforms encourage data sharing, cross-referencing and communication in cyberspaces, free information disseminations to some extent can motivate and sustain user participations (Goodchild, 2007b; Goodchild, 2008).
Management	Managed by organizations established for the purpose whether as public or commercial bodies. There may be collaboration between organizations but on the basis of legal agreements including commercial contracts.	Managed either by people who established VGI communities (as gate-keepers) or by community members themselves (collective management) (Goodchild and Li, 2012).
Source	Collected by professional and certified people who are paid to do so.	VGI not only is the product of authoritative agencies but also the product of the broader and amateur communities, often through contributions to collaborative activities, without monetary remuneration (Coleman *et al.*, 2009; Foody *et al.*, 2013).
Collection	Based on well-established methods, standards, specifications, and practices for focused data collection, can ensure the data completeness for targeted study site, through continuous observations over time or observations at a fixed temporal interval.	Based on random, pervasive, offhand, and real-time observations from contributors equipped with their particular personal or local knowledge (Tulloch, 2008; Goodchild, 2009). May encounter appearing and disappearing coverage in observation, but could be ubiquitous in spatiotemporal coverage and rich in data volume (Kuhn, 2007; Mashhadi and Capra, 2011).
Quality	Quality assured to varying degrees during the production of the data and supplied with some information, however basic, on the quality of the data.	Lack of quality control in the data collection processes, more prone to be erroneous, even artificial (Girres and Touya, 2010). Often lack of explicit metadata regarding the data quality (Brando and Bucher, 2010).
Licensing	Protected by some forms of copyright and governed by formal agreements or licenses.	VGI has higher shareability. (Ballatore and Bertolotto, 2011), but user privacy and security and the related legal issues are attracting attention (Song and Sun, 2010; Blatt, 2013; Scassa, 2013).
Access	Access limited, in many cases, to only certain organizations or individuals for reasons of security, data protection, or commercial advantage.	Higher accessibility, in some cases (e.g., OpenStreetMap), users are free to copy, distribute, transmit, and adapt VGI data, as long as they credit the VGI community and its contributors (Goodchild, 2007b).

For tactical IPM, empirical statistics have been widely adopted for predicting crop pest emergencies. For example, Lam *et al.* (2001) adopted multiple regression model for predicting the population fluctuations of bean leaf beetle in soybean; Gumpertz *et al.* (2000) used logistic regression to predict southern pine beetle outbreaks. Artificial intelligence is another popular approach for predicting crop pest emergencies. For example, Yang *et al.* (2009) predicted the population dynamics of paddy stem borer based on artificial neural network; Tripathy *et al.* (2011) used Naïve Bayes to model Thrips pest population dynamics. However, these approaches suited to analyzing conventional pest surveillance data become less appropriate for analyzing VGI. Because they require regular and systematic data collections over very long periods of time (e.g., over years) to establish the correlations between population dynamics and environmental variables. Heat-driven phenology models such as degree-day model has also been used for pest emergency prediction (Herms, 2004). For example, Jones *et al.* (2008) used degree-model to predict codling moth emergence. This approach may not need data collections over very long periods of time, as its focus is the mechanisms associated with the occurrence time of specific phenological events of pests rather than the correlation mentioned above. However, it is challenging for individual non-professional pest observers to correctly identify and report the occurrence time. Therefore, in order to make sense of VGI to possibly generate meaningful outcomes, special treatments, or interdisciplinary methods suited to the characteristics of VGI are needed.

For strategic IPM, two approaches are becoming popular for understanding potential pest distributions: physiological model (Pilkington and Hoddle, 2006; Hartley *et al.*, 2010) and ecological niche model (ENM) (Beddow *et al.*, 2010; Wang *et al.*, 2010; Aragón *et al.*, 2013). Using a physiological model for such a purpose requires large amount of experimentation to accurately estimate various physiological parameters of pest developments, which are beyond the reach of VGI efforts (at least for so far). VGI seems fit the input requirement of ENM which simply needs location-based species occurrence records collected anytime and anywhere. As a spatially ubiquitous source of data, VGI can be valuable data for ENM, especially given that ENM for mega areas (e.g., globe) is often constrained by the scarcity of species occurrence records (Urbina-Cardona and Loyola, 2008). An empirical study is therefore needed to utilize VGI for large-area ecological niche modelling, demonstrating whether meaningful sense making outcomes can be generated.

This thesis therefore aimed to generate insights into the sense making of such unconventional data for both tactical and strategic IPM.

2.3.3 VGI quality assurance

The differences between conventional datasets and VGI datasets summarized in Table 2.2 also contribute to the concerns over VGI quality. How to assure VGI quality is also a major challenge in enhancing IPM through VGI. The following section will first review how the quality of VGI is assured by the approaches from existing work. It will also illustrate the related shortcomings, which pointed out the direction of my research on developing a novel method for assuring the quality of volunteered pest surveillance data.

2.3.3.1 Geospatial data quality and its assessment elements

In GIS community, data quality has been studied for more than 20 years, but the meanings of the data quality remain context-dependent (Devillers *et al.*, 2005; Stark *et al.*, 2011). There are at least two kinds of quality of geospatial data (Devillers and Jeansoulin, 2010): (1) internal data quality, referring to the assessment of the difference between a dataset and the reality it represents; and (2) external data quality, referring to the fitness for use, or the extent to which a dataset can be a good fit for its different uses. External data quality can be considered complementary to internal data quality (Poser and Dransch, 2010).

Indeed, the classification of geospatial data quality are typically perceived from the producer and the user perspectives (Devillers *et al.*, 2005). The producer perspective focuses on internal quality while the user perspective focuses on both internal and external quality. For VGI, because the boundary between data providers and users is no longer clear-cut, the two perspectives of data quality are quickly converging. VGI should fit for a wide variety of purposes from different users (i.e., external data quality), but the data quality also requires to be assured prior to externalization (i.e., internal data quality) by the users themselves as they also play the role of data provider.

Unlike tangible products of which the quality can be measured through their physical properties (e.g., life span of a laptop screen), geospatial data have no such physical characteristics that allow quality to be easily assessed. Instead, its quality is a function of five intangible elements (i.e., completeness, logical consistency, positional accuracy, temporal accuracy, and thematic accuracy), and three overview elements (i.e., purpose, usage, and lineage) (ISO-19113, 2002). These elements form the standards of assessing both internal and external data qualities.

2.3.3.2 Direct and indirect approaches of VGI quality assurance

Based on the data quality elements mentioned above, several existing studies on VGI quality assurance have adopted a direct approach which compares a VGI dataset to an authoritative gold standard reference dataset. For example, Zielstra and Zipf (2010) examined the completeness of a German OpenStreetMap dataset in comparison to a TeleAtlas MultiNet dataset. Haklay (2010) compared a London OpenStreetMap dataset with an Ordnance Survey dataset based on positional accuracy and completeness. More comprehensively, Girres and Touya (2010) extended the work of Haklay (2010) by comparing a French OpenStreeMap dataset with a BD TOPO® IGN dataset based on a larger set of spatial data quality elements.

Such a direct approach can be seen as an adoption of the traditional data quality assessment method that focuses on internal data quality (Devillers *et al.*, 2005). It, however, has limited applicability for assuring the quality of VGI as there is generally an absence of authoritative gold standard reference datasets for VGI applications (Bishr, 2007; Kuhn, 2007). For example, in the case of utilizing VGI for species surveillances (pest species in this thesis), voluntary observations are often conducted in sparsely populated, rural, or less explored areas of the world. In such a case the gold standard reference datasets are often lacking. In addition, VGI dataset is often more up-to-date than authoritative dataset and thus may be more accurate than the so called gold standard reference dataset (Goodchild and Li, 2012). To cope with this issue, indirect approaches relying on surrogate criteria were proposed. Five mainstream indirect approaches are described as follows:

(1) The user review approach (Maué and Schade, 2008; Goodchild and Li, 2012):

This approach is user-driven and relies on Linus' Law which assumes that "given enough eyes all bugs are shallow". Based on Linus' Law, user contributions converge on a truth through an iterative error correction process, either in terms of attributive error or positional error, or both. If one user commits an error, the error can be detected or corrected by the other users. Haklay *et al.* (2010) have applied this approach to OpenStreetMap and suggested its applicability to VGI in general.

(2) Targeted recruitment (Seeger, 2008; Cinnamon and Schuurman, 2013):

This approach relies on recruiting participants with the abilities that meet pre-defined criteria to contribute geospatial information within an established geographic extent, or even relies on participant training to ensure high quality data can be volunteered (See Section 2.1.2 for more details).

(3) The provenance approach (Trame and Keßler, 2011; Celino, 2013):

This approach relies on the history of volunteered information. Requesting or tracing the history of a VGI dataset (e.g., who are the data providers?) is helpful in better understanding and assessing its quality.

(4) The geographic approach (Goodchild and Li, 2012):

This approach is based on Tobler's first law of geography, which assumes things that are closer are more related than things that are farther apart (Tobler, 1970). A VGI contribution should fit its geographic context, e.g., a report of a species occurrence is more likely to be true if many similar reports exist nearby. In addition, more credit can be given to a VGI if it is volunteered by a local resident who is physically close to the site of the VGI event and is familiar with the local environment (Seeger, 2008).

(5) The trust approach (Bishr, 2007; Bishr and Mantelas, 2008; Bishr and Janowicz, 2010):

It uses trust as a proxy of quality to establish a link between VGI quality and VGI contributors' authority based on subjective evaluations. It rests on the extent to which a VGI contributor has provided honest and accurate information. Trusted VGI contributors tend to provide more trustworthy information compared to less trusted ones. The criteria for evaluating the trustworthiness of VGI replace traditional quality measures of geospatial information (e.g., completeness, logical consistency, and positional accuracy). Indeed, the information asymmetry and imperfection of a VGI environment can lead to social uncertainties in VGI consumptions (Sniezek and Van Swol, 2001). When high social uncertainties exist, trust appears to be particularly important as it reduces social uncertainties by confining the range of behavior expected from another (Sniezek and Van Swol, 2001).

2.3.3.3 Challenges in using the indirect approaches for pest surveillance applications

Among the indirect approaches, the user review approach works well for those VGI that are more traceable, such as those in Wikimapia and OpenStreetMap. However, it is problematic for pest surveillance applications because the objects being recorded are often highly mobile or persist for only a short period of time. It is hardly possible to go back to the reported locations to verify every user surveillance report and therefore it is not peer-reviewable. Goodchild and Li (2012) also pointed out that this approach works less well for obscure phenomena, including those short-lived ones. Conducting the review process for time-critical issues (e.g., pest outbreaks) is also impossible because the process is generally time-consuming. Additionally, Linus' Law sometimes fails. In a crowdsourcing-based cropland capture game, Salk *et al.* (2015) demonstrated that the majority agreement among volunteers cannot fully substitute the quality assessment by experts on crowdsourced tasks. In addition, the target recruitment approach only applies to projects based on facilitated VGI creations. For projects based on active VGI creations with unconfined geographic extents and more heterogeneous participants, this approach simply does not work (See Section 2.1.2 about active and facilitated VGI creation).

The provenance approach, geographic approach, and trust approach appear to be more applicable. However, when used alone, all three approaches fall short in fully describing VGI data quality.

The provenance approach considers VGI provenance, including data contributors' expertise. What is challenging, though, is how to appropriately incorporate provenance of user expertise as the expertise level of a VGI contributor is difficult to collect (Keßler *et al.*, 2009a). There are also resistances in providing such information due to the concerns on personal privacy and security (Song and Sun, 2010). According to Coleman *et al.* (2009), VGI contributors can be classified into five types: (1) neophyte, (2) interested amateur, (3) expert amateur, (4) expert professional, and (5) expert authority. Normally, people are inclined to trust contributors who are expert professional and expert authoritative. However, a contributor considered to be an expert may understand a project's specification very well but lack the knowledge of local history or attributes. A contributor considered as either a neophyte or interested amateur may know little about the professional part of a VGI project but is very familiar with the characteristics and details of his or her current location. In short, the boundary between non-expert amateur and expert professional is quickly blurring in VGI environments where the expertise of a contributor cannot be simply judged based on contributor type.

As for the geographic approach, considering only fitness of geographic context tends to be less effective if a user report fits surrounding geographic context well but actually is a false observation.

Regarding the trust approach, how trust as a proxy of quality can be effectively realized in VGI contexts is problematic. It demands appropriate methods to evaluate and quantify the trustworthiness of VGI. In Bishr and Mantelas (2008) an approach combining the trust approach and the geographic approach was proposed to assure VGI quality. Their work does provide valuable insights into the usage of the proxy. First, indeed, the four indirect approaches reviewed here are not mutually exclusive. For instance, some of the elements of trust fall under the geographic approach, i.e., the trustworthiness of VGI can be assessed based on geographic contexts. Second, their approach leverages crowd's dual roles in VGI creation–contributing locational data and ascertaining the reliability of data (i.e., user trust rating). The second role can be helpful in evaluating the trustworthiness of VGI. Despite these insights, in the combined approach, fuzziness that is inherent in trust (Chang *et al.*, 2005) is not well accounted for. Assessing the quality of VGI based on trust requires special attentions to the fuzzy nature of trust.

Novel approaches thus are called for to synthesize the advantages and minimize the disadvantages of the approaches mentioned above to assure the quality of VGI, with a better way to account for user expertise, geographic context, and fuzziness involved in trust judgment.

2.3.4 The roles of volunteer participants and participation incentives

Apart from the research gaps mentioned above, it is also important to learn what a volunteer participant can do in VGI-based IPM, as maximizing the value of volunteer participants by understanding their roles in management issues contributes to the realization of the corresponding management goals. Previous studies have investigated the roles of VGI in various application domains such as landscape planning and site design process (Seeger, 2008), urban management (Song and Sun, 2010), outdoor recreation planning (Parker *et al.*, 2013), and emergency management (Li and Goodchild, 2012). However, the specific roles that volunteer participants can play in VGI-based IPM are seldom understood, which should differ from those in traditional participatory IPM (e.g., workshop participants, see Section 2.2.1). This necessitates an exploration of this study to understand the practical and potential roles of volunteer participants in VGI-based IPM.

With the understanding of the roles, motivating and sustaining volunteer participants to play such roles is also important. The success of a socially networked community depends on the degree to which members will stay and keep participating (Jin *et al.*, 2009), especially for those based on facilitated VGI (Seeger, 2008; Cinnamon and Schuurman, 2013). Coleman *et al.* (2009) studied the motivations of VGI contributors in general. Hall and Graham (2004) examined the user motivations to exchange knowledge in Yahoo! e-group online community. Gouveia and Fonseca (2008) explored the motivations of VGI creators for environmental monitoring issues. The motivations of OpenStreetMap data contributors have been even more thoroughly investigated (Budhathoki, 2010; Lin, 2011; Steinmann *et al.*, 2013). The identified motivations may range from an individual agenda such as professional or personal interest, enjoyment, reputation, anticipated reciprocity, to more altruistic ones. Despite these previous studies trying to understand what reasons have initiated people to contribute their data, little research so far focused on exploring proper incentive strategies to motivate and sustain volunteer participation in VGI platforms. VGI contribution cannot simply be stimulated through strategies from traditional participatory projects (e.g., material or monetary incentives) given the volunteer nature of VGI. Indeed, without proper incentives, a large number of useful VGI contributions are usually made by a few enthusiastic individuals (Goodchild and Li, 2012). Similar observation was reported for OpenStreetMap, merely 5% of all registered users productively contributed to the project; and 81% of registered users contributed nothing or never returned after making only a few contributions with various fleeting motivations (Neis and Zipf, 2012). Martella *et al.* (2015) proposed using gamification as a strategy to keep participant engaged. However, the applicability of this approach is limited. In the case of crop pest surveillance, for example, the crop fields in rural areas may not be ideal sites for gaming. This thesis thus aimed to explore incentive strategies to motivate and sustain the volunteer participation in VGI-based IPM.

The current chapter has identified several research gaps related to achieving VGI-based IPM. The next chapter will address the first research gap, i.e., proposing an enhanced conceptual framework of VGI-based IPM.

3 An enhanced conceptual framework of VGI-based IPM

The social element of GIS has long been implicated by one of its early definitions, namely GIS is "a system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing, and disseminating information about areas of the earth" (Dueker and Kjerne, 1989). Such social participation aspect of GIS has only become prominent since the advent of public participation GIS (PPGIS) in the 1990s (Frančula, 2011) and VGI in the 2000s (Goodchild, 2007a).

Traditional GIS has often been argued as a contradictory technology that empowers the elite few while also marginalizes most people and communities (Stewart *et al.*, 2008). It has been criticized for its exclusive characteristics due to its high accessing barriers, e.g., its high costs of hardware, proprietary software, high complexity, and high level of expertise required for implementation (Manuel, 1996; Harris and Weiner, 1998; Obermeyer, 1998; Ghose, 2001; Hoyt *et al.*, 2005; Sieber, 2006). GIS experts are able to produce professional geospatial products and conduct complex spatial analyses whereas non-experts are only able to view the information from final GIS outputs produced by experts. Traditional GIS has thereby been seen as the "cartography for the few" (Cartwright, 2009), in which far less attention has been given to the use of GIS for grassroot groups, unprivileged communities, ordinary citizens, and activists alike to benefit their everyday lives. These criticisms have sparked changes to traditional uses of GIS (Elwood, 2009); PPGIS and VGI have therefore featured prominently.

As mentioned in Section 1.2, the VGI-based IPM framework from Deng and Chang (2012) is still primitive. A good conceptual framework of VGI-based IPM should emphasize the unique advantages or features of VGI in enhancing IPM. It appears that PPGIS, as a close relative of VGI (Cinnamon and Schuurman, 2013), actually also can be adopted to enhance IPM because it also engages public involvement and interactive data collection and dissemination. But distinctions between VGI and PPGIS do exist and should be identified (Tulloch, 2008; Cinnamon and Schuurman, 2013; Brown *et al.*, 2014). In addition, it is necessary to enhance the framework by incorporating components regarding the methods of VGI collection, quality assurance, and sense making into it. Therefore, this chapter aims to propose such an enhanced framework of VGI-based IPM for guiding the related research, and thus the development of more comprehensive VGI-based IPM.

The following sections of this chapter will first discuss the characteristics of traditional GIS and the related critics that drive the development of PPGIS and VGI phenomena (Section 3.1). The characteristics of PPGIS and VGI will be discussed next, where a comparison and contrast between them from theoretical perspectives will follow

(Section 3.2 and Section 3.3). An enhanced VGI-based IPM framework will be proposed in the last section (Section 3.4).

3.1 Traditional GIS

It is often argued that traditional GIS rests on and amplifies an essentially positivist philosophical perspective (Warf and Sui, 2010). According to Kwan (2002), traditional GIS has been largely understood as a positivist or empiricist science, which is rooted in the quantitative revolution of geography and as such inherits the corresponding positivism or empiricism.

Ontologically, positivism recognizes one reality that can be known with certain probability; there is a universal law independent from spatiotemporal contexts (Mertens, 1999). Epistemologically, positivism is associated with subject-object dualism, in which the knower is thought to be value-free and separated from the reality (Mantoura and Potvin, 2013). Methodologically, positivism drives quantitative approaches. The research process is largely deductive in that it focuses on testing theories rather than developing theories (Cupchik, 2001; Higginbottom and Lauridsen, 2014). This privileges the quantitative and the observable which are context-free rather than issue-driven. The qualitative and the non-observable are underprivileged. Space in traditional GIS is represented as a Cartesian coordinate system defined by Euclidean geometry. It follows Newton who views spatiality as absolute conceptualizations, representing space as independent spatial features (e.g. discrete vector features or raster cells), rather than Einstein and Leibnitz who view space as relational (Sheppard, 2005).

Therefore, traditional GIS has often been criticized for its inadequacy in representing relational spaces of social power and subjective differences amongst its analyzed objects (Kwan, 2002; Schuurman and Pratt, 2002). It lacks the power to enable researchers to understand neighborhood-level knowledge about lived experiences of local people, or social ties and attachments of local people to their communities (Pavlovskaya, 2009). Therefore, researchers and decision-makers alike lack well-grounded, rich descriptive, communicative, and explanative local contexts for approaching realities so as to make the best decisions.

3.2 PPGIS

To address the critiques on traditional GIS, a range of qualitative GIS have thrived in the postmodern era (Warf and Sui, 2010), among which PPGIS is perhaps the most well-known (Pavlovskaya, 2009; Yeager and Steiger, 2013; Schoepfer and Rogers, 2014). The development of qualitative GIS reveals that GIS can also be employed within a non-positivism paradigm. Many kinds of qualitative materials and situated perspectives (e.g., photographs, sketch maps, grounded visualizations, videos, personal experiences, preferences and perceptions, and narratives) can be incorporated into GIS.

PPGIS has risen since 1996 through the coalescing of GIS and participatory action research (Frančula, 2011). This occurred more than ten years earlier before the coining of VGI by Goodchild (2007a). PPGIS was originally defined as “a variety of approaches to make GIS and other spatial decision-making tools available and accessible to all those with a stake in official decisions” (Sieber, 2006). It is typically targeted at enhancing public participation in planning and policy issues (Brown and Kyttä, 2014). A growing number of projects have been implemented under the banner of PPGIS (Stewart *et al.*, 2008), which supposedly reduced or dismantled the barriers of traditional GIS mentioned above. For example, Barnett *et al.* (2016) used a PPGIS approach that enabled fishermen to tell their perceptions of risk in marine debris mitigation in the Bay of Fundy, Canada. More comprehensively, Brown (2012) reviewed and discussed 15 empirical PPGIS research which allowed people to contribute their opinions, preferences, personal experiences to assist regional and environmental planning.

Differing from traditional GIS, PPGIS is rooted in constructivist philosophy (Sieber, 2006; Michanowicz *et al.*, 2012). In this paradigm, people are motivated to produce their own GIS outputs based on public available GIS tools or settings. This denies the positivistic GIS’ reliance on the correspondence theory of truth which corresponds to objective realities and assumes that representation is not a social process (Warf and Sui, 2010). Qualitative GIS in general and PPGIS in particular in this sense are closer to the consensus theory of truth (Warf and Sui, 2010). In this theory, truth is constructed. Communications among individuals and communities of interest are central to the social process of truth construction.

In constructivism, ontologically, there is no universal law or absolute truth, but there are multiple realities that are socially constructed; it embraces an absolute relativism, i.e., no perspective is any privileged than any other perspective (Pound *et al.*, 2003; Dunn, 2007; Higginbottom and Lauridsen, 2014). Epistemologically, it is subjective and value-laden, but equal validity are granted to all personal values; social interaction is thought to be essential for knowledge creation (Mertens, 1999; Cupchik, 2001). Methodologically, it is dominated by qualitative approaches and the research process is largely inductive that focuses on developing theories rather than testing theories (Cupchik, 2001; Higginbottom and Lauridsen, 2014).

Based on constructivism, PPGIS projects are thus strongly influenced by and highly contingent on the social context in which the project are positioned and the related spatial decisions are made (Ghose and Elwood, 2003). In such a sense, GIS becomes more reflective, where people collaboratively construct knowledge and make decisions through social interactions.

3.3 VGI

Unlike PPGIS, VGI is more diverse. VGI is not limited to planning and policy issues. It also has been observed that VGI interaction is neither limited to qualitative dominated approach nor to quantitative dominated approach. For example, OpenStreetMap mainly includes traditional quantitative map information; Flickr mainly involves qualitative information, i.e., spatial context-associated photographs; GeoCommons (http://geocommons.com/) includes both quantitative data such as the U.S. Unemployment Rate map (http://geocommons.com/maps/206016), and qualitative data such as Binders Full of Women which maps the geotagged Tweets responding to the U.S. presidential debate (http://geocommons.com/overlays/284513); and Wikimapia includes quantitative-qualitative-combined information, i.e., traditional quantitative map features added with people's qualitative descriptions and comments about the mapped features. On the contrary, the information collected through PPGIS projects are generally qualitatively dominated (see Section 3.2).

Therefore, I argue that one considerable relevance to VGI is the transformative paradigm (Mertens, 1999; Mertens, 2007; Mertens *et al.*, 2009). Transformative paradigm can be seen as an "emerging paradigm", which draws from and combines both the positivist and constructivist perspectives (Pound *et al.*, 2003). Transformative paradigm values marginalized individuals, groups, and communities (Sweetman *et al.*, 2010). Ontologically speaking, transformative paradigm acknowledges that knowledge are influenced by human interests; there are multiple socially constructed realities, but it recognizes the influence of differences in personal values in determining what is real (Mertens *et al.*, 2009). This is in contrast to the absolute relativism of constructivist point of view in which all perspectives have an equal legitimacy (Mertens, 1999). Indeed, determining the best or most trusted version of "reality" through data quality enhancement processes is among the core issues of any VGI program (Bishr and Mantelas, 2008; Goodchild and Li, 2012). From an epistemological angle, transformative paradigm seeks a balanced and complete view of a phenomenon to achieve accurate knowledge, which requires not only interactions but also in-depth interactions with the communities on which the program have impacts (Mertens, 1999). Lastly, in methodological terms, transformative paradigm may involve quantitative, qualitative, or combined methods (no single type of approach is always dominated). Knowers need not to prescribe a specific methodological orientation (Mertens, 1999). It might include the use of deduction, induction, abduction, or a combination of the aforementioned approaches.

3.4 An enhanced conceptual framework

To enhance IPM, utilizing VGI seems more appropriate than utilizing PPGIS, due to the diversity of VGI in comparison to PPGIS mentioned above. A more effective participatory IPM needs to be more diverse to empower large amounts of heterogeneous participants from everywhere to exchange diverse information (e.g., quantitative pest surveillance information for pest risk predictions, qualitative perspectives from farmers for facilitating IPM planning and policy makings, or the combination of both). Based on such a diversity of VGI compared to PPGIS, an enhanced framework of VGI-based IPM is proposed here (Figure 3.1). The framework incorporates single or combined methods, suited in a transformative paradigm. It serves as a framework of reference for enhancing IPM based on VGI as an alternative to the positivist paradigm adopted in traditional IPM (Douthwaite *et al.*, 2003), and for the development of more comprehensive VGI-based IPM.

In the context of VGI, data collection and sense making can be conducted simultaneous (in a real-time manner). The VGI retrieval component and VGI sense making component are therefore combined in the enhanced framework, represented using a bigger circle. Due to the diversity of VGI, a variety of choices of ways for creating and analyzing VGI should be supported in a more comprehensive VGI-based IPM. The VGI creations can be in active or passive, or facilitated forms (See Section 2.1.2). VGI collection and sense making methods for tactical and strategic IPM (see Section 1.2) can be purely quantitative or qualitative, or can be the combination of quantitative and qualitative methods in sequential or interactive (or arbitrary) order through the process. When quantitative methods are used with qualitative methods, qualitative information can provide contexts for the patterns generated by quantitative information, and vice versa (Johnson and Onwuegbuzie, 2004). Qualitative and quantitative methods can be utilized with equal weights or with different priorities. In addition, data quality assurance is highlighted in the VGI collection and sense making component, due to its importance in VGI uses. In Chapter 6, an expert system to assure the quality of quantitative VGI) will be presented.

Regarding the information dissemination component, personalized information dissemination (i.e., the usefulness of information dissemination to individual participants) needs to be stressed. It has been shown that the usefulness of the information received from a virtual community positively affects the willingness for one to continue participating in the virtual community; and the usefulness is identified by the extent to which users think that information is relevant, timely, accurate, and complete (Chen, 2007; Jin *et al.*, 2009). Enhancing the usefulness of information and information dissemination is therefore important to attracting participants at a broad geographic scale and to encouraging their continual contribution of VGI, thus the sustainability of VGI projects. Note that users from different regions may have different needs. An active participant who is satisfied by existing information disseminated may still discontinue his/her participation someday due to his/her varying needs (disenchantment discontinuance) or due to an availability of better alternatives (replacement discontinuance) (Peshin *et al.*, 2009b). Hence the usefulness enhancement mentioned above should also be adaptive.

Based on this enhanced framework, the next chapter will introduce a case study conducted for this thesis to contribute to the filling of the research gaps pertaining to VGI sense making for tactical IPM, participant roles in tactical IPM, and participation incentive issue. As mentioned in Section 1.5, this thesis focused on quantitative approaches (quantitative pest surveillance data), the qualitative approaches (e.g., the cognitive ability of volunteer participants in pest management) are left for future research.

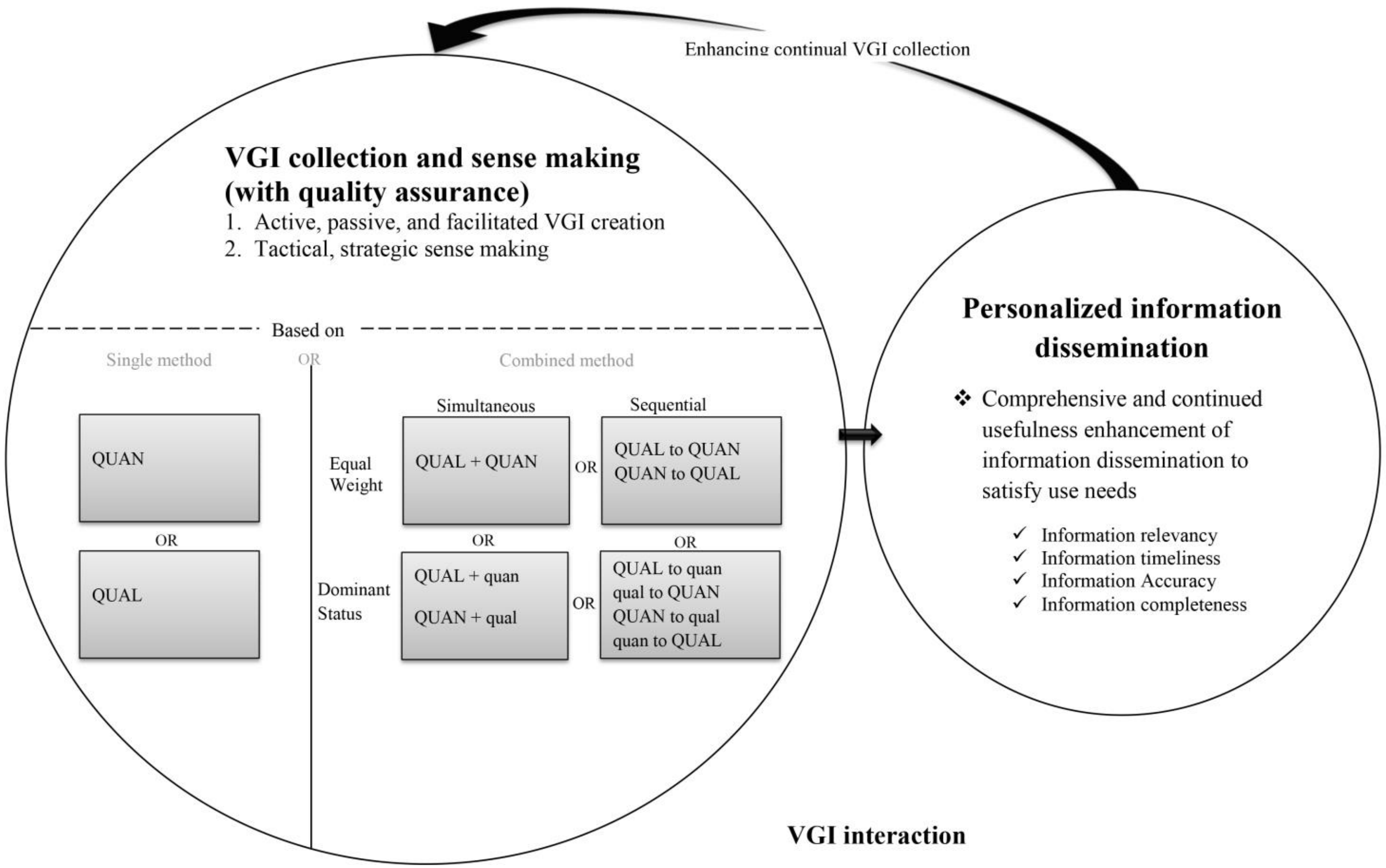


Figure 3.1 An enhanced framework of VGI-based IPM. In developing this framework, it was probably most influenced by Mertens (1999), Johnson and Onwuegbuzie (2004), Deng and Chang (2012), Peshin *et al.* (2009b), and Jin *et al.* (2009). *Note*. In the square boxes, "qual" stands for qualitative, "quan" stands for quantitative, "+" stands for simultaneous, "to" stands for sequential, capital letters denote high priority or weight, and lower case letters denote lower priority or weight.

4 Enhancing tactical IPM through facilitated VGI and the participation incentive for a VGI-based IPM

The traditional linear, top-town, and non-interactive crop pest management approaches have resulted in the limited success of the past IPM (Douthwaite *et al.*, 2003) (see Section 2.2.1 for the review). Therefore, this chapter explores a bottom-up IPM approach based on VGI, aiming to contribute to the filling of the research gap as to whether VGI sense making can indeed generate meaningful outcomes to enhance tactical IPM. It also aims to explore how to motivate and sustain the volunteer participation in VGI-based IPM. Meanwhile, it contributes to the understanding of the practical and potential roles of volunteer participants in tactical VGI-based IPM.

To serve the stated aim, a case study was conducted at Shuibian town, Jiangxi province, China. Based on the enhanced framework of VGI-based IPM proposed in Chapter 3, in the case study the real-time surveillance information regarding crop pest infestations was collected, analyzed, and disseminated. The VGI collected in this case study is relatively short term, forecasting-oriented, timely, and regionalized, which fit the concept of tactical IPM. The case study adopted a facilitated VGI creation method (see Section 2.1.2 about facilitated VGI creation) to manage the overwintering infestations of striped rice stem borers (*Chilo suppressalis*) in the study area. As indicated by the enhanced conceptual framework in Chapter 3, quality assurance is an important component of VGI-based IPM. Since VGI-based IPM is in its early developmental stage where lacks concrete VGI quality assurance techniques, utilizing facilitated VGI appeared to be more appropriate as the process of facilitated VGI collection would be more controllable. The VGI interaction in this case study was enabled through mobile phones and Web 2.0-based GeoWeb.

The following sections of this chapter will first present the background, methods, and results of this case study (Section 4.1 to Section 4.4). Section 4.5 will discuss the results of the sense making, the participation incentive analysis, and the roles of volunteer participants in tactical IPM.

4.1 Study area and background of the case study

This case study was conducted in Shuibian town at Jiangxi province, China (Figure 4.1). In every spring of Shuibian town, there is an intensive overwintering occurrence (outbreak) of striped rice stem borers. This pest can cause severe economic losses to various crops such as corn, sugarcane, oilseed rape, and especially rice, the major crop planted in Shuibian town. Pest management experts from the local agricultural department has made great effort to reduce the infestations caused by this pest.

The pest management experts investigate the pest occurrences typically by deploying pest traps and conducting pest surveys. However, due to limited resources, crucial real-time data for revealing the spatiotemporal characteristics of the infestations of the pest outbreak have barely been collected. On the one hand, the existing local expert-led pest surveillances have never discovered any spatial pattern of the pest outbreak across the whole town due to the limited areas of the pest surveillances they could conduct. On the other hand, the existing local expert-led pest surveillances could identify outbreak of adult emergence of the pest using pest traps, but are not able to identify any temporal pattern about the actual infestations of the pest across the whole town. The infestations are caused by larvae rather than adults of the pests, while the traditionally used pest traps can only trap adult pests. A systematic survey regarding infestations caused by the larvae across a large area was also time consuming, laborious, and costly. Therefore, such spatiotemporal information is lacking from the existing local expert-led pest surveillance outcomes. Accordingly, this case study investigated whether patterns of the infestations of the pest outbreak could be revealed through volunteer efforts.

This case study facilitated the local farmers to conduct a collaborative pest surveillance from 15 April to 27 May 2015. They were asked to report the stem borer infestation incidents during their routine farming work in real-time. Since this collaborative pest surveillance involved real-time user participations and information disseminations, user participation incentive was also investigated.

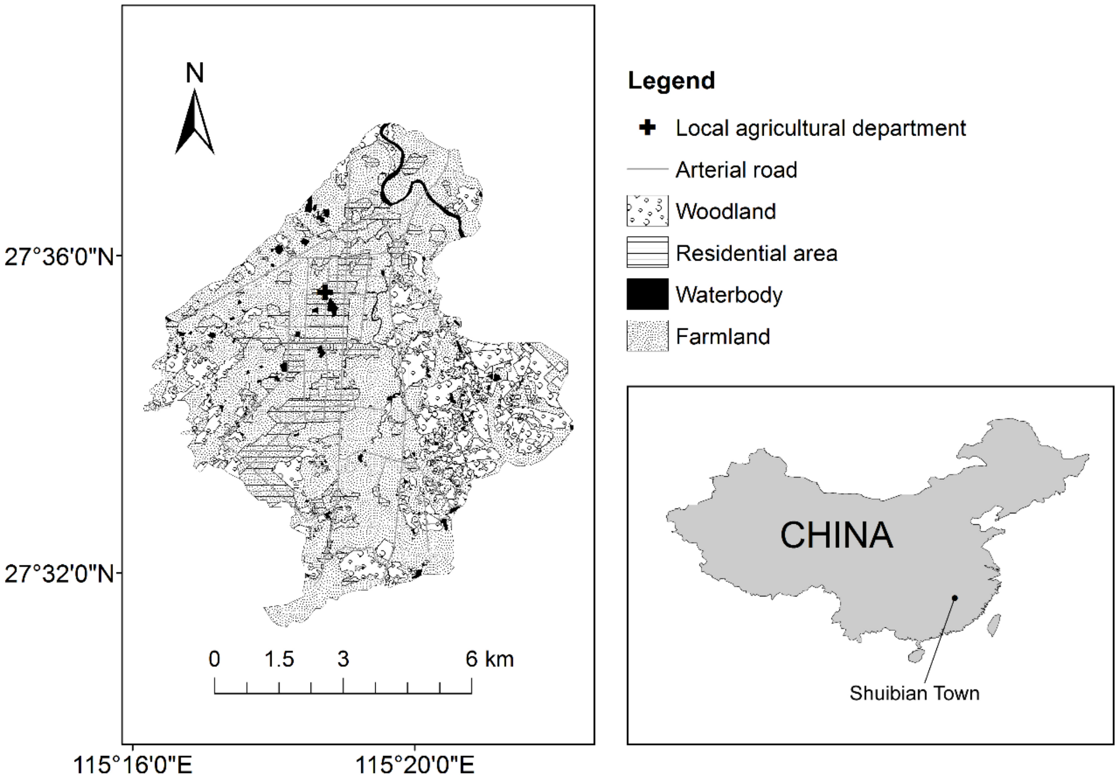


Figure 4.1 Study area: Shuibian town of Xiajiang prefecture, Jiangxi province, China.

4.2 Study setup: Participant recruitment and pest data collection

A local agricultural extension worker was hired to recruit participants several months prior to the specified pest surveillance period. In order to facilitate the local farmers to participate and ensure that the surveillance was driven by the farmers' own interests, the rationale, goals, and process of the project were explained. Brief negotiations were conducted by telling that, with their cooperation in the surveillance, they would receive personalized feedback for themselves to manage the pest infestations. Therefore, they would be willing to take time and effort to share accurate pest observation information. The farmers were neither paid nor given material compensation for their participations.

Farmers were randomly approached initially, and only the farmers with rice stem borer observation experiences and expressed willingness of participation were recruited. The strategy adopted here was referred to as targeted recruitment which is a participant recruitment strategy of facilitated VGI creation (Cinnamon and Schuurman, 2013). The crop fields of the recruited farmers randomly distributed across the study area. In the end, 233 participants were recruited. Table 4.1 shows the participant demographic characteristics. Note that most of the participants were aged between 26 and 40 (45.5%). This was because many elders showed less interest in the project than the relatively younger farmers, and many young people (25 or younger) in the study area have moved to big cities to make their livings. Hence the resulting analysis outcome may have certain bias towards the ages between 26 and 40. The participants' crop fields were coded. The locations of the crop fields were pre-collected using the Trimble® GeoXT handheld GPS device, which delivered 50 cm positioning accuracy.

Table 4.1 Participant demographic characteristics.

Characteristic	**Percentage (%)**
Age	
25 or younger	18.5
26 to 40	45.5
41 to 55	24
56 or older	12
Gender	
Female	38.2
Male	61.8
Education	
Primary school or below	16.7
Junior high school	58.5
Senior High School	23.2
Secondary technical school	10.7
Diploma	0.9
Bachelor's degree or above	0
Farming experience	
5 years or less	18
5 to10 years	20.6
10 to 20 years	27.5
20 to 30 years	15.9
30 to 40 years	9.4
40 years or more	8.6

The participants could report pest infestation incidents through two ways: (1) using short message service (SMS), for those who had no smartphones and Internet access; (2) using a GeoWeb application for the crowdsourcing of pest infestation observations, built based on ArcGIS Online (https://www.arcgis.com/home/). Trainings were provided to ensure that the participants would generate the reports correctly. A report should include the code of the crop field in which a pest infestation incident is observed and name of the observer. The participants were only reimbursed for the SMS fees.

4.3 Analysis

4.3.1 Spatiotemporal analysis

A total of 318 infestation incident reports were collected from the farmer participants. Among these reports, those with incomplete or invalid information (i.e., the reports which lacked crop field codes, or lacked observer names, or lacked both; the reports with invalid crop field codes or observer names) were removed. A total of 293 infestation incident reports remained after the data cleansing, which distributed across the study area and were used as the inputs of the VGI sense making.

To investigate the spatial characteristics of the pest outbreak, spatial distribution of the reported pest infestation incidents was examined using Getis-Ord Gi* statistic (Ord and Getis, 1995). Getis-Ord Gi* was adopted because of its ability to test the statistical significance of the results (Panteras *et al.*, 2015). Following Bruce *et al.* (2014), a symmetric one/zero spatial weight matrix (i.e., the spatial weight between a given feature and each of its surrounding features is one if the distance between them is within an assigned distance band, and is zero if otherwise) was applied to generating the Gi* statistics using fixed distance band weighting. This ensured the same scale of analysis across the entire study area. A distance band of 2120 m was specified using incremental spatial autocorrelation (Global Moran's *I*) tool of ArcGIS 10.1 (ESRI Products, Redlands, CA). It reflected the most pronounced spatial autocorrelation of the dataset.

A cluster of hot or cold spots detected by the statistic can be further highlighted using a one standard deviational ellipse polygon which encloses approximately 68% of the spots in the cluster. The ellipse can be used to measure the distribution of a cluster, exhibiting its orientation or spatial structure. The cluster geometric center can be calculated to exhibit the central location of a cluster.

In addition, a change point analysis (Taylor, 2014) was conducted to explore the time series of the reported pest infestation incidents. This analysis method is capable of determining whether there is any significant change in the number of daily reports, and the time of change (if any). One thousand bootstrap samples were used in this analysis. A change point can indicate a phenological event of the pest (e.g., pest outbreak) (Herms, 2004). Once a change point was detected, the degree-days accumulated from 1 January till the detected change point was calculated. The concept of degree-day is rooted in the theory that insect development is directly related to temperature and time (Zou *et al.*, 2004). Single sine wave method with a horizontal cut-off was used to calculated the degree-days (Allen, 1976). In this study, the upper developmental temperature threshold (UDTT) and the lower developmental temperature threshold (LDTT) for the stem borer were 30 ℃ and 12.9 ℃ (Jiao *et al.*, 2006). The daily temperature data used for the degree-day calculation was obtained from the local weather station.

Except for the Getis-Ord Gi* statistics computed for the entire dataset of the incidents reports collected during the whole pest surveillance period, Getis-Ord Gi* statistics were also computed for the incident reports collected until the change points, if detected. Moreover, the standard deviational ellipse polygon and geometric center location can be generated for a cluster to observe if there is any structural or locational change of the cluster over time.

4.3.2 Information dissemination and participation incentive analysis

Three types of information were disseminated to the participants via SMS:

1. Type A: Whether a crop field was identified as a hot (cold) spot or non-hot (cold) spot after the pest surveillance.
2. Type B: The change point analysis was conducted on a daily basis. If any change point was detected to indicate significant change in number of the daily reports, occurrence date of the detected change point was disseminated.
3. Type C: Pesticide selection and how to correctly use pesticide (e.g., dosage).

A questionnaire survey was conducted to understand the effects of the information dissemination on the enthusiasm of users' participation. After the specified pest surveillance period when the participants had received the disseminated information, a questionnaire was handed out to each of the participants. The questionnaire first sought anonymous responses to three interrelated questions (or survey construct items) regarding the perceived usefulness of the overall disseminated information (Table 4.2). The questionnaire then sought anonymous responses to two additional questions regarding to what extent a participant is willing to continue participating if the pest surveillance were to continue (Table 4.2). The questionnaire employed a nine-point Likert scale (1 = strongly disagree, 9 = strongly agree).

Table 4.2 Survey constructs.

Construct	Item	Wording	Source
Perceived usefulness (PU)	PU1	The overall information dissemination is useful	Adapted from (Chee Wei *et al.*, 2006)
	PU2	The overall information dissemination is beneficial	
	PU3	The overall information dissemination is advantageous	
Intention to share (ITS) information	ITS1	If the collaborative pest surveillance were to continue, my willingness to keep sharing my pest observation information is high	Adapted from (Kim *et al.*, 2012)
	ITS2	If the collaborative pest surveillance were to continue, the likelihood that I will keep sharing my pest observation information is high	

In the participant recruitment stage, the farmers were told that they would receive personalized feedback to manage the pest infestations. This was related to the protection or enhancement of personal investments (Coleman *et al.*, 2009). It was therefore hypothesized that:

H1: *Participants' perceived usefulness of the disseminated information has a positive effect on their intention to continue sharing their pest observation information.*

H1 was tested using partial least squares (PLS) regression technique with Smart PLS 2.0.M3 (http://www.smartpls.de/). PLS regression is a component-based structural equation modelling. This method was selected because it analyzes measurement and structural models with multi-item constructs (which is the case of the questionnaire survey of this study) and works without requirement on large sample size and distributional assumptions (Haenlein and Kaplan, 2004). Before testing H1, convergent and discriminant validity of the survey questions were assessed. Convergent validity was assessed based on standardized path loadings of all survey questions, composite reliability (CR), Cronbach's alpha, and the average variance extracted (AVE) of constructs (Kim *et al.*, 2012). Discriminant validity is supported when the square root of AVE of each construct is higher than its correlations with any other construct (Kim *et al.*, 2012). Four additional factors (i.e., age, gender, education, and farming experience) were considered in the PLS regression as control variables related to the dependent variable.

Moreover, the questionnaire sought a user rating regarding the usefulness of each of the three types of information dissemination, using three ordinal-scale scores (1 = least useful, 2 = medium useful, 3 = most useful). Based on the user ratings, the overall usefulness of each of the three types of information dissemination was ranked. ANOVA (Larson, 2008) with post hoc Tukey's HSD analysis (Tukey, 1953) was performed to test if the mean scores obtained by the three different types of information dissemination differed from each other significantly.

4.4 Results

4.4.1 Spatiotemporal analysis results

The analysis revealed spatial patterns of the infestations of the pest outbreak that have never been discovered by the existing local expert-led pest surveillances. The results showed that the incident reports exhibited local clusters (Figure 4.2). A large hot spot cluster was detected in the east part of the study area, which exhibited a structure that showed slightly elongation in a roughly east–west direction. A large proportion of the hot spots were found having high statistical significance (z-score $\geqslant$2.58). The hot spots were far from the local agricultural department and closer to the main woodlands. A cold spot cluster was detected in the northwest part of the study area, which had a structure that extended roughly in a southwest–northeast direction. A large proportion of the cold spots were with 95% confidence level. No cold spot was observed at 99% confidence level. The cold spots were closer to the local agricultural department and the main residential areas.

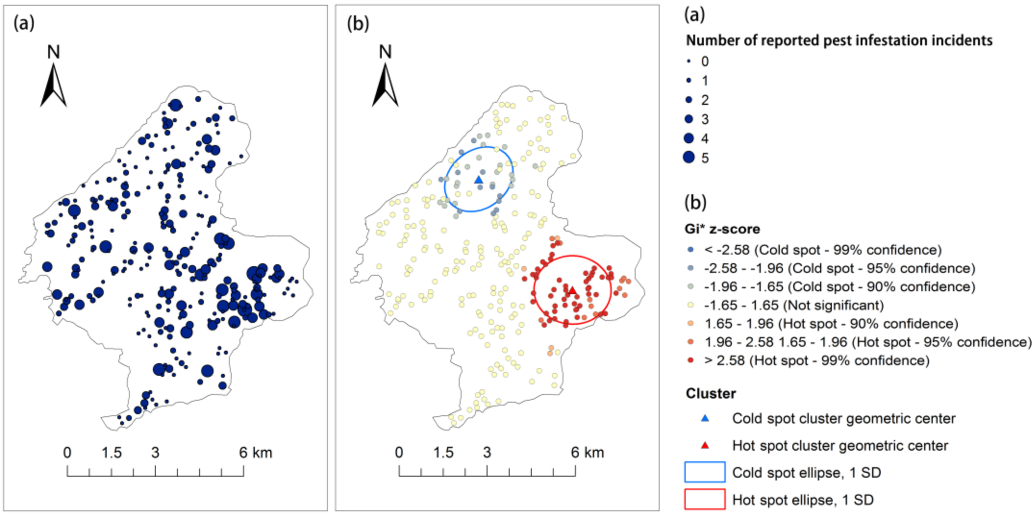


Figure 4.2 (a) Pest infestation incidents reported by the participants. (b) The corresponding Getis-Ord Gi* statistics (z-scores) and the hot and cold spot clusters highlighted using standard deviational ellipses, with size capped at one standard deviation (1 SD).

The change point analysis also revealed temporal patterns of the pest infestations that have never been recorded by the existing local expert-led pest surveillances. Three statistical significant (*P* < 0.01) change points were detected (Figure 4.3, Table 4.3). The first change point occurred on 1 May (350.85 DD). Prior to the first change point the average of the daily report amounts was 1.4 while after that it was 7.6. The second change point occurred on 6 May (397.35 DD). The average of the daily report amounts increased from 7.6 to 13.9 around the second change point. The third change point was observed on 20 May (551.61 DD) with the average of the daily report amounts decreased from 13.9 to 4.6. According to the local pest management experts and the past local experiences about the pest outbreak, the three change points may indicate three phenological events, i.e., the onset of the pest infestations outbreak, beginning of peak period of the outbreak, and end of the outbreak period, respectively.

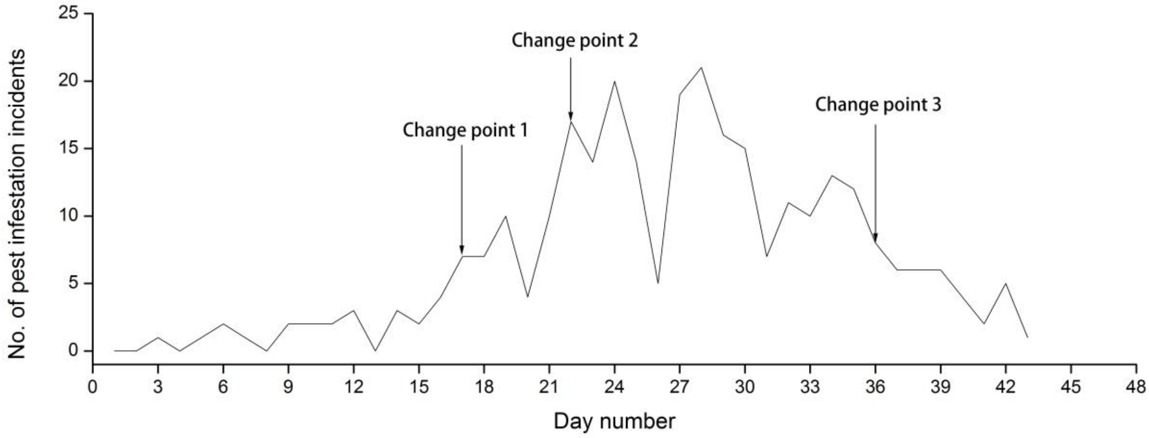


Figure 4.3 Time series of the farmer reported pest infestation incidents, and the three detected change points.

Table 4.3. A summary about the detected change points.

Change point	Date	Accumulated degree-days (DD)	Phenological event
1	1/5/2015	350.85	Onset of the pest infestations outbreak
2	6/5/2015	397.35	Beginning of the peak period of the outbreak
3	20/5/2015	551.61	End of the outbreak period

The analysis further revealed notable spatiotemporal patterns of the process of the pest infestations that are lacking from the existing local expert-led pest surveillances (Figure 4.4). The Getis-Ord Gi* statistics were computed for two change points, i.e., 1 May 2015 and 6 May 2015. Hot spot clusters and cold spot clusters were detected for both the change points. It was observed that the hot spot cluster was expanding eastward and the cold spot cluster was expanding southwestward, in terms of both their directional structures and locations of geometric centers.

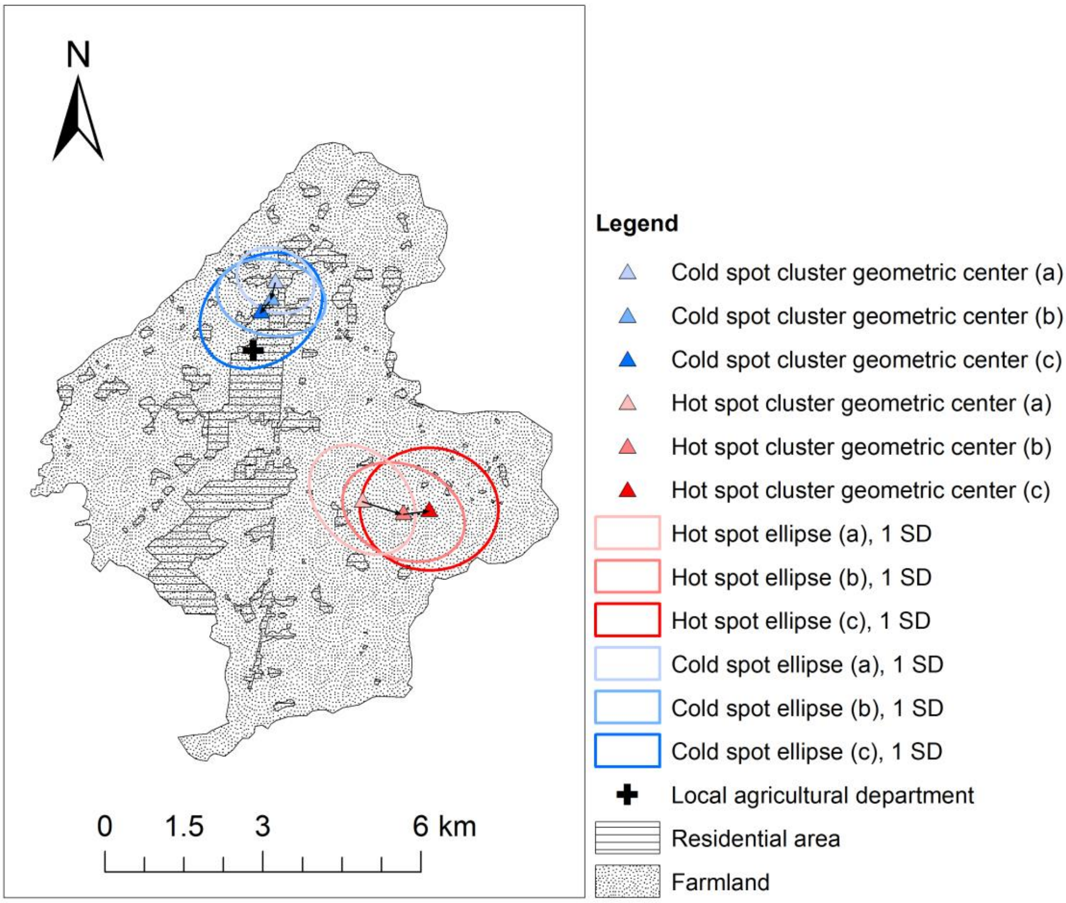


Figure 4.4 Structural and locational changes of the hot spot cluster and cold spot cluster over time. "a" represents 1 May 2015; "b" represents 6 May 2015; "c" represents 27 May 2015 (the last day of the pest surveillance).

4.4.2 Outcomes related to the information dissemination

The standardized path loadings of all the survey questions were statistically significant ($P < 0.01$) and were higher than 0.7; the CR and Cronbach's alpha for both of the survey constructs exceeded 0.7; the AVE for both of the survey constructs were greater than 0.5 (Table 4.4). Each test result therefore met its threshold criterion, and the convergent validity was supported (Kim *et al.*, 2012). Discriminant validity was also supported (Table 4.5).

Table 4.4 Results of convergent validity tests.

Construct	**Item**	**Standard path loading**	**AVE**	**CR**	**Cronbach's alpha**
Intention to share (ITS)	ITS1	0.96**	0.93	0.97	0.93
	ITS2	0.97**			
Perceived usefulness (PU)	PU1	0.97**	0.93	0.98	0.96
	PU2	0.96**			
	PU3	0.97**			

Note. ** represents a statistical significance at 99% confidence level, i.e., $P < 0.01$. AVE is average variance extracted. CR is composite reliability.

Table 4.5 Descriptive statistics and correlations.

Construct	**Items**	**Mean**	**Standard deviation**	**Intention to share (ITS)**	**Perceived usefulness (PU)**
Intention to share (ITS)	2	7.76	1.06	**0.96**	
Perceived usefulness (PU)	3	7.32	1.37	0.351	**0.96**

Note. Leading diagonal shows the squared root of AVE of each construct (i.e., ITS and PU). The off-leading diagonal value (0.351) is the correlation between the two constructs.

H1 was supported (Figure 4.5). None of the control variables was found to be significant ($P > 0.1$, Figure 4.5), meaning that H1 was robust across variation in the control variables.

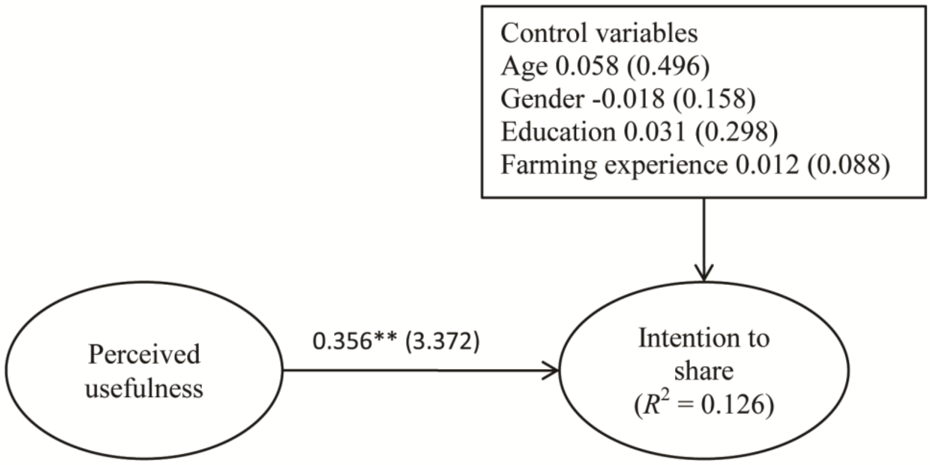


Figure 4.5 Testing results of H1 and effects of the control variables. *Note*. ** represents a statistical significance at 99% confidence level. The numbers outside the brackets are the path coefficients, and the numbers inside the brackets are the corresponding *t*-values indicating significant levels.

In addition, the user rating showed that type B information dissemination occupied the largest portion (63.9%) of score-3, type A information dissemination occupied the largest portion (57.5%) of score-2, and type C information dissemination occupied the largest portion (67.4%) of score-1 (Figure 4.6). Therefore, in general, type B was ranked as the most useful, type A was the medium useful, and type C was the least useful. This was supported by the significant different mean rating scores obtained by the three types of information dissemination according to the ANOVA (F = 135. 80, P < 0.01) and post hoc Tukey's HSD analysis (P < 0.01), of which type B had the highest mean score at 2.57 and type C had the lowest one at 1.52 (Table 4.6).

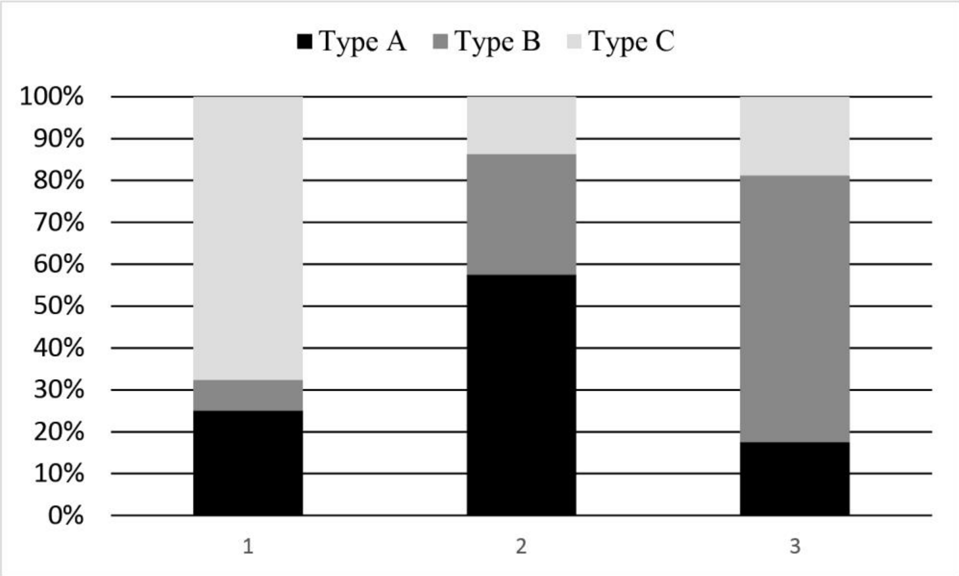


Figure 4.6 User-rated usefulness of the three types of information dissemination. Horizontal axis represents the three scores provided for the rating purpose. Vertical axis represents to what percentage each type of the information disseminations occupies a given score.

Table 4.6. Means and standard deviations of the scores obtained by the three types of information dissemination.

Type	Mean	Standard deviation
A	1.93	0.65
B	2.57	0.63
C	1.52	0.79

4.5 Discussion

4.5.1 The outcomes and implications

4.5.1.1 Spatiotemporal patterns of the pest outbreak

This chapter is first concerned with whether VGI sense making can indeed generate meaningful outcomes for enhancing tactical IPM. Therefore, using the VGI collected in the case study, spatiotemporal sense making about the pest infestations was conducted using interdisciplinary methods (GIS spatiotemporal data mining and phenology methods). The analysis revealed the occurrences of both hot and cold spot clusters of pest infestations and their directional structures. Specifically, the results showed that the detected hot spot cluster was located in an area with a high woodland coverage, a low coverage of residential area, and the area was relatively far away from the local agricultural department. On the contrary, the detected cold spot cluster was located in an area with a low woodland coverage, but a high coverage of residential area, and the area was relatively nearer to the local agricultural department. The analyses also found that the development of the hot spot had an eastward direction towards high woodland density areas, while the development of the cold spot had a southwestward direction towards the local agricultural department and the main residential areas.

The abovementioned observations suggested that the woodland coverage, coverage of residential area, and distance to the local agricultural department may be related to the occurrences of the hot and cold spots, which has indeed been statistically supported by an additional analysis based on ordinary least squares (OLS) regression (Appendix 1). Further investigations are still needed to confirm what social or physical factors or mediators (e.g., spatial differences in farmers' tillage methods, microclimate effects) had led to such relations. Several possible preliminary explanations are:

(1) an area with a high woodland coverage may provide a cozy microclimate and diverse overwintering sites for the stem borers, where the pest's overwintering larvae was difficult to manage;

(2) conducting straw burning after autumn harvest to kill overwintering larvae for preventing spring outbreak is easier in an area with a low woodland coverage;

(3) closeness to residential area and the agricultural department implies higher accessibility to necessary pest management resources and information from the markets or government, and thus the pest can be better monitored and managed.

Despite that the driven factors require further investigations, these discovered pest distributional patterns could still be beneficial to both the local farmers and pest management authorities. They would be able to know which areas should be tactically prioritized, in order for pest infestation reductions. In fact, spatial patterns discovered from the overwintering pest generation was important to the management of subsequent generations of the pest. For example, an area found with serious infestations caused by the overwintering generation was also likely to have serious infestations from the subsequent generations.

In addition, the change point analysis generated real-time information related to the pest infestations outbreak. Such information was also beneficial to the stakeholders' decision-making, which provided timely alerts for controlling the infestations. The temporal analysis also offered a detection method to match past experiences and knowledge of experts (i.e., phenological event matching). The degree-day computations associated with the phenological events also provided an important basis for future predictions about the pest infestations. Degree-day modelling has a strong physiological basis. It reflects the mechanisms of pest growths, which are therefore robust at extrapolating to future years (Herms, 2004).

Generalized from this case study, the above sense making results implicate that a proper use of VGI can reveal spatiotemporal patterns of pest infestations that are lacking from traditional pest surveillance means. Even with a limited amount of data from contributors (e.g., 293 reports in this case study), meaningful sense making results can be generated, with the premise that the contributors are properly facilitated, recruited, and trained. With the information discovered from VGI, stakeholders can target the allocation of limited pest management resources (e.g., pesticide) to narrowly defined geographic areas and directions (e.g., those areas with a high woodland coverage in the case study, the direction where a hot spot cluster extends), and temporal points (e.g., those change points with significance increases in number of the daily infestation reports). Therefore, this case study showcased two merits of VGI in the enhancement of tactical IPM, the creation of previously unrecorded spatial data for the discovery of previously unknown knowledge, and the creation of critical pest management information in the timeliest manner.

4.5.1.2 Participation incentive

This chapter also sought to investigate how to motivate and sustain volunteer participation, which was concerned with the sustainability or long-term viability of VGI-based IPM. The results showed that the more the participants found the disseminated information of use to them, the more they were willing to participate further. Therefore, usefulness of information dissemination was identified as an important factor contributing to a continued and more engaged user participation, which echoes the proposed VGI-based IPM in Chapter 3. Indeed, as mentioned in Section 3.4, it has been shown that users' need satisfaction with the information of a virtual community positively affects their intentions to continue their participation in the virtual community, which is influenced by the usefulness of the information (Chen, 2007; Jin *et al.*, 2009).

In this case study, three types of information were disseminated. It was found that type B was perceived by the farmer participants to be the most useful one. Type B information provided the participants with timely alerts for any significant increase in the number of daily pest infestation incident reports, meaning that farmers tend to concern temporal patterns of pest outbreaks most (i.e., specific timings of pest risk management). The farmers could combine such information with their own farming experiences to determine the best timing for pesticide spraying throughout the pest outbreak period. Therefore, in this case, enhancing type B information dissemination can be an incentive strategy to ensure farmer participations. In VGI-based IPM, such an incentive strategy is better than the one through gamification proposed by Martella *et al.* (2015) (see Section 2.3.4 for the review), because crop fields in rural areas may not be ideal sites for gaming (e.g., too hot during summer). In fact, farmers concern more about their cropping productivity, and hence providing information more useful for them to maximize their profits tends to be more encouraging.

It is important to note that IPM information dissemination is not limited to the three types taken in this case study. For any other VGI-based IPM projects with different contexts, in order to retain or encourage continued user participation, probably the first step is to interact actively with participants to learn what kinds of information they need most. Also, as mentioned in Section 3.4, farmers from different regions may have different needs; and an active farmer participant who is satisfied by existing information disseminated may still discontinue his/her participation someday due to his/her varying needs (disenchantment discontinuance) or due to an availability of better alternatives (replacement discontinuance) (Peshin *et al.*, 2009b). Hence the usefulness enhancement mentioned above should also be a comprehensive and continuance process, meaning that researchers have to keep adjusting a project's information dissemination to always meet its users' needs which may vary over space and time.

4.5.2 Potential improvement

While this study explored VGI for tactical IPM, the participants' role was limited to the provision of pest surveillance data. Haklay (2013) introduced four ascending levels (ladders) of participation and engagement in citizen science projects. This case study was based on the first level, i.e., crowdsourcing, in which cognitive engagement in IPM was minimal. A comprehensive VGI-based IPM should ascend to higher levels to fully utilize participants' cognitive engagement in IPM, i.e., farmers' knowledge, attitudes, and practices (Tait and Napompeth, 1987). Researchers can facilitate participants to also play the roles of problem definers, data analyzers, and knowledge interpreters. At the highest level, participants could actively involve themselves in the whole process of the project, and are encouraged to be inquisitive and innovative. Therefore, exploring high level VGI-based IPM approaches is suggested for future research.

Another point to be stressed is the quality of VGI. The quality of VGI in this case study was improved by using a targeted recruitment strategy suited to facilitated VGI creation. However, VGI-based IPM may not be limited to facilitated VGI creation in which the VGI collection process is relatively more manageable. In fact, diverse forms of VGI-based IPM community can be established based on different types of VGI creation (e.g., active VGI creation, see Section 2.1.2 about the topology of VGI creation forms) and with more heterogeneous participants. More robust VGI quality assurance approaches are therefore needed to satisfy diverse scenarios of VGI-based IPM, and Chapter 6 will present such an approach.

While this chapter mainly focuses on tactical IPM and participation incentive, in the next chapter, a second case study focusing on strategic IPM will be presented.

5 Exploring the potential distributional changes of invasive crop pest species associated with global climatic change: a VGI-based strategic IPM

As mentioned in Section 2.2.2, strategic IPM deals with relatively long-term and planning-oriented issues which are also with broader geographic extents. This chapter presents a case study pertaining to strategic IPM, which explored potential future pest problems rather than pest problems in the near-term. Specifically, the potential distributional changes of invasive crop pest species associated with global climatic change were investigated using ecological niche model (ENM). It aims to explore whether VGI sense making can indeed generate meaningful outcomes to enhance strategic IPM. Meanwhile, it contributes to the understanding of the practical and potential roles of volunteer participants in strategic VGI-based IPM.

VGI from citizens were adopted as an important source of data in this case study. Compared with the facilitated VGI used in Chapter 4, the VGI (including both active VGI and facilitated VGI) used in this chapter are with broader geographic extents, fitting the concept of strategic IPM in general and the need of large spatial-coverage studies using ENM in specific. The VGI can be valuable complements to traditional spatial data. Without the VGI, the case study is not possible to be completed due to the lack of species occurrence records from traditional data collection means.

The following sections of this chapter will first present the background, sense making methods and results of this case study (Section 5.1 to Section 5.3). Section 5.4 will discuss the sense making results and the roles of volunteer participants in strategic IPM.

5.1 Background of the case study

The world's human population is estimated to increase by 70 million per annum (Popp *et al.*, 2012). As the human population grows, so will the problem of the global food supply. Sustainable agricultural productions are therefore critical. However, agriculture in its diverse forms and large extent across the globe are highly sensitive to climate change (Howden *et al.*, 2007). The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) has suggested that global mean surface temperature is expected to increase maximally 4.8℃ by the end of the 21th century (IPCC, 2013). Precipitation pattern will change, and extreme weather will become more frequent. Climate change can influence crop yields through effects mediated by changes in crop pest distributions, especially those invasive pest species (Estay *et al.*, 2009; Ziska *et al.*, 2010; Bebber *et al.*, 2013). Managing invasive species for securing cropping activities is challenging because such species can survive in diverse environments, mature quickly, and compete with local species for resources to affect ecosystem functioning

(Bradley, 2009; Hoddle, 2014). The dispersal of invasive crop pest species is also likely to be incremental due to global warming (Thomson *et al.*, 2010; Ziska *et al.*, 2010).

Among the invasive crop pests, poikilothermic (cold-blooded) species are recognized for the large area of agricultural losses that they cause. Temperature has been considered as the most important abiotic limiting factor that governs their distributions (Pimentel *et al.*, 2000; Ziska *et al.*, 2010; Fand *et al.*, 2014). Poikilothermic pest species can thrive in warmer environments, so climate warming can potentially enhance their overwintering survivability (Maxmen, 2013) and facilitate the accumulation of degree-days required by their growths (Herms, 2004). However, when temperature exceeds the upper threshold of a species' tolerance, it can result in a decreased growth and reproduction, thereby increasing the mortality for the species (Netherer and Schopf, 2010). Precipitation can also influence the species distributions as it alters ambient humidity, affecting the moisture needed for pest growths. Furthermore, precipitation extremes can negatively impact on pest growths, such that the resulting floods can wash off eggs and larvae and drown young pests (Kobori and Amano, 2003).

Climate change thus have both positive and negative effects on the pests. Numerous studies have examined the possible climate change effects on the distributions of various species at regional levels, such as plants (Garcia *et al.*, 2013), mammals (Alamgir *et al.*, 2015), birds (Tingley *et al.*, 2009), reptiles (Araújo *et al.*, 2006), amphibians (Thomas *et al.*, 2004), and insects (Hongoh *et al.*, 2012). Prior work has suggested that many species are likely to move poleward in latitude or upward in elevation as the climate warms (Jepsen *et al.*, 2008; Estay *et al.*, 2009; Tingley *et al.*, 2009). Regarding crop pests, recent effort has described the patterns and trends in their global spread with a temporal focus on the second half of the 20th century (Bebber *et al.*, 2014). Nevertheless, little is known if the overall global distributions of poikilothermic species of invasive crop pests will expand or contract as a result of future climate change, and to what extent will climate change influence the species richness across different regions. Research is thus needed to analyze the possible consequences of future climate change on the global distributions of the invasive crop pest species.

To estimate potential range and direction of distributional changes for invasive species, ENM has been suggested as a useful tool (Jeschke and Strayer, 2008). ENMs fall into two classes: mechanistic (or process-based) approach assesses physiological aspects of species; and correlative approach considers correlations between observed species distributions and environmental variables (Morin and Thuiller, 2009). Although both approaches have demonstrated their predictive capabilities (Thuiller *et al.*, 2005; Kriticos *et al.*, 2013), correlative ENMs are frequently used in analyzing multiple species for at least three reasons. First, the lack of relevant knowledge on the physiological tolerance for all species of concern compromises the precision of mechanistic models (Wiens *et al.*, 2009; Mokany and Ferrier, 2011). Second, species presence data for correlative ENMs are widely available, while data based on which mechanistic ENMs can be parameterized are often lacking (Morin and Thuiller, 2009). Third, the complexity of mechanistic models is high, requiring a relatively long time to fit a mechanistic model with appropriate data (Elith, 2015).

A correlative modelling program commonly applied to construct ENMs is maximum entropy modelling, or Maxent (Phillips *et al.*, 2006). It is a presence-only machine learning algorithm that estimates the probability distribution of species based on occurrence records and randomly generated background points by finding the maximum entropy. Prior work has demonstrated its better performance than other modelling methods in predicting invasive species distributions outside their native ranges (Bidinger *et al.*, 2012). This study thus uses species presence data to construct the correlative ENMs in Maxent in order to understand the impacts of climate change on the global distributions of poikilothermic invasive crop pest species. Specifically, this case study addresses three research questions. First, will the overall global distribution of the pest species expand or contract as a result of climate change? Second, what are the spatial patterns of distributional changes in pest species richness? Third, how will temperature and precipitation variations across different regions affect the distributional changes of the pest species? Insights into the possible direction and range of the distributional changes of invasive crop pests are essential for adapting the current agricultural systems to climate change, so as to prevent world food insecurity and long-term nutritional emergencies.

5.2 Materials and methods

5.2.1 Study area and species records

This study focused on the current global cropland extent to examine the responses of the invasive crop pest species to climate change within the boundary of existing global cropping activities. The spatial extent of the global cropland (Figure 5.1) was obtained from Pittman *et al.* (2010) as the boundary of the study area. It is uncertain if the extent will increase or decrease for 2050 and 2070 (IPCC, 2013). This study hence focused on calibrating the ecological niche models and examining the responses of the pest species to climate change within this cropland extent so as to provide a reference for how decisions and planning can be made for adapting the existing agriculture systems within the current global cropland extent to climate change with regard to invasive crop pest risks. Future species distributional changes outside this extent was not taken into account in the analysis.

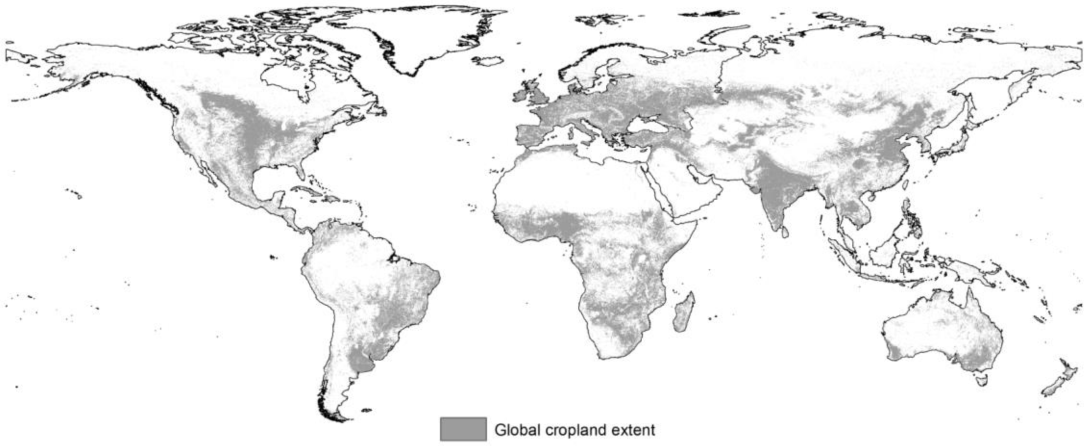


Figure 5.1 Study area. White color areas denote no cropland.

Analyzing the complete cropping ecological system is almost impossible because of the large number of species. For example, just the arthropods in alfalfa alone consist of approximately 1500 species, although most of them do not cause economic loss to agriculture and can thus be ignored (Gutierrez *et al.*, 2007). For the purpose of this study, invasive crop pest species were selected based on their economic importance or quarantine significance to major food and cash crops (CABI, 2016). This resulted in a total of 76 species covering all climatic zones across the globe (Appendix 2).

The occurrence records of the 76 species were extracted from multiple data sources. These included field surveys, bibliographic information and a range of pest species repositories that involve large amounts of crowd sourced (i.e., VGI) and expert-generated species occurrence records as follows: the Invasive Species Compendium and the Distribution Maps of Plant Pests of Centre for Agricultural Bioscience International (CABI) (http://www.cabi.org/), the Global Database of European and

Mediterranean Plant Protection Organization (EPPO) (https://gd.eppo.int/), Global Biodiversity Information Facility (GBIF) (http://www.gbif.org/), Lifemapper (http://lifemapper.org/), Invasive Species Specialist Group (ISSG) (http://www.issg.org/), and Village Tree (http://cosmic.nus.edu.sg/). Records in CABI and EPPO were compiled and validated by experts, and widely used by plant health organizations (Pasiecznik *et al.*, 2005). GBIF, Lifemapper, ISSG and Village Tree were also valuable databases for invasive species studies (Gallien *et al.*, 2012; Bellard *et al.*, 2013; Suen *et al.*, 2014), garnering already validated pest occurrence records sourced from citizens (i.e., VGI, including both active VGI and facilitated VGI, see Section 2.1.2 about the different forms of VGI creation), researchers and various institutions. The occurrence records of the 76 species extracted from these multiple data sources were cross-checked to remove redundancies. A total of 101,662 occurrence records were finally used to construct the ENMs, of which over 90% were VGI.

5.2.2 Future climatic projections and environmental variables

For future climatic projections, a conservative and a less conservative greenhouse gas emission scenario, i.e., RCP2.6 (BCC-CSM1-1) and RCP4.5 (BCC-CSM1-1), were used. The increases of global mean surface temperatures in these two scenarios will be rapid during the first half of the 21th century, but will slow down in the second half which estimate that the likely global mean surface temperature increases will be maximally 1.6℃ and 2.0℃ till 2065, respectively; and will be 1.7℃ and 2.6℃ till 2100, respectively (IPCC, 2013). This tends to be more realistic since various mitigation strategies in response to global warming are expected to be taken (Houghton, 2005). Indeed, the 2015 United Nations Climate Change Conference negotiated the Paris Agreement (COP21, 2015), a global agreement on the reduction of climate change, which sets out a global action plan to avoid dangerous climate change by limiting global warming to well below 2℃. Because the trend of temperature increase will slow down after 2050 in these two greenhouse gas scenarios, the species distribution models were projected to the two scenarios for two future time periods, 2050 and 2070. Consequently, a total of four future climatic projections (i.e., 2050-RCP2.6, 2050-RCP4.5, 2070-RCP2.6 and 2070-RCP4.5) were analyzed.

To construct the ENMs, 19 bioclimatic variables (for 2050 and 2070) and three topographical variables (assumed to remain the same) were first derived from WorldClim database (http://www.worldclim.org/) as potential predictors (Table 5.1). All of the environmental variables had a spatial resolution of 2.5 arc-minutes, resulting in 2.1×10^6 pixels for the study area. Next, for each species, a pair-wise Pearson correlation matrix was calculated from all these variables based on the species occurrence locations. Then, following Bellard *et al.* (2013), only variables that were not collinear ($r < 0.75$) were kept in the subsequent analyses.

Table 5.1 Explanatory variables derived from WorldClim (http://www.worldclim.org/) for potential use in the ecological niche modelling.

Variable	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly [maximum temperature – minimum temperature])
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation×100)
BIO5	Maximum Temperature of Warmest Month
BIO6	Minimum Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter
	Elevation
	Slope
	Aspect

5.2.3 Ecological niche modelling with Maxent

Maxent version 3.3.3k (Phillips et al., 2006) was used to construct the ENMs. For each species, 75% of its occurrence records were randomly selected as the training dataset and the remaining 25% were set aside for validation (Moffett *et al.*, 2007). The logistic output format was enabled in the modelling, which produced a predicted probability of species presence (between 0 and 1) on each pixel, indicating the degree to which the environmental conditions of the pixel were suitable for the species.

For model evaluation, two tests were performed. First, the area under the curve (AUC) was used, the curve here refers to receiver operating characteristic (ROC) curve. The ROC curve is a plot of the sensitivity against 1-specificity (commission error) at all possible threshold probabilities for a positive prediction that ranges from 0 to 1. A model with an AUC above 0.8 was considered having "good" discrimination abilities (Swets, 1988). Second, a binomial probability test of omission was conducted to determine the statistical significance of the prediction of each model (Moffett *et al.*, 2007).

5.2.4 Examining impacts of climate change on species distributions

5.2.4.1 Overall change in mean probability of presence

With the 76 species distributions modelled under the current climate and the four future climatic projections, three analyses were conducted in ArcGIS 10.0 (ESRI Products, Redlands, CA). First, future changes in the mean probability of presence across all pixels of the study area were calculated for each species. Second, frequency distributions of species showing the projected future change in the mean probability of presence were analyzed. Third, future changes in the average of the mean probabilities of presence were computed.

Additionally, these 76 species were classified into three categories based on their main host plants/plants affected (Appendix 2). Category "A" species mainly causes high economic losses to major food grains, such as cereal crops and legumes. Category "B" species mainly causes high economic losses to major cash crops, such as sugar-yielding crops, oil-bearing crops and raw materials crops. Category "C" species can cause high economic losses to both major food grains and major cash crops. To evaluate whether climate change has different impacts on the global distributions of these three categories of pest species in terms of their changes in species mean probability of presence, Nonparametric Kruskal-Wallis tests were performed.

5.2.4.2 Spatial patterns of species richness change

Following Wiens *et al.* (2009), a pixel-level species richness index was obtained by summing the pixel-level species probabilities of presence of all 76 species. The species richness index of a pixel represents the overall occurrence risk of the pest species at that pixel. The changes of species richness under the four future climatic projections were calculated and mapped to assess the direction of the potential distributional changes of the species.

As differences in distributional changes among species will change species co-occurrence, increase or decrease in species richness does not reflect the change of species composition within a pixel. In this study the ENMs were constructed to predict

b_i,

the responses of individual species, independently of one another, to climate change. Species that do not co-occur at certain pixels under the current climate may co-occur at those pixels in the future, or vice versa. Therefore, for each of the four future climatic projections, a modified Jaccard index (J_i) of community similarity (Wiens *et al.*, 2009) was first computed for each pixel to characterize the potential change in community composition. The index was calculated for each pixel i of the output raster files (Eq. 5.1):

$$J_i = \frac{c_i}{a_i + b_i - c_i}, \tag{5.1}$$

where a_i (Eq. 5.2) and b_i (Eq. 5.3) are indices of current and future species richness, respectively, over 76 species at pixel i, and c_i (Eq. 5.4) is an index of the species overlap between the current and future time periods at pixel i:

$$a_i = \sum_{j=1}^{76} a_{i,j}, \tag{5.2}$$

where $a_{i,j}$ is the predicted probability of presence for species j at pixel i in the current climate;

$$b_i = \sum_{j=1}^{76} b_{i,j}, \tag{5.3}$$

where $b_{i,j}$ is the predicted probability of presence for species j at pixel i in the future climate; and

$$c_i = \sum_{j=1}^{76} \min(a_{i,j},\ b_{i,j}). \tag{5.4}$$

Next, to highlight the patterns of community dissimilarity as opposed to similarity, (1-J_i) was computed for each pixel. The value signified species turnover, and ranged from 0 to 1. A high value represented a high community dissimilarity or high species turnover, indicating a high change in species composition. Lastly, to assess the uncertainty of the projected species turnovers, the coefficient of variation of the model projections was calculated (Wiens *et al.*, 2009). This analysis calculated the coefficient of variation of model projections for species turnovers between the two climate change scenarios of the same year (i.e., 2050-RCP2.6 vs. 2050-RCP4.5; 2070-RCP2.6 vs. 2070-RCP4.5) for each pixel, so as to visualize areas of greater uncertainties in the predictions.

5.2.4.3 Distributional change in relation to temperature and precipitation

The pixels showing the projected changes in species richness were divided into two groups, i.e., increase in species richness and decrease in species richness, to examine the species richness distributional changes in relation to temperature and precipitation variations. To do so, frequency distributions of the two groups of pixels for the four future climatic projections were plotted against the current annual mean temperature values (i.e., BIO1 variable in Table 5.1) and the current annual precipitation values (i.e., BIO12 variable in Table 5.1), respectively. This allowed the understanding of whether certain temperature and precipitation ranges might encounter higher increases or decreases in species richness.

5.3 Results

5.3.1 Model performance

All the AUC values of the models for the 76 species were greater than 0.8 with a mean of 0.903. The highest AUC value was 0.994 and the lowest was 0.817. Additionally, all the binomial tests on the models for the 76 species yielded significant results at a 99% confidence interval. These values indicated very good model performances.

5.3.2 Potential changes in species distributions

5.3.2.1 Overall change in mean probability of presence

Changes in species distributions quantified using the mean probability of presence revealed that the number of species predicted to increase (50, 42, 51, 49 species for 2050-RCP2.6, 2050-RCP4.5, 2070-RCP2.6, 2070-RCP4.5, respectively) was greater than that predicted to decrease (26, 34, 25, 27 species for 2050-RCP2.6, 2050-RCP4.5, 2070-RCP2.6, 2070-RCP4.5, respectively) under all four future climatic projections

(Figure 5.2a). Some species were predicted to increase their mean probabilities of presence as high as 0.20, while the largest decrease was at -0.06 (Figure 5.2a, Appendix 2).

Averages of the mean probabilities of presence of the 76 species were all predicted to increase, with the increases in mean and median respectively being 2.2% and 3.6% for 2050-RCP2.6, 1.7% and 3.5% for 2050-RCP4.5, 2.4% and 3.9% for 2070-RCP2.6, and 2.1% and 4.0% for 2070-RCP4.5 (Figure 5.2b). No statistically significant differences ($P > 0.05$) with regard to the predicted future changes in mean probability of presence were found for Categories A-C species classified based on their main host plants/plants affected.

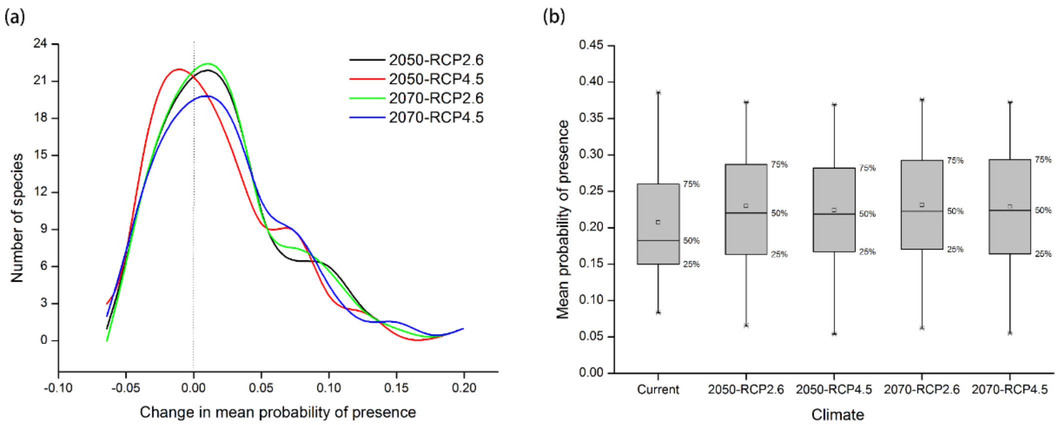


Figure 5.2 (a) Frequency distribution of species and (b) box plot showing projected changes in mean probability of species presence in the study area, based on four future climatic projections (2050-RCP2.6, 2050-RCP4.5, 2070-RCP2.6, and 2070-RCP4.5). The squares insides the boxes and the horizontal lines dividing the boxes respectively represent the means and the medians of the datasets.

5.3.2.2 Spatial patterns of species richness change

Assessment of species richness index for the current climate showed high species richness in Eastern and Southern India, Southwest China, Southeast Asia, Mexico and Eastern United States (Figure 5.3). The predicted future species richness change exhibited similar spatial patterns across the four climatic projections (Figure 5.4). Those regions currently with high species richness, along with some other tropical and subtropical regions (e.g., Central Africa, Florida and some areas in South America), were predicted to experience reductions in species richness as a result of climate change (Figure 5.4). Alternatively, low species richness under the current climate was mainly found in higher latitudes, such as Northern and Northeast Asia, Eastern Europe and the northern part of North America (Figure 5.3). Species richness for these regions were mostly predicted to increase based on future climatic projections (Figure 5.4). In addition, it was observed that many regions at relatively lower altitudes were predicted to experience decreases in species richness (e.g., North China Plain, European Plain, Indo-Gangetic Plain and the southeast coastal areas of the United States, Figure 5.4), while regions at relatively higher altitudes were predicted to have increases in species richness (e.g., Southwest China, Western United States, Northwest Ethiopia, Southeast Brazil, Pakistan and the northeast part of South Africa, Figure 5.4).

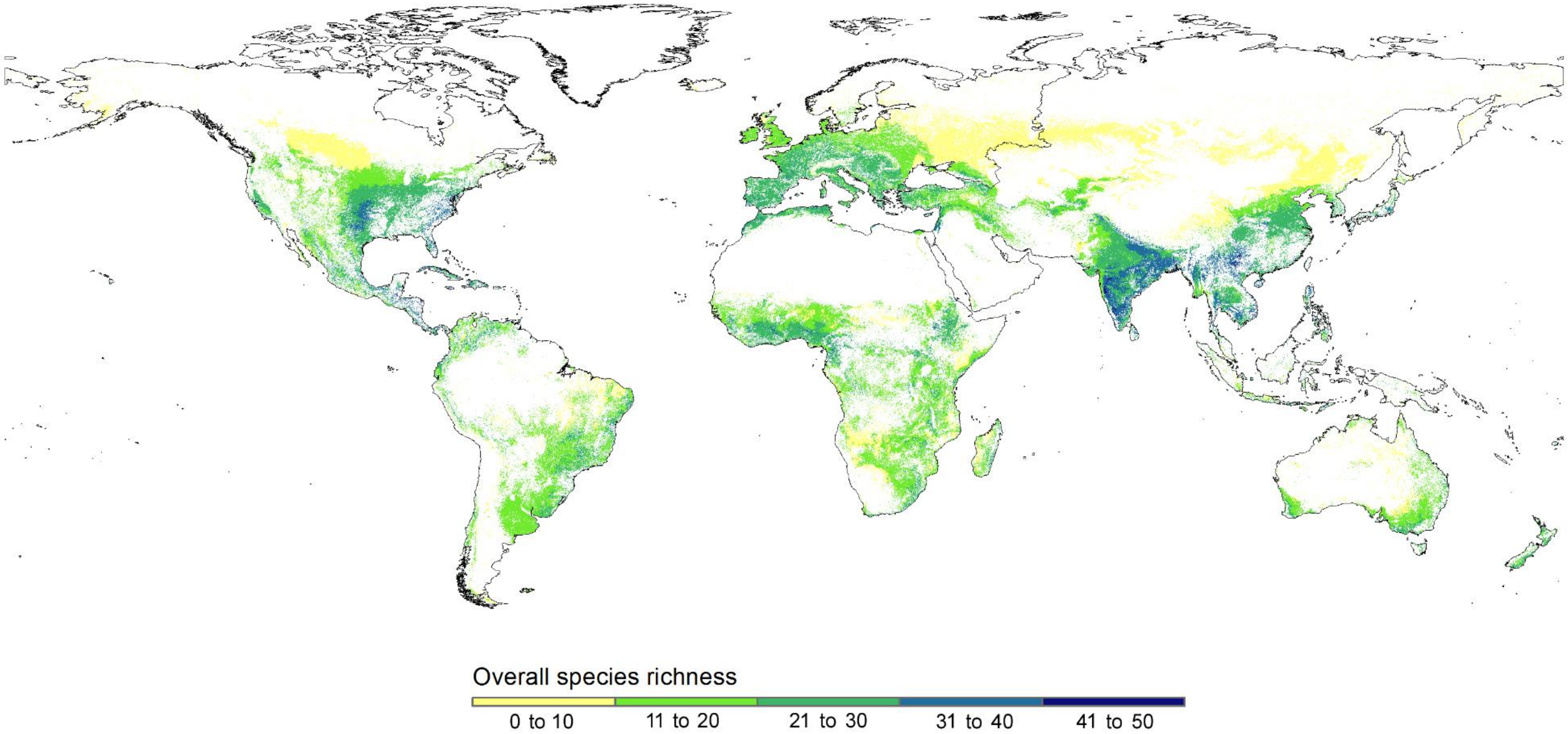


Figure 5.3 Invasive crop pest species richness under the current climate.

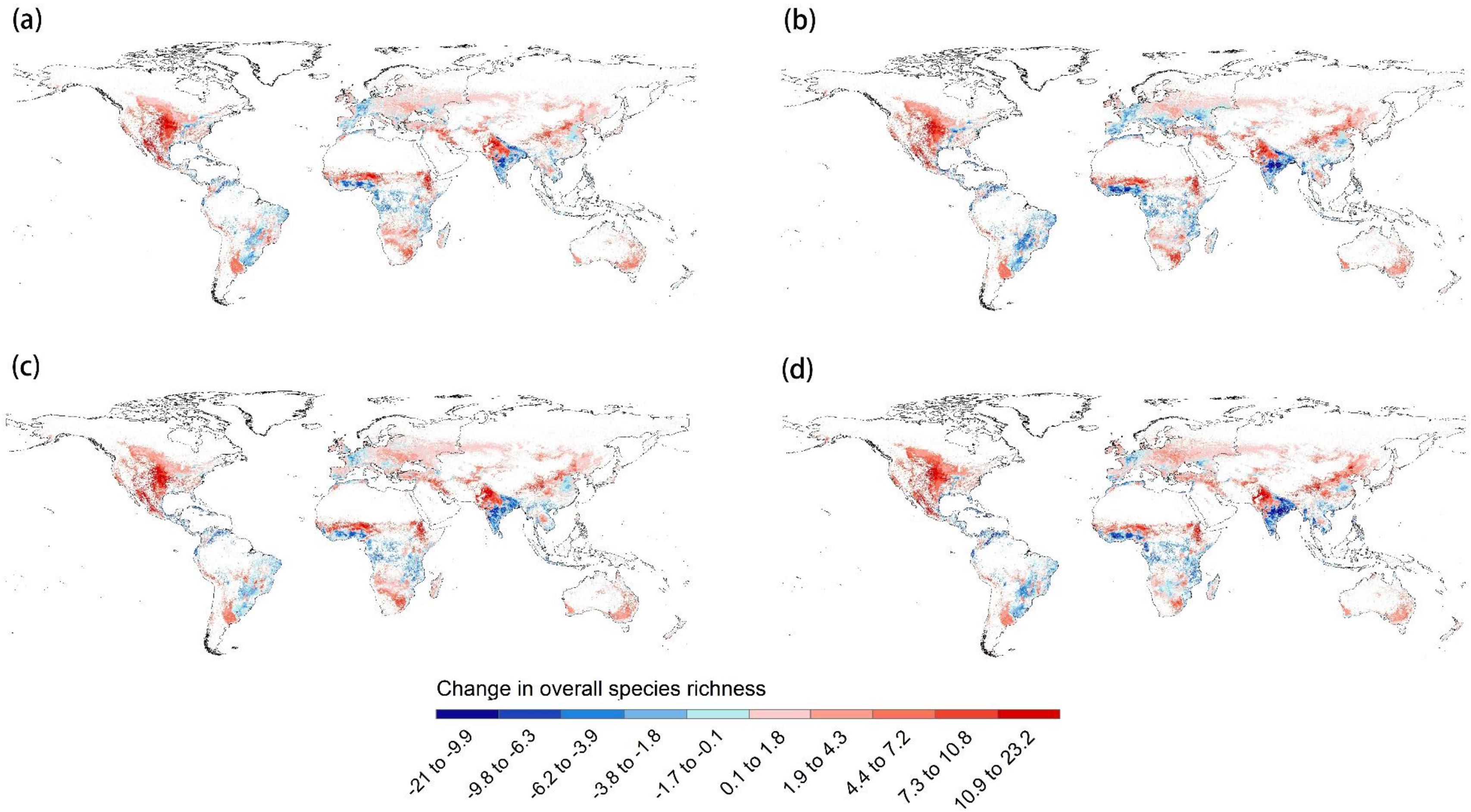


Figure 5.4 Changes in the invasive crop pest species richness predicted based on future climatic projections: (a) 2050-RCP2.6; (b) 2050-RCP4.5; (c) 2070-RCP2.6; (d) 2070-RCP4.5.

Quantification of community dissimilarity between the current and future climatic projections revealed that high species turnovers mainly occurred at the regions where their species richness were predicted to increase (e.g., Central United States, Northern and Northeast Asia and Northwest India, cf. Figure 5.4 and Figure 5.5). Conversely, low species turnovers were found in regions where their species richness were predicted to decrease (e.g., Eastern China, Central Europe, cf. Figure 5.4 and Figure 5.5). Additionally, the coefficients of variation of model projections for species turnovers in 2050 and 2070 showed that the uncertainties related to the use of different climate change scenarios were relatively low in most parts of the study area (Figure 5.6). Greater uncertainties were observed in Canada, Eastern India, Eastern Europe, and northeast China, mostly associated with higher species turnovers. Furthermore, higher values of coefficients of variation were observed for 2070 than for 2050, suggesting greater uncertainties in the predictions.

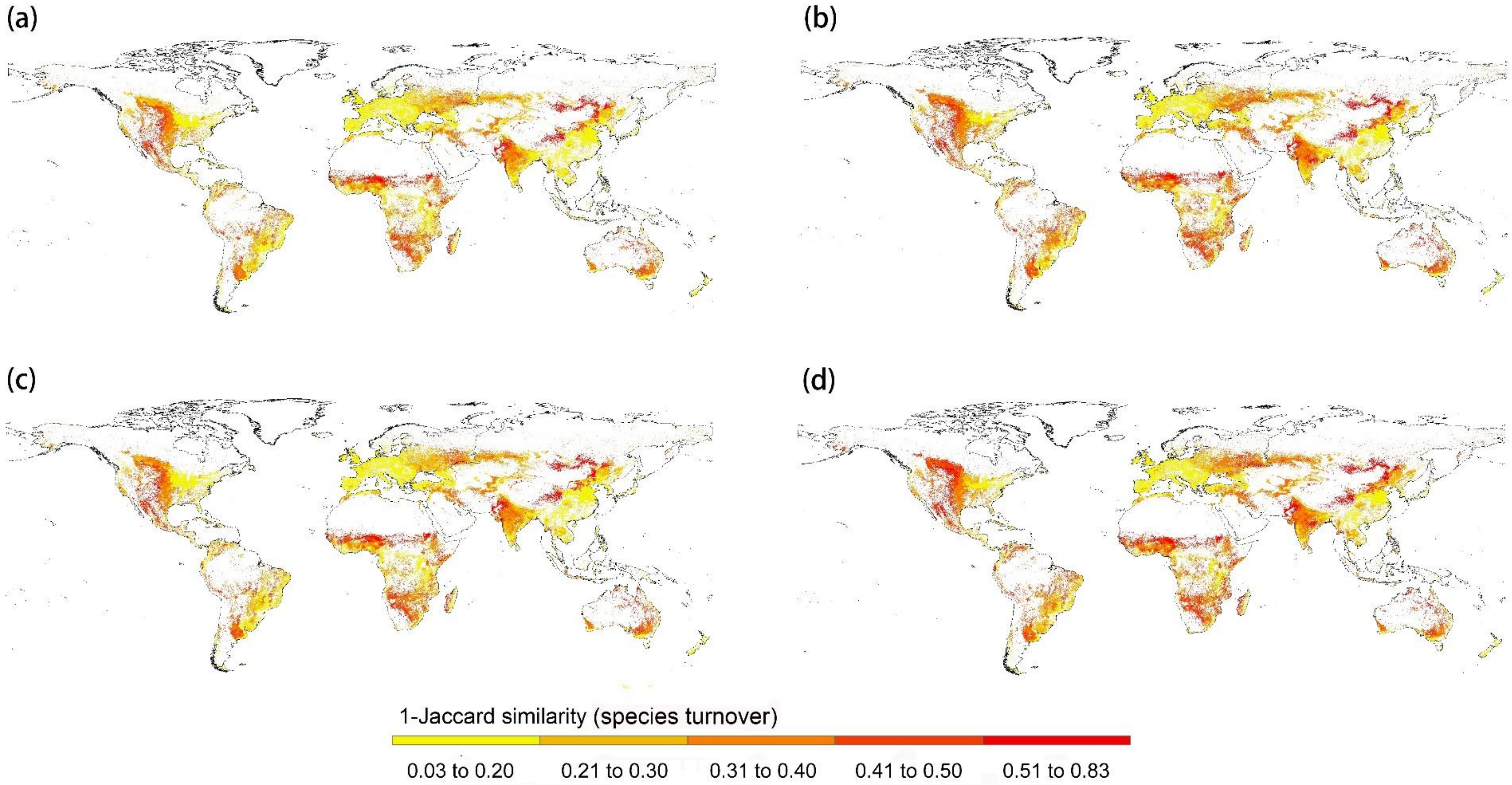


Figure 5.5 Community dissimilarity of invasive crop pest species between the current and future climatic projections: (a) 2050-RCP2.6; (b) 2050-RCP4.5; (c) 2070-RCP2.6; (d) 2070-RCP4.5. Higher "1-Jaccard similarity" (species turnover) values represent greater community dissimilarities, indicating more changes in species composition.

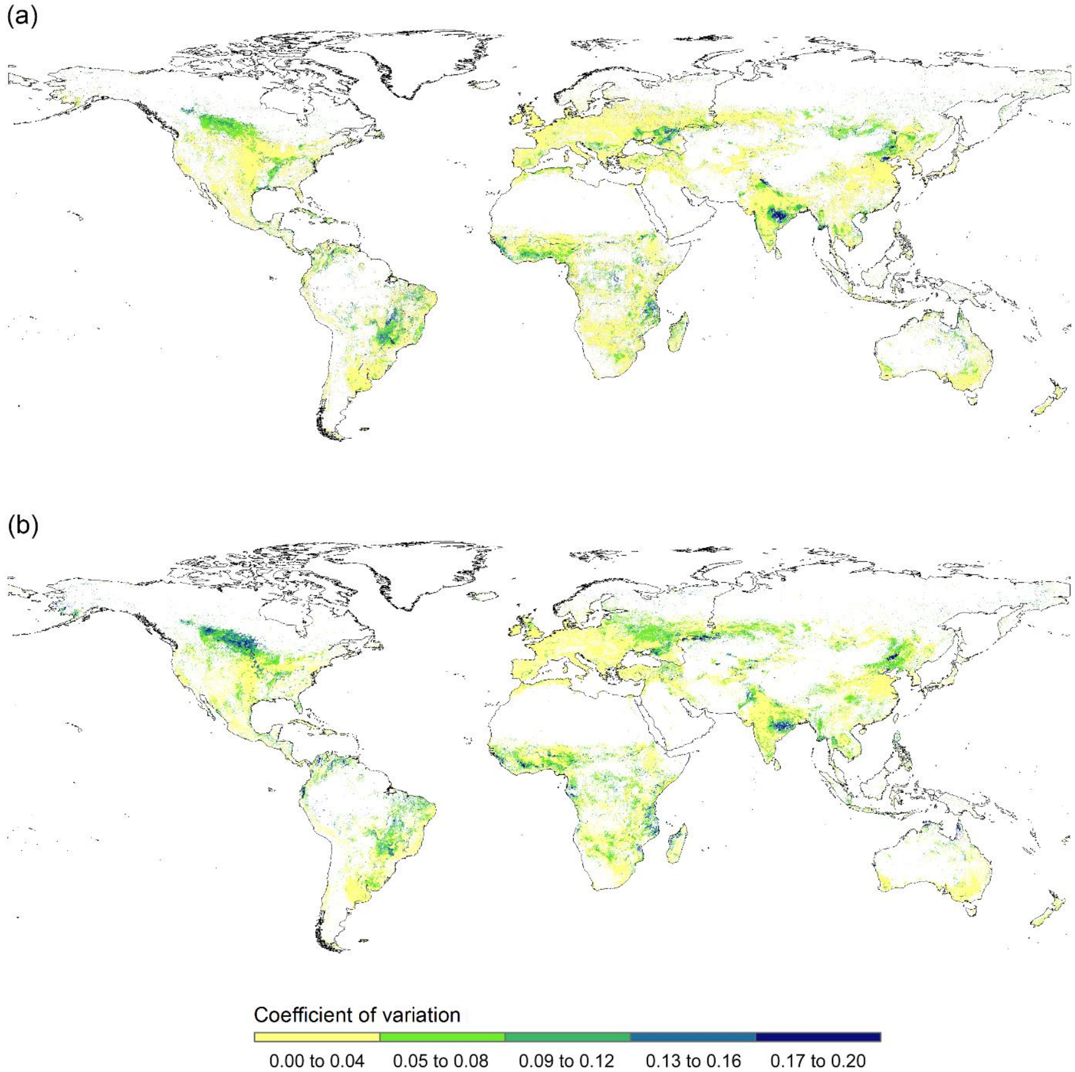


Figure 5.6 Uncertainties of model projections for future species turnovers, expressed as the pixel-level coefficients of variation, between the two climate change scenarios of the same year: (a) 2050-RCP2.6 vs. 2050-RCP4.5; (b) 2070-RCP2.6 vs. 2070-RCP4.5.

5.3.2.3 The distributional change in relation to temperature and precipitation

Investigation of the relationship between species richness change and current annual mean temperature showed that areas with lower current annual mean temperatures below approximately 21℃ were predicted to have more pixels with increases in species richness (Figure 5.7a). Pixels with increases in species richness were mainly distributed in areas with current annual mean temperatures between 0℃ and 10℃, with a peak at approximately 7.5℃. Alternatively, areas with higher current annual mean temperatures above approximately 22℃ were predicted to have more pixels with decreases in species richness. Pixels with decreases in species richness were mostly predicted to occur in areas with current annual mean temperatures between 21℃ and 28℃, with a peak at around 27℃.

Analysis of the relationship between species richness change and current annual precipitation illustrated similar trends of the four future climatic projections (Figure 5.7b). Areas with current annual precipitations below approximately 1100 mm were predicted to have more pixels with increases in species richness, with most changes predicted to occur in areas with current annual precipitation between 400 mm and 700 mm. In contrast, in areas with current annual precipitations above 1100 mm, species richness decrease was predicted to occur more frequently than species richness increases.

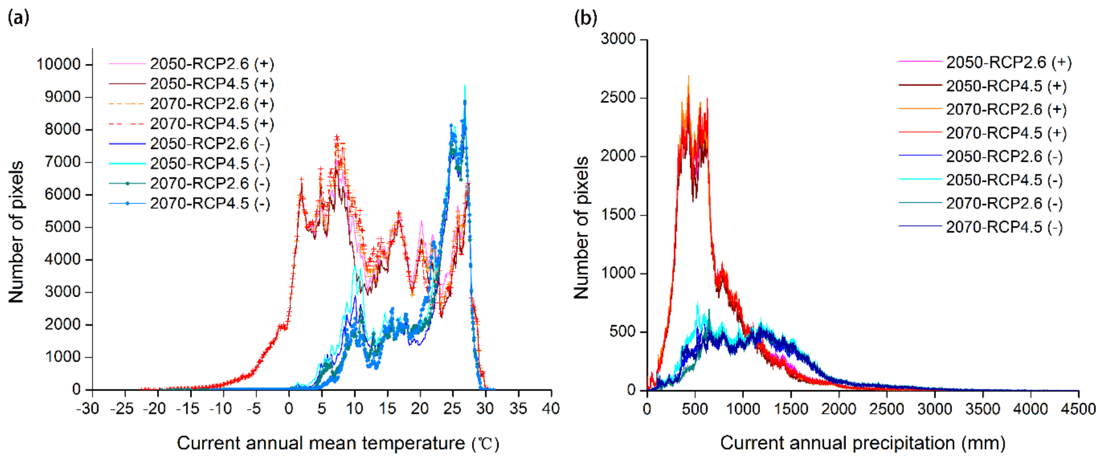


Figure 5.7 The number of pixels with increases in species richness (+) and decreases in species richness (-), predicted based on four future climatic projections (i.e., 2050-RCP2.6; 2050-RCP4.5; 2070-RCP2.6; 2070-RCP4.5), against (a) the current annual mean temperature values of the pixels and (b) the current annual precipitation values of the pixels.

5.4 Discussion

5.4.1 Potential distributional changes

The predicted outcome for increases in averages of the mean probabilities of presence of the 76 invasive crop pest species (Figure 5.2) suggests that the overall global distribution of the pest species could be expanded by climate change. These changes have no significant differences among the three categories of pest species, indicating that major food grains and cash crops are likely to be all affected. Potential pest species distributional changes associated with climatic change thus pose a considerable threat to world food security.

The projected changes in species distributions vary across the globe. Regions of relatively higher latitudes or higher altitudes are projected to experience more increases in species richness (Figure 5.4). Such finding confirms Bebber *et al.* (2014) that climate change is likely to affect the future distributions of crop pests, and echoes previous studies (Jepsen *et al.*, 2008; Estay *et al.*, 2009; Tingley *et al.*, 2009) that species are likely to spread poleward in latitude or upward in elevation as the climate warms. The findings of this study further reveals that areas with current annual mean temperatures at around 7.5℃ (Figure 5.7a) or with annual precipitations below 1100 mm (Figure 5.7b) could be the most vulnerable because of their predicted increases in invasive crop pest species richness.

In contrast, most of the regions at the tropics, lower altitudes, or with current annual precipitation above approximately 1100 mm are predicted to experience decline in species richness (Figure 5.4 and Figure 5.7b). The largest decline in species richness is expected to occur in regions with high current annual mean temperatures at around 27℃ (Figure 5.7a). The declines in species richness are probably positive news for pest species control and eradication, as the overall pest occurrence risks and economic damages to cropping activities may be lessened. However, as the projected community dissimilarity in these regions are generally low (cf. Figure 5.4 and Figure 5.5) denoting small changes in species composition, it is unlikely that climate change can result in the removal of many of the existing pest species from these regions.

5.4.2 Physiological explanations

Climate change can influence the pest physiology through a number of inter-related processes, subsequently affecting their distributions. Most of the regions projected to experience both increased species richness and high species turnovers are located at higher latitudes or higher altitudes, where the current annual mean temperatures are generally low (Figure 5.4 and Figure 5.5). Pest species in such regions are often living in environments that are presently much cooler than their physiological optima (Deutsch *et al.*, 2008). Increases in temperature within their physiological optimum range can hence have at least two advantages in enhancing pest fitness.

First, small increases in temperature can accumulate into a large number of degree-days in the physiological development times of the pest species. A pest needs to accumulate certain amount of degree-days to complete its life-cycle (Herms, 2004). As the globe warms, some regions currently do not provide sufficient degree-days within a calendar year for the development of some pest species may then be able to provide sufficient degree-days, and subsequently becoming habitable for those species. Regions currently already habitable for some pest species may provide even more degree-days for the development of additional generations of those species within a calendar year. Second, warming climate may also improve the overwintering survival of the pest species in regions currently with freezing winters. Freezing winter temperatures act as nature barriers to the dispersal of crop pest species (Maxmen, 2013), but this situation may change with the warming climate. For example, low temperature of freezing winter is the dominant lethal factor for the overwintering aphids if it drops below the aphids' cold tolerance. Increased winter temperatures improve their winter survival, make them more dangerous crop pests, and possibly lead to increased damages to agricultural activities (Cannon, 1998).

A pest species' adaptability to the climate regime of a region can be reduced if temperatures exceed its physiological optimum range (Netherer and Schopf, 2010). This may explain why most of the tropics and low altitude areas, of which their current annual mean temperatures are relatively high, are predicted to experience reductions in species richness (Figure 5.4). Further warming may result in that the temperatures exceed the physiological optima of the pest species in these regions, thereby affecting their physiological functions and resource competitiveness (Bellard *et al.*, 2013).

Similarly, changes in precipitation patterns can influence species distributions. In a warmer climate, average precipitation is expected to increase, and extreme weather will become more frequent (IPCC, 2013). Increased precipitation can wash off the eggs and drown the larvae of soil-dwelling insects (Kobori and Amano, 2003; Ziska *et al.,* 2010), and potentially push pest species spreading into areas that are of lower precipitation. Drought, as a precipitation extreme, can increase plant carbohydrate concentrations, making host plants more attractive to insects (Ziska *et al.*, 2010). Nevertheless, compared to the relationship between species richness change and temperature (Figure 5.7a), the relationship between species richness change and precipitation does not exhibit multiple distinguishable peaks (Figure 5.7b), implying that pest species distributions are less sensitive to precipitation variation than to temperature variation. Indeed, prior work has suggested a relatively lesser impact of precipitation on the pests than temperature change (Herms, 2004).

5.4.3 Limitations

Although the coefficient of variation regarding the model projections for species turnover revealed small uncertainties in most parts of the study area (Figure 5.6), large uncertainties were estimated in areas with higher species turnovers (Figure 5.5 and Figure 5.6). This indicates that the models are less certain in predicting the extent to which high species turnovers will occur. The predictions for 2070 also showed greater uncertainties than those for 2050 (Figure 5.6), signifying a less confidence for the predicted species turnovers for the “far” future than for the “near” future. The lower agreements between climate change scenarios in the predictions may be because the differences between the extents to which climate will change estimated from different greenhouse gas emission scenarios will increase with time (IPCC, 2013). Uncertainties associated with the modelling analysis remain with regard to the different future climate change scenarios being considered. Despite these uncertainties, the findings of this study offer insights into the potential influences of climate change on the crop pest species distributions, at a global scale.

Concerning the species occurrence records used in the ENMs, this study has focused on common, highly harmful, widespread, and easily detected species because they are of economic importance to major food and cash crops, frequently spotted and reported, and capable of establishing and spreading into new ecosystems (CABI, 2016). Rare species were not included in this study due to the potential accuracy issues involved in the observations of such pests. Nonetheless, certain secondary pests may become primary pests and expand their distributions when, for example, the environmental conditions become suitable (Heinrichs and Mochida, 1984; Kogan, 1998); the impacts of these crop pest species on future agricultural systems should hence not be overlooked. Future research on the influences of climate change on the distribution of such species would be desirable.

It should be noted that this study has assumed species evolutionary stasis based on the concept of evolutionary conservatism of ecological niches (Wiens *et al.*, 2009), although recent work has found that species, such as the red-legged earth mite in Australia, would develop higher thermal resistance to respond to novel climate (Hill *et al.*, 2013). Also, this study has not considered the effects of biotic interactions for the correlative ENMs used and the global extent examined. Prior work has suggested that climate tends to have a dominant control over species distributions, while the effects of biotic interactions appear to be less important in coarse-resolution, broad spatial extent analyses (Pearson and Dawson, 2003). However, the assumption of evolutionary stasis and the exclusion of biotic interactions may affect the estimates of species distributional changes (Araújo and Luoto, 2007). Fusing correlative and mechanistic ENMs, which can take more ecological interactions and evolutionary assumptions into consideration (Mokany and Ferrier, 2011), is likely to address these issues of uncertainty.

Lastly, this study held a relatively conservative view about climatic change. Despite of Paris Agreement, it might be problematic to assume that global warming is limited to well below 2℃ due to the voluntary nature of the international cooperation. It will be necessary to explore the two relatively extreme climatic change scenarios (i.e., RCP6.0 and RCP8.5), so as to better understand the potential consequences to the pest species in extreme climatic change scenarios. For example, this study found that the relatively conservative climatic changes will expand the overall global distribution of the pest species. However, it is questionable whether stronger climatic change will diminish the overall global distribution of the pest species, as overheated climate tends to surpasses the heat tolerances of many species (see Section 5.4.2).

5.4.4 Recommendations

The results of this study reinforce the prior suggestions that renewed efforts to monitor the occurrence of the crop-destroying organisms are important (Estay *et al.*, 2009; Ziska *et al.*, 2010; Bebber *et al.*, 2013). Controlling crop pest species invasions associated with climate change will be critical in alleviating the potential threats to global food security that may lead to long-term nutritional emergencies and food crisis (Flood, 2010). Measures to early detect and restrain the invasive crop pest species are crucial to prevent colonization, naturalization and further spreading of the pests (Walther *et al.*, 2009), because once the species become established, they are often extremely difficult to eradicate (Gallien *et al.*, 2012).

Regions predicted to experience both increases in species richness and high species turnovers should be focused. This includes high latitude countries such as the United States and Europe where agricultural productivity per unit land area is highest (Bebber *et al.*, 2013). Prior work suggests that climate change is possibly beneficial to agricultural systems of high latitudes as present agriculture at those regions is largely constrained by low temperatures (Cannon, 1998). It is, however, crucial to take into account the influence of pest distributional change because, as demonstrated in this study, climate change can pose negative effects of pest invasions to those regions. The threats of potential crop pest invasions to food security may therefore be further complicated, presenting a challenge to invasive pest species management.

Likewise, for regions with predicted species richness reductions, their generally low species turnovers suggest that crop pest to food security will still be an issue of concern. Many less developed countries are at low latitudes and they are less able to monitor and manage the pests. Consequently, continuous strengthening pest control, management, quarantine, and eradication in these regions remains important.

Detecting pest species invasions and monitoring their diversities at the early stage of their invasions are difficult. Conventional approaches of biological control, chemical control, and quarantine and monitoring programs have to be enhanced to reduce the potential invasion risks. Additionally, considering the limited financial and human resources, particularly in many remote rural areas, approaches to leverage on observations from citizen scientists (e.g., amateur entomologist, experienced farmers who possess useful local and indigenous pest knowledge) are desirable. VGI not only is a valuable data source for strategic IPM sense making, as the study of this chapter showcased, VGI can also be adopted in ubiquitous pest invasion surveillance for strategic pest control. The involvement of non-professional observers on the ground allows early detection and reporting of incidents of pest species invasions or anomalies in a real-time manner. Image recognition technology carried on smartphone application can help non-professional observers accurately identify pest species through photo taking (Suen *et al.*, 2014). Better information communication and interaction among citizen scientists, pest experts, and policy makers will be useful for monitoring and managing the risks from pest species distributional changes.

Two case studies pertaining to tactical IPM, user participation incentive analysis (Chapter 4), and strategic IPM (the current chapter) have been presented so far. In the next chapter, another study specifically about VGI quality assurance will be presented, which will involve the development of a novel fuzzy expert system to systematically assure the quality of VGI.

6 Utilizing fuzzy set theory to assure the quality of VGI

As mentioned in Chapter 1 and Chapter 2, it remains challenging to assure the quality of VGI to ensure that valuable intelligence for managing pest risks can be derived. In the case studies of the previous chapters, tactical and strategic VGI sense making were conducted (see Chapter 4 and Chapter 5), in which the VGI were used with caution with regards to their quality. VGI (facilitated VGI) quality in Chapter 4 was mainly assured through targeted recruitment (for details see Chapter 4) and VGI quality in the case study of Chapter 5 was mainly assured by only collecting already validated historical VGI from citizens (for details refer to Chapter 5). However, on the one hand, the targeted recruitment approach only applies to projects based on facilitated VGI creations. On the other hand, already validated historical VGI data do not work for time-critical issues at present (e.g., pest outbreaks), which need real-time quality assurance. Therefore, more robust VGI quality assurance approaches are needed to satisfy diverse scenarios of VGI-based IPM, which is the research gap that this chapter aims to fill. Meanwhile, this chapter aims to contribute to the understanding of the practical and potential roles of volunteer participants in VGI quality assurance.

Specifically, in this chapter a fuzzy expert system for assuring VGI quality for the purposes pest surveillances is proposed. The system takes advantage of fuzzy set theory to handle data uncertainty and ambiguity inherent in VGI contributions, incorporating explicitly the unique property of VGI–trust. To demonstrate the usefulness of the fuzzy system in handling VGI quality, the system was adopted to measure the quality of a set of volunteered crop pest surveillance reports collected in Xiajiang prefecture of Jiangxi province, China.

The following sections of this chapter will introduce the fuzzy system first (Section 6.1 and Section 6.2), followed by a discussion on the features of the fuzzy system, the roles of VGI providers in the system, and the future directions of this line of research (Section 6.3).

6.1 A fuzzy system

Section 2.3.3 has reviewed how the quality of VGI is assured by the approaches from existing work. It also illustrated the related shortcomings. The rule-based fuzzy system to be presented in this section for assuring the quality of user-generated species (pest) surveillance reports has been informed by the literature review. The system uses trust as a proxy of quality, considering both the track record of the VGI contributors (i.e., provenance of user expertise) and the fitness of geographic context as defining factors of the trust.

6.1.1 Fuzziness in geospatial data quality

As mentioned in Section 2.3.3.1, traditionally, geospatial data quality is categorized into internal quality or external quality (Devillers *et al.*, 2005). Evaluating geospatial data quality of both kinds involves using realities or fitness as the baseline for comparison. The result of the comparison is clear-cut, or "crisp", as they can either be meeting or failing to meet the standard.

From a user perspective, VGI quality may be considered by users to be meeting the standard or slightly below standard, implying a transition between all levels of quality. It is extremely limiting to treat a VGI that is slightly below the standard in the same way as another VGI that virtually fails to meet the standard. Yongting (1996) proposed the concept of fuzzy quality to account for such a transition by expressing the quality with a fuzzy set instead of Boolean logic. In addition, given the role of trust in evaluating VGI data quality and the nature of trust being inherently fuzzy (Chang *et al.*, 2005), adopting fuzzy set theory to assess the quality of VGI is likely to capture more accurately the whole assessment process.

6.1.2 Fuzzy set theory

Fuzzy set was first introduced by Zadeh (1965) to model continuous phenomena. It generalizes conventional crisp sets by allowing their elements to have degrees of membership. The membership is defined by mapping every element x from a universe of discourse X to an interval [0, 1], representing the degree to which x is an element of a fuzzy set, expressed as Eq. 6.1.

$$\mu_A(x): X \rightarrow [0,1], \text{ where}$$

$$\mu_A(x) = 1 \textit{ if x is totally in } \text{the } \textit{fuzzy set};$$

$$\mu_A(x) = 0 \textit{ if x is not in the fuzzy set};$$

$$0 < \mu_A(x) < 1 \textit{ if x is partly in the fuzzy set}. \tag{6.1}$$

Fuzzy set is often used for modelling subjective human reasoning using natural languages in which many expressions have vague or imprecise meanings (Caha *et al.*, 2012). It is therefore a prominent alternative to more traditional modelling paradigms for addressing complex, ill-defined, and less tractable systems (Manca and Curtin, 2012). In geography, fuzzy set has been applied to modelling the uncertainty inherent in spatial datasets (Al-kheder *et al.*, 2008; Zhang *et al.*, 2014).

6.1.3 System development

To introduce the system development based on fuzzy set theory, the following sections will first describe its core fuzzy inference method. Then the two input variables (i.e., provenance of user expertise and fitness of geographic context), one output variable (i.e., the trustworthiness of user reports), and fuzzy rules of the system will be defined. Lastly, the system usage will be introduced.

6.1.3.1 Fuzzy inference

Mamdani-style fuzzy inference is adopted in the system as it is better suited to handling fuzziness and data uncertainty and it works better with human inputs (Power *et al.*, 2001). The inference requires the developer to create both input and output membership functions from linguistic interpretations of a subject. It generates output values through compositional inference rules and a defuzzification algorithm. Details about Mamdani-style fuzzy inference can be found in Mamdani (1974) and Negnevitsky (2005). A brief workflow showing how my fuzzy system derives the quality (trustworthiness) of a user report based on Mamdani-style is given in Figure 6.1, which has four steps as follows:

Step 1. Fuzzification: Fuzzifying the crisp inputs of the system against appropriate linguistic fuzzy sets and generating membership degrees based on given membership functions.

Step 2. Rule evaluation: Applying a fuzzy rule set to infer fuzzy trustworthiness outputs.

Step 3. Aggregation of the rule outputs: Aggregating the output of each rule into a single fuzzy set for the overall fuzzy output.

Step 4. Defuzzification: Defuzzifying the aggregate output fuzzy set into a final crisp trustworthiness score using the center of gravity (COG) algorithm. The algorithm finds the point (COG) where a vertical line would slice the aggregate set, on the interval [a, b], into two equal masses using Eq. (6.2).

$$COG = \frac{\sum_{x=a}^{b} \mu_{aggregate}(y)y}{\sum_{x=a}^{b} \mu_{aggregate}(y)}. \qquad (6.2)$$

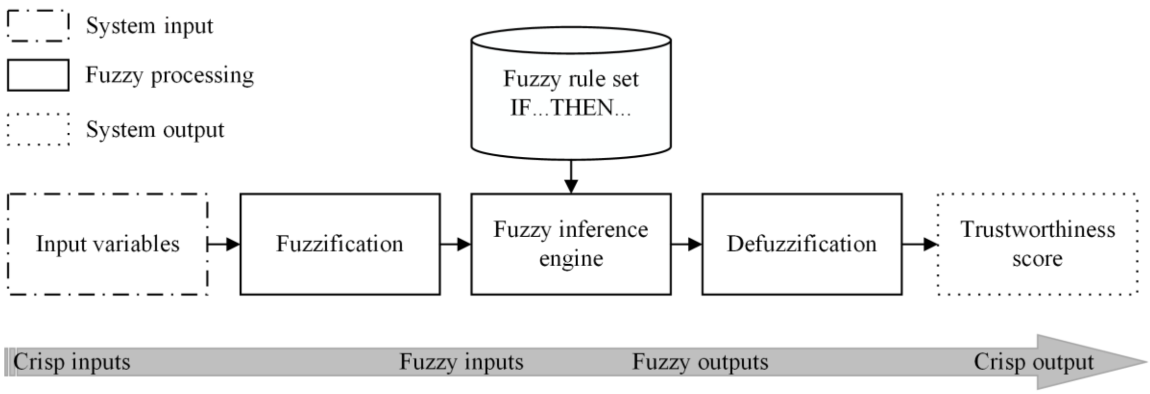


Figure 6.1 Workflow of the Mamdani-style trustworthiness score inference.

6.1.3.2 Input variable one: Provenance of user expertise

The proposed system adopts user confidence, the strength to which a person believes that a piece of information is the best available (Peterson and Pitz, 1988), as a surrogate to represent provenance of user expertise because it has been shown that confidence can be a valid cue to information accuracy (Sniezek and Van Swol, 2001). This piece of information, specifically the level of confidence about the correctness of a user report, is provided by the user who has generated the report. It contributes to the willingness to accept a piece of information, especially when other materials about the information providers are unavailable (Sniezek and Buckley, 1995; Cofta, 2007). Indeed, confidence has been utilized to automatically evaluate the expertise of the volunteers in performing tasks such as land cover map validation (Foody *et al.*, 2013) and galaxy classification (Bordogna *et al.*, 2014b).

My fuzzy system requires users to choose a value from a ten-point Likert scale to report their confidence levels. The value provides a measure of self-evaluation to VGI quality. Following the four-level fuzzy confidence adopted in Yu and Tsai (2006), four linguistic fuzzy sets–*Not Confident* (NC), *Somewhat Confident* (SC), *Confident* (C), and *Very Confident* (VC)–are defined for the input user confidence levels, using standard triangular and left/right trapezoidal shapes. The corresponding membership functions are defined by Eq. 6.3 and illustrated in Figure. 6.2a. Note that the four fuzzy sets are not symmetric around the median value of the universe of discourse (i.e., five) (Figure. 6.2a) for the following reason. As the confidence declared by non-expert VGI contributors tend to be less reliable as the confidence declared by experts because some contributors may be somewhat overconfident about their expertise (Pulford, 1996), the membership functions representing moderate to relatively high levels of user confidence (i.e., SC, C, and VC) are shifted closer to the right end of the universe of discourse to compensate for over-confidence. The left starting point of VC is kept at 7.5, allowing the values between 7.5 and 8 to have certain low degrees of membership to VC.

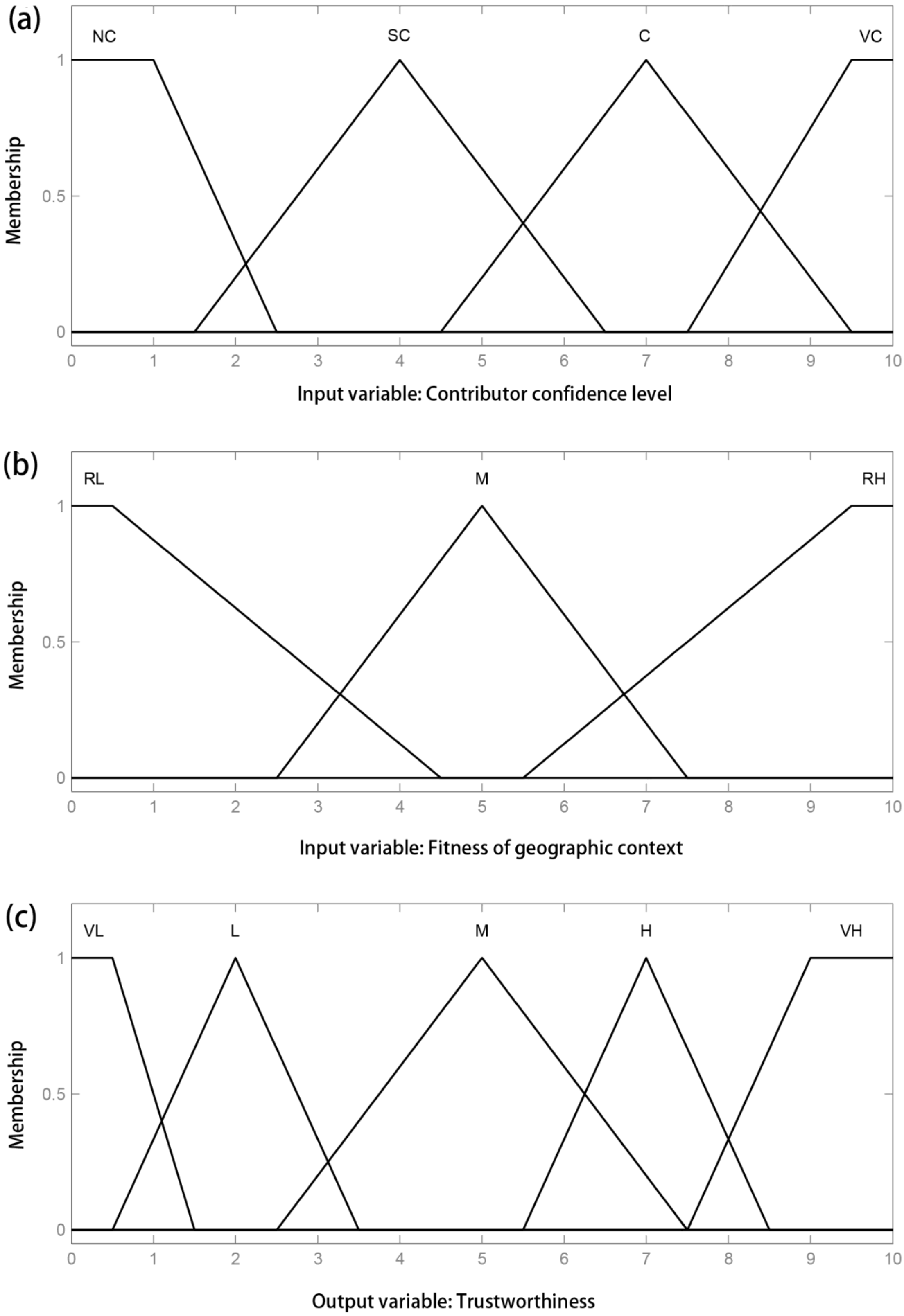


Figure. 6.2 Membership functions of (a) contributor confidence level, (b) fitness of geographic context, and (c) trustworthiness.

$$
\mu_C(c)=\begin{cases}
Not\ Confident\begin{cases}1 & 0\le c\le 1\\ (-\frac{2}{3}c+\frac{5}{3}) & 1\le c\le 2.5\\ 0 & c\ge 2.5\end{cases}\\
Somewhat\ Confident\begin{cases}0 & c\le 1.5\\ \frac{2}{5}c-\frac{3}{5} & 1.5\le c\le 4\\ -\frac{2}{5}c+\frac{13}{5} & 4\le c\le 6.5\\ 0 & c\ge 6.5\end{cases}\\
Confident\begin{cases}0 & c\le 4.5\\ \frac{2}{5}c-\frac{9}{5} & 4.5\le c\le 7\\ -\frac{2}{5}c+\frac{19}{5} & 7\le c\le 9.5\\ 0 & c\ge 9.5\end{cases}\\
Very\ Confident\begin{cases}0 & c\le 7.5\\ \frac{1}{2}c-\frac{15}{4} & 7.5\le c\le 9.5\\ 1 & c\ge 9.5\end{cases}
\end{cases}
\tag{6.3}
$$

6.1.3.3 Input variable two: Fitness of geographic context

Species (pest) occurrences usually form clusters. Therefore, fitness of geographic context is evaluated using spatial clustering analysis. According to Tobler’s Law, it is highly possible that a species (pest) can be observed at its habitat center (i.e., cluster center) and the possibility decreases with increasing distance away from the habitat center. Therefore, if a cluster of species (pest) surveillance reports is contributed by users, its fitness of geographic context is evaluated using its spatial proximity to the center of the cluster.

The fuzzy system uses DBSCAN clustering algorithm (Ester *et al.*, 1996) to locate VGI clusters. DBSCAN can effectively distinguish noise points (i.e., outliers) and discover clusters with arbitrary shapes. Fitness of geographic context is quantified based on an inverse hyperbolic sine function (Eq. 6.4). The equation captures precisely the characteristics of the fitness of geographic context–it decays with the distance departing from the center of a VGI cluster (i.e., inverse relation with distance) by generating a value between 0 (zero fitness of geographic context) to 10 (perfect fitness of geographic context). Outliers identified by DBSCAN are assigned zero fitness of geographic context.

$$Fitness\ of\ \ geographic\ context = \left\{ 1 - \ln\left[\frac{Dist_{rtc}}{Dist_{max}} + \sqrt{\left(\frac{Dist_{rtc}}{Dist_{max}} \right)^2 + 1} \right] \right\} \times 10, \qquad (6.4)$$

where $Dist_{rtc}$ is the distance from a user report to its corresponding cluster center, $Dist_{max}$ is the distance between the cluster's outermost user report and the cluster's center.

Following the three-level fuzzy proximity adopted in Al-kheder *et al.* (2008), three linguistic fuzzy sets–*Relatively Low* (RL), *Medium* (M), and *Relatively High* (RH)–are defined for the input fitness of geographic context, using standard triangular and left/right trapezoidal shapes. The corresponding membership functions are defined by Eq. 6.5 and illustrated in Figure. 6.2b. The fuzzy sets are symmetric around the median value of the universe of discourse (i.e., five) (Figure. 6.2b).

$$
\mu_F(f)=\begin{cases}
Relatively\ Low\begin{cases}1 & 0\le f\le 0.5\\ (-\frac{1}{4}f+\frac{9}{8}) & 0.5\le f\le 4.5\\ 0 & f\ge 4.5\end{cases}\\
Medium\begin{cases}0 & f\le 2.5\\ \frac{2}{5}f-1 & 2.5\le f\le 5\\ -\frac{2}{5}f+3 & 5\le f\le 7.5\\ 0 & f\ge 7.5\end{cases}\\
Relatively\ High\begin{cases}0 & f\le 5.5\\ \frac{1}{4}f-\frac{11}{8} & 5.5\le f\le 9.5\\ 1 & f\ge 9.5\end{cases}
\end{cases}
\tag{6.5}
$$

6.1.3.4 Output variable: Trustworthiness

Following the five-level fuzzy trustworthiness in Song *et al.* (2004), five linguistic fuzzy sets–*Very Low* (VL), *Low* (L), *Medium* (M), *High* (H), and *Very High* (VH)–are defined for the output trustworthiness using standard triangular and left/right trapezoidal shapes. The corresponding membership functions with a universe of discourse from 0 to 10 are defined by Eq. 6.6 and illustrated in Figure. 6.2c. The five fuzzy sets are asymmetric around the median value (i.e., five) (Figure. 6.2c) for the following reasons. Goodchild and Li (2012) suggested that greater weights can be assigned to similar reports that are spatially clustered than to a single report. This system assesses clustered reports which already have relatively greater weights. Therefore, the fuzzy sets representing relatively poor data quality (i.e., VL and L) are placed closer to the left end of the universe of discourse, meaning that a trustworthiness can be linguistically interpreted as low or very low only when it is associated with a sufficiently low value. The peak of VL is not placed at zero to ensure that the peak value stays the same over a certain range (Zhang *et al.*, 2014). Additionally, the wider range of M can maintain sufficient overlap in adjacent fuzzy sets (especially L and M) for the system to respond smoothly (Negnevitsky, 2005).

$$\mu_{\mathrm{T}}(t)=\begin{cases} \textit{Very Low}\begin{cases} 1 & 0\le t\le 0.5 \\ (-t+\frac{3}{2}) & 0.5\le t\le 1.5 \\ 0 & t\ge 1.5 \end{cases} \\ \textit{Low}\begin{cases} 0 & t\le 0.5 \\ \frac{2}{3}t-\frac{1}{3} & 0.5\le t\le 2 \\ -\frac{2}{3}t+\frac{7}{3} & 2\le t\le 3.5 \\ 0 & t\ge 3.5 \end{cases} \\ \textit{Medium}\begin{cases} 0 & t\le 2.5 \\ \frac{2}{5}t-1 & 2.5\le t\le 5 \\ -\frac{2}{5}t+3 & 5\le t\le 7.5 \\ 0 & t\ge 7.5 \end{cases} \\ \textit{High}\begin{cases} 0 & t\le 5.5 \\ \frac{2}{3}t-\frac{11}{3} & 5.5\le t\le 7 \\ -\frac{2}{3}t+\frac{17}{3} & 7\le t\le 8.5 \\ 0 & t\ge 8.5 \end{cases} \\ \textit{Very High}\begin{cases} 0 & t\le 7.5 \\ \frac{2}{3}t-5 & 7.5\le t\le 9 \\ 1 & t\ge 9 \end{cases} \end{cases} \tag{6.6}$$

6.1.3.5 Fuzzy rules

The full IF-THEN fuzzy rule set defined for this system is shown in Figure 6.3, using a conjunction, AND, for all the rules (e.g., IF confidence level is SC AND fitness of geographic context is RL THEN trustworthiness is VL). The conjunctions in the fuzzy rules are evaluated using the fuzzy operation *intersection* (Negnevitsky, 2005). Assuming that A and B are two fuzzy sets membership functions μ_A and μ_B, respectively, the fuzzy operation *intersection* for creating the intersection of the two fuzzy sets is expressed as Eq. 6.7.

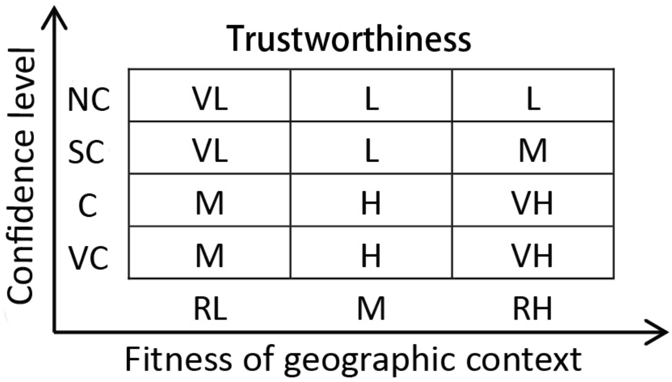


Figure 6.3 Fuzzy rule set defined for the system, using a conjunction, AND, for all the rules.

$$\mu_{A \cap B}(x) = min[\mu_A(x), \mu_B(x)]. \qquad (6.7)$$

6.1.3.6 System output surface

To evaluate the performance of a Mamdani-style fuzzy system, I used its three-dimensional output surface following the suggestion by Negnevitsky (2005). A satisfactory system building is achieved through empirical tunings until the system generates a gradual changing surface which appropriately emulates subjective human reasoning regarding how the interactions of the system's inputs influence its output in the context the problem is viewed.

The output surface of my system is shown in Figure 6.4. The membership functions and the fuzzy rules mentioned above are decided based by assessing this surface. The general trend should be that higher user confidence levels and higher fitness of geographic context lead to higher trustworthiness, while certain special considerations should be appropriately reflected on the surface. For example, if a report has an extremely low user confidence level (meaning a very low user expertise), its trustworthiness should be very low even if its fitness of geographic context is high. Conversely, even if a report has an extremely low fitness of geographic context, its trustworthiness should be moderate if its user confidence level is very high.

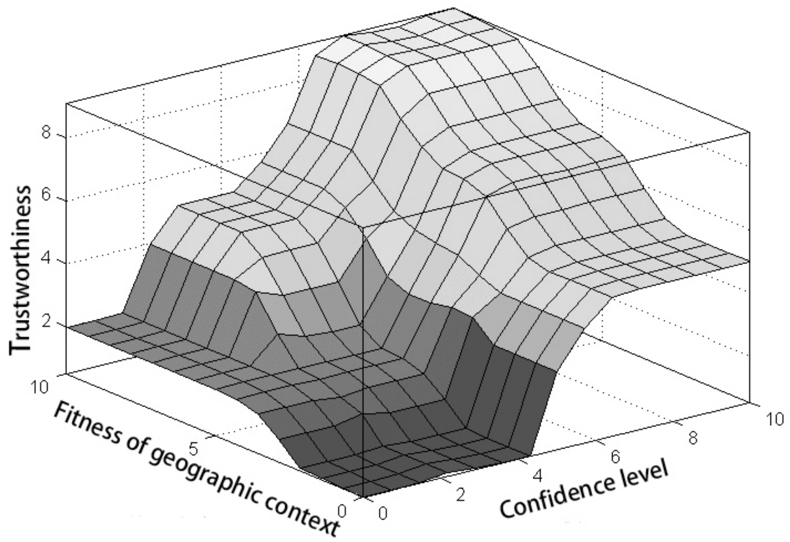


Figure 6.4 Output surface of the fuzzy system.

6.1.3.7 System usage with a running example

Figure 6.5 shows an example of generating trustworthiness of a reporting (7.68) with two crisp inputs of confidence level (8) and fitness of geographic context (6.5). The red vertical line through the aggregate output fuzzy set depicts location of the COG.

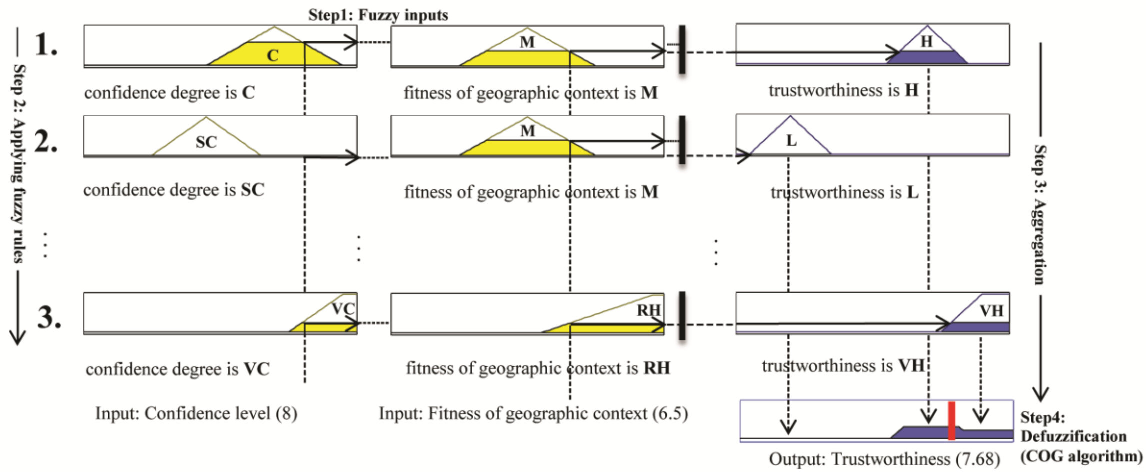


Figure 6.5 Fuzzy inference process for an assumed user report.

Once the system generated the trustworthiness scores for a VGI dataset, a user-preferred threshold is used to reject or accept the reports. Non-outlier reports with trustworthiness scores lower than the threshold will be rejected and will be accepted if otherwise. Outlier reports should be specially treated. Outlier reports with trustworthiness scores lower than an assigned threshold can be simply discarded. However, outlier reports above the threshold should be treated with caution. It should be reserved or held for further observations, i.e., to see whether or not similar reports will be reported nearby to confirm it.

The selection of threshold is context-dependent and subject to the accuracy requirements of specific projects. Setting a higher threshold can reduce the number of false positives (FP), but it will inevitably increase the number of false negatives (FN). Setting a lower threshold can reduce the number of FN, while it will increase the number of FP. In the context of VGI, FN is actually better than FP. Because rejecting good quality VGI incorrectly is actually better than accepting poor quality VGI incorrectly. Certainly one can choose a very high threshold to only collect VGI with extremely high trustworthiness scores, and ignoring FN.

This system has been implemented using the following tools. DBSCAN algorithm was integrated to ArcGIS (ESRI Products, Redlands, CA) as an extension through Python. The fuzzy logic toolbox of MATLAB was used for performing the fuzzy inference. Figure 6.6 illustrates the architecture of the implemented system.

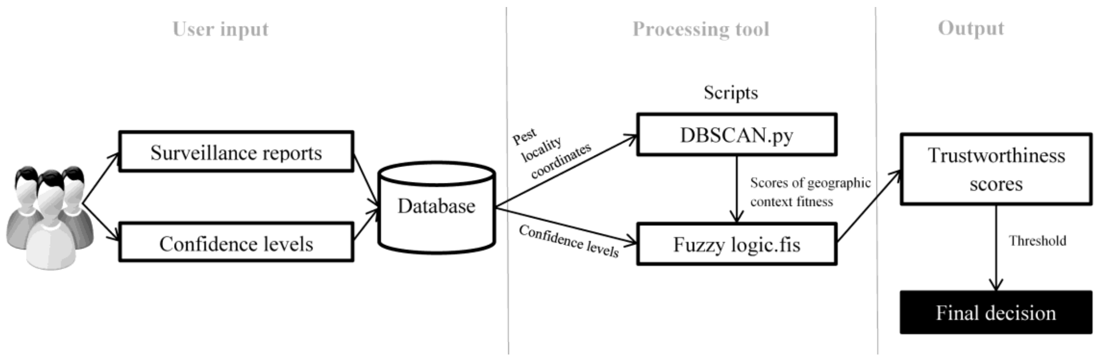


Figure 6.6 Architecture of the system implementation.

6.2 Case study: A VGI-based crop pest surveillance

In order to demonstrate the usefulness of the fuzzy system in handling VGI quality, the system was adopted to measure the quality of a set of crop pest surveillance reports collected in Xiajiang prefecture of Jiangxi province, China.

6.2.1 Study design and data analysis

6.2.1.1 VGI collection and quality assessment using the fuzzy system

The major crop type cultivated in Xiajiang prefecture was rice which accounted for around 90% of the total cropland of the prefecture (216km^2). Two hundred local rice farmers distributed across the prefecture were recruited to conduct a rice pest surveillance.

The pest surveillance was conducted by the farmers from 15 to 25 August, 2014. They reported rice pest incidents (pest occurrences, damages, or both) observed during their daily farming activities. To report an observed pest incident, the observer inserted a flat bamboo chip firmly into the soil of the rice paddy where the pest incident was observed. The observer also recorded the species name, observation time, and confidence level on the bamboo chip. No other coaching to the farmers was conducted to ensure a minimum intervention to the user contributions. Hence the volunteered crop pest surveillance data can be considered as active VGI (see Section 2.1.2 about the forms of

VGI creation). There was no pre-defined criteria confining the participant recruitment. That is, differing from those participants recruited for the study in Chapter 4, here the abilities of the participants in pest recognition were not known beforehand. After the pest surveillance, I collected the geographic coordinates of the inserted bamboo chips using Trimble® GeoXT handheld GPS devices which delivered a 50 cm positioning accuracy.

Various rice pest incidents were reported, the species included mainly rice stem borers, rice leaf rollers, rice plant hoppers, rice water weevils, and mole cricket. Of the species, the rice stem borers' scope of activity was relatively fixed over time. It thus would be easier to conduct post-surveys to verify the actual presences of the reported rice stem borer incidents. Rice stem borer incident reports were therefore used to evaluate the usefulness of the system. During the pest surveillance period, 209 rice stem borer incident reports were collected.

The quality of the 209 incident reports were assessed using the fuzzy system. A threshold should be assigned to the generated trustworthiness scores of the reports to determine whether or not a report should be accepted. As mentioned above, one can set a high threshold to only collect VGI with extremely high trustworthiness scores and ignore FN. In this case study, however, I intended to preserve as many reports as possible. Thus, a moderate threshold is more appropriate. I used a range of thresholds, from 4 to 6 with an increment of 0.2, to evaluate the performance differences. Outliers, if any, were specially treated according to the method stated in Section 6.1.3.7. The system generated categorical results, i.e., accepted, rejected, and withheld.

6.2.1.2 Ground truth data collection

From 26 August to 1 September, 2014 (immediately after the pest surveillance conducted by the farmers), a field pest survey was conducted by the pest management experts from the local agricultural department to verify the actual presences of the 209 reported rice stem borer incidents. The experts scrutinized the evidences including feeding wounds or holes, larval frass, egg masses, damage symptoms, and pupas of the stem borers within a two-meter buffer zone (considering the mobility of the borers) surrounding each bamboo chip. If none of these evidences could be detected within the buffer zone of a report, the report was rejected by the experts; and was accepted if otherwise. The pest survey thus generated categorical results, i.e., reports being accepted and reports being rejected. Since the survey was conducted by experienced experts, the results of which were considered as accurate ground truth data.

6.2.1.3 Conformity tests

Subsequently, conformity tests were conducted. For each threshold within the interval [4, 6], a Cohen's kappa statistic (Viera and Garrett, 2005) showing the degree of agreement between the fuzzy system-generated results and the pest survey results was calculated. The sensitivity (Eq. 6.8) and specificity (Eq. 6.9) were also calculated, respectively. Note that the reports in withheld status were not included in the calculations. The system-generated results corresponding to the highest kappa value were mapped for visualization, for which a confusion matrix was provided to show the details about the degree of agreement.

$$Sensitivity = TP / (TP + FN), \tag{6.8}$$

$$Specificity = TN / (TN + FP), \tag{6.9}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

To evaluate the impacts of sample sizes on the system performance, randomization tests were performed. Using the threshold that corresponded to the highest kappa value and following the method mentioned above, 30 rounds of conformity test using 30 groups of different subsets from the whole sample set were conducted. That is, ten groups of 90%-subset, ten groups of 60%-subset, and ten groups of 30%-subset were randomly extracted from the whole dataset to conduct the conformity tests. Cohen's kappa statistic was calculated for each run of the tests. Mean and standard deviation were calculated for the Cohen's kappa statistics from each of the same percentile subsets.

6.2.2 Results

The fuzzy system identified from the 209 rice stem borer incident reports eight clusters and 16 outliers (Figure 6.7a) and trustworthiness scores from 0.51 to 9.10 (Figure 6.7b).

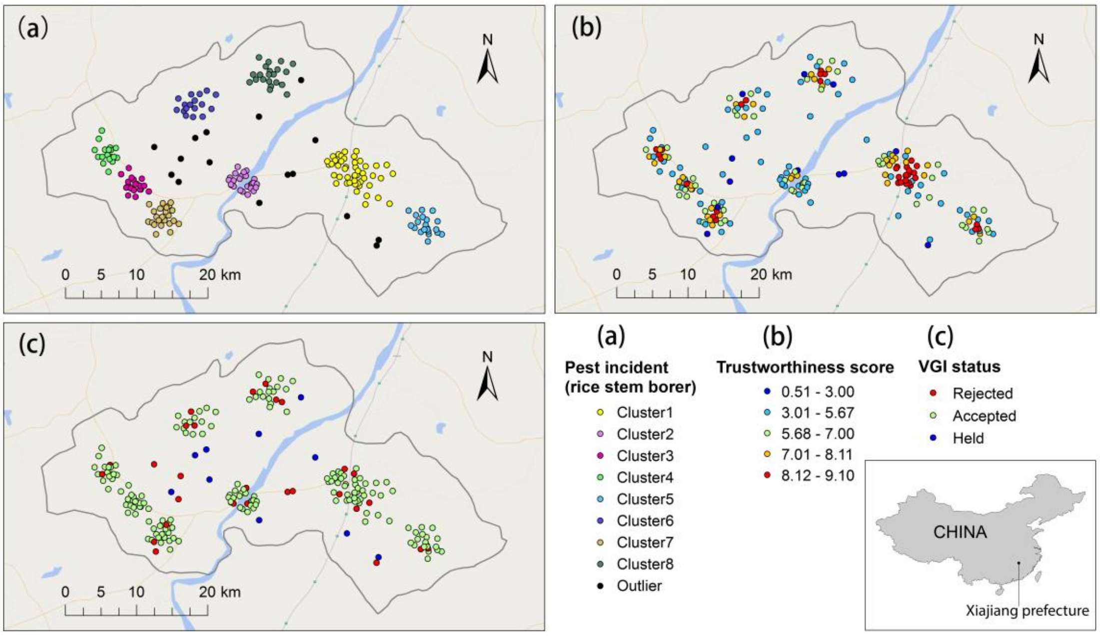


Figure 6.7 Maps showing the VGI quality assessment results generated by the fuzzy system. Background map is the road map of Bing Maps. Thumbnail on the lower right corner shows the relative location of Xiajiang prefecture in China. (a to c) Grey polyline represents boundary of Xiajiang prefecture. (a) 209 reported rice stem borer incidents. (b) Inferred trustworthiness scores of the reports. (c) Final statuses of the reports based on a threshold five.

The results of the conformity tests using different thresholds are shown in Table 6.1. The sensitivity and specificity values confirm that elevating and lowering the thresholds can increase the numbers of FN and FP, respectively, leading to lower kappa values. The highest kappa value (0.67) corresponds to the thresholds 4.8 and five. Therefore, for the purpose of this case study, the integer five was adopted as the threshold for further analyses, although the threshold 4.8 obtained a same kappa value as the threshold five did.

Table 6.1 Results of the conformity tests using different thresholds.

Threshold	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0
Kappa value	0.59	0.59	0.64	0.62	0.67	0.67	0.37	0.35	0.34	0.34	0.29
Sensitivity	0.98	0.98	0.98	0.96	0.96	0.96	0.76	0.73	0.70	0.69	0.65
Specificity	0.53	0.53	0.59	0.60	0.66	0.66	0.73	0.76	0.81	0.81	0.81

With a threshold five, see Figure 6.7c, the fuzzy system rejected 29 reports, including six outliers with trustworthiness scores lower than five (red circles), and accepted 170 reports (green circles). Ten outlier reports were held (blue circles) due to their relatively high user confidence levels associated. The conformity test showed that around 91% of the fuzzy system-generated results agreed to the survey results with a corresponding kappa value 0.67. Details are visualized by the confusion matrix shown in Figure 6.8. Regarding the ten pest incident reports that were held, eight of which had in fact suffered infestations according to the survey results.

Sensitivity: 0.96 Specificity: 0.66		Model results		Total
		Rejected	Accepted	
Survey results	Rejected	23 (TN)	12 (FP)	35
	Accepted	6 (FN)	158 (TP)	164
Total		29	170	199

Figure 6.8 Confusion matrix visualizing the degree of agreement.

Lastly, using the threshold five, it was observed that the system performed better with larger sample sizes, as the mean kappa values increased with the increase of sample size (Figure 6.9). The standard deviations also decreased with the increase of sample size (Figure 6.9).

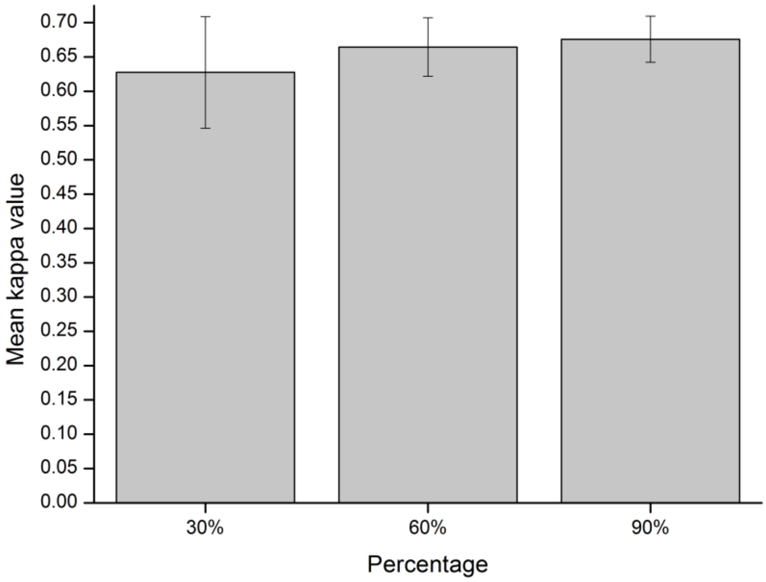


Figure 6.9 Means and standard deviations (shown as whiskers) calculated for the Cohen's kappa statistics from the three groups of percentile subsets extracted for testing the impacts of sample sizes.

6.3 Discussion

6.3.1 Features of the system

Using the pest surveillance data, it is demonstrated that fuzzy set theory can lead to desired VGI quality assurance results in a near real-time manner. The fuzzy system was developed based on the idea that the quality of VGI can be assured based on its geographic context and provenance of user expertise, and trust can be used as a proxy of quality. Fuzziness involved in trust judgement requires special attentions, and quality itself is also inherently fuzzy. Therefore, fuzzy set theory was adopted as the key to the system design, which easily incorporates semantic knowledge into the quality assessment. Bordogna *et al.* (2014a) promote a linguistic decision-making approach to assess the quality of VGI. The system developed here extends their work by demonstrating the utility of fuzzy set theory in assessing the quality of user-contributed species (pest) surveillance reports in particular.

The system design echoes the view of Van Exel *et al.* (2010) that assessing VGI quality must consider not only feature quality and user quality but also the interdependency between them. To account for fitness of geographic context (feature quality), DBSCAN clustering was used for identifying VGI clusters. As pointed out by Goodchild and Li (2012), quality measures of VGI can arise from the data themselves. More credit can be given to a clustering of similar reports than to a single report, in which case one can develop metrics of quality based on the clustered reports. In the current system, the metric is based on the proximities of user reports to their corresponding cluster centers. It resembles Gao *et al.* (in press) in which a distance-decay function is used to measure the memberships of a cluster of VGI points assigned to Harvard University campus. The closer a point was to the campus core area, the higher membership the point obtained. Similarly, Liu *et al.* (2010) used an interpolation procedure to measure the weights of candidate point locations assigned to South China region. The closer a location was to the core area of South China region, the higher weight the point obtained. In addition, confidence was used as a surrogate to represent provenance of user expertise (user quality). The case study confirmed that requesting the volunteers to self-evaluate the correctness of their observations was useful for assessing the quality of the generated information (i.e., the role in VGI metadata creation). By considering the interdependency between the feature quality and user quality, the fuzzy system could detect those VGI which seemed to fit geographic context well but virtually were of high uncertainty. Note that this system satisfies both active VGI (participant capability is unknown beforehand) and facilitated VGI creation (participant capability is known beforehand). For active VGI creation, fitness of geographic context and a participant' confidence level provide indicators for the post-assessment of VGI quality; while for facilitated VGI creation, assessing fitness of geographic context and the confidence level helps enhance the quality of VGI further.

Through the compositional inference rules (Figure 6.3), the system can appropriately emulate human reasoning about how the interactions of the two system input variables influence the system output (Figure 6.4). Using a simple linear method, for example, combining the values of the two variables by summation does not have the same capability, which will be demonstrated by three exemplar input-output combinations in Table 6.2. For Combinations 1 and 2, the simple linear method obtains two identical values (i.e., 12), while the fuzzy system generates two different values (i.e., 3.6 and 5). This is an advantage of using rule-based fuzzy system. In Combination 1, although the fitness of geographic context is perfect, the fuzzy system treats the report as being less trustworthy than that of Combination 2 due to its overly low user confidence. In Combination 2, although the fitness of geographic context is relatively low, the system ranks the report with higher credibility as the user confidence is very high. In Combinations 1 and 3, using the simple linear method obtains two different values (i.e., 12 and 7.9), while the fuzzy system generates two identical values (i.e., 3.6). The former method gives more credit to Combination 1. However, for human judgement, it appears to be less appropriate due to its overly low user confidence level. Therefore, a fuzzy system with appropriately defined system parameters (e.g., membership functions) can deal with such complicated non-linear cases through mimicking human thinking.

Table 6.2 Three different input-output combinations.

Combination	**C**	**F**	**T**	**Sum**
1	2	10	3.6	12
2	10	2	5	12
3	5	2.9	3.6	7.9

C, F, and T denote user confidence level, fitness of geographic context, and system-derived trustworthiness, respectively.

Furthermore, VGI datasets are often large in volume as VGI contributors on the ground are ubiquitous. In the case study, although the entire dataset was not large, a trend was observed that the system's performance improved with increasing sample size, rather than the opposite (Figure 6.9). The system is also robust at handling non-clustered VGI (i.e., outlier VGI detected by DBSCAN). Such VGI can be processed with two options, i.e., discarding (for reports with low user confidence levels) and holding (for reports with high user confidence levels), than simply being rejected as poor quality VGI. In the case study, the survey results showed that eight of the ten outlier reports with holding statuses had in fact suffered pest infestations. This supports my thought that outlier VGI should not be simply discarded.

Lastly, since user histories (provenance) can be harvested from VGI and fitness of geographic context can be determined from VGI, the approach also has a potential to be adapted and applied to similar VGI applications in different contexts, e.g., VGI-based earthquake casualty surveillances.

6.3.2 Potential future improvement

The fuzzy system should be extended in ways that generalize its applicability. Fuzzy logic enables tools to model the inherent fuzziness that would otherwise be neglected by traditional crisp logic, while it also introduces subjectivity into the modelling process. Imprecision related to subjectivity has often been cited as a limitation in conventional fuzzy systems (Al-kheder *et al.*, 2008; Adhikari and Li, 2013). In this study, although the selection and turnings of the system parameters are justified, they are still the results of a subjective process. Therefore, optimizing system parameters is perhaps the most important.

Taking the case study for example, the highest kappa value obtained was 0.67 on a 91% agreement. Although 0.67 is considered a substantial agreement (Viera and Garrett, 2005), the specificity (0.66) was not as good as the sensitivity (0.96) (Table 6.1). In order to preserve as many reports as possible while reducing false positive value, one solution is to further improve the system through parameter calibrations (e.g., membership function calibrations) based on sensitivity analyses using the pest survey data (ground truth data) collected in the case study. After the calibrations, the system can be generalized to larger spatiotemporal extents with greater reliability. Another solution is to use the consensus approach in which system parameters are optimized based on the subjective opinions of multiple decision-makers (Zhang *et al.*, 2014). However, both solutions are often laborious and time-consuming. To remediate this problem, a machine learning approach, which involves the use of artificial neural network to determine the appropriate system parameters automatically, seems more promising as demonstrated in studies using the combined neuro-fuzzy systems for understanding environmental quality issues, such as Carnevale *et al.* (2009) and Yan *et al.* (2010). It will be interesting to investigate how such systems can be utilized to better assure VGI quality.

Moreover, in calculating fitness of geographic context, the spatial extent of a cluster is subject to the VGI points within the cluster. A small number of false contributions in a VGI cluster would not significantly affect the cluster's spatial extent (the spatial extent affects the memberships of the points within the cluster), especially not when point density of the clusters is high. However, if the majority of the contributions in a VGI cluster are false contributions, my approach will be less effective because the uncertainty about the spatial extent of the cluster is high. This problem points to the need to incorporate a user reputation database to my fuzzy system to exclude contributions from contributors with lower reputation before my system performs a refined data quality assurance work. The idea is similar to that in a facilitated VGI creation system suggested in Cinnamon and Schuurman (2013).

This thesis so far has proposed an enhanced conceptual framework of VGI-based IPM (Chapter 3), conducted VGI sense making (Chapter 4 and 5) and developed this novel VGI quality assurance (the current Chapter). The next chapter will draw the conclusions of this thesis.

7 Concluding remarks

The studies introduced in the preceding chapters explored VGI for the enhancement of IPM. Four specific research objectives were aimed to be achieved: (1) proposing an enhanced conceptual framework of VGI-based IPM; (2) exploring VGI sense making to enhance IPM; (3) developing an approach to assure the quality of VGI; and (4) exploring the roles of volunteer participants in VGI-based IPM, and by extension, how to motivate and sustain the volunteer participation. This chapter will draw the conclusions for this thesis in response to these research objectives. It will first summarize the major findings (Section 7.1). The related implications of the research will be discussed next (Section 7.2). Lastly, several recommendations for future work will be given (Section 7.3).

7.1 Major findings

In the following four sub-sections, the major findings of this research will be summarized by following the order of the research objectives listed in Section 1.3.

7.1.1 Objective one: Enhanced VGI-based IPM framework

In response to the first research objective, which was to propose an enhanced conceptual framework of VGI-based IPM, I proposed such a framework that incorporates single or combined methods, suited to the transformative paradigm. The framework also incorporates several new elements, including the typology of VGI creation forms, the typology of IPM sense making, the quality assurance of VGI, and the participation incentive of VGI interaction. It serves as a framework of reference for enhancing IPM based on VGI as an alternative to the positivist paradigm adopted in traditional IPM, and for the development of more comprehensive VGI-based IPM.

7.1.2 Objective two: VGI sense making for VGI-based IPM

In response to the second research objective, which was to explore VGI sense making to enhance IPM, studies were conducted with regard to both tactical and strategic IPM.

First, VGI was integrated to tactical IPM in a real case, namely the surveillance of overwintering striped rice stem borers by volunteer farmers (Chapter 4). Valuable spatiotemporal characteristics of the overwintering pest outbreak were indeed discovered from VGI provided by volunteer farmers, including the hot and cold spot clusters and their structural and locational changes over time, and the detected phenological events. Regarding strategic IPM, a case study was conducted to investigate the potential global distributional shifts of poikilothermic invasive crop pest species associated with climate change (see Chapter 5). The sense making results through ecological niche modellings provided valuable strategic insights into the possible direction and range of the distributional changes of invasive crop pests.

Note that using methods that can handle the characteristics of VGI is an important step of VGI sense making. The sense making methods used in the case studies have been shown to be quite suited to doing so. The hot spot analysis (Getis-Ord Gi* statistic) (Chapter 4) and the ecological niche modelling (Chapter 5) simply need random, offhand species occurrence observation data. In other words, generating data on a regular basis is unnecessary. In addition, the degree-day model used in the phenological analysis (Chapter 4) does not need continuous data collection over very long periods of time as needed by other pest analysis models such as regression and artificial neural network (Yan *et al.*, 2015). Although it is challenging for individual non-professional pest observers to correctly identify and report the occurrence time of phenological events (see Section 2.3.3.2), collectively, phenological information may be derived (through change point analysis) as long as pest incidents are intensively reported during a certain period of time (e.g., pest outbreak period).

7.1.3 Objective three: VGI quality assurance for VGI-based IPM

To achieve the third research objective of this thesis, which was to develop an approach to assure the quality of VGI, Chapter 6 presented an expert system developed for assuring the quality of VGI based on fuzzy set theory. It was found that the degree to which the quality measurement results generated by the expert system conformed to our actual field survey results was satisfactory. In addition, qualitatively speaking, various features of the system have been shown in the study of Chapter 6, including mainly its advantages in terms of linguistic fuzziness handling, geographic context measuring, provenance acquiring, and outlier treating. Note that the results further found that the measurement results are threshold-dependent, selecting an appropriate threshold for different contexts is essential for achieving the best quality assurance results.

7.1.4 Objective four: Participant roles and incentives in VGI-based IPM

The last research objective of this study was to investigate the roles that volunteer participants can play in VGI-based IPM, and the ways of motivating and sustaining volunteer participation.

From this study, three types of roles for volunteer participants were identified. The first type of roles showcased in this study is to do with the basic geospatial data that volunteer participants can contribute for knowledge discovery and decision-making (see Table 7.1, and also Chapter 4 and 5). For this type of roles, participants play an active role in the surveillances of domestic and exotic pests in the field. They provide their real-time and ubiquitous pest surveillance information for the updating of pest databases, knowledge, and maps, and for facilitating farmer-to-farmer or farmer-to-researcher communications regarding pest risks. Note that the geospatial information provided by the participants for tactical IPM can be reused or repurposed for strategic IPM, and vice versa. For instance, the VGI collected for managing the overwintering outbreak of the striped rice stem borers in Shuibian town of China (Chapter 4) can be reutilized as species occurrence locality records for the ecological niche modelling of Chapter 5; alien species occurrences provided for strategic IPM can also be reutilized for immediate responses for tactical IPM on eradication of the alien species invasions. However, this type of roles, is limited to the provision of pest surveillance data.

The second type of participant roles showcased in this study involves participants in the creation of metadata (see Table 7.1, and also Chapter 6). Chapter 6 revealed the crowd's dual roles in geospatial data collection and in ascertaining the reliability of data. Metadata data such as provenance of users' expertise is an important source of information for evaluating the quality of VGI, which can be obtained from VGI contributors directly.

The third type of roles is to do with the even higher level contributions that volunteer participants may make in VGI-based IPM. This involves cognitive engagements of volunteer participants in VGI-based IPM. These high level participant roles have been conceptualized and discussed in this study as a potentiality for achieving even better VGI-based IPM (see Table 7.1, and also Section 4.5.2 and Section 3.4).

Table 7.1 The roles of volunteer participants in VGI-based IPM.

Participants' practical and potential role		Description	Example
1	Basic geospatial data contributions for knowledge discovery and decision-making (Explored)	Identifying and collecting real-time and ubiquitous data. Geo-referencing data. Reporting data. Commitments of individual volunteers may vary from random involvements to continuous involvements. Cognitive engagement is minimal.	Observing and identifying pest occurrences; collecting, geo-referencing, and reporting pest occurrence incidents. Invasive pest species surveillances and reporting.
2	Metadata creations (Explored)	Creating historical information about data.	A data provider's expertise, time, location, purpose of data creation.
3	Higher level contributions (Conceptualized)	Cognitive abilities of participants are to be utilized. Defining problems. Generating quantitative, qualitative, or quantitative-qualitative-combined information. Processing and analysing data. Disseminating and exchanging information. Participants are even expected to conduct the whole process.	Sharing pest surveillance information. Sharing views and experiences among stakeholders. Communication among participants and the authoritative communities. Social networking. Promotion of teamwork. Field experimenting and data analyses.

In terms of the second aspect of this research objective, which is to suggest ways to motivate and sustain volunteer participation in VGI-based IPM, results of this study showed that the dissemination of appropriate information to participants can motivate their participation (see Chapter 4). An important factor that was identified contributing to a continued and more engaged user participation was the usefulness of information disseminated to individual participants. The more the individual participants found the information of use to them, the more willing they were to contribute VGI. In the case study of Chapter 4, the questionnaire survey results showed that, for the farmer participants, information dissemination regarding temporal patterns of pest outbreaks (i.e., specific timings of pest risk management) tends to be more useful than spatial patterns (i.e., specific areas of pest risk management), so that temporal information dissemination should be enhanced to better satisfy user needs so as to keep them engaged.

7.2 Implications

In the following two sub-sections, the implications of this research pertaining to VGI-based IPM will be discussed. The implications will be divided into two parts, namely the research implications and the practical implications.

7.2.1 Research implications

Before the explorations of this thesis, VGI-based IPM was only a concept, being limited to the primitive conceptual framework proposed by Deng and Chang (2012). This study proposes an enhanced framework for developing VGI-based IPM using single or combined methodology under the transformative paradigm. The thesis has also demonstrated the value of VGI sense making for IPM enhancement, and identified several appropriate sense making methods. The results obtained and the experience shared here can increase the confidence of researchers on VGI-based IPM and enable them to develop and explore VGI-based IPM further. This thesis has also pinpointed and conceptualized the roles of volunteer participants in VGI-based IPM; it has explored participants' participation incentives (see Chapter 3 to Chapter 6). These research outcomes provide theoretical support for the establishment and further development of VGI-based IPM, and methodologically enhance the existing participatory IPM.

In addition to the implications for the establishment and development of VGI-based IPM, the specific sense making results of this research contribute to a better understanding of the characteristics of crop pests. With regard to the VGI sense making conducted for tactical IPM (Chapter 4), the knowledge discovered from the collected VGI are informative about the spatiotemporal distributions and phenological characteristics of the local rice pests. As for the VGI sense making conducted for strategic IPM (Chapter 5), the knowledge reflected from the collected VGI provides a theoretical basis for an enhanced understanding of the impacts of climatic change on the distributions of invasive crop pest species.

Furthermore, by developing a novel approach for assuring the quality of farmers' pest surveillance reports in the form of the fuzzy system (Chapter 6), this study has made a significant contribution to the research on VGI quality assurance. An interpretative insight pertaining to this contribution is that it is important to account for the fuzziness of trust judgement if trust is used as a proxy of quality. Data quality itself is also inherently fuzzy. Fuzzy logic provides a promising solution for addressing fuzziness, which is therefore adopted as a key to the system design. The fuzzy system can generate satisfactory VGI quality assessment results, which demonstrate that the use of fuzzy set theory is indeed appropriate. Additionally, data contributor's expertise is accounted for by the fuzzy system based on user self-evaluation, which has been demonstrated to be useful for VGI quality assessment through the case study conducted in Chapter 6. Since individuals can provide VGI with varying levels of accuracy, it is important to take into account their strengths and weaknesses when assessing VGI quality.

7.2.2 Practical implications

When it comes to its practical value, the implications that can be drawn from this study are manifold. As discussed in the previous chapters, this study has implications for the enhancement of IPM's effectiveness, for preventing both short-term and long-term risks in agricultural productions, and by extension, for securing world food security (see Chapter 3 to Chapter 6). The ultimate goal of developing VGI-based IPM is to revolutionize the traditional landscapes of IPM in order to maximize agricultural productivities. In VGI-based IPM, traditional information consumers are enabled to become information producers as well; experts are alleviated from data collection to focus on data analyses. A successfully implemented VGI-based IPM allows for the seamless integration of people's daily experiences with virtual services, communities, and databases.

In addition, this research provides empirical guidelines for the design of tools for the implementation of VGI-based IPM. For example, to assure more accurate data collections, incorporating a component into the tools that collects user confidence levels will be very helpful, as user confidence levels can account for their expertise (see Chapter 6). Change point analysis algorithm and hot spot analysis algorithm also can be easily integrated into the tools in order to achieve near-real time VGI sense making (see Chapter 4). For high level VGI-based IPM, enabling the tools to collect both quantitative and qualitative information is important, as this will stimulate users' cognitive abilities in pest management (see Section 4.5.2, and Section 7.1.4). In terms of information dissemination, this study suggests that information disseminators should put more effort on disseminating information useful to individual users, so as to attract more contributions from them (see Chapter 4).

In addition to the implications for the tool design, the specific sense making results of this research also have benefits for pest managerial investments. For both the tactical and strategic IPM, the sense making results provide guidelines for the allocations of limited pest managerial resources (see Chapter 4 and Chapter 5). This can make the pest managerial investment more economic, efficient, and effective.

Lastly, the VGI data quality assurance approach developed in this study may be generalizable. It has a potential to be adapted and applied to other VGI applications with similar structures but different contexts, e.g., VGI-based earthquake casualty surveillances (see Chapter 6). Therefore, this research also has practical implications for designing tools or methods for assuring the quality of VGI collected in different contexts.

7.3 Recommendations for future work

This research does not merely make significant contributions to the existing research and practices but also points to the directions for future studies to enhance VGI-based IPM further. Given the topic specificity of the key chapter of this study (Chapter 4, Chapter 5, and Chapter 6), limitations and recommendations for future research that emerged from the individual studies of those chapters have already been discussed earlier. The purpose of this section is to make recommendations for future research that are common to the overall research of this thesis. These concerns are related to: (1) the limited use of VGI; (2) cognitive abilities of participants in IPM; and (3) user privacy in data sharing.

7.3.1 The limited use of VGI

Due to the constrains from the current developmental status of VGI-based IPM, this study suffers from the limited uses of VGI. In Chapter 4, a facilitated VGI approach was used, where the participants participated with their own interest in crop protection without obtaining any monetary remuneration or material compensation. However, as mentioned in Section 2.1.2, facilitated VGI creation approach may have limited ability to truly foster the empowerment of marginalized communities, compared to active and passive VGI creations, due to its inherent constrains on participant abilities, data contribution freedom, and geographic extents. In Chapter 5, the study depicted that ecological niche modelling can be used to make sense of the VGI to manage pests strategically. The study was with a global scale, partially due to the sparseness of the existing pest occurrence records. If unlimited species occurrence records can be contributed, finer resolution (e.g., continental level, country level, or state level) ecological niche analyses can be conducted so as to better manage pests. Therefore, strengthening VGI-based IPM platforms (e.g., mobile phone platforms) to encourage drastic VGI creations is suggested as a future work, and hence big VGI data related to IPM are expected in the future. It has been revealed by this thesis that VGI sense making can indeed general beneficial knowledge for IPM, but more interesting and meaningful knowledge may be discovered through big data analyses. To handle big VGI data, the sense making methods adopted in this thesis can still be explored. However, a variety of big data sense making methods can also be explored with big data. For example, Steiger *et al.* (2016) proposed a geographic, hierarchical self-organizing map to analyze the spatial, temporal, and semantic characteristics of georeferenced tweet data. Zheng *et al.* (2012) exploited the concept of mobility entropy and Markov chain model to mine travel patterns from geotagged Flickr photos. Hagenauer and Helbich (2012) mined urban land-use patterns from OpenStreetMap using artificial neural networks and genetic algorithms. Although these VGI sense makings were not in the context of IPM, the data mining methods adopted in these studies could be insightful for the VGI sense making for enhancing IPM. In addition, for the fuzzy system of VGI quality assurance developed in Chapter 6, it will be necessary to test the system with big data as inputs to examine its performance. Therefore, collecting, storing, analyzing, and disseminating big VGI data will be an important future task.

7.3.2 Cognitive abilities of participants

Although this study explored VGI-based IPM, the information that the participants provided was only limited to location-based quantitative pest observation information and the related metadata. i.e., simple crowdsourcing in which cognitive ability is minimal.

In the future research, it will be necessary to explore the cognitive abilities of participants for VGI-based IPM based on the enhanced framework of VGI-based IPM proposed in Section 3.4. Higher level user participations require more effective consultations, deeper involvements and engagements, and closer collaborations in VGI-based IPM, in which participants' knowledge, perceptions, constraints, objectives, and their complex demands can be fully dealt with. Additionally, cognitive abilities can also be utilized to assure the quality of VGI. For example, discussion forums can be built to discuss the quality of various information contributed by others. Furthermore, referring back to Chapter 3, since VGI is diverse, various methodologies can be employed in VGI research, involving collection and analysis of either quantitative data, or qualitative data, quantitative-qualitative-combined data. Therefore, corresponding supporting methods, protocols, infrastructures, and strong guidance should be developed to fulfill VGI-based IPM that have more profound impacts than simple crowdsourcing of pest surveillances. Perhaps, learning the valuable experiences gained from farmer-field-school (see Section 2.2.1), or incorporating certain successful components of farmer-field-school into VGI-based IPM would be promising.

7.3.3 User privacy

A commonality shared among various VGI application domains is that participants' data are prepared for others' uses. This, of course, as all data on human subjects "inevitably raises privacy and security issues, and the real risks of abusing such data are difficult to quantify" (Boyd and Crawford, 2012). Roick and Heuser (2013) identified four specific categories of privacy threats: identity privacy; location privacy; absence privacy (e.g., not at home); and co-location privacy (e.g., inferring the location of a user according to the locations of the other users). The fuzzy system developed in Chapter 6 does not need to collect user particulars to assess user expertise, which reduces the concern with user identity privacy. However, the overall research in this thesis does not have any component to reduce the potential infringements of location privacy, absence privacy, and co-location privacy.

Privacy and security are always top issues in technology communities (Goranson *et al.*, 2013), and they are even amplified in VGI environments (Ricker *et al.*, 2014). Since there is no explicit data manager in VGI environments, individual information uploaded to a virtual space mostly become available to anyone and can be used for any purpose. The vast amount of information about individuals could even be leveraged by ill-disposed attempts. Although a VGI author can request administrators to delete his/her information permanently for preventing potential infringing acts, his/her personal information may have already been transmitted by other users before the deletion.

Considering the potential surveillance capability and intrusiveness of VGI, many people are reluctant to make their information available to the general public, as well as to governments (Goodchild, 2007a). For instance, some farmers may not be willing to reveal that their farms suffer more pest infestations than other farms do, due to certain commercial considerations. In this case, perhaps developing a VGI visualization method using relative relationships based on topological maps rather than absolute relations will reduce such concerns. Facilitated approaches to VGI may be able to avoid these concerns, because such approaches may not produce geospatial information that are accessible to the general public (Cinnamon and Schuurman, 2013). But when using facilitated VGI, one also needs to be cautious about potential risks of user information leakages. There is therefore a necessity to explore more robust ways to protect user privacy during data collection, analysis, and dissemination.

Bibliography

Adhikari, B. & Li, J. 2013. Modelling Ambiguity in Urban Planning. *Annals of GIS,* **19**, 143-152.

Al-kheder, S., Wang, J. & Shan, J. 2008. Fuzzy Inference Guided Cellular Automata Urban-Growth Modelling Using Multi-Temporal Satellite Images. *International Journal of Geographical Information Science,* **22**, 1271-1293.

Alamgir, M., Mukul, S. A. & Turton, S. M. 2015. Modelling Spatial Distribution of Critically Endangered Asian Elephant and Hoolock Gibbon in Bangladesh Forest Ecosystems under a Changing Climate. *Applied Geography,* **60**, 10-19.

Allen, J. C. 1976. A Modified Sine Wave Method for Calculating Degree Days. *Environmental Entomology,* **5**, 388-396.

Aragón, P., Coca-Abia, M. M., Llorente, V. & Lobo, J. M. 2013. Estimation of Climatic Favourable Areas for Locust Outbreaks in Spain: Integrating Species' Presence Records and Spatial Information on Outbreaks. *Journal of Applied Entomology,* **137**, 610-623.

Araújo, M. B. & Luoto, M. 2007. The Importance of Biotic Interactions for Modelling Species Distributions under Climate Change. *Global Ecology and Biogeography*, **16**, 743-753.

Araújo, M. B., Thuiller, W. & Pearson, R. G. 2006. Climate Warming and the Decline of Amphibians and Reptiles in Europe. *Journal of Biogeography,* **33**, 1712-1728.

Augustin, S., Boonham, N., De Kogel, W. J., Donner, P., Faccoli, M., Lees, D. C., Marini, L., Mori, N., Petrucco Toffolo, E., Quilici, S., Roques, A., Yart, A. & Battisti, A. 2012. A Review of Pest Surveillance Techniques for Detecting Quarantine Pests in Europe. *EPPO Bulletin,* **42**, 515-551.

Bakshy, E., Hofman, J. M., Mason, W. A. & Watts, D. J. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. *Proceedings of the fourth ACM international conference on Web search and data mining*, Hong Kong, China, 65-74.

Ballatore, A. & Bertolotto, M. 2011. Semantically Enriching Vgi in Support of Implicit Feedback Analysis. *In:* Tanaka, K., Fröhlich, P. & Kim, K.-S. (eds.) *Web and Wireless Geographical Information Systems.* Springer Berlin Heidelberg.

Barfield, C. S. & Swisher, M. E. 1994. Integrated Pest Management: Ready for Export? Historical Context and Internationalization of Ipm*. *Food Reviews International*, **10**, 215-267.

Barnett, A. J., Wiber, M. G., Rooney, M. P. & Curtis Maillet, D. G. 2016. The Role of Public Participation Gis (Ppgis) and Fishermen's Perceptions of Risk in Marine Debris Mitigation in the Bay of Fundy, Canada. *Ocean & Coastal Management,* **133**, 85-94.

Bebber, D. P., Holmes, T. & Gurr, S. J. 2014. The Global Spread of Crop Pests and Pathogens. *Global Ecology and Biogeography*, **23**, 1398-1407.

Bebber, D. P., Ramotowski, M. A. T. & Gurr, S. J. 2013. Crop Pests and Pathogens Move Polewards in a Warming World. *Nature Clim. Change*, **3**, 985-988.

Beddow, J. M., Kriticos, D., Pardey, P. G. & Sutherst, R. W. 2010. Potential Global Crop Pest Distributions Using Climex: Harvestchoice Applications. *HarvestChoice Working Paper*.

Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M. & Courchamp, F. 2013. Will Climate Change Promote Future Invasions? *Global Change Biology*, **19**, 3740-3748.

Bidinger, K., Lötters, S., Rödder, D. & Veith, M. 2012. Species Distribution Models for the Alien Invasive Asian Harlequin Ladybird (Harmonia Axyridis). *Journal of Applied Entomology*, **136**, 109-123.

Bishr, M. 2007. Weaving Space into the Web of Trust: An Asymmetric Spatial Trust Model for Social Networks. *Proceedings of the 1st Conference on Social Semantic Web*, Leipzig, Germany, 35-46.

Bishr, M. & Janowicz, K. 2010. Can We Trust Information?-the Case of Volunteered Geographic Information. *Proceedings of the Workshop "Towards Digital Earth: Search, Discover and Share Geospatial Data" at Future Internet Symposium*, Berlin, Germany, 11-16.

Bishr, M. & Mantelas, L. 2008. A Trust and Reputation Model for Filtering and Classifying Knowledge About Urban Growth. *GeoJournal,* **72**, 229-237.

Blatt, A. J. 2013. Technological Changes in Maps and Cartography. *Journal of Map & Geography Libraries,* **9**, 361-367.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M. & Rampini, A. 2014a. A Linguistic Decision Making Approach to Assess the Quality of Volunteer Geographic Information for Citizen Science. *Information Sciences,* **258**, 312-327.

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M. & Rampini, A. 2014b. On Predicting and Improving the Quality of Volunteer Geographic Information Projects. *International Journal of Digital Earth,* **9**, 134-155.

Boyd, D. & Crawford, K. 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society,* **15**, 662-679.

Bradley, B. A. 2009. Regional Analysis of the Impacts of Climate Change on Cheatgrass Invasion Shows Potential Risk and Opportunity. *Global Change Biology,* **15**, 196-208.

Brando, C. & Bucher, B. 2010. Quality in User Generated Spatial Content: A Matter of Specifications. *Proceedings of the 13th AGILE International Conference on Geographic Information Science,* Guimarães, Portugal, 11-14.

Brown, G. 2012. Public Participation Gis (Ppgis) for Regional and Environmental Planning: Reflections on a Decade of Empirical Research. *Journal of Urban and Regional Information Systems Association*, **25**, 7-18.

Brown, G., Kelly, M. & Whitall, D. 2014. Which 'Public'? Sampling Effects in Public Participation Gis (Ppgis) and Volunteered Geographic Information (Vgi) Systems for Public Lands Management. *Journal of Environmental Planning and Management*, **57**, 190-214.

Brown, G. & Kyttä, M. 2014. Key Issues and Research Priorities for Public Participation Gis (Ppgis): A Synthesis Based on Empirical Research. *Applied Geography*, **46**, 122-136.

Bruce, E., Albright, L., Sheehan, S. & Blewitt, M. 2014. Distribution Patterns of Migrating Humpback Whales (*Megaptera Novaeangliae*) in Jervis Bay, Australia: A Spatial Analysis Using Geographical Citizen Science Data. *Applied Geography*, **54**, 83-95.

Budhathoki, N. R. 2010. *Participants' Motivations to Contribute Geographic Information in an Online Community*. PhD Thesis, University of Illinois at Urbana-Champaign.

CABI. 2016. *Invasive Species Compendium* [Online]. Available: http://www.cabi.org/isc [Accessed 16 March 2016].

Caha, J., Tuček, P., Vondráková, A. & Paclíková, L. 2012. Slope Analysis of Fuzzy Surfaces. *Transactions in GIS,* **16**, 649-661.

Cannon, R. J. C. 1998. The Implications of Predicted Climate Change for Insect Pests in the Uk, with Emphasis on Non-Indigenous Species. *Global Change Biology,* **4**, 785-796.

Carnevale, C., Finzi, G., Pisoni, E. & Volta, M. 2009. Neuro-Fuzzy and Neural Network Systems for Air Quality Control. *Atmospheric Environment,* **43**, 4811-4821.

Cartwright, W. 2009. Small Voices Magnified: Using Web 2.0 for Mapping Alternative Australian Viewpoints. *Proceedings of the 2009 Surveying & Spatial Sciences Institute Biennial International Conference*, Adelaide, SA, Australia, 469-480.

Castells, M. 2000. Toward a Sociology of the Network Society. *Contemporary Sociology,* **29**, 693-699.

Celino, I. 2013. Human Computation Vgi Provenance: Semantic Web-Based Representation and Publishing. *Geoscience and Remote Sensing, IEEE Transactions on,* **51**, 5137-5144.

Chang, E., Thomson, P., Dillon, T. & Hussain, F. 2005. The Fuzzy and Dynamic Nature of Trust. *In:* Katsikas, S., López, J. & Pernul, G. (eds.) *Trust, Privacy, and Security in Digital Business: Second International Conference, Trustbus 2005,* *Copenhagen, Denmark, August 22-26, 2005. Proceedings.* Springer Berlin Heidelberg.

Chee Wei, P., Sutanto, J., Kankanhalli, A., Yan, L., Tan, B. C. Y. & Hock-Hai, T. 2006. Senior Citizens' Acceptance of Information Systems: A Study in the Context of E-Government Services. *Engineering Management, IEEE Transactions on,* **53**, 555-569.

Chen, I. Y. 2007. The Factors Influencing Members' Continuance Intentions in Professional Virtual Communities-a Longitudinal Study. *Journal of Information Science,* **33**, 451-467.

Cinderby, S. & Forrester, J. 2005. Facilitating the Local Governance of Air Pollution Using Gis for Participation. *Applied Geography,* **25**, 143-158.

Cinnamon, J. & Schuurman, N. 2013. Confronting the Data-Divide in a Time of Spatial Turns and Volunteered Geographic Information. *GeoJournal,* **78**, 657-674.

Cofta, P. 2007. Confidence, Trust and Identity. *BT Technology Journal,* **25**, 173-178.

Coleman, D. J., Georgiadou, Y. & Labonte, J. 2009. Volunteered Geographic Information: The Nature and Motivation of Produsers. *International Journal of Spatial Data Infrastructures Research,* **4**, 332-358.

Coote, A. & Rackham, L. 2008. Neogeographic Data Quality–Is It an Issue. *Annual Conference of the Association for Geographic Information*, Stratford-upon-Avon, UK, 1-17.

COP21. 2015. *2015 United Nations Climate Change Conference* [Online]. Available: http://www.cop21.gouv.fr/en/ [Accessed 10 October 2016].

Craglia, M., Ostermann, F. & Spinsanti, L. 2012. Digital Earth from Vision to Practice: Making Sense of Citizen-Generated Content. *International Journal of Digital Earth*, **5**, 398-416.

Cupchik, G. 2001. Constructivist Realism: An Ontology That Encompasses Positivist and Constructivist Approaches to the Social Sciences. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, **2**, Art. 7.

De Longueville, B., Annoni, A., Schade, S., Ostlaender, N. & Whitmore, C. 2010. Digital Earth's Nervous System for Crisis Events: Real-Time Sensor Web Enablement of Volunteered Geographic Information. *International Journal of Digital Earth*, **3**, 242-259.

Deng, Y., Chang, K., Sha, X. & Tan, B. 2012. Mobile Ict and Knowledge Sharing in Underserved Communities. *AMCIS 2012 Proceedings*, Seattle, Washington, USA, 1-10.

Deng, Y. & Chang, K. T. 2012. A Design Framework for Event Recommendation in Novice Low-Literacy Communities. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, **6**, 999 - 1004.

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C. & Martin, P. R. 2008. Impacts of Climate Warming on Terrestrial Ectotherms across Latitude. *Proceedings of the National Academy of Sciences,* **105**, 6668-6672.

Devillers, R., Bédard, Y. & Jeansoulin, R. 2005. Multidimensional Management of Geospatial Data Quality Information for Its Dynamic Use within Gis. *Photogrammetric Engineering & Remote Sensing,* **71**, 205-215.

Devillers, R. & Jeansoulin, R. 2010. Spatial Data Quality: Concepts. *In:* Devillers, R. & Jeansoulin, R. (eds.) *Fundamentals of Spatial Data Quality.* ISTE.

Dlott, J., Altieri, M. & Masumoto, M. 1994. Exploring the Theory and Practice of Participatory Research in Us Sustainable Agriculture: A Case Study in Insect Pest Management. *Agriculture and Human Values,* **11**, 126-139.

Douthwaite, B., Kuby, T., van de Fliert, E. & Schulz, S. 2003. Impact Pathway Evaluation: An Approach for Achieving and Attributing Impact in Complex Systems. *Agricultural Systems,* **78**, 243-265.

Dueker, K. J. & Kjerne, D. 1989. *Multipurpose Cadastre: Terms and Definitions*, American Society for Photogrammetry and Remote Sensing.

Dunn, C. E. 2007. Participatory Gis—a People's Gis? *Progress in Human Geography,* **31**, 616-637.

Elith, J. 2015. *Predicting Distributions of Invasive Species* [Online]. Available: http://arxiv.org/abs/1312.0851 [Accessed 10 January 2016].

Elwood, S. 2009. Multiple Representations, Significations and Epistemologies in Community-Based Gis. *In:* Cope, M. & Elwood, S. (eds.) *Qualitative Gis: A Mixed Methods Approach.* Sage.

Elwood, S., Goodchild, M. F. & Sui, D. Z. 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers,* **102**, 571-590.

Estay, S. A., Lima, M. & Labra, F. A. 2009. Predicting Insect Pest Status under Climate Change Scenarios: Combining Experimental Data and Population Dynamics Modelling. *Journal of Applied Entomology,* **133**, 491-499.

Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, USA, 226-231.

Fand, B. B., Choudhary, J. S., Kumar, M. & Bal, S. K. 2014. Phenology Modelling and Gis Applications in Pest Management: A Tool for Studying and Understanding Insect-Pest Dynamics in the Context of Global Climate Change. *In:* Gaur, R. K. & Sharma, P. (eds.) *Approaches to Plant Stress and Their Management.* Springer India.

Flood, J. 2010. The Importance of Plant Health to Food Security. *Food Security,* **2**, 215-231.

Foody, G. M. & Boyd, D. S. 2012. Using Volunteered Data in Land Cover Map Validation: Mapping Tropical Forests across West Africa. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, Munich, Germany, 6207-6208.

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C. & Boyd, D. S. 2013. Assessing the Accuracy of Volunteered Geographic Information Arising from Multiple Contributors to an Internet Based Collaborative Project. *Transactions in GIS,* **17**, 847-860.

Fowler, A., Whyatt, J. D., Davies, G. & Ellis, R. 2013. How Reliable Are Citizen-Derived Scientific Data? Assessing the Quality of Contrail Observations Made by the General Public. *Transactions in GIS,* **17**, 488-506.

Frančula, N. 2011. Public Participation Geographic Information Systems. *Kartografija i Geoinformacije (Cartography and Geoinformation),* **10**, 178.

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E. & Thuiller, W. 2012. Invasive Species Distribution Models – How Violating the Equilibrium Assumption Can Create New Insights. *Global Ecology and Biogeography,* **21**, 1126-1136.

Gao, S., Li, L., Li, W., Janowicz, K. & Zhang, Y. in press. Constructing Gazetteers from Volunteered Big Geo-Data Based on Hadoop. *Computers, Environment and Urban Systems*.

Garcia, K., Lasco, R., Ines, A., Lyon, B. & Pulhin, F. 2013. Predicting Geographic Distribution and Habitat Suitability Due to Climate Change of Selected Threatened Forest Tree Species in the Philippines. *Applied Geography,* **44**, 12-22.

Ghose, R. 2001. Use of Information Technology for Community Empowerment: Transforming Geographic Information Fystems into Community Information Systems. *Transactions in GIS,* **5**, 141-163.

Ghose, R. & Elwood, S. 2003. Public Participation Gis and Local Political Context: Propositions and Research Directions. *URISA Journal,* **15**, 17-22.

Girres, J.-F. & Touya, G. 2010. Quality Assessment of the French Openstreetmap Dataset. *Transactions in GIS,* **14**, 435-459.

Goodchild, M. 2009. Neogeography and the Nature of Geographic Expertise. *Journal of Location Based Services,* **3**, 82-96.

Goodchild, M. F. 2007a. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal,* **69**, 211-221.

Goodchild, M. F. 2007b. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0 *International Journal of Spatial Data Infrastructures Research,* **2**, 24-32.

Goodchild, M. F. 2008. Commentary: Whither Vgi? *GeoJournal,* **72**, 239-244.

Goodchild, M. F. & Li, L. 2012. Assuring the Quality of Volunteered Geographic Information. *Spatial Statistics,* **1**, 110-120.

Goranson, C., Thihalolipavan, S. & di Tada, N. 2013. Vgi and Public Health: Possibilities and Pitfalls. *In:* Sui, D., Elwood, S. & Goodchild, M. (eds.) *Crowdsourcing Geographic Knowledge.* Springer Netherlands.

Gouveia, C. & Fonseca, A. 2008. New Approaches to Environmental Monitoring: The Use of Ict to Explore Volunteered Geographic Information. *GeoJournal,* **72**, 185-197.

Gumpertz, L., Wu & Pye, M. 2000. Logistic Regression for Southern Pine Beetle Outbreaks with Spatial and Temporal Autocorrelation. *Forest Science,* **46**, 95-107.

Gutierrez, A. P., Ponti, L., d'Oultremont, T. & Ellis, C. K. 2007. Climate Change Effects on Poikilotherm Tritrophic Interactions. *Climatic Change,* **87**, 167-192.

Haenlein, M. & Kaplan, A. M. 2004. A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics,* **3**, 283-297.

Hagenauer, J. & Helbich, M. 2012. Mining Urban Land-Use Patterns from Volunteered Geographic Information by Means of Genetic Algorithms and Artificial Neural Networks. *International Journal of Geographical Information Science,* **26**, 963-982.

Haklay, M. 2010. How Good Is Volunteered Geographical Information? A Comparative Study of Openstreetmap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design,* **37**, 682-703.

Haklay, M. 2013. Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. *In:* Sui, D., Elwood, S. & Goodchild, M. (eds.) *Crowdsourcing Geographic Knowledge.* Springer Netherlands.

Haklay, M., Basiouka, S., Antoniou, V. & Ather, A. 2010. How Many Volunteers Does It Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal,* **47**, 315-322.

Haklay, M., Singleton, A. & Parker, C. 2008. Web Mapping 2.0: The Neogeography of the Geoweb. *Geography Compass,* **2**, 2011-2039.

Hall, H. & Graham, D. 2004. Creation and Recreation: Motivating Collaboration to Generate Knowledge Capital in Online Communities. *International Journal of Information Management,* **24**, 235-246.

Han, S. 2012. *Web 2.0,* Routledge.

Harris, T. & Weiner, D. 1998. Empowerment, Marginalization, and "Community-Integrated" Gis. *Cartography and Geographic Information Systems,* **25**, 67-76.

Hartley, S., Krushelnycky, P. D. & Lester, P. J. 2010. Integrating Physiology, Population Dynamics and Climate to Make Multi-Scale Predictions for the Spread of an Invasive Insect: The Argentine Ant at Haleakala National Park, Hawaii. *Ecography,* **33**, 83-94.

Heinrichs, E. A. & Mochida, O. 1984. From Secondary to Major Pest Status: The Case of Insecticide-Induced Rice Brown Planthopper, Nilaparvata Lugens, Resurgence. *Protection Ecology,* **7**, 201-218.

Herms, D. A. 2004. Using Degree-Days and Plant Phenology to Predict Pest Activity. *Ipm (Integrated Pest Management) of Midwest Landscapes.* Minnesota Agricultural Experiment Station Publication.

Higginbottom, G. & Lauridsen, E. I. 2014. The Roots and Development of Constructivist Grounded Theory. *Nurse Researcher,* **21**, 8-13.

Hill, M. P., Chown, S. L. & Hoffmann, A. A. 2013. A Predicted Niche Shift Corresponds with Increased Thermal Resistance in an Invasive Mite, Halotydeus Destructor. *Global Ecology and Biogeography,* **22**, 942-951.

Hoddle, M. 2014. *Frequently Asked Questions About Invasive Species* [Online]. Available: https://cisr.ucr.edu/invasive_species_faqs.html [Accessed 17 January 2016].

Hongoh, V., Berrang-Ford, L., Scott, M. E. & Lindsay, L. R. 2012. Expanding Geographical Distribution of the Mosquito, Culex Pipiens, in Canada under Climate Change. *Applied Geography,* **33**, 53-62.

Houghton, J. 2005. Global Warming. *Reports on Progress in Physics,* **68**, 1343-1403.

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M. & Meinke, H. 2007. Adapting Agriculture to Climate Change. *Proceedings of the National Academy of Sciences,* **104**, 19691-19696.

Hoyt, L., Khosla, R. & Canepa, C. 2005. Leaves, Pebbles, and Chalk: Building a Public Participation Gis in New Delhi, India. *Journal of Urban Technology,* **12**, 1-19.

IPCC 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press.

ISO-19113 2002. *Geographic Information -- Quality Principles*, Technical report, International Organization for Standardization.

Jepsen, J. U., Hagen, S. B., Ims, R. A. & Yoccoz, N. G. 2008. Climate Change and Outbreaks of the Geometrids Operophtera Brumata and Epirrita Autumnata in Subarctic Birch Forest: Evidence of a Recent Outbreak Range Expansion. *Journal of Animal Ecology,* **77**, 257-264.

Jeschke, J. M. & Strayer, D. L. 2008. Usefulness of Bioclimatic Models for Studying Climate Change and Invasive Species. *Annals of the New York Academy of Sciences,* **1134**, 1-24.

Jiao, X.-G., Xuan, W.-J. & Sheng, C.-F. 2006. A Forecasting Model for Emergence and Flight Pattern of the Overwintering Generation of *Chilo Suppressalis* (Lepidoptera: Pyralidae) Based on Pheromone Trapcatches and Degree-Days in Northeastern China. *Acta Entomologica Sinica,* **49**, 705 - 709.

Jin, X.-L., Cheung, C. M. K., Lee, M. K. O. & Chen, H.-P. 2009. How to Keep Members Using the Information in a Computer-Supported Social Network. *Computers in Human Behavior,* **25**, 1172-1181.

Johnson, R. B. & Onwuegbuzie, A. J. 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher,* **33**, 14-26.

Jones, C. E. & Weber, P. 2012. Towards Usability Engineering for Online Editors of Volunteered Geographic Information: A Perspective on Learnability. *Transactions in GIS,* **16**, 523-544.

Jones, V. P., Doerr, M. & Brunner, J. F. 2008. Is Biofix Necessary for Predicting Codling Moth (Lepidoptera: Tortricidae) Emergence in Washington State Apple Orchards? *Journal of Economic Entomology,* **101**, 1651-1657.

Keßler, C., Janowicz, K. & Bishr, M. 2009a. An Agenda for the Next Generation Gazetteer: Geographic Information Contribution and Retrieval. *Proceedings of* *the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* Seattle, Washington, USA, 91-100.

Keßler, C., Maué, P., Heuer, J. & Bartoschek, T. 2009b. Bottom-up Gazetteers: Learning from the Implicit Semantics of Geotags. *In:* Janowicz, K., Raubal, M. & Levashkin, S. (eds.) *Geospatial Semantics.* Springer Berlin Heidelberg.

Kim, H.-W., Chan, H. C. & Kankanhalli, A. 2012. What Motivates People to Purchase Digital Items on Virtual Community Websites? The Desire for Online Self-Presentation. *Information Systems Research,* **23**, 1232-1245.

Kobori, Y. & Amano, H. 2003. Effect of Rainfall on a Population of the Diamondback Moth, *Plutella Xylostella* (Lepidoptera: Plutellidae). *Applied Entomology and Zoology,* **38**, 249-253.

Kogan, M. 1998. Integrated Pest Management: Historical Perspectives and Contemporary Developments. *Annual Review of Entomology,* **43**, 243-270.

Kriticos, D. J., Morin, L., Leriche, A., Anderson, R. C. & Caley, P. 2013. Combining a Climatic Niche Model of an Invasive Fungus with Its Host Species Distributions to Identify Risks to Natural Assets: Puccinia Psidii Sensu Lato in Australia. *PLoS ONE,* **8**, e64479.

Kuhn, W. 2007. Volunteered Geographic Information and Giscience. *NCGIA and Vespucci Workshop on Volunteered Geographic Information*, Santa Barbara, CA, 13-14.

Kwan, M.-P. 2002. Is Gis for Women? Reflections on the Critical Discourse in the 1990s. *Gender, Place and Culture: A Journal of Feminist Geography,* **9**, 271-279.

Lam, W.-K. F., Pedigo, L. P. & Hinz, P. N. 2001. Population Dynamics of Bean Leaf Beetles (Coleoptera: Chrysomelidae) in Central Iowa. *Environmental Entomology,* **30**, 562-567.

Langley, S. A., Messina, J. P. & Grady, S. C. 2013. Utilizing Volunteered Information for Infectious Disease Surveillance. *International Journal of Applied Geospatial Research (IJAGR),* **4**, 54-70.

Larson, M. G. 2008. Analysis of Variance. *Circulation,* **117**, 115-121.

Lewis, W. J., van Lenteren, J. C., Phatak, S. C. & Tumlinson, J. H. 1997. A Total System Approach to Sustainable Pest Management. *Proceedings of the National Academy of Sciences,* **94**, 12243-12248.

Li, L. & Goodchild, M. F. 2012. The Role of Social Networks in Emergency Management: A Research Agenda. *In:* Jennex, M. E. (ed.) *Managing Crises and Disasters with Emerging Technologies: Advancements.* IGI Global.

Lin, Y.-W. 2011. A Qualitative Enquiry into Openstreetmap Making. *New Review of Hypermedia and Multimedia,* **17**, 53-71.

Litsinger, J., Libetario, E. & Canapi, B. 2009. Eliciting Farmer Knowledge, Attitudes, and Practices in the Development of Integrated Pest Management Programs for

Rice in Asia. *In:* Peshin, R. & Dhawan, A. (eds.) *Integrated Pest Management: Dissemination and Impact.* Springer Netherlands.

Liu, Y., Yuan, Y., Xiao, D., Zhang, Y. & Hu, J. 2010. A Point-Set-Based Approximation for Areal Objects: A Case Study of Representing Localities. *Computers, Environment and Urban Systems*, **34**, 28-39.

Luther, G. C., Harris, C., Sherwood, S., Gallagher, K., Mangan, J. & Gamby, K. T. 2005. Developments and Innovations in Farmer Field Schools and the Training of Trainers. *In:* Norton, G. W. (ed.) *Globalizing Integrated Pest Management: A Participatory Research Process.* Blackwell Publishing.

Mamdani, E. H. 1974. Application of Fuzzy Algorithms for Control of Simple Dynamic Plant. *Proceedings of the Institution of Electrical Engineers*, **121**, 1585-1588.

Manca, G. & Curtin, K. 2012. Fuzzy Analysis for Modeling Regional Delineation and Development: The Case of the Sardinian Mining Geopark. *Transactions in GIS*, **16**, 55-79.

Mantoura, P. & Potvin, L. 2013. A Realist–Constructionist Perspective on Participatory Research in Health Promotion. *Health Promotion International*, **28**, 61-72.

Manuel, C. 1996. *The Rise of the Network Society*, John Wiley & Sons.

Martella, R., Kray, C. & Clementini, E. 2015. A Gamification Framework for Volunteered Geographic Information. *In:* Bacao, F., Santos, Y. M. & Painho,

M. (eds.) *Agile 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities.* Springer International Publishing.

Mashhadi, A. J. & Capra, L. 2011. Quality Control for Real-Time Ubiquitous Crowdsourcing. *Proceedings of the 2nd International Workshop on Ubiquitous Crowdsouring*, Beijing, China, 5-8.

Matyas, S., Kiefer, P., Schlieder, C. & Kleyer, S. 2011. Wisdom About the Crowd: Assuring Geospatial Data Quality Collected in Location-Based Games. *In:* Anacleto, J. C., Fels, S., Graham, N., Kapralos, B., Saif El-Nasr, M. & Stanley, K. (eds.) *Entertainment Computing – Icec 2011.* Springer Berlin Heidelberg.

Maué, P. & Schade, S. 2008. Quality of Geographic Information Patchworks. *Proceedings of the 11th AGILE International Conference on Geographic Information Science*, Girona, Spain, 1-8.

Maxmen, A. 2013. Crop Pests: Under Attack. *Nature*, **501**, S15-S17.

McCall, M. K. 2008. *Participatory Mapping and Participatory Gis (Pgis) for Cra, Community Drr and Hazard Assessment*, ITC Enschede Netherlands

McMaugh, T. 2005. *Guidelines for Surveillance for Plant Pests in Asia and the Pacific*, Australian Centre for International Agricultural Research Canberra.

Mertens, D. M. 1999. Inclusive Evaluation: Implications of Transformative Theory for Evaluation. *American Journal of Evaluation*, **20**, 1-14.

Mertens, D. M. 2007. Transformative Paradigm Mixed Methods and Social Justice. *Journal of Mixed Methods Research,* **1**, 212-225.

Mertens, D. M., Holmes, H. & Harris, R. L. 2009. Transformative Research and Ethics. *In:* Mertens, D. M. & Ginsberg, P. E. (eds.) *Handbook of Social Research Ethics.* Sage.

Michanowicz, D., Malone, S., Kelso, M., Ferrar, K., Christen, C. & Volz, C. D. 2012. A Participatory Geographic Information System (Pgis) Utilizing the Geoweb 2.0: Filling the Gaps of the Marcellus Shale Natural Gas Industry. *Journal of Systemics, Cybernetics and Informatics,* **10**, 45-53.

Miller, H. & Goodchild, M. 2015. Data-Driven Geography. *GeoJournal,* **80**, 449-461.

Moffett, A., Shackelford, N. & Sarkar, S. 2007. Malaria in Africa: Vector Species' Niche Models and Relative Risk Maps. *PLoS One,* **2**, e824.

Mokany, K. & Ferrier, S. 2011. Predicting Impacts of Climate Change on Biodiversity: A Role for Semi-Mechanistic Community-Level Modelling. *Diversity and Distributions,* **17**, 374-380.

Morin, X. & Thuiller, W. 2009. Comparing Niche- and Process-Based Models to Reduce Prediction Uncertainty in Species Range Shifts under Climate Change. *Ecology,* **90**, 1301-1313.

Morse, S. & Buhler, W. 1997. Ipm in Developing Countries: The Danger of an Ideal. *Integrated Pest Management Reviews,* **2**, 175-185.

Negnevitsky, M. 2005. *Artificial Intelligence: A Guide to Intelligent Systems (Second Edition),* England, Pearson Education.

Neis, P. & Zipf, A. 2012. Analyzing the Contributor Activity of a Volunteered Geographic Information Project — the Case of Openstreetmap. *ISPRS International Journal of Geo-Information,* **1**, 146.

Netherer, S. & Schopf, A. 2010. Potential Effects of Climate Change on Insect Herbivores in European Forests—General Aspects and the Pine Processionary Moth as Specific Example. *Forest Ecology and Management,* **259**, 831-838.

Norton, G. W., Rajotte, E. G. & Luther, G. C. 2005. The Paricipatory Integrated Pest Management (Pipm) Process. *In:* Norton, G. W., Heinrichs, E. A., Luther, G. C. & Irwin, M. E. (eds.) *Globalizing Integrated Pest Management: A Participatory Research Process.* Blackwell Publishing.

O'reilly, T. 2007. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, Sebastopol, CA, USA, 17-37.

Obermeyer, N. J. 1998. The Evolution of Public Participation Gis. *Cartography and Geographic Information Systems,* **25**, 65-66.

Ord, J. K. & Getis, A. 1995. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, **27**, 286-306.

Ostermann, F. O. & Spinsanti, L. 2011. A Conceptual Workflow for Automatically Assessing the Quality of Volunteered Geographic Information for Crisis Management. *Proceedings of the 14th AGILE International Conference on Geographic Information Science*, Utrecht, The Netherlands, 1-6.

Panteras, G., Wise, S., Lu, X., Croitoru, A., Crooks, A. & Stefanidis, A. 2015. Triangulating Social Multimedia Content for Event Localization Using Flickr and Twitter. *Transactions in GIS,* **19**, 694-715.

Parker, C. J., May, A. & Mitchell, V. 2013. The Role of Vgi and Pgi in Supporting Outdoor Activities. *Applied ergonomics,* **44**, 886-894.

Pasiecznik, N. M., Smith, I. M., Watson, G. W., Brunt, A. A., Ritchie, B. & Charles, L. M. F. 2005. Cabi/Eppo Distribution Maps of Plant Pests and Plant Diseases and Their Important Role in Plant Quarantine. *Eppo Bulletin,* **35**, 1-7.

Pavlovskaya, M. 2009. Non-Quantitative Gis. *In:* Cope, M. & Elwood, S. (eds.) *Qualitative Gis: A Mixed Methods Approach.* Replika Press Pvt Ltd.

Pearson, R. G. & Dawson, T. P. 2003. Predicting the Impacts of Climate Change on the Distribution of Species: Are Bioclimate Envelope Models Useful? *Global Ecology and Biogeography,* **12**, 361-371.

Peshin, R., Bandral, R., Zhang, W., Wilson, L. & Dhawan, A. 2009a. Integrated Pest Management: A Global Overview of History, Programs and Adoption. *In:*

Peshin, R. & Dhawan, A. (eds.) *Integrated Pest Management: Innovation-Development Process.* Springer Netherlands.

Peshin, R. & Dhawan, A. K. 2009a. *Integrated Pest Management: Dissemination and Impact*, Springer Netherlands.

Peshin, R. & Dhawan, A. K. 2009b. *Integrated Pest Management: Innovation-Development Process*, Springer Netherlands.

Peshin, R., Vasanthakumar, J. & Kalra, R. 2009b. Diffusion of Innovation Theory and Integrated Pest Management. *In:* Peshin, R. & Dhawan, A. (eds.) *Integrated Pest Management: Dissemination and Impact.* Springer Netherlands.

Peterson, D. K. & Pitz, G. F. 1988. Confidence, Uncertainty, and the Use of Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14**, 85-92.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum Entropy Modeling of Species Geographic Distributions. *Ecological Modelling,* **190**, 231-259.

Pilkington, L. J. & Hoddle, M. S. 2006. Use of Life Table Statistics and Degree-Day Values to Predict the Invasion Success of Gonatocerus Ashmeadi (Hymenoptera: Mymaridae), an Egg Parasitoid of Homalodisca Coagulata (Hemiptera: Cicadellidae), in California. *Biological Control,* **37**, 276-283.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. 2000. Environmental and Economic Costs of Nonindigenous Species in the United States. *BioScience,* **50**, 53-65.

Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V. & Justice, C. O. 2010. Estimating Global Cropland Extent with Multi-Year Modis Data. *Remote Sensing*, **2**, 1844-1863.

Popp, J., Pető, K. & Nagy, J. 2012. Pesticide Productivity and Food Security. A Review. *Agronomy for Sustainable Development*, **33**, 243-255.

Poser, K. & Dransch, D. 2010. Volunteered Geographic Information for Disaster Management with Application to Rapid Flood Damage Estimation. *Geomatica,* **64**, 89-98.

Pound, B., Snapp, S., McDougall, C. & Braun, A. 2003. *Managing Natural Resources for Sustainable Livelihoods: Uniting Science and Participation*, London: Earthscan Publications.

Power, C., Simms, A. & White, R. 2001. Hierarchical Fuzzy Pattern Matching for the Regional Comparison of Land Use Maps. *International Journal of Geographical Information Science,* **15**, 77-100.

Pretty, J. N. 1995. Participatory Learning for Sustainable Agriculture. *World Development*, **23**, 1247-1263.

Pulford, B. D. 1996. *Overconfidence in Human Judgement.* PhD Thesis, University of Leicester.

Pultar, E., Raubal, M., Cova, T. J. & Goodchild, M. F. 2009. Dynamic Gis Case Studies: Wildfire Evacuation and Volunteered Geographic Information. *Transactions in GIS,* **13**, 85-104.

Ricker, B., Schuurman, N. & Kessler, F. 2014. Implications of Smartphone Usage on Privacy and Spatial Cognition: Academic Literature and Public Perceptions. *GeoJournal,* **80**, 637-652.

Roick, O. & Heuser, S. 2013. Location Based Social Networks – Definition, Current State of the Art and Research Agenda. *Transactions in GIS,* **17**, 763-784.

Salk, C. F., Sturn, T., See, L., Fritz, S. & Perger, C. 2015. Assessing Quality of Volunteer Crowdsourcing Contributions: Lessons from the Cropland Capture Game. *International Journal of Digital Earth,* **9**, 410-426.

Scassa, T. 2013. Legal Issues with Volunteered Geographic Information. *The Canadian Geographer/Le Géographe canadien,* **57**, 1-10.

Schoepfer, I. & Rogers, S. R. 2014. A New Qualitative Gis Method for Investigating Neighbourhood Characteristics Using a Tablet. *Cartographica: The International Journal for Geographic Information and Geovisualization,* **49**, 127-143.

Schuurman, N. & Pratt, G. 2002. Care of the Subject: Feminism and Critiques of Gis. *Gender, Place & Culture,* **9**, 291-299.

Seeger, C. J. 2008. The Role of Facilitated Volunteered Geographic Information in the Landscape Planning and Site Design Process. *GeoJournal*, **72**, 199-213.

Sheppard, E. 2005. Knowledge Production through Critical Gis: Genealogy and Prospects. *Cartographica: The International Journal for Geographic Information and Geovisualization*, **40**, 5-21.

Sieber, R. 2006. Public Participation Geographic Information Systems: A Literature Review and Framework. *Annals of the Association of American Geographers,* **96**, 491-507.

Sniezek, J. A. & Buckley, T. 1995. Cueing and Cognitive Conflict in Judge-Advisor Decision Making. *Organizational Behavior and Human Decision Processes,* **62**, 159-174.

Sniezek, J. A. & Van Swol, L. M. 2001. Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes,* **84**, 288-307.

Song, S., Hwang, K. & Macwan, M. 2004. Fuzzy Trust Integration for Security Enforcement in Grid Computing. *In:* Jin, H., Gao, G. R., Xu, Z. & Chen, H. (eds.) *Network and Parallel Computing: Ifip International Conference, Npc 2004, Wuhan, China, October 18-20, 2004. Proceedings.* Springer Berlin Heidelberg.

Song, W. & Sun, G. 2010. The Role of Mobile Volunteered Geographic Information in Urban Management. *18th International Conference on Geoinformatics*, Beijing, China, 1-5.

Stark, H., Car, A., Griesebner, G. & Strobl, J. 2011. Quality Assessment of Volunteered Geographic Information Using Open Web Map Services within Openaddresses. *Proceedings of the Geoinformatics Forum Salzburg*, Salzburg, Austria, 101-110.

Stefanidis, A., Crooks, A. & Radzikowski, J. 2012. Harvesting Ambient Geospatial Information from Social Media Feeds. *GeoJournal,* **78**, 319-338.

Steiger, E., Resch, B. & Zipf, A. 2016. Exploration of Spatiotemporal and Semantic Clusters of Twitter Data Using Unsupervised Neural Networks. *International Journal of Geographical Information Science,* **30**, 1694-1716.

Steinmann, R., Gröchenig, S., Rehrl, K. & Brunauer, R. 2013. Contribution Profiles of Voluntary Mappers in Openstreetmap. *Proceedings of Action and Interaction in Volunteered Geographic Information (ACTIVITY) Workshop at AGILE*, Leuven, Belgium, 1-6.

Stewart, E. J., Jacobson, D. A. N. & Draper, D. 2008. Public Participation Geographic Information Systems (Ppgis): Challenges of Implementation in Churchill, Manitoba. *Canadian Geographer / Le Géographe canadien,* **52**, 351-366.

Suen, R. C. L., Chang, K. T. T., Wan, M. P.-H., Ng, Y. C. & Tan, B. C. Y. 2014. Interactive Experiences Designed for Agricultural Communities. *CHI'14*

Extended Abstracts on Human Factors in Computing Systems, Toronto, ON, Canada, 551-554.

Sweetman, D., Badiee, M. & Creswell, J. W. 2010. Use of the Transformative Framework in Mixed Methods Studies. *Qualitative Inquiry*, **16**, 441-454.

Swets, J. 1988. Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285-1293.

Tait, J. & Napompeth, B. 1987. *Management of Pests and Pesticides: Farmers' Perceptions and Practices*, London, Westview Press.

Taylor, W. 2014. *Change-Point Analysis: A Powerful New Tool for Detecting Changes* [Online]. Available: http://www.variation.com/cpa/tech/changepoint.html [Accessed 29 April 2015].

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L. & Williams, S. E. 2004. Extinction Risk from Climate Change. *Nature*, **427**, 145-148.

Thompson, W. 2013. *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters*, Island Press.

Thomson, L. J., Macfadyen, S. & Hoffmann, A. A. 2010. Predicting the Effects of Climate Change on Natural Enemies of Agricultural Pests. *Biological Control*, **52**, 296-306.

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. 2005. Climate Change Threats to Plant Diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245-8250.

Tingley, M. W., Monahan, W. B., Beissinger, S. R. & Moritz, C. 2009. Birds Track Their Grinnellian Niche through a Century of Climate Change. *Proceedings of the National Academy of Sciences*, **106**, 19637-19643.

Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240.

Trame, J. & Keßler, C. 2011. Exploring the Lineage of Volunteered Geographic Information with Heat Maps. *Proceedings of GeoViz: Linking Geovisualization with Spatial Analysis and Modeling*, Hamburg, Germany, 10-12.

Tripathy, A. K., Adinarayana, J., Sudharsan, D., Merchant, S. N., Desai, U. B., Vijayalakshmi, K., Reddy, D. R., Sreenivas, G., Ninomiya, S., Hirafuji, M., Kiura, T. & Tanaka, K. 2011. Data Mining and Wireless Sensor Network for Agriculture Pest/Disease Predictions. *2011 World Congress on Information and Communication Technologies (WICT)*, Mumbai, India, 1229-1234.

Tukey, J. W. 1953. The Problem of Multiple Comparisons. Princeton University, Princeton, New Jersey.

Tulloch, D. L. 2008. Is Vgi Participation? From Vernal Pools to Video Games. *GeoJournal,* **72**, 161-171.

Urbina-Cardona, J. N. & Loyola, R. D. 2008. Applying Niche-Based Models to Predict Endangered-Hylid Potential Distributions: Are Neotropical Protected Areas Effective Enough. *Tropical Conservation Science,* **1**, 417-445.

Van den Berg, H. & Jiggins, J. 2007. Investing in Farmers—the Impacts of Farmer Field Schools in Relation to Integrated Pest Management. *World Development,* **35**, 663-686.

van den Berg, H., Senerath, H. & Amarasinghe, L. 2002. *Participatory Ipm in Sri Lanka: A Broad-Scale and in-Depth Impact Analysis,* Wageningen, FAO Programme for Community IPM in Asia.

Van Exel, M., Dias, E. & Fruijtier, S. 2010. The Impact of Crowdsourcing on Spatial Data Quality Indicators. *Proceedings of the 6th GIScience International Conference on Geographic Information Science,* Zurich, Switerland, 1-4.

Vaz, E. & Jokar Arsanjani, J. 2015. Crowdsourced Mapping of Land Use in Urban Dense Environments: An Assessment of Toronto. *The Canadian Geographer / Le Géographe canadien,* **59**, 246-255.

Viera, A. J. & Garrett, J. M. 2005. Understanding Interobserver Agreement: The Kappa Statistic. *Fam Med,* **37**, 360-363.

Walther, G.-R., Roques, A., Hulme, P. E., Sykes, M. T., Pyšek, P., Kühn, I., Zobel, M., Bacher, S., Botta-Dukát, Z., Bugmann, H., Czúcz, B., Dauber, J., Hickler, T., Jarošík, V., Kenis, M., Klotz, S., Minchin, D., Moora, M., Nentwig, W., Ott, J., Panov, V. E., Reineking, B., Robinet, C., Semenchenko, V., Solarz, W., Thuiller, W., Vilà, M., Vohland, K. & Settele, J. 2009. Alien Species in a Warmer World: Risks and Opportunities. *Trends in Ecology & Evolution,* **24**, 686-693.

Wang, Y., Watson, G. W. & Zhang, R. 2010. The Potential Distribution of an Invasive Mealybug Phenacoccus Solenopsis and Its Threat to Cotton in Asia. *Agricultural and Forest Entomology,* **12**, 403-416.

Warf, B. & Sui, D. 2010. From Gis to Neogeography: Ontological Implications and Theories of Truth. *Annals of GIS,* **16**, 197-209.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A. & Snyder, M. A. 2009. Niches, Models, and Climate Change: Assessing the Assumptions and Uncertainties. *Proceedings of the National Academy of Sciences,* **106**, 19729-19736.

Yan, H., Zou, Z. & Wang, H. 2010. Adaptive Neuro Fuzzy Inference System for Classification of Water Quality Status. *Journal of Environmental Sciences,* **22**, 1891-1896.

Yan, Y., Feng, C.-C., Wan, M.-H. & Chang, K.-T. 2015. Multiple Regression and Artificial Neural Network for the Prediction of Crop Pest Risks. *In:* Bellamine Ben Saoud, N., Adam, C. & Hanachi, C. (eds.) *Information Systems for Crisis Response and Management in Mediterranean Countries.* Springer International Publishing.

Yanenko, O. & Schlieder, C. 2012. Enhancing the Quality of Volunteered Geographic Information: A Constraint-Based Approach. *In:* Gensel, J., Josselin, D. & Vandenbroucke, D. (eds.) *Bridging the Geographic Information Sciences.* Springer Berlin Heidelberg.

Yang, L.-n., Peng, L., Zhang, L.-m., Zhang, L.-l. & Yang, S.-s. 2009. A Prediction Model for Population Occurrence of Paddy Stem Borer (*Scirpophaga Incertulas*), Based on Back Propagation Artificial Neural Betwork and Principal Components Analysis. *Computers and Electronics in Agriculture,* **68**, 200-206.

Yeager, C. D. & Steiger, T. 2013. Applied Geography in a Digital Age: The Case for Mixed Methods. *Applied Geography,* **39**, 1-4.

Yongting, C. 1996. Fuzzy Quality and Analysis on Fuzzy Probability. *Fuzzy Sets and Systems,* **83**, 283-290.

Yu, Z. & Tsai, J. J. P. 2006. Fuzzy Model Tuning for Intrusion Detection Systems. *In:* Yang, L. T., Jin, H., Ma, J. & Ungerer, T. (eds.) *Autonomic and Trusted*

Computing: Third International Conference, Atc 2006, Wuhan, China, September 3-6, 2006. Proceedings. Springer Berlin Heidelberg.

Zadeh, L. A. 1965. Fuzzy Sets. *Information and Control*, **8**, 338-353.

Zalucki, M. P., Adamson, D. & Furlong, M. J. 2009. The Future of Ipm: Whither or Wither? *Australian Journal of Entomology*, **48**, 85-96.

Zalucki, M. P., Furlong, M. J., Schellhorn, N. A., Macfadyen, S. & Davies, A. P. 2015. Assessing the Impact of Natural Enemies in Agroecosystems: Toward "Real" Ipm or in Quest of the Holy Grail? *Insect Science*, **22**, 1-5.

Zhang, Z., Demšar, U., Rantala, J. & Virrantaus, K. 2014. A Fuzzy Multiple-Attribute Decision-Making Modelling for Vulnerability Analysis on the Basis of Population Information for Disaster Management. *International Journal of Geographical Information Science*, **28**, 1922-1939.

Zheng, Y.-T., Zha, Z.-J. & Chua, T.-S. 2012. Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **3**, 56.

Zielstra, D. & Zipf, A. 2010. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. *AGILE 2010: 13th AGILE International Conference on Geographic Information Science*, Guimarães, Portugal, 1-15.

Ziska, L. H., Blumenthal, D. M., Runion, G. B., Hunt, E. R. & Diaz-Soltero, H. 2010. Invasive Species and Climate Change: An Agronomic Perspective. *Climatic Change*, **105**, 13-42.

Zou, L., Stout, M. J. & Ring, D. R. 2004. Degree-Day Models for Emergence and Development of the Rice Water Weevil (Coleoptera: Curculionidae) in Southwestern Louisiana. *Environmental Entomology,* **33**, 1541-1548.

Appendices

Appendix 1 An additional analysis about the spatial distribution of the detected hot and cold spots.

Based on Getis-Ord Gi*, it was observed that the pest hot spot discovered based on the entire VGI dataset was located in an area with a high woodland coverage, a low coverage of residential area, and the area was relatively far away from the local agricultural department. On the contrary, the cold spot was located in an area with a low woodland coverage, a high coverage of residential area, and the area was relatively nearer to the local agricultural department. Several possible explanations for these observations have been given in Section 4.5.1. Therefore, I hypothesized that the woodland coverage and distance to the local agricultural department were positively related to the occurrences of the hot spots (i.e., the occurrences of higher z-scores). On the contrary, the coverage of residential area was hypothesized to be negatively related to the occurrences of the hot spots.

In order to seek statistical supports, the land use/land cover of the study area was first digitized based on Google Earth using satellite imagery @ 2015 CNES/Astrium. An ordinary least squares (OLS) multiple linear regression was conducted using ArcGIS 10.0 (ESRI Products, Redlands, CA) by following the procedures as follows:

Step 1. Interpolating the Getis-Ord Gi* statistics or z-scores over the study area to generate a raster surface with 250 m cell size (970 cells in total). The 250 m cell size was determined based on average nearest neighbor distance (Bruce *et al*., 2014).

Step 2. A fishnet vector layer containing 970 polygons (250 m by 250 m) was generated from the raster surface. Each polygon has a normalized z-score attribute (from 0 to 1) as the dependent variable (Appendix 1.1a), using linear normalization (Eq. A1):

$$Linear\ normalized(y_i) = \frac{y_i - y_{min}}{y_{max} - y_{min}}, \tag{A1}$$

where y_{min} is the minimum value of the dependent variable, y_{max} is the maximum value of dependent variable.

Step 3. The percent coverage of residential area in each polygon was calculated as one of the independent variables (Appendix 1.1b).

Step 4. The percent coverage of woodlands in each polygon was calculated as the second independent variable (Appendix 1.1c).

Step 5. The distance from each cell center to location of the local agricultural department was calculated as the third independent variable (Appendix 1.1d).

Step 6. Possible mathematical transformation was conducted if any assumption of the regression was violated.

The following regression equation (Eq. A2) was the result obtained from the OLS regression analysis:

$$\text{z-score} = -0.087\text{RESI}^{**} + 0.423\text{WOOD}^{**} + 0.042\text{DIST}^{**} + 0.141^{**}, \qquad \text{(A2)}$$

where RESI denotes the percent coverage of residential area, WOOD denotes the percent coverage of woodlands, DIST denotes the distance to the agricultural department, ** represents a statistical significance at 99% confidence level, i.e., both P and robust $P < 0.01$. All the three independent variables were statistically significant. There was no multicollinearity among the independent variables (Variance Inflation Factors < 7.5). The overall model was significant according to the joint Wald statistic ($P < 0.01$). The three variables together explained a meaningful proportion of the variance ($R^2 = 0.43$). However, the Jarque-Bera statistic was significant ($P < 0.01$), indicating that the residuals (model over and under predictions) were not normally distributed. In addition, the homoscedasticity assumption (the residuals should have a constant variance) was violated.

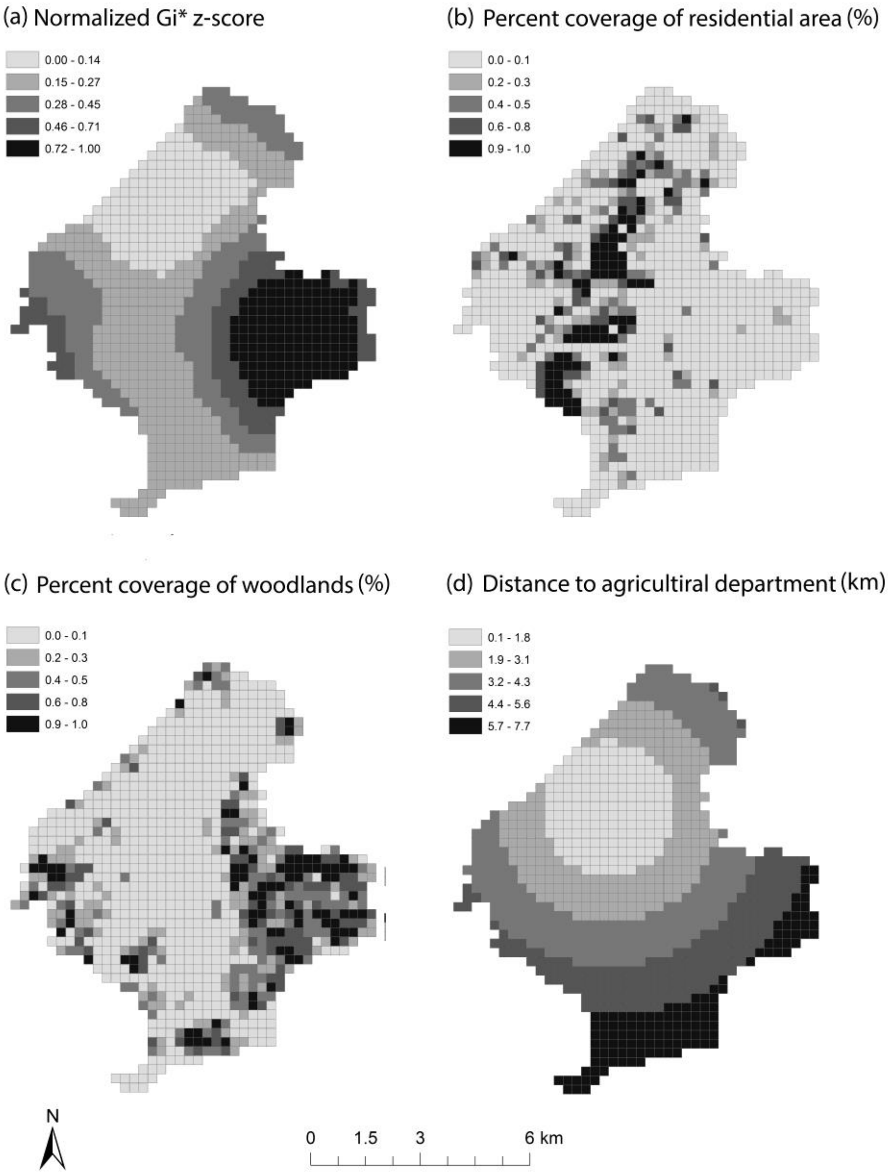
The problems mentioned above were solved by taking the square root of the dependent variable, i.e., predicting the square root of the dependent variable from the independent variables. Lastly, the following regression equation (Eq. A3) was obtained:

$$\text{Sqrt (z-score)} = -0.077\text{RESI}^{**} + 0.311\text{WOOD}^{**} + 0.043\text{DIST}^{**} + 0.357^{**} \quad \text{(A3)}$$

Where Sqrt (z-score) denotes the square root of a z-score. The overall model was also significant according to the joint Wald statistic ($P < 0.01$) and R^2 equaled to 0.45. All the assumptions of the regression were satisfied. The coefficients had the expected sign. The woodland coverage and distance to the local agricultural department significantly related to the square root of the clustering index (Getis-Ord Gi* statistics or z-scores) in a positive way, and the coverage of residential area significantly related to the square root of the clustering index in a negative way (both P and robust $P < 0.01$). Therefore, the hypotheses were supported.

However, spatially autocorrelated residuals was observed, which was confirmed by the Global Moran's I (I = 0.85, z = 36.37, P = 0.00). Statistically significant spatial autocorrelation is often a symptom of misspecification (one or more key variable is missing from the model). Indeed, the variance explained by the three independent variables was not quite high (R2 = 0.45). Therefore, further investigation is needed to fully understand all possible influencing factors contributing to the pest distribution. After that, spatial regression analysis (e.g., geographically weighted regression) may be adopted to better understand specifically how these factors contribute to the observed pest distribution pattern in different parts of the study area, so as to better manage the pest infestations.

Appendix 1.1 (a) Dependent and (b-d) independent variables of the OLS multiple linear regression.



Appendix 2 List of the 76 invasive pest species with their scientific names, common names, species taxonomies, and predicted future changes in the mean probability of presence under four climatic projections. The 76 species were selected based on their economic importance or quarantine significance to major food and cash crops recorded in CABI (2016). In addition, these species are common, highly harmful, widespread, easily detected, and frequently spotted and reported. They have been recorded in CABI (2014) as being capable of establishing and spreading into new ecosystems. The distributions of these 76 species together cover all climatic zones and all the continents, except for Antarctica, thereby representing a wide spectrum of the locations across the globe. According to Oerke et al. (2012), CABI (2016), and Agropages (2015), the 76 pest species were classified into three categories based on their main host plants/plants affected. Category 'A' species mainly causes high economic losses to major food grains including cereal crops, tuber crops, and legumes. Category 'B' species mainly causes high economic losses to major cash crops including sugar-yielding crops, oil-bearing crops, fibres, fruits, vegetables, floral crops, melon crops, nuts, raw materials crops, aromatic crops, tobacco, tea, coffee, and cocoa. Category 'C' species can cause high economic losses to both major food grains and major cash crops.

Scientific name	Common name	Species taxonomy	Host plants/ plants affected	Change in mean probability of presence			
				2050-RCP2.6	2050-RCP4.5	2070-RCP2.6	2070-RCP4.5
Agrotis ipsilon	black cutworm	Insecta	C	-0.007	-0.017	0.001	-0.010
Anthonomus grandis	Mexican cotton boll weevil	Insecta	B	0.007	0.008	0.006	0.020
Aphis gossypii	cotton aphid	Insecta	C	-0.017	-0.032	-0.012	-0.026
Bactrocera cucurbitae	melon fly	Insecta	C	0.070	0.064	0.070	0.062
Bactrocera dorsalis	Oriental fruit fly	Insecta	B	0.122	0.124	0.123	0.133
Bactrocera oleae	olive fruit fly	Insecta	B	0.003	0.002	0.004	0.006

Bemisia tabaci	tobacco whitefly	Insecta	C	-0.005	-0.012	0.002	-0.010
Brevipalpus phoenicis	false spider mite	Arachnida	B	0.068	0.065	0.071	0.068
Callosobruchus chinensis	Chinese bruchid	Insecta	A	0.091	0.070	0.075	0.048
Callosobruchus maculatus	cowpea weevil	Insecta	A	0.001	-0.007	0.008	-0.001
Cerataphis lataniae	palm aphid	Insecta	B	0.003	0.021	0.008	0.004
Ceratitis cosyra	mango fruit fly	Insecta	B	-0.029	-0.041	-0.033	-0.040
Chilo partellus	spotted stem borer	Insecta	C	0.018	0.007	0.009	0.013
Chilo suppressalis	striped rice stem borer	Insecta	A	0.040	0.044	0.054	0.059
Chrysodeixis chalcites	golden twin-spot moth	Insecta	C	-0.009	-0.013	-0.009	-0.015
Eriosoma lanigerum	woolly aphid	Insecta	B	0.000	-0.001	0.006	0.005
Cornu aspersum	common snail	Gastropoda	C	-0.010	-0.016	-0.009	-0.012
Helicoverpa armigera	cotton bollworm	Insecta	C	0.016	0.003	0.018	0.009
Crocidosema plebejana	cotton tipworm	Insecta	B	0.039	0.033	0.043	0.037
Cylas formicarius	sweet potato weevil	Insecta	A	0.005	-0.010	-0.002	-0.012
Deroceras reticulatum	grey field slug	Gastropoda	C	-0.009	-0.021	-0.012	-0.011

Diabrotica virgifera virgifera	western corn rootworm	Insecta	A	0.004	-0.004	0.006	0.009
Dinoderus minutus	bamboo borer	Insecta	B	0.088	0.069	0.085	0.064
Diuraphis noxia	Russian wheat aphid	Insecta	A	0.001	-0.006	-0.003	0.003
Ferrisia virgata	striped mealybug	Insecta	C	-0.014	-0.025	-0.013	-0.021
Frankliniella occidentalis	western flower thrips	Insecta	C	-0.040	-0.047	-0.037	-0.039
Grapholita molesta	oriental fruit moth	Insecta	B	0.042	0.035	0.047	0.048
Helicoverpa zea	American cotton bollworm	Insecta	C	-0.025	-0.038	-0.026	-0.033
Heliothrips haemorrhoidalis	black tea thrips	Insecta	B	0.011	-0.003	0.014	0.006
Heterodera glycines	soybean cyst nematode	Heteroderidae	A	0.014	0.013	0.017	0.024
Heterodera goettingiana	pea cyst eelworm	Heteroderidae	A	0.015	0.014	0.018	0.021
Heteropsylla cubana	leucaena psyllid	Insecta	B	0.090	0.073	0.086	0.074
Hypothenemus hampei	coffee berry borer	Insecta	B	-0.002	-0.004	-0.001	0.000
Leptinotarsa decemlineata	Colorado potato beetle	Insecta	C	0.009	0.004	0.011	0.013
Liriomyza sativae	vegetable leaf miner	Insecta	C	-0.039	-0.060	-0.036	-0.053
Lissachatina fulica	giant African land snail	Gastropoda	C	0.094	0.085	0.093	0.090

Lissorhoptrus oryzophilus	rice water weevil	Insecta	A	-0.002	-0.005	0.004	-0.001
Listroderes costirostris	vegetable weevil	Insecta	C	0.007	0.001	0.009	0.002
Maconellicoccus hirsutus	pink hibiscus mealybug	Insecta	C	0.058	0.052	0.056	0.054
Mononychellus tanajoa	cassava green mite	Arachnida	C	-0.017	-0.026	-0.019	-0.046
Mythimna unipuncta	rice armyworm	Insecta	A	0.019	0.018	0.020	0.030
Nezara viridula	green stink bug	Insecta	C	0.030	0.022	0.029	0.031
Nipaecoccus nipae	spiked mealybug	Insecta	C	0.003	0.026	0.027	0.003
Oryctes rhinoceros	coconut rhinoceros beetle	Insecta	C	0.112	0.114	0.113	0.124
Ostrinia nubilalis	European maize borer	Insecta	A	-0.004	-0.006	0.002	0.001
Papilio demoleus	chequered swallowtail	Insecta	B	0.085	0.091	0.083	0.098
Pectinophora gossypiella	pink bollworm	Insecta	B	-0.004	-0.015	-0.003	-0.005
Phenacoccus solenopsis	cotton mealybug	Insecta	B	-0.009	-0.023	-0.003	-0.015
Phthorimaea operculella	potato tuber moth	Insecta	C	0.003	-0.005	0.001	0.004
Planococcus kenyae	coffee mealybug	Insecta	C	-0.004	-0.009	-0.007	-0.005
Plutella xylostella	diamondback moth	Insecta	B	0.001	-0.011	0.005	-0.007

Polyphagotarsonemus latus	broad mite	Arachnida	C	0.024	0.017	0.032	0.029
Pomacea canaliculata	golden apple snail	Gastropoda	C	-0.024	-0.033	-0.023	-0.031
Prostephanus truncatus	larger grain borer	Insecta	A	-0.003	-0.013	-0.010	0.015
Raoiella indica	red palm mite	Arachnida	B	-0.002	-0.009	-0.006	0.003
Rhynchophorus ferrugineus	red palm weevil	Insecta	B	0.043	0.046	0.017	0.074
Rhyzopertha dominica	lesser grain borer	Insecta	A	-0.007	-0.011	-0.003	-0.006
Scirtothrips dorsalis	chilli thrips	Insecta	C	0.191	0.190	0.188	0.199
Sesamia cretica	greater sugarcane borer	Insecta	C	-0.008	-0.011	-0.008	-0.004
Sitobion miscanthi	indian grain aphid	Insecta	A	0.144	0.142	0.143	0.136
Sitophilus granarius	grain weevil	Insecta	A	-0.002	0.001	0.004	0.002
Sitophilus zeamais	greater grain weevil	Insecta	A	0.021	0.023	0.023	0.025
Sitotroga cerealella	grain moth	Insecta	A	-0.009	-0.014	-0.004	-0.002
Solenopsis geminata	tropical fire ant	Insecta	C	0.010	0.005	0.010	0.010
Spodoptera exempta	black armyworm	Insecta	C	0.081	0.068	0.066	0.060
Spodoptera littoralis	cotton leafworm	Insecta	C	-0.006	-0.010	-0.003	-0.011

Spodoptera litura	taro caterpillar	Insecta	C	0.059	0.054	0.086	0.059
Spoladea recurvalis	Hawaiian beet webworm	Insecta	C	0.017	0.005	0.018	0.009
Sternochetus mangiferae	mango seed weevil	Insecta	B	0.037	0.028	0.038	0.036
Tetranychus urticae	two-spotted spider mite	Arachnida	C	-0.041	-0.051	-0.037	-0.043
Thrips palmi	melon thrips	Insecta	C	0.070	0.060	0.073	0.077
Trialeurodes vaporariorum	greenhouse whitefly	Insecta	C	-0.053	-0.064	-0.047	-0.055
Trogoderma granarium	khapra beetle	Insecta	C	0.060	0.064	0.058	0.074
Tryporyza incertulas	yellow stem borer	Insecta	A	0.033	0.034	0.031	0.040
Xyleborus perforans	island pinhole borer	Insecta	B	0.050	0.072	0.075	0.044
Xylosandrus compactus	shot-hole borer	Insecta	B	0.089	0.085	0.086	0.089

References for appendices

Agropages. 2015. *World Agrochemical Yellow Page Online* [Online]. Available: http://cn.agropages.com/Bcc/ [Accessed 14 January 2015].

Bruce, E., Albright, L., Sheehan, S. & Blewitt, M. 2014. Distribution Patterns of Migrating Humpback Whales (*Megaptera Novaeangliae*) in Jervis Bay, Australia: A Spatial Analysis Using Geographical Citizen Science Data. *Applied Geography*, **54**, 83-95.

CABI 2016. *Invasive Species Compendium* [Online]. Available: http://www.cabi.org/isc [Accessed 16 March 2016].

Oerke, E. C., Dehne, H. W., Schönbeck, F. & Weber, A. 2012. *Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crops,* Amsterdam, the Netherlands., Elsevier.