Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software 22 (2007) 442–448
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Fuzzy expert systems and GIS for cholera health risk prediction in southern Africa

Gavin Fleming*, Marna van der Merwe, Graeme McFerren

CSIR, Box 395, Pretoria, 0001, South Africa

Received 11 July 2005; received in revised form 10 October 2005; accepted 21 December 2005
Available online 10 March 2006

Abstract

Cholera (*Vibrio cholerae*) is endemic in southern Africa and frequently breaks out in epidemics along the eastern seaboard. Extensive resources are directed at combating cholera yet it remains a significant problem. Limited resources could better be directed to prevent outbreaks if it were possible to assess the risk of an outbreak in space and time. The CSIR in South Africa is investigating technologies to predict health risk in line with national priorities. This paper describes an early warning GIS prototype tool aimed at identifying favourable preconditions for cholera outbreaks. These preconditions were defined using an expert system approach. The variables thus identified were input into a spatial fuzzy logic model that outputs risks. The model is based on the assumption that endemic reservoirs of cholera occur and that environmental conditions, especially algal blooms, trigger *Vibrio* growth in the natural environment. If the preconditions are met, the subsequent spread of cholera depends mainly on socio-economic factors such as human behaviour and access to safe water supply and sanitation. This paper focuses on the environmental preconditions. The methodology described relies on capturing expert knowledge and historic data that integrate climatic and biophysical parameters with epidemiological data to produce a fuzzy surface of cholera outbreak risk potential.

Keywords: Cholera; Health risk prediction; Algal bloom; Expert system; Fuzzy logic

1. Introduction

Cholera is a severe diarrhoeal disease caused by the bacterium *Vibrio cholerae*. A patient may lose up to a litre of fluid per hour, which can lead to death in fewer than 24 h if untreated. It is a highly infectious disease and transmission to humans is mainly through exposure to contaminated water or food.

The severity and extent of cholera is recognised worldwide. It is listed as one of three internationally quarantinable diseases by the World Health Organisation (WHO), along with plague and yellow fever (WHO, 2000). It is the first disease that received organised modern public health surveillance and reporting of worldwide incidence. Cholera is one of the most researched communicable diseases, yet it still has devastating effects on local communities (Collins, 2003). Most of the new cases reported since 2000 are from eastern and southern Africa.

South Africa experienced major cholera outbreaks during 1980 to 1984 and further outbreaks have occurred since August 2000. From August 2000 to February 2002, the disease infected at least 114 000 people in KwaZulu-Natal (more than 70% of the total cases reported in the country) and claimed at least 259 lives in the province (Cottle and Deedat, 2002). Initial reports of the cholera outbreak in 2000 came from the largely rural and impoverished communities near Empangeni town. The source of the epidemic was traced to the uMhlathuze River, also in the northern part of the KwaZulu-Natal Province (Fig. 1). Subsequent outbreaks have occurred in the Eastern Cape and Mpumalanga provinces. Hot and humid climatic conditions and socio-economic conditions are risk factors, with most cases reported during the 'hot' months of the year,

* Corresponding author. Fax: +27 86 6164820.
E-mail address: gfleming@mailbox.co.za (G. Fleming).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2005.12.008

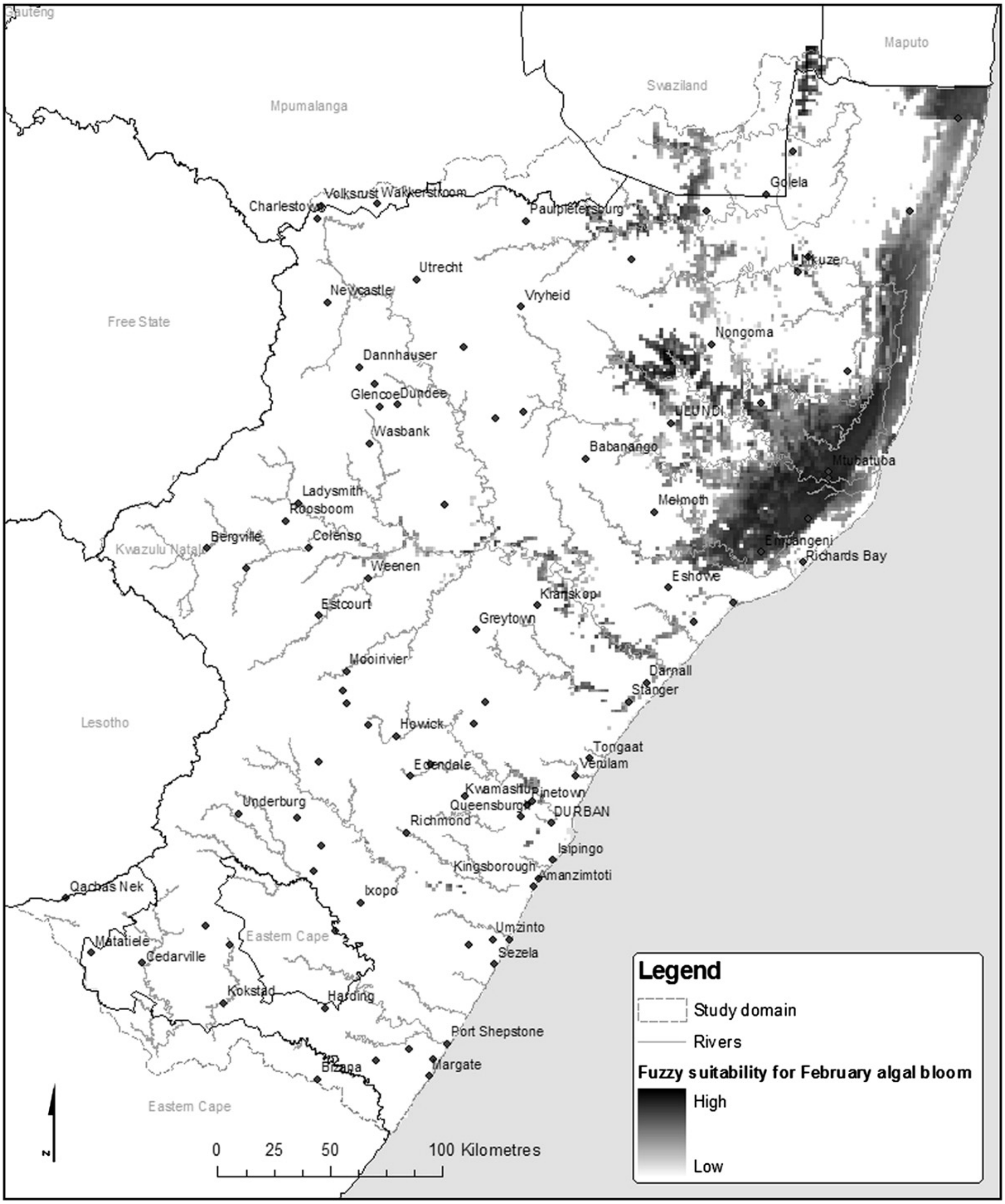

Fig. 1. Map showing the KwaZulu-Natal river basins, major rivers, locations of reported cholera cases and a preliminary fuzzy suitability surface for February algal bloom. The river thick grey river exiting south of Richards Bay is the uMhlathuze.

February being one of these. (Source: Cholera cases – GIS unit, KwaZulu-Natal Department of Health.) The model we develop is based on the assumption that endemic reservoirs of cholera exist and that environmental conditions trigger algal blooms which in turn trigger *Vibrio* growth. Once the cholera is present in water in sufficient numbers, any outbreak and spread of cholera depends mainly on socio-economic factors such as human behaviour and access to safe water supply and sanitation, as well as proximity to surface water, high population density and poor educational level (Ali et al., 2002). This paper focuses on the environmental preconditions.

Our aim was to develop an integrative modelling approach per se and to demonstrate its applicability to a real problem like cholera. We developed a model based on knowledge captured from literature reviews and discussions with experts together with historical data on cholera outbreaks and environmental conditions in South Africa. The model is based on assumptions about the conditions that trigger a cholera outbreak. We chose KwaZulu-Natal province as the study area since most of the cholera cases reported during the 2000 outbreak occurred there. KwaZulu-Natal is also the only province for which a GIS-based surveillance system has been developed (operated by the GIS unit of the KwaZulu-Natal Department of Health: Private Bag X9051, Pietermaritzburg, 3200, South Africa).

The focus of this study was to investigate and develop an integrative prediction model that ultimately can form part of an integrative approach to reduce the number of cholera cases in the country, through applying a range of intervention types and policy decisions. An integrative prediction model uses multiple inputs. An integrative case reduction approach includes a range of interventions and policy decisions.

2. Problem statement

Cholera is a complex health problem to manage because of a combination of factors. These vary from biophysical to social and economic. No single solution has been identified that effectively can reduce or eliminate the impact of cholera. We therefore sought to assist with the management objective of allocating resources to the right places, early enough to prevent a cholera outbreak by:

- Understanding the environmental preconditions for a cholera outbreak;
- Assessing the risk of an outbreak based on these environmental preconditions.

Predicting cholera has proven difficult because of a number of factors including multiple strains of the bacteria, high levels of case fatalities in some regions of the world and a complex of influences on its distribution in people and the environment. Just a slight improvement in information could improve resource allocation and coordination between departments dramatically.

3. Approach and methodology

The prediction model described herein is based on the application of a simulation model to parameterise an expert system, which is then implemented with fuzzy logic in a GIS. An hierarchical approach was followed to understand the ecology and possible spread of *V. cholerae* in the environment (Fig. 2). The integrative modelling approach is based on this hierarchical structure.

A flow diagram to illustrate the application of the different techniques is provided in Fig. 3. The model output is the predicted, or estimated, risk of an algal bloom that potentially could trigger a cholera outbreak.

The simulation model and expert systems were used to capture and simulate the knowledge gained from the literature review and discussions with cholera experts. Subjective expert input can be relied upon if handled correctly (Kawano et al., 2005). The simulation model simulated the changes in bacterial numbers in the natural environment based on the combined impact of changes in certain environmental conditions and provided some inputs for the expert system. The simulation model was based on quoted values for the different variables in the literature and no real data were used. Fig. 4 is a flow diagram that illustrates the simulation model, which was developed in Powersim. The outputs of the expert system were used to establish the high-level structure and flow of the integrated model. The expert system could equally have been derived from a complex decision tree or binary model. Fig. 5 illustrates the expert system model, which was developed in the ERDAS Imagine Knowledge Engineer. The outputs were then integrated and implemented as a fuzzy logic model in

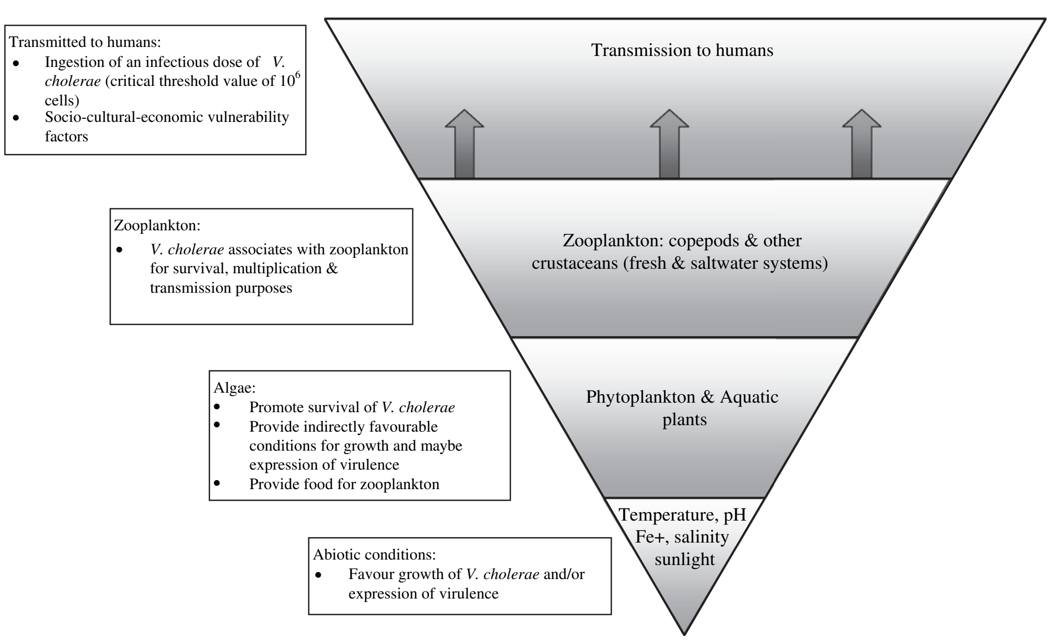

Fig. 2. The hierarchical approach to describe the 'ecology' and spread of cholera in the environment (modified from Lipp et al., 2002).

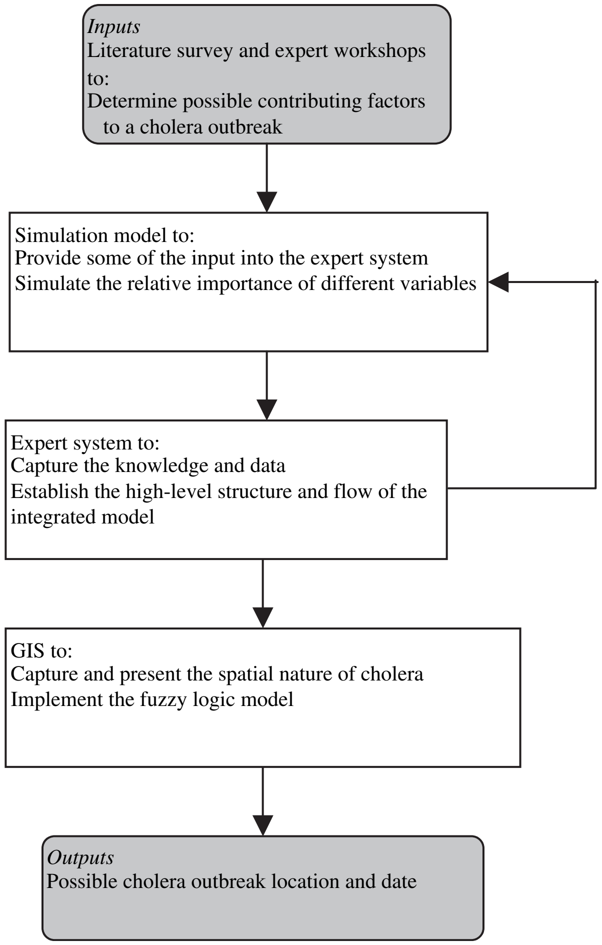

Fig. 3. Flow diagram to illustrate the application of techniques used.

a GIS, namely ArcInfo GRID, coded in AML (Arc Macro Language).

The optimal values or ranges for the individual variables identified and used as part of the simulation model and expert system are provided in Table 1. The flow diagram provided to illustrate the expert system (Fig. 5) must be interpreted together with Table 1.

A large number of the environmental variables have high uncertainties. It is possible, using fuzzy logic techniques, to take these uncertainties into account when attempting to predict these variables in space or time (Iliadis, 2005; Metternicht and Gonzalez, 2005), fuzzy set membership functions were developed for average annual rainfall, mean maximum daily temperature on a monthly basis and 'month of first rains' per pixel. The 'month of first rains' was included because it is thought that the flush of salts into the rivers from the first rains of the season contributes to the salinity requirements of a cholera bloom. A fuzzy gamma coefficient of 0.75 was used in the generation of a fuzzy surface representing membership of the set 'at risk of developing an algal bloom' for the month with the highest number of reported cases, namely February (Fig. 1). The expert system (Table 1 and Fig. 5) provided the rules for decision points in the model, i.e. the cut-off points and ranges for fuzzy membership functions. The fuzzy membership functions were linear functions within the ranges in Table 1. Fuzzy logic or a combination of fuzzy sets was applied to conduct the model inference, matching as closely as possible the logic represented in the expert system.

A large literature exists regarding the prediction of algal blooms by using artificial neural networks. The use of an expert system together with fuzzy logic as described above is more in line with the work done by Marsili-Libelli (2004) on the development of a Sugeno fuzzy inferential engine that predicts algal blooms. More regularly monitored data on the diurnal pattern of dissolved oxygen and oxidation–reduction potential, and to a lesser extent pH and water surface temperature are needed for the rivers in South Africa, especially in the areas where cholera cases are reported. These physico-chemical water parameters contribute to water conditions that are favourable for algal blooms. Cholera blooms and subsequent outbreaks are assumed to be linked to algal blooms.

The prediction model is based on the assumption that endemic reservoirs of *V. cholerae* occur in the study area and

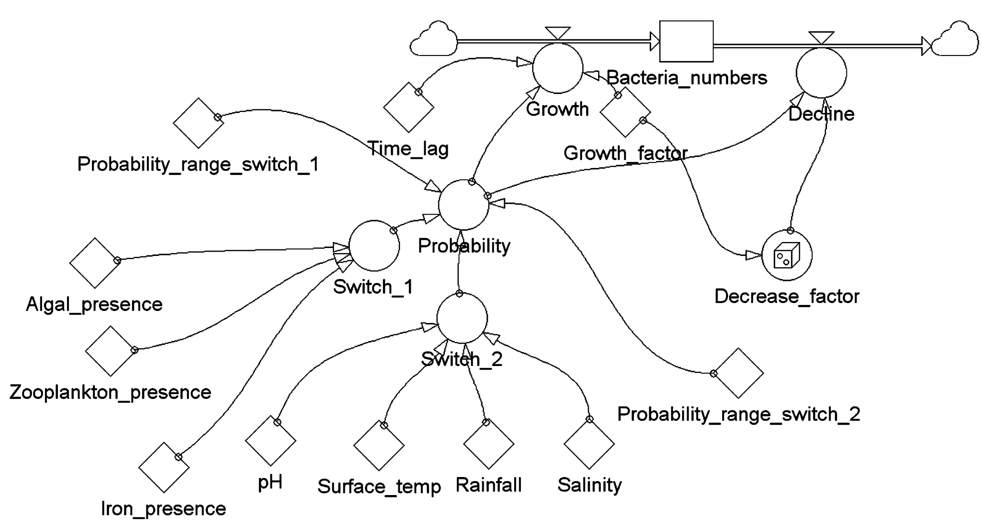

Fig. 4. Flow diagram to illustrate the simulation model.

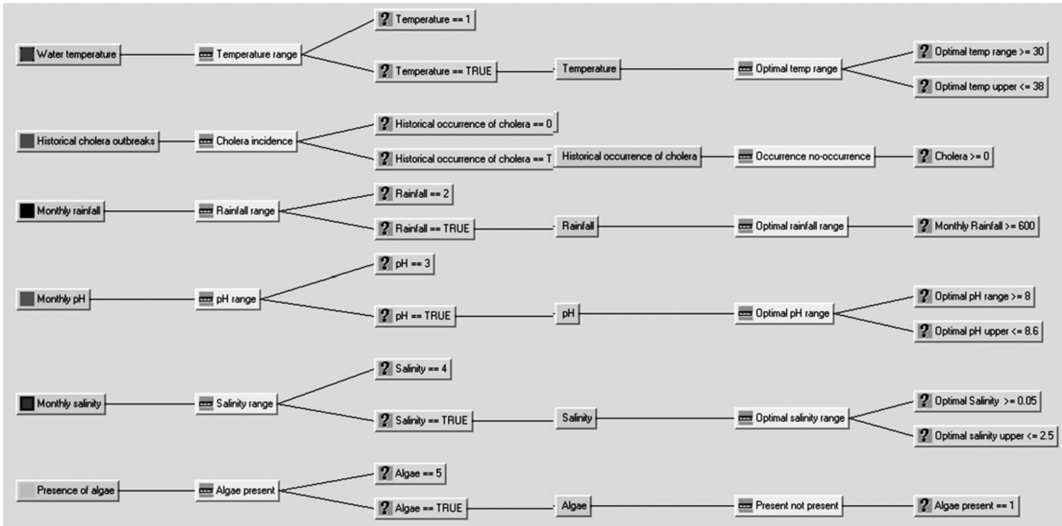

Fig. 5. Expert system design diagram from Erdas Knowledge Engineer.

that the bacteria are closely associated with the occurrence of algae in the water. It was also assumed that environmental factors, for example an increase in water temperature and rainfall, are linked, directly or indirectly, to cholera blooms. The relative importance of all the variables used was assumed equal. The rate and magnitude of spread of cholera after the first case is reported, depends mainly on socio-economic factors. The rational for these assumptions are based on the results of the literature review. From the literature it became clear that changing environmental conditions favour disease

Table 1
Optimal values and ranges for variables identified to be important for an algal or cholera bloom to occur in the natural environment

Variable	Range	Optimal value
Occurrence of cholera in the past		In epidemic form, isolated cases are poor indication of endemic reservoir in the natural environment
Average rainfall (mm/month)	>600	
Mean maximum daily surface temperature (°C/day)	30–38 °C	37 (temperatures below 15 °C reduce growth and survival rates significantly)
Number of consecutive 'hot' months overlapping with the rainy season	1–4	>1 month
Salinity for growth purposes (total salts, %)	0–45	Values between 5 and 25% considered to be optimal
Salinity for expression of toxicity (total salts, %) (Häse and Barquera, 2001). Derive from conductivity.	0.05–2.5	Values between 2 and 2.5% considered to be optimal
pH	8–8.6	8.2 (value below 4.6 and when combined with low temperatures reduce growth and survival rates significantly)
Fe^{+} (soluble or insoluble form)	Must be present (moderate amounts)	Low < 0.1 Moderate = 0.1 to 0.5 High > 0.5
Presence of phytoplankton and algae	Direct and indirect link with phytoplankton: factors driving algal growth during the growth season are similar to bacterial growth and survival factors. The availability of dissolved organic material can support sub-optimal or poor survival conditions, e.g. close to 0% or 45% salinity, increase in pH due to the uptake of carbon during photosynthesis	
Presence of zooplankton	The simple presence of crustacean copepods enhances the survival of *V. cholerae* 01	
Dissolved oxygen daily cycles for every month of the year (mg/l)	Daily fluctuations provide a preliminary indication of algal blooms	
Oxidation–reduction potential daily cycles for every month of the year		

outbreaks throughout the world by either increasing the prevalence and virulence of existing diseases or facilitating new diseases (Colwell and Huq, 1994; Harvell et al., 1999; Codeço, 2001; Huq et al., 2001; Sack et al., 2003). For example, climate variability and human activity appear to play major roles in driving disease outbreaks and epidemics by undermining the host resistance and facilitating pathogen transmission. The biophysical environment, for example changes in climatic conditions, affects ecosystem functioning over extended temporal and spatial scales. Pathogens are transported globally through human activity.

4. Results

The bio-complex nature of cholera together with the lack of detailed environmental data makes it difficult to accurately predict cholera outbreaks in a specific geographic area. The use of a 'simple' simulation model that was built on the understanding gained from discussions with experts and the literature review helped to develop a framework for the expert system. The simulation model proved useful in building scenarios based on changes in the values quoted in the literature for the variables identified. Non-linear functions to link the relative importance of the different variables were used to create these different scenarios.

Interim results produced by the Expert–GIS model show relative long-term risk. The model currently does not predict locations and times based on actual environmental conditions. Further work will incorporate remote sensing data that can supply input surfaces for some of the variables, for example surface temperatures of water sources large enough to be detected on satellite images, phytoplankton and algal blooms and spread. Field measurements will be needed for daily temperature and rainfall data as well as diurnal data on dissolved oxygen, oxidation–reduction potential, salinity levels, pH and the presence of the bacteria.

5. Discussion and conclusions

Once daily weather data are incorporated, predictions of past cholera outbreaks will be tested against reported statistics from those outbreaks. Once operational, the model will be run daily to forecast algal blooms. If an algal bloom is forecast, the time lag between the time of an algal bloom and the date of the first case of cholera reported will be built into the prediction model to predict the date of a possible outbreak in a given area. The time-lag factor will be based on historical data and derived from the difference in time between an algal bloom and the date of the first case reported where it is sure or known that human contact was not the only possible source of cholera. A time-lag factor varying between 14 days to about 84 days was assumed for the simulation model. Socio-economic models will be triggered that can highlight the communities that are most likely to be at risk. These socio-economic models will be applied together with scenario analysis models to estimate the effects of various interventions.

An integrative approach using different modelling techniques is an effective way to model complex integrated biological and socio-economic problems. The expert system proved an effective tool to capture expert information and modelled or simulated results. The spatial characteristics of an environmental disease such as cholera are best modelled in a GIS environment combined with the inputs generated by other modelling techniques such as expert systems and fuzzy logic.

The outputs generated by any model or combination of models are as good as the data that were used as inputs. Further research work involving field verification and the use of remote sensing technology will fill some or most of the current data gaps.

Intensive data collection is needed to populate, test and verify the outcomes of the model. This will form part of future research efforts to update the model and make it more applicable to specific areas. Based on the success of the research conducted on cholera in Bangladesh, future research work will include the use of remote sensing to detect *V. cholerae* by indirect measurement (Lobitz et al., 2000). Remote sensing data will be used as input into the fuzzy GIS–Expert system model. Funding has been awarded to include the coastline of Mozambique and Beira as part of the study area and to incorporate remotely sensed data and field measurements of various factors. Ultimately the prediction model, if successfully developed, will be linked with spread and risk models, i.e. models that incorporate the social risk factors that contribute to the spread of cholera from the natural source of cholera to humans and from one area to the next.

Acknowledgements

The project was part of a study funded internally by the CSIR, for which the authors would like to express their sincere thanks, which are extended to the GIS unit, KwaZulu-Natal Department of Health, Pietermaritzburg for the use of their epidemiological data.

References

Ali, M., Emch, M., Donnay, J.P., Yunus, M., Sack, R.B., 2002. Identifying environmental risk factors for endemic cholera: a raster GIS approach. Health & Place 3, 201–210.

Codeço, C.T., 2001. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. BMC Infectious Diseases 1 (1).

Collins, A.E., 2003. Vulnerability to coastal cholera ecology. Social Science and Medicine 57, 1397–1407.

Colwell, R.R., Huq, A., 1994. Environmental reservoir of *Vibrio cholerae*, the causative agent of cholera. Annals of the New York Academy of Sciences 740, 44–54.

Cottle, E., Deedat, H., 2002. The cholera outbreak: a 2000–2002 case study of the source of the outbreak in the Madlebe Tribal Authority areas, uThungulu region, KwaZulu-Natal. Health Systems Trust, <http://www.hst.org.za>.

Harvell, C.D., Kim, K., Burkholder, J.M., Colwell, R.R., Epstein, P.R., Grimes, D.J., Hofmann, E.E., Lipp, E.K., Osterhaus, A.D.M.E., Overstreet, R.M., Porter, J.W., Smith, G.W., Vasta, G.R., 1999. Emerging marine diseases – climate links and anthropogenic factors. Science 285 (5433), 1505–1510.

Häse, C.C., Barquera, B., 2001. Role of sodium bioenergetics in *Vibrio cholerae*. Biochimica et Biophysica Acta 1505, 169–178.

Huq, A., Sack, B., Colwell, R.R., 2001. Cholera and global environmental changes. In: Aron, J.L., Patz, J.A. (Eds.), Ecosystem Change and Public Health: a Global perspective. John Hopkins University Press, Baltimore, Maryland, pp. 327–352 (Chapter 11).

Iliadis, L.S., 2005. A decision support system applying an integrated fuzzy model for long-term forest fire risk estimation. Environmental Modelling & Software 20 (5), 613–621.

Kawano, S., Huynh, V.N., Ryoke, M., Nakamori, Y., 2005. A context-dependent knowledge model for evaluation of regional environment. Environmental Modelling & Software 20 (3), 343–352.

Lipp, E.K., Huq, A., Colwell, R.R., 2002. Effects of global climate on infectious disease: cholera model. Clinical Microbiology Reviews 15 (4), 757–770.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faroque, A.S.G., et al., 2000. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. Proceedings of the National Academy of Sciences, USA 97, 1438–1443.

Marsili-Libelli, S., 2004. Fuzzy prediction of the algal blooms in the Orbetello lagoon. Environmental Modelling & Software 19 (9), 799–808.

Metternicht, G., Gonzalez, S., 2005. FUERO: foundations of a fuzzy exploratory model for soil erosion hazard prediction. Environmental Modelling & Software 20 (6), 715–728.

Sack, R.B., Siddique, A.K., Longini Jr., I.M., Nizam, A., Yunus, Md, Islam, M.S., Morris Jr., J.G., Ali, A., Huq, A., Nair, G.B., Qadri, F., Faruque, S.M., Sack, D.A., Colwell, R.R., 2003. A 4-year study of the epidemiology of *Vibrio cholerae* in four rural areas of Bangladesh. The Journal of Infectious Diseases 187, 96–101.

World Health Organisation (WHO), 2000. WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases. WHO/CDS/CSR/ISR. 2000.1, Geneva. Web version <http://www.who.int/csr/resources/publications/surveillance/CSR_ISR_2000_1web/en/>.