FUZZY EXPERT SYSTEMS AND GIS FOR CHOLERA HEALTH RISK PREDICTION IN SOUTHERN AFRICA

Authors

Graeme McFerren , Marna van der Merwe , Gavin Fleming, Kerry Murphy,

Abstract

The South African Council for Scientific and Industrial Research (CSIR) is investigating technologies to predict health risk in line with national priorities. Cholera cases are commonly reported in the region, despite the funnelling of extensive resources into ameliorating the problem. It is possible that this results from a lack of foresight as to the likely outbreak locations and times. CSIR has been designing an early warning GIS prototype tool using ArcGIS and Erdas Imagine, aimed at identifying favourable pre-conditions for cholera outbreaks. The model behind the tool is driven by boolean algebra and fuzzy logic approaches that interrogate data patterns of spatial layer attributes. These attribute values determine the kind of algorithm to utilise and to directly interpret the risks. This tool relies on the capture of expert knowledge and historic data that integrates climatic and biophysical parameters with socio-economic parameters to produce a fuzzy surface of cholera outbreak risk potential.

Paper Body

1. Introduction

This paper provides relatively detailed background information about the subject of cholera in South Africa to describe the characteristics of the disease in the region. With this context in place, this paper explores the application of the various techniques and approaches behind the model.

2. Background

Project description

The main aim of the project from which this paper output stems, was to develop an integrative modelling approach or concept using GIS and remote sensing software products form ESRI and Leica Geosystems and to demonstrate its applicability to a 'real' world problem like cholera. A model was developed based on the capturing of knowledge gained from literature reviews and discussions with experts together with historical data on cholera outbreaks and environmental conditions in South Africa. 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). It is envisaged that remote sensing data will be used as input data into the fuzzy GIS-Expert system model. Funding processes are currently being explored to include the coastline of Mozambique, particularly Maputo Bay, as part of the study area and to incorporate remote sensing data. This paper focuses on the prediction model for environmental eruptions of *V. cholerae* and will be linked with spread and risk models in future iterations, 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.

The project did not have as its aim to prove or disprove any hypotheses or theories regarding the occurrence and extent of cholera or an endemic reservoir of cholera in the study area. The model is based on certain assumptions about the conditions that trigger a cholera outbreak. Only one province in South Africa, namely KwaZulu-Natal, was included as part of the study as most of the cholera cases reported during the 2000 outbreak occurred in this province. KwaZulu-Natal is also the only province for which a GIS-based surveillance system was developed. This system is operated by the GIS unit of the Department of Health in KwaZulu-Natal.

Cholera described

Cholera (frequently called Asiatic cholera or epidemic cholera) is a severe diarrhoeal disease caused by the bacterium *Vibrio cholerae*. A patient may lose up to a litre of fluid per hour for several hours leading to extreme dehydration and even to death within fewer than 24 hours if left untreated. It is a highly infectious disease and transmission to humans is mainly through the consumption of and by other exposure to contaminated water and/or food.

History of cholera

Cholera epidemics have affected human populations throughout history. Recorded evidence of cholera epidemics in India dates back more than 440 years. In the nineteenth century, cholera spread from its apparent ancestral site in the Orient to other parts of the world, causing pandemics in Europe. The first pandemic was recorded in 1817. In the seventh pandemic, which has not yet receded, cholera type O1 El Tor originated in Indonesia during 1961. It spread rapidly to Asia, Europe and Africa and reached South America in 1991. Cholera is now endemic in many parts of the world.

The importance of cholera as a public health issue on a global and national scale

The severity and extent of cholera is recognised on a worldwide scale by governments and international organisations such as the World Health Organisation (WHO). It is listed as one of three internationally quarantinable diseases in the WHO's International Health Regulations, 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. It is also one of the most researched infectious diseases in the world (Collins, 2003).

Estimates based on recorded and estimated non-recorded values put the total number of cholera cases in the world during the 1990's in the order of 6 million infectees (Collins, 2003). Most of the new cases reported since the start of the new millennium are from Africa. Countries in the eastern and southern parts of the continent, for example South Africa, Mozambique, Zambia and Zimbabwe are those mainly affected.

As early as 1971, certain parts of South Africa were considered to be at risk because of hot and humid climatic conditions and socio-economic factors. South Africa experienced major cholera outbreaks during 1980 to 1984 and further outbreaks have occurred since August 2000. More than 22,000 people were infected by a cholera outbreak during the 1980's in the KwaZulu-Natal province. In the period between August 2000 and February 2002, the disease infected 113,966 people (more than 70% of the total cases reported in the country) and claimed 259 lives in the province (Cottle and Deedat, 2002). These numbers are most likely to be underestimates as not all cases are reported. The 2000 outbreak was one of the worst cholera epidemics in the country's recent history. Initial reports of the cholera outbreak 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 (see Figure1). The northern KwaZulu-Natal cholera outbreak soon spread to eight of South Africa's nine provinces with varying levels of intensity. Since the start of 2003, most reported cholera cases occurred in the Eastern Cape province, south of KwaZulu-Natal. Most of the 2004 cases have been reported from the Mpumalanga province, close to the Mozambican border.

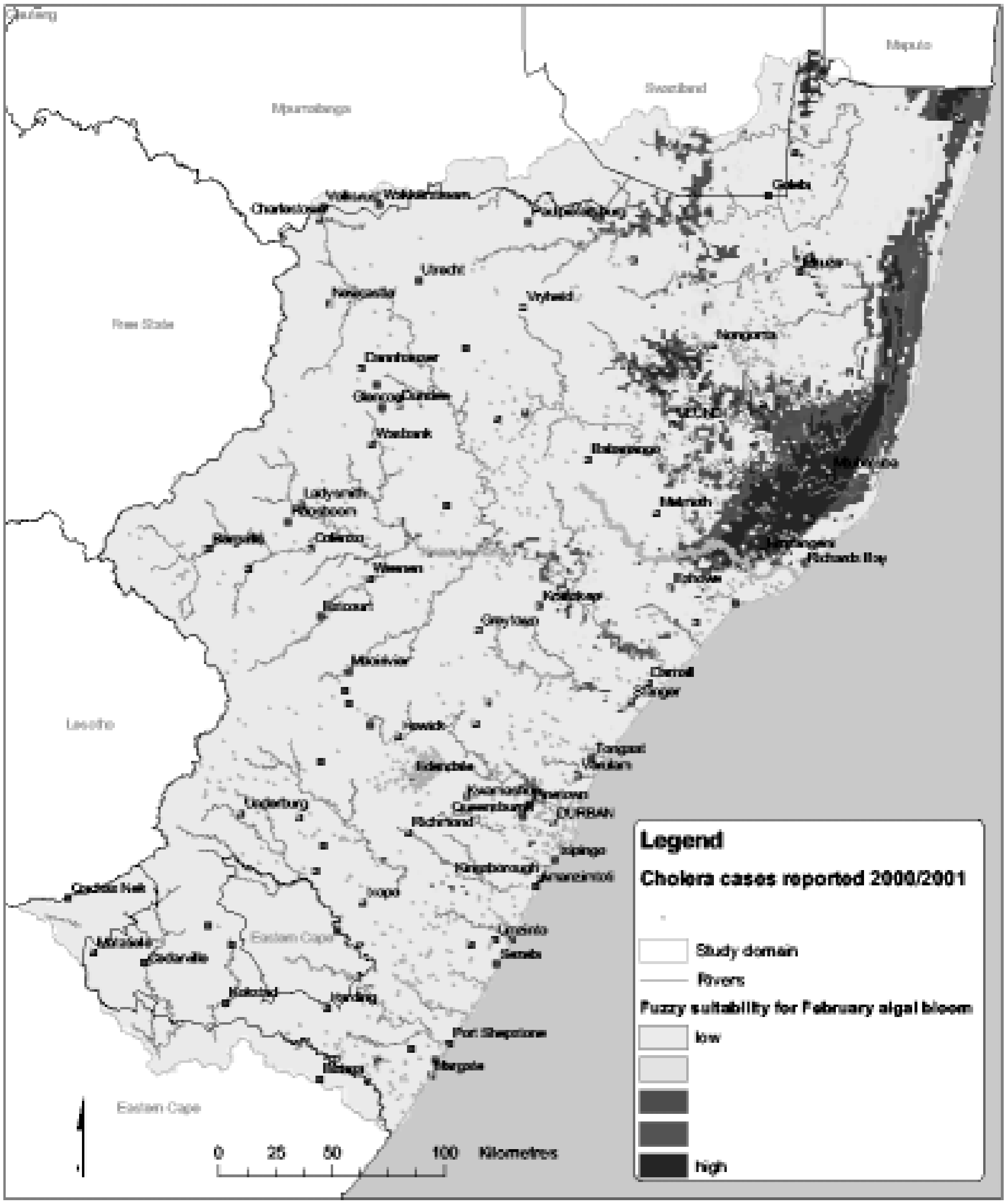

Figure 1. Map showing the KwaZulu Natal river basins, the major rivers, locations of reported cholera cases and a preliminary fuzzy suitability surface for February algal bloom. The river marked in thick grey and exiting south of Richards Bay, is the uMhlatuze river. A large number of cases were reported during the 'hot' months of the year, e.g. February.

(Source: Cholera cases – KZN Department of Health, GIS-unit)

3. Problem statement

Cholera is a complex health problem to manage because of a combination of factors. These vary from biophysical to socio-economic and include conditions that contribute directly and/or indirectly to the occurrence and severity of outbreaks. These factors often have interwoven spatial aspects, which can be examined using spatial models and various forms of spatial data. Cholera is one of the most researched communicable diseases in the world, nevertheless it still has devastating effects on local communities and areas (Collins, 2003). No single solution has been identified yet that can effectively reduce or eliminate the impact of cholera.

Predictions of potential outbreaks are still made on a rough or low-resolution scale. Improved predictions have not been possible because of a number of factors including multiple strains of the bacteria, high levels of case fertilities in some regions of the world and a complex of influences on disease distribution in people and the environment. Predictive models are thus complex.

Progress towards providing cost-effective solutions is also hindered by the amount of uncertainty associated with the question of which is the most effective array of interventions (Collins, 2003). The general approach in many countries to address the problem of cholera is to provide safe water and sanitation to as many people as possible. In case of an outbreak, the focus is to provide care for the sick and to supply safe water where necessary. This intervention can prevent many deaths, but the lack of strong preventative measures means that the disease is still a major public health problem in approximately 50 resource-poor countries (WHO, 2004).

Despite substantial intervention programmes by the South African government, cases of cholera are still reported in different provinces and the total number of cases is increasing. Based on statements made by the WHO, individuals in the national Department of Health as well as published articles and statistics, it became clear that no integrated early warning system for southern Africa exists. Ideally, such a system should provide decision makers with information on the most likely date and location of a potential outbreak before it occurs.

4. Solution

A combination of interventions ranging from safe water supply and sanitation, education and awareness programmes, changes in social behaviour patterns, poverty reduction, to providing decision makers with reliable early warnings on the most likely date and location of potential outbreaks, is needed to reduce the impact of cholera effectively.

The focus of this study was to investigate and develop an integrative prediction model that ultimately can form part of an integrative approach (i.e. applying a range of different intervention types and policy decisions) to reduce the number of cholera cases in the country. The prediction model described in this paper is based on the combination and application of several techniques: a simulation model, an expert system designed in Erdas Knowledge Engineer, spatial analysis performed in the ArcGIS environment and fuzzy logic approaches. This combination is directed towards the environmental health problem of cholera.

Decision makers can use the outcomes of the prediction model to direct and focus their effort ahead of time on the communities that are most likely to be affected by an increase in the total number of *V.cholerae* bacteria in the natural environment, i.e. when bacterial numbers exceed the threshold value for human disease threat.

5. Approach and methodology

Hypothesis

A model calibrated or trained with long-term environmental data and cholera case reporting time series should identify areas where there is a high probability of algal blooms. The same model using daily environmental indicators should be able to predict specific periods and locations of future blooms. When coupled with social components such as the location of communities in close proximity to affected rivers or water reservoirs, probabilities of cholera outbreaks stemming from those algal blooms should be predictable.

Approach

A hierarchical approach was followed in order to understand the ecology and possible spread of *Vibrio cholerae* in the environment (Figure 2). The integrative modelling approach is based on this hierarchical structure.

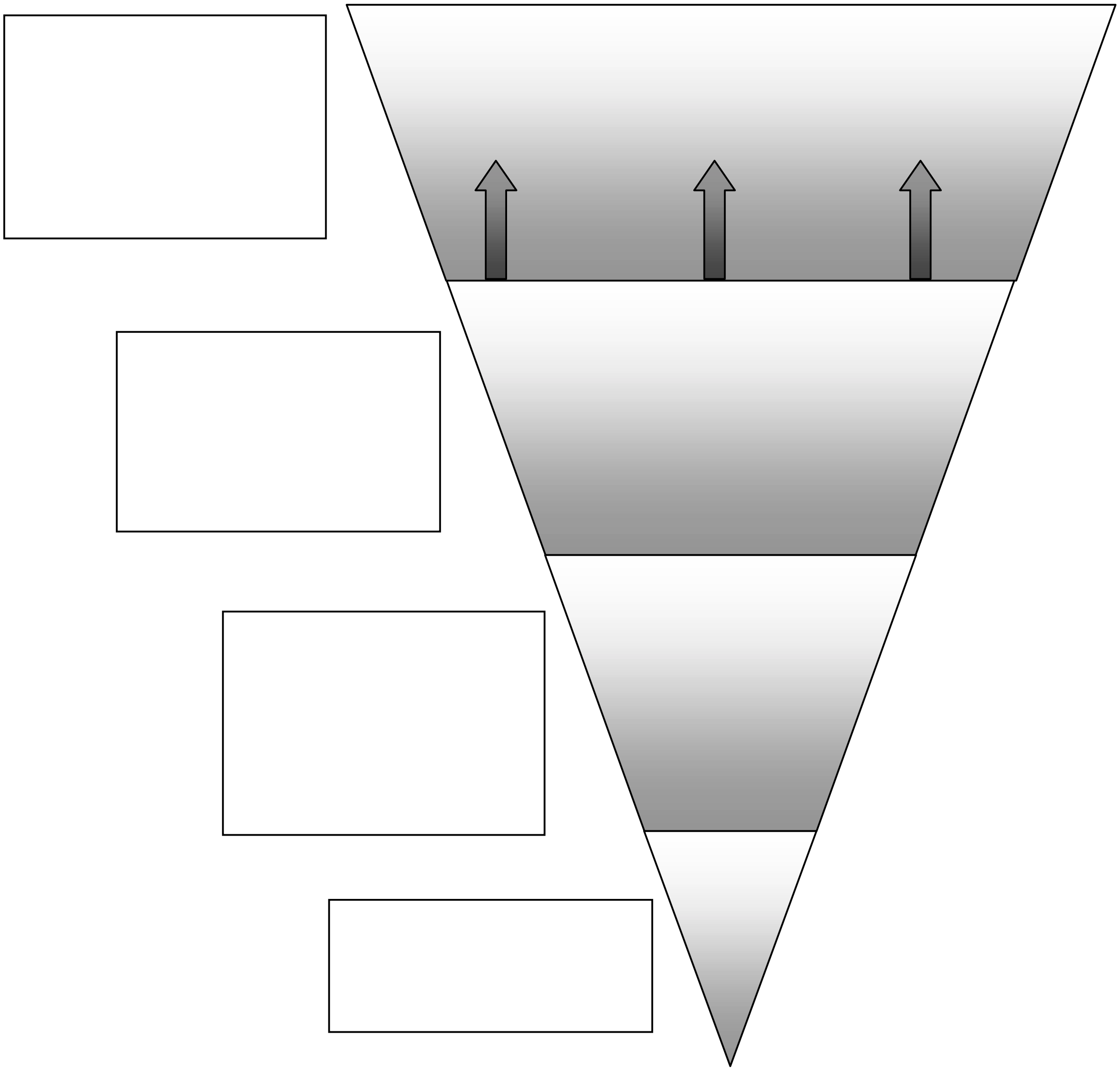

Figure 2. The hierarchical approach to describe the 'ecology' and spread of cholera in the environment (modified from Lipp, Huq and Colwell, 2002)

Methodology

The prediction model is based on the application of a range of different modelling techniques, namely a simulation model, an expert system, a GIS model and fuzzy logic. A flow diagram to illustrate the application of the different techniques is provided in figure 3.

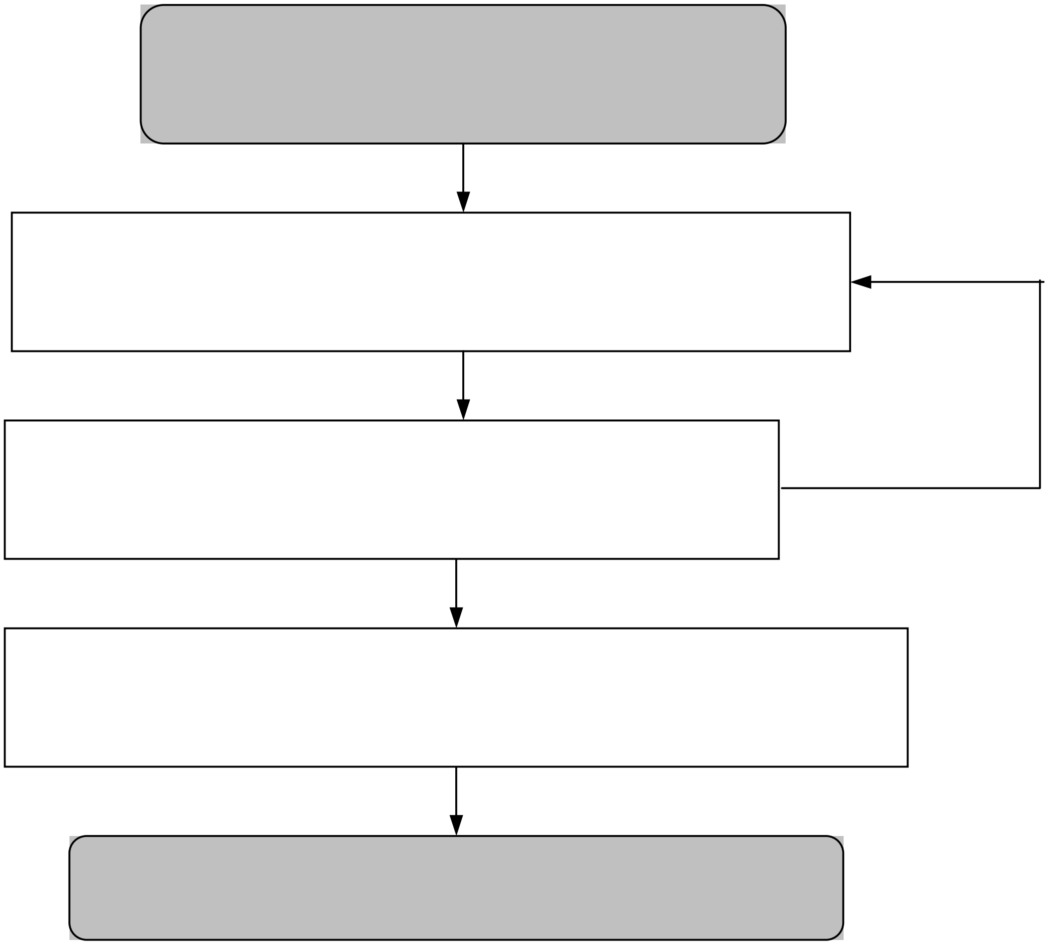

Figure 3. Flow diagram to illustrate the application of techniques used.

The simulation model and an expert system were used to capture and simulate the knowledge and information gained from the literature review and discussions with experts in the field of cholera. The simulation model was developed to ~~provide~~ define some of the rules used in the expert system~~some of the inputs into the expert system~~. The simulation model was based on quoted values for the different variables in the literature. The main aim of the simulation model was to simulate the combined impact of changes in the values quoted in the literature on the growth of cholera bacteria in the environment, i.e. no real data were used. A flow diagram that illustrates the simulation model is provided in figure 4. The simulation model was also used to simulate the relative importance of the different factors that create favourable environmental conditions for algal blooms and hence, cholera blooms to occur.

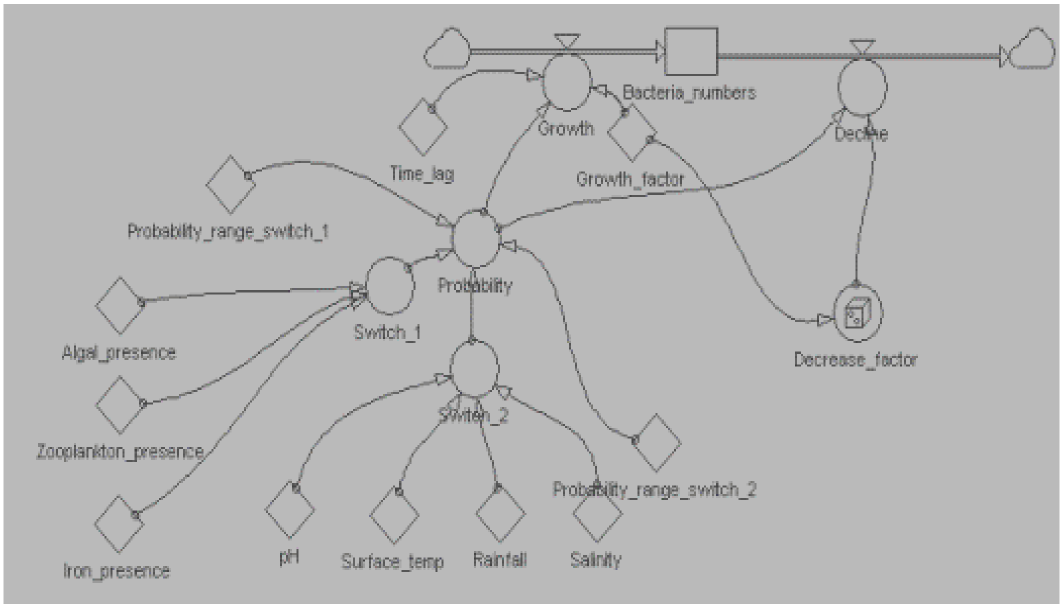

Figure 4. Flow diagram to illustrate the simulation model.

The expert system is built on a number of hypotheses and sub-hypotheses, derived from the outputs of the simulation model and knowledge captured from the literature review. Spatial data (actual data when available and modelled data where no measured data were available) on variables such as water temperature, pH, salinity, monthly rainfall, the presence of algae and finally the location and number of cholera cases were used to proof these hypotheses within in the expert model. The historical occurrence of cholera in a given area was also built into the expert system. This factor is considered important but need to be interpreted together with information whether cholera was brought into the area by cholera infectees and how often cholera cases have been reported in the same area over a period of time. The outputs of the expert system were used to establish the high-level structure and flow of the integrated model. The expert system was developed using the facilities of Erdas Knowledge Engineer and a model design diagram of the expert system is provided in figure 5. The expert system outputs were integrated and implemented in an ArcGIS environment as the final link in the chain, to allow for fuzzy logic characterisation, visualisation and overlay of contextual information. ArcGIS was favoured because of the explicitly spatial nature of the study as well as the ability of the spatial model to efficiently capture the logic followed and structures developed.

The threshold ranges that were used as guidelines in the expert system were derived from the simulation model, literature review and discussions with experts in the field of cholera and algal blooms. Actual data were not available for each of the variables listed in Table 1~~identified.~~ The optimal values or ranges for ~~the individual~~all the ~~-~~variables identified and the ones used as part of the simulation model and expert system are provided in Table 1. ~~The model design diagram of the expert system (Figure 5) must be interpreted together with Table 1.~~

Table 1. Optimal values and ranges for variables identified to be important for an algal and/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	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 5-25% considered to be optimal
Salinity for expression of toxigenity (total salts, %) (Häse and Barquera, 2001). Derive from conductivity.	0.05-2.5	Values between 2-2.5% 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 and/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,	

Variable	Range	Optimal value
	e.g. close to 0% or 45% salinity, increase in pH levels 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		

A large number of the environmental variables have high uncertainties. Fuzzy set membership functions were derived for average annual rainfall and the mean maximum daily temperature on a monthly basis per pixel area. A fuzzy gamma coefficient of 0.75 was created for the month of February (Figure 1). The expert system provided the rules for decision points in the model, i.e. the cut-off points for fuzzy membership functions. Fuzzy logic or a combination of fuzzy sets was applied to conduct the model inference.

A large amount of 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 (2003) 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 (with a certain amount of time lag) are assumed to be linked to algal blooms.

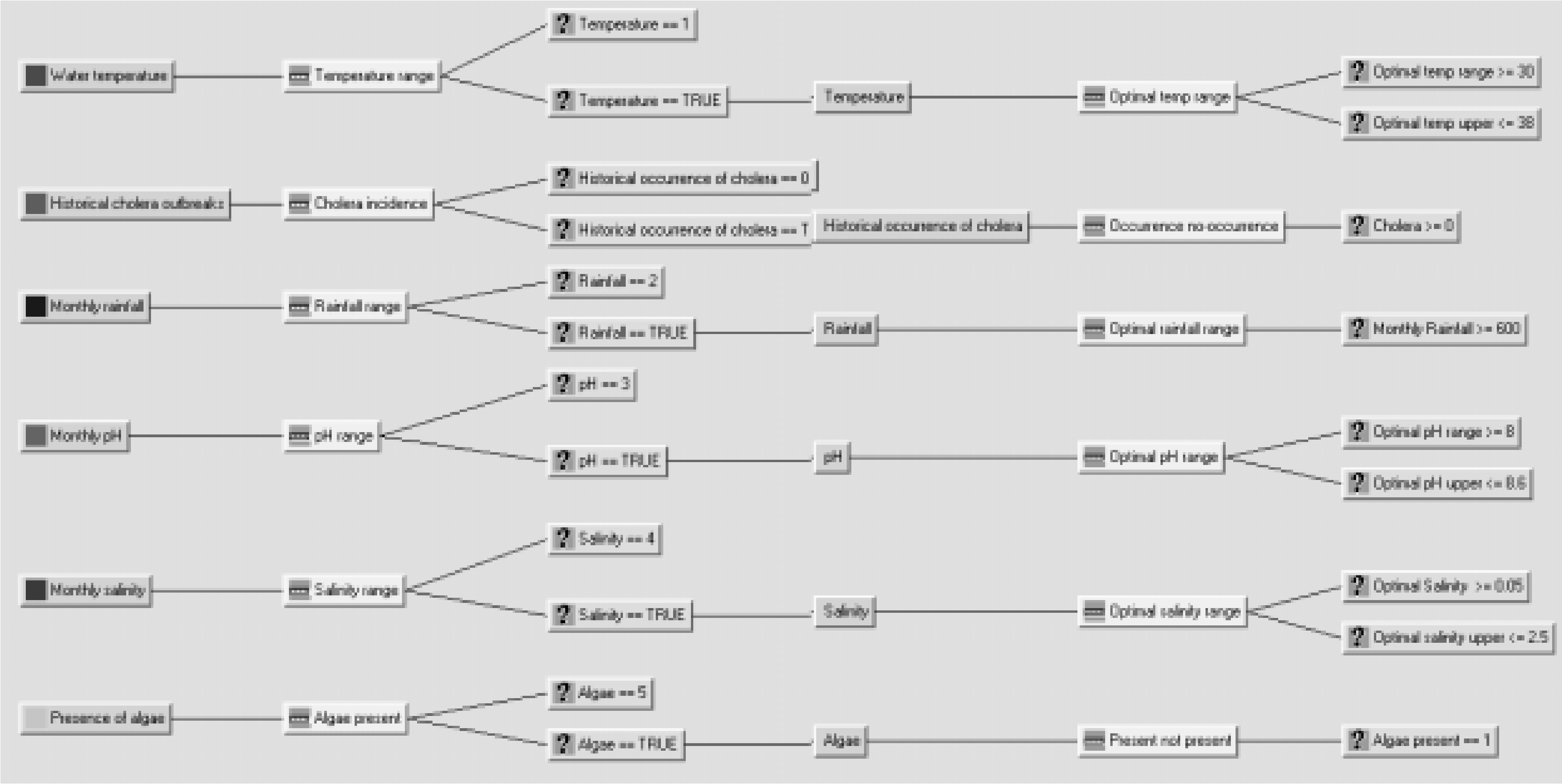

Figure 5. Expert System Design Diagram from Erdas Knowledge Engineer

Assumptions

The prediction model is based on the assumption that endemic reservoirs of *V. cholerae* occur in the study area and that the bacteria is 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, changes in salinity and pH levels 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 rationale for these assumptions is based on the results of the literature review. From the literature it became clear that changing environmental conditions favour disease 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 a*l, 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.

6. Results

The bio-complexity nature of cholera together with the lack of detailed environmental data make 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 to build 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 due to the lack of measured data for some of the variables. 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, and pH.

7. Discussion

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 the amount of 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. Studies are currently conducted in Mozambique and South Africa to collect data and information on the ecology of *V.cholerae* and its survival and growth mechanisms in the natural environment. The outcomes of these studies will provide more information on the magnitude of the time lags that occur between significant changes in the most important environmental variables and the reporting of the first case(s) of cholera in a specific geographic area. These results will be used to verify the time lags calculated using historical data. A time-lag factor varying between14 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.

8. Conclusions

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. Spatial software such as ArcGIS and ERDAS proofed very useful to model, link and visualise the spatial aspects of cholera.

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

Acknowledgements

The project was done as part of a study funded internally by the CSIR. The authors would like to express their sincere thanks to the following groups and people:

- The CSIR managers of the EKM thrust and Terrestrial programme who provided funds to conduct this preliminary study
- GIS unit, KwaZulu-Natal Department of Health, Pietermaritzburg for the use of their epidemiological data

References

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

Collins, A.E. 2003. Vulnerability to coastal cholera ecology. Social science and medicine, Vol 57, pp 1397-1407

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

Cottle, E., and 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. Published by Health Systems Trust: http://www.hst.org.za (2004/02/25)

GIS unit, KwaZulu-Natal, Department of Health, P/Bag X9051, Pietermaritzburg
3200, South Africa

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. Vol 285, Issue 5433, pp1550-1510.

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

Huq, A., Sack, B., and 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, Chapter 11, pp 327-352.

Lipp, E.K., Huq, A., and Colwell, R.R. 2002. Effects of global climate on infectious disease: cholera model. Clinical Microbiology reviews, Oct. 2002, Vol 15, No 4, pp 757-770.

Lobitz B, Beck L, Huq A, Wood B, Fuchs G, Faroque ASG, et al. 2000. Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. Proc National Academy Sciences. Vol 97, pp1438-43.

Marsili-Libelli, S. 2003. Fuzzy prediction of the algal blooms in the Orbetello lagoon. Environmental Modelling and Software (article in press)

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., and 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, 2003:187, pp 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/) (2004/02/12)

World Health Organisation (WHO). 2004. *Mozambique: WHO Launches Oral Vaccination Project Against Cholera*. Posted to the web on 15 January, 2004. http://allafrica.com/stories/200401150558.html (2004/02/20)

Author Information

Title	Forename	Surname	Address	E-mail	Tel	Fax
Mr	Graeme	McFerren	CSIR, Box 395, Pretoria, 0001, South Africa	gmcferren@csir.co.za	27 12 841 3188	27 12 841 2689
Ms	Marna	Van der Merwe	CSIR, Box 395, Pretoria, 0001, South Africa	Mvdmerwe2@csir.co.za	27 12 841 3000	27 12 841 2689
Mr	Gavin	Fleming	CSIR, Box 395, Pretoria, 0001, South Africa	gfleming@csir.co.za	27 12 841 2489	27 12 841 2689
Mr	Kerry	Murphy	CSIR, Box 320, Stellenbosch, 7599, 0001, South Africa	kmurphy@csir.co.za	27 21 888 2506	27 21 888 2682
~~Ms~~	~~Marna~~	~~Van der Merwe~~	~~CSIR, Box 395, Pretoria, 0001, South Africa~~	~~Mvdmerwe2@csir.co.za~~	~~27 12 841 3000~~	~~27 12 841 2689~~