Environmental Modelling & Software 100 (2018) 278–290

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Distributed data collection and web-based integration for more efficient and informative groundwater pollution risk assessment

Jon Sege [a], Marwan Ghanem [b], Waseem Ahmad [b], Hanadi Bader [b], Yoram Rubin [a, *]

[a] University of California, Berkeley, Department of Civil and Environmental Engineering, Berkeley, CA, USA
[b] Birzeit University, Geography Department, P.O.Box 14, Birzeit, Palestine

ARTICLE INFO

Article history:
Received 18 January 2017
Received in revised form
13 October 2017
Accepted 16 November 2017

Keywords:
Distributed data collection
Web-based GIS
Risk assessment
Groundwater
Pollution

ABSTRACT

We present a web-based software platform for assimilation of field data on groundwater vulnerability and assessment of groundwater pollution risk. Groundwater vulnerability and risk assessments couple hydrogeologic characterization and pollution source information to provide a relative measure of risk to a groundwater supply. Vulnerability is based solely on hydrogeologic factors, while risk also considers pollution hazards. Previous studies have identified two key limitations in the collection of pollution hazards information: detailed investigations throughout a study area are extremely time-consuming, and pollution inventories are not regularly updated to reflect changes in land use. This paper presents a methodology to address these two limitations by allowing broad-based effort to collect pollution hazard parameters in the field using GPS-tagged smartphone forms and incorporating the new information rapidly into a web-based risk modelling framework. To demonstrate this approach, a case-study is presented from the Natuf Basin in the Ramallah-Al Bireh Governorate of Palestine.

1. Introduction

This paper focuses on the development and application of a networked software platform supporting efficient and flexible groundwater pollution risk modelling. This section will introduce key principles of groundwater vulnerability and risk assessments, and motivate the use of new software tools to improve these models. Intrinsic groundwater vulnerability can be defined as “an intrinsic property of a groundwater system that depends on the sensitivity of that system to human and/or natural impacts” (Vrba and Zaporožec, 1994). In this study, vulnerability was considered to relate to features of local hydrology, soil, and geologic layers that may mitigate contaminants introduced by human activities. Vulnerability maps portray the effectiveness of natural filtration processes in certain areas relative to others within a geographic study domain. Such maps are commonly used by land use planners to evaluate the potential relative impact of human development on groundwater quality in different areas (Vrba and Zaporožec, 1994). Vulnerability is solely a function of hydrogeologic factors characterizing the overlying soil and geological materials, and does not consider human activities (Doerfliger et al., 1999).

Vulnerability assessments can be expanded to obtain overall groundwater pollution risk measures, which consider both natural and anthropogenic factors. In general, risk is defined as the likelihood of certain events occurring and the magnitude of their possible consequences (Simpson et al., 2014). This study considers a rather narrow definition of risk, namely that human activities at or near the land surface will introduce contaminants to the groundwater supply, which may be harmful to human health. This definition was adopted for simplicity, as our primary goal is to demonstrate the concepts and ideas behind the risk platform. The risk maps we consider portray a relative measure of the risk of aquifer contamination resulting from human activities in the study area. These maps are obtained by combining intrinsic vulnerability with an inventory of potential groundwater pollution hazards. Hazards are generally assessed using a combination of information about the type of potential contaminant release, the extent of a potential release, and the likelihood of such a release given containment measures in place at the site.

Published groundwater pollution risk studies (Zwahlen, 2003; Simpson et al., 2014) have identified limitations in current methodologies, particularly related to creating and updating the hazard inventory. For one, hazard inventories are generally time-consuming and labor-intensive to collect. Detailed on-site investigations throughout the study area, which are ideal for hazard data collection, are extremely time consuming and are therefore rarely done (Simpson et al., 2014). Instead, hazard inventories are often created from existing land use data, and may represent general, rather than specific, assessments. In such assessments, basic land use data for

---

* Corresponding author.
E-mail address: rubin@ce.berkeley.edu (Y. Rubin).

https://doi.org/10.1016/j.envsoft.2017.11.027

each type, such as activity type, size, and age are used to estimate a generalized hazard value. In some cases, more specific information about certain point hazards is added to land use data (Simpson et al., 2014).

The second main limitation in many hazard analyses is that hazard inventories are produced during the data collection phase of the risk assessment, and reflect conditions only as they were at that point in time. However, changes in land use may introduce new hazards, remove old hazards, or change the nature of certain hazards (Zwahlen, 2003). Some of these changes can occur on a daily or weekly time scale. When subsurface residence times are short, as in Karst systems, new hazards can quickly impact the water supply and should result in rapid changes in the risk assessment (Zwahlen, 2003). However, without ongoing hazard data collection, these changes are not reflected in the risk map being used by planners. Risk values may no longer reflect the current land use distribution and decisions based on this information may be affected.

This study addresses the two main limitations of hazard inventories outlined above: A methodology for distributing the data collection burden, collecting and evaluating hazard information in-situ, and introducing temporal variation into a risk assessment algorithm, is developed. In this approach, information about pollution hazards is stored in an online database, which is available, selectively, to multiple users. Up-to-date information about hazards can be collected by multiple users, GPS-tagged in the field using a smartphone, and submitted to the database, where it can be analyzed and shared in real time (Rubin and Michaelis, 2017). New additions, changes, or deletions in the database are reflected in the risk model, which is recalculated daily or as otherwise directed.

Previous work, for example by Granell et al. (2010), has introduced decentralized geospatial processing tools in order to reduce tedious or repetitive tasks being carried out by researchers during the modelling process. This approach extends that work to also include decentralized collection and integration of model data from the field. It applies decentralized data collection principles, such as those discussed in Horsburgh et al. (2011), directly to an environmental modelling problem, resulting in a framework that allows for decentralized collection and integration of new data into the model, and also automates the task of updating the model over time. By collecting, storing, and geoprocessing model data all within the same system, this approach avoids interoperability issues and challenges integrating data and modelling services (Peckham and Goodall, 2013; Castronova et al., 2013).

This methodology is applied in a case study and the results are discussed. We show that the risk model changes significantly as new hazard information is added by researchers in the field and that the proposed software platform supports rapid updating through community participation. In a developed model, this new information could represent previously non-existent hazards (such as new garbage dump sites or chemical spills) or the removal or modification of previously existing ones.

2. Methods

A simplified risk modelling problem definition has been adopted here, in order to focus our presentation on the networked software platform's ability to overcome key risk modelling limitations. A similar approach can be taken with more complex risk modelling methodologies and in different settings. The procedure was designed to allow flexibility in updating the model on an ongoing basis, using data transmitted through multiple widely-used mobile devices. To accomplish this, the risk assessment was split into components such that parameters that change over relatively long time scales were assessed offline using ESRI ArcGIS desktop and Quantum GIS (QGIS), while those changing over relatively short time scales were analyzed on a web-based, GIS-powered environmental information management system (EIMS). Putting the shorter time scale analyses on the web allowed information to be submitted by mobile devices in use by field researchers. This introduced time-dependence in the form of new information with which to rapidly update the model. The general scheme of the analysis is depicted in Fig. 1, whose terms will be discussed in the following sub-section.

2.1. Vulnerability assessment

Several commonly-used groundwater vulnerability assessment methods were reviewed, including DRASTIC (Aller et al., 1987), GOD (Foster, 1987), and AVI (Van Stempoort et al., 1993). Since the application area for this study overlies a karstified aquifer, the PI method, which is designed for Karst terrains, was chosen. The PI method considers two factors to determine vulnerability: protective cover (P) and infiltration conditions (I) (Goldscheider, 2005). The other vulnerability assessment methods considered do not provide tools for Karst terrains, or need to be modified for such environments (Ravbar and Goldscheider, 2009). The PI method specializes in karstified aquifers by introducing a factor to describe the degree to which protective cover is bypassed as a result of lateral surface and subsurface flows through Karst conduits. Furthermore, in a comparative study, Ravbar and Goldscheider (2009) found that the PI method more closely matched multi-tracer tests, at least during low-flow conditions, compared with other vulnerability methods designed for Karst aquifers (Ravbar and Goldscheider, 2009). Since the study area is located in Karst terrain in a semi-arid climate with a distinct seasonal precipitation pattern, the PI method was selected in order to best represent prevailing low-flow conditions during the dry season.

The vulnerability assessment was carried out in ArcGIS by overlaying thematic layers of soil cover, aquifer geology, elevation, slope, vegetation, and recharge, following the PI methodology outlined in the European Commission's COST Action 620 Final Report (Zwahlen,

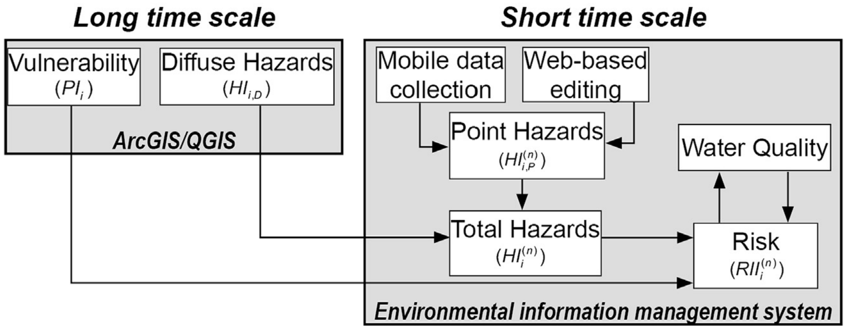

Fig. 1. Data flow for risk assessment using long time scale (ArcGIS/QGIS) and short time scale (EIMS) components.