Geovisualisation of Uncertainty in a National Immunisation Register

Julian Kardos, Antoni Moore & George Benwell

Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
Phone: +64 3 479-8301
Email: jkardos@infoscience.otago.ac.nz

Presented at SIRC 2002 – The 14th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 3-5th 2002

ABSTRACT

The New Zealand Ministry of Health (MoH) is currently creating a National Immunisation Register (NIR). The main objective of the NIR is to have a central repository of individual immunisation history for health service providers to access. The MoH are striving towards having 95% of New Zealand children fully vaccinated by 2 years of age. Immunisation coverage rates are a standard way to determine if the 95% target rate has been achieved. Data uncertainties in the NIR could skew coverage rates and should be taken into consideration. Fundamentally, these uncertainties have a spatial component. The aim of immunisation is to prevent the spread of disease, but immunisation coverage rates do not show if the spread of disease has been controlled. Furthermore, coverage only shows the number of people that have been immunised for a particular vaccine. However, the concept of herd immunity (the more people vaccinated, the harder it is for a disease to spread through non-vaccinated individuals in an environment) is used in vaccine programmes to prevent and eventually eradicate the spread of infectious disease. This paper proposes the use of geovisualisation to highlight these uncertainties. A NIR geovisualisation tool portraying uncertainties in immunisation coverage rates and herd immunity for disease has the potential to be a powerful policy making tool.

Keywords and phrases: geovisualisation, uncertainty, immunisation, coverage rates, NIR, MoH

1.0 INTRODUCTION

A goal of the Ministry of Health (MoH) is to create a National Immunisation Register (NIR), a database designed to hold vaccine status records for children. The objective of the NIR is to help increase immunisation coverage rates and localise immunisation data. Similar initiatives in other countries have proven effective (National Health Committee. 1999). Currently, there are a number of problems contributing to low immunisation coverage rates. Some examples are data fragmentation, high family mobility and children living in geographically isolated areas. These are problems with a spatial component. More problems will be outlined later in this paper. The MoH desires to obtain an immunisation coverage rate of 95% for all New Zealand children 2 years of age by 2005 (Herbert 2002). This paper proposes to use geovisualisation on the NIR to help ascertain if this coverage rate is being achieved, through the highlighting of uncertainty.

Geographic visualisation (Geovisualisation (MacEachren 1995a)) can seek important information in sets of data through spatial representations promoting visual thinking and visual communication (MacEachren 1994a) (see section 4.0). Geovisualisation is visualisation of geographic or spatial data stored digitally

(Andrienko et al. 2000) and can be used to turn raw data into a geo-visual representation, ideally assisting in the decision making process. When data are geo-referenced, geovisualisation can help prompt hypotheses about patterns and clusters in a spatial context, and therefore the role of location in human and environmental processes (MacEachren 1995b). Geovisualisation can help identify gaps in immunisation coverage, and help evaluate the efficiency of the current method used to increase immunisation coverage rates.

The aim of immunisation is to reduce the spread of a disease and eventually eliminate it. Coverage rates do not show if the spread of disease has been controlled, coverage simply shows the number of people that have been immunised versus those not, for a vaccine. Vaccine programmes are designed to achieve herd immunity; this will help prevent the spread of disease. Herd immunity is defined as:

> *"... the resistance of a group of people to attack by a disease to which a large proportion of the members are immune, thus lessening the likelihood of a patient with a disease coming into contact with a susceptible individual."*

(Snowdon 2000)
(http://vaxchoice.crosswinds.net/definitions/herd_immunity.html)

Vaccines protect more than the vaccinated individual. They protect society as well. When immunisation levels in a community are high, the few who cannot be vaccinated, (those who have legitimate reasons to avoid immunisation due to allergic reactions, medical conditions etc.,) are often indirectly protected through herd immunity. Vaccinated people, who are protected against a disease, create herd immunity by surrounding unvaccinated individuals minimising the ability for a disease to spread (Anderson and May 1991; Iowa Department of Public Health. 1997). Mass immunisation decreases the amount of hosts, and the reproduction rate for a disease. As an indirect consequence, it is not important (or possible) to immunise every individual in a population to eradicate a disease. Therefore, once vaccine coverage has reached a crucial level, a disease will be unable to maintain a reproductive rate to continue existence in a community (Anderson and May 1991). Anderson and May (1991) have created equations defining the critical proportion of a population to be immunised based a number of factors defining herd immunity. These formulae, used in conjunction with spatial uncertainty modelling in the NIR dataset, will provide an effective overview of the NIR data. This will enable the verification of the immunisation programme and pin point possible disease outbreak areas.

Creating a digital spatial dataset from the original ground data will produce a generalisation of the truth. Spatial data error can propagate due to technological shortcomings of the sampling technique; typically real phenomena are continuous data points, but only discrete data are collected forming a subset of the actual phenomena. The collected data is typically moulded into a form so it can be displayed through a digital medium, like a geovisualisation. Immunisation data possibly is discrete, but interpolation will produce a generalisation of the information. Therefore, data inherently has uncertainty and this needs to be expressed in the geovisualisation technique.

Uncertainty can manifest itself in three ways: location, attribute and temporal deviations. The focus of current research is to establish the existence, nature and magnitude of spatial uncertainty in the NIR dataset and then present a geovisualisation method for exploring the uncertainty. Before the uncertainty information can be shown, the uncertainty issues associated with the NIR need to be discovered. These are indicated by some of the issues raised above (data fragmentation, high family mobility, geographically isolated children, etc.) and in sections 2.1 and 2.2. Once these issues are understood an enumeration measure of the uncertainty is needed. With an account of the uncertainty available it becomes possible to generate an optimal geovisualisation technique to represent the uncertainty and highlight immunisation deficient areas. This could be done through a geovisualisation of the NIR dataset showing coverage rates and the herd immunity of a disease including possible uncertainties. It is hoped that there will be differences in the coverage rate and herd immunity geovisualisations, emphasising that herd immunity is a more viable way to understand immunisation data.

This paper is designed to educate the reader in uncertainty issues surrounding spatial data in the context of a NIR. Eventually policy creators will use the NIR data to make decisions; uncertainty issues should be known and could be geovisualised when making these decisions. The benefits of geovisualisation are also discussed, and finally current techniques to illustrate uncertainty are examined.

2.0 DATA LIMITATIONS AND UNCERTAINTIES

Goodchild (1991) describes spatial data accuracy as the relationship between a measurement and the reality that it is designed to represent. Reality (r) is the actual ground truth, not the source document. The source document (s) is a model of reality and allows the creation of spatial information. The accuracy of the data can therefore be described through the equation s/r. An equation to show uncertainty would therefore be $1-(s/r)$. When dealing with large datasets finding an accurate measure of r becomes difficult. The most accurate and complete population dataset is the New Zealand Census information, but this is only collected every five years. As most of the immunisation schedule is within the first two years of a child's life mismatches are likely.

There are also temporal issues associated with comparing immunisation data and Census data. The NIR aims to collect immunisation data on a day-to-day basis whereas Census data are collected every five years. This should be taken into consideration if a comparison of these datasets were to occur.

2.1 Statistics New Zealand Data

A comparison between Statistics New Zealand childbirth data and the NIR could provide a realistic population rate to measure the NIR against, taking into consideration immigration. Intra-migration, or the movement of individuals to different geographic areas within the country is another factor to take into account. Statistics New Zealand collates data from Internal Affairs about New Zealand registered births by quarter for urban areas, geographic regions, and territorial authorities; ethnic data is also collected. In New Zealand, a child's personal data under the Births Deaths and Marriages Registration Act 1995 is collected by appropriate persons and sent to Identity Services (part of Internal Affairs) within five days of the birth (Statistics New Zealand. 2002b). This takes into consideration home births and hospital births. Immigration data is also available and this could be assembled with the current statistical childbirth data providing a realistic birth rate for New Zealand in its geographic regions. Statistical methods are used to collate the information and would typically include computational error techniques (Statistics New Zealand. 2002a). This would provide a realistic population figure for young children in New Zealand to measure against the immunisation count in the NIR.

2.2 Uncertainty in the NIR

Uncertainty in the NIR can come from database errors including, entry of data that is in a valid format but contains incorrect information, the inclusion of multiple records for an individual, or the address of a child not being updated causing their record to be shown in the wrong location. Uncertainty can also come from providers not frequently sending results to the NIR, therefore reducing the consistency and amount of data. Also, if a child does not participate in the immunisation programme, uncertainty will surround their immunisation status. Errors also arise from the lack of an accurate and up to date population data source (e.g. census data) to reference the NIR data to; by comparing an accurate population figure to the NIR data an accurate coverage estimate is obtained. Immunisation data in New Zealand is historically uncertain, it ought to be noted that data accuracy should increase with an immunisation registry; but encompassing 100% accuracy is still a long way off, if even possible. Therefore, instead of viewing the data as if it is correct, uncertainty techniques will help people view the data with its uncertainties.

Data for the NIR will come from births recorded in hospitals. Before a mother is discharged from the hospital, staff will collect and check the relevant child details. The child's mother/guardian will be asked to provide details about the child(s), including who the child's primary health care provider (GP or any other Well Child Provider (WCP)) will be. Currently, an option to opt-off the NIR is planned. This means that the child's immunisation status will not be maintained by the NIR, but their demographic information will be retained. The MoH hopes to retain this demographic data to provide an accurate population measure in the NIR. The caregiver's demographic information for a child will also be retained. This information is passed into the NIR and a message is generated informing the primary health care provider of the birth (Ministry of Health. 2002).

Home births are an issue of uncertainty in the NIR. Since the NIR data only comes from hospitals, there will be a proportion of births that are not recorded. The MoH hope to obtain this data through lead maternity carers (LMC), or midwives that perform home births. The relationship between midwives and GPs has been weak of late, creating barriers between primary health care providers. The MoH are progressing towards creating policy that will ensure LMC will be obliged to provide this information (Ministry of Health. 2002).

Migrant children will be entered into the NIR when they go for their first immunisation event. The child's immunisation status is unknown up until this point (Ministry of Health. 2002).

By informing parents about the benefits of immunisation an increased coverage can be obtained (Prislin et al. 1998). Immunisation information will be given to a pregnant woman booked in with a maternity unit (Ministry of Health. 2002). When an expectant mother is not booked with a maternity unit then she may not receive any information on the benefits of immunisation, therefore potentially decreasing possible immunisation events.

2.3 Relevant Historic Problems Contributing to Low Coverage Rates

Past information about immunisation is fragmented. Estimates are often used, and there is no one reliable source from which to obtain this information. This makes it hard to form any reliable conclusions about past coverage. Some estimates are based on immunisation benefit claim data that Health Benefits (HB) collects when a claim is made. Vaccinators complete a benefit claim form, but this process provides limited data quality. Also, some vaccinators do not claim, due to the low cost of the benefit compared to charging for the service. Attempts to adjust the data for practices that do not claim are made without accurate child population data. A vaccinator has no incentive to provide accurate data, since the primary function of the immunisation benefit claim form is to claim for money, not to provide accurate data, and therefore, vaccine specific information may be incomplete (ESR:Communicable Disease Centre. et al. 2000). Some individual practices, or a collection of practices, gather immunisation information for their local areas. A Rotorua practice association reported 90% immunisation coverage, but the estimate was calculated on children enrolled with providers rather than all children in the area (ESR:Communicable Disease Centre. et al. 2000).

Maori and Pacific children consistently have low immunisation coverage rates. A 1996 immunisation coverage survey in Auckland and Northland concluded that by age two, 44.6% Maori and 53.1% Pacific children were fully immunised compared to 72.3% other children. An immunisation questionnaire outlined several factors contributing to low immunisation in Maori and Pacific. High mobility rates were significant, with Maori recorded as having the greatest mobility with over 30% of families moving house more than once in 2-3 years. Maori and Pacific respondents agreed that 'immunisation injections are too upsetting and painful for young children', and the cost of getting to the doctor is an important factor in getting children immunised. 46.8% of interviewees would rather the child be immunised when visiting the child health nurse instead of at a doctors surgery (ESR:Communicable Disease Centre. et al. 1998). International studies have concluded that socio-economic status is a contributing factor to whether a child is immunised. The studies showed a correlation between children not being immunised on time according to a vaccine schedule, if at all, and poverty (The National Vaccine Advisory Committee. 1999). Also, geographically isolated children are contributing to low immunisation coverage rates due to excessive travelling distances and costs involved in accessing the resource (Edwards 2002).

Children are not being immunised on time according to the immunisation schedule, thus reducing the overall uptake of fully vaccinated kids. Failure to receive the first immunisation on time is strongly associated to the child not being fully immunised at 2 years. This first immunisation is essential; trends show immunisation percentage rates increase later in the schedule (ESR:Communicable Disease Centre. et al. 1998). This correlation is consistent with findings from other studies (Hanna et al. 1994; Northern

Regional Health Authority. 1997). Therefore it is essential that vaccinators are able to identify a newborn child as early as possible and encourage immunisation (ESR:Communicable Disease Centre. et al. 1998).

A lack of parent immunisation education can also be a retardant to coverage rates. If the parent doesn't know or understand the importance of immunisation, there is little drive to have their child vaccinated. An American study found that parents well educated in immunisation and vaccines are more compliant towards vaccinators and medical professionals, consequently producing higher immunisation rates among their children (Prislin et al. 1998). In New Zealand, health promoters are employed to help aid in immunisation education for specific ethnicities, Tiru Ora for example provide Maori healthcare programmes and establish a link of trust between whanau and mainstream service providers (National Health Committee. 1999). The National Health Committee (1999) found that relatively few parents are opposed to immunisation, and in Australia almost all parents believe in immunisation even though they felt uninformed about vaccine side effects. However, improving access, rather than changing parental attitudes, appeared to be the key to increasing compliance. This means that, although people might not take positive steps towards being included in an immunisation program, they also would not object to being included (Edwards 2002).

3.0 SPATIAL DATA FOR DECISION MAKING

When decisions need to be made based on digital information it is helpful to use an information accuracy measure to back up the data used to make the decisions. This will increase the trust associated with digital information, and highlight areas where information is uncertain, creating an entry point to address data accuracy problems. Two types of uncertainties can be identified: accuracy linked to data such as positional, attribute, and temporal error (MacEachren 1994b; Davis and Keller 1997; Hunter 2002); and the accuracy linked to the processing of the data including visualisation, interpolation, and modelling (Heuvelink 1998). These factors need to be considered when the MoH make decisions using the NIR dataset.

3.1 Spatial Data Uncertainty Models

Spatial data uncertainty models can be utilised to help determine the error associated with data, ideally by providing an uncertainty measure. A spatial data uncertainty model will be used to gauge the level of uncertainty in the NIR dataset. Different spatial data uncertainty models are used for different purposes. Some models are ideal when using soil data; like fuzzy set theory (Goodchild 1994b), other models like Monte Carlo are good when dealing with random factors of error (Longley et al. 2001). Geographic data can be represented in many ways, so choosing the right model for the data is important. There are also various models that can represent geographic data uncertainty and it can be challenging to determine the best model to use, also there are no measures to determine if one model is of a higher quality to the next (Ehlschlaeger 2002).

A number of statistical models have been created to describe the uncertainty in information including but not limited to; root mean square error (rmse), percent correctly classified (pcc), mean deviation, and standard deviation. These models can measure the magnitude of the error (Lilburne and North 2002) and are classed as statistical probability models (Davis and Keller 1997). Fuzzy set theory can be applied to represent spatial uncertainty (Goodchild 1994b; Davis and Keller 1997); this allows more than one class to be represented at a particular coordinate mimicking a higher order of truth.

Below is an explanation of three techniques to model spatial uncertainty: Taylor Series and Monte Carlo followed by fuzzy set theory. These are possible models that could be used for the NIR dataset.

3.1.1 *Taylor series procedure*

The Taylor Series method is an approximate uncertainty model and it is good to use as an initial exploratory method. The theory of Taylor Series is based on statistical error propagation, producing a simple, fast and computationally efficient uncertainty model. The down side is there is little control over whether approximations are acceptable. Taylor Series determines an output error for an entity based on a function associated with each input attribute and its error. All of the input values are linked through an arithmetic relationship, normally addition. Once the function is applied an error figure can be obtained, this can vary depending on the arithmetic relationship used (e.g. addition, multiplication, subtraction, division, exponentiation) and if the variables are correlated or not. If the variables are correlated a function variable can be added to derive a new error value. (Burrough and McDonnell 1998)

3.1.2 *Monte Carlo procedure*

The Monte Carlo method is a simple approach to modelling statistical errors in data using ordinary statistics and random variables. Monte Carlo simulations can help identify the error associated with an attribute using an arithmetical operation. The assumption is that each attribute's error has a Gaussian (normal) probability distribution function (PDF), and by continuously repeating the arithmetical operation the PDF variation will be removed. The normal bell shaped distribution is a standard approach to measuring variation; other distribution functions could also be employed. The final results for an attribute after running a Monte Carlo procedure will be a mean error and standard deviation value (Burrough and McDonnell 1998; Fisher and McGwire 2001). This procedure can provide information about how possible errors in the data can affect the results of numerical operations in different parts of a geographic area (Burrough and McDonnell 1998). Monte Carlo is good for obtaining high accuracy and is used for exact values, although it is time expensive due to the iterations involved for each entity, normally more than one hundred. A Monte Carlo simulation would be a good approach to analysing error in the NIR because of its straight statistical methodology, and the use of a Gaussian distribution is ideal when dealing with a large number of random factors contributing to error (Longley et al. 2001).

3.1.3 *Fuzzy set theory*

Fuzzy set theory uses numbers between zero and one to represent a degree of membership to a particular class (Zadeh 1965). Partial memberships allowed in fuzzy sets are well suited to represent uncertainty, and allow for more than one class to be represented in terms of a membership (Goodchild 1994b; Davis and Keller 1997). For example a fuzzy index of 0.75 at a particular coordinate on an immunisation coverage map (immunised/not immunised) may indicate that the data gathering technique is not absolutely certain, or that the immunisation status at that location shows most characteristics as “immunised” and some towards “not immunised”.

Incorporating a spatial data uncertainty model into the geovisualisation procedure can allow an uncertainty figure to be derived. Next is an explanation on why geovisualisation is an excellent way to collate the NIR data, and some benefits geovisualisation can render.

4.0 GEOVISUALISATION

A lot of data used in important decision-making is stored as text; the reader must perform their own visualisation if they wish to discover relationships or patterns in the data. Using geovisualisation spatial patterns and trends that are not noticeable when the data is viewed as text can be quickly discovered. Geovisualisation is primarily used for viewing groups of data, such as a collection of records or a large amount of search results (Fabrikant 2000). An analyst strives to understand the data; a graphical presentation of data can prompt the discovery of important traits, patterns, trends and relationships without any prior assumptions (Andrienko et al. 2000).

The human's primary medium for information input is visual perception (Ware 2000). Information should be represented in a form that best matches human capabilities (Fabrikant 2000). For example, a passage where an author explains the details of a character:

> *“His face looked weathered and was hard as leather under the oilskin hat. He had a large frame, one comparable to a washed up wrestler”.*

The reader will invoke some imagination and create a mental picture of the character, but never to the extent an image of the individual can portray (Ware 2000). The ability for a human to rapidly perceive imagery makes geovisualisation an excellent tool to discover knowledge from a variety of data sources. It has been shown that humans have powerful visual thinking abilities, and physical, spatial, and visual representations are easier to learn, understand and communicate than abstract numeric or textual information (Fabrikant 2000).

> *Even the simplest information retrievals can become overwhelming. Most of the information comes in text form, in different styles, for various purposes. The contents often have fuzzy and incomplete messages, are interrelated but not directly, and provide a literal pile of paper to read.*

(Pagh 1999)
(http://www.pnl.gov/infoviz/spire/spire.html)

Literature indicates that spatial capability is an effective and efficient medium when performing common information retrieval tasks (Egan 1988; Vicente and Willeges 1988; Hook et al. 1996).

> *Visualization provides an ability to comprehend huge amounts of data. The important information from more than a million measurements is immediately available.*
>
> *Visualization often enables problems with the data itself to become immediately apparent. …Visualizations can be invaluable in quality control.*

(Ware 2000) p.2

> *If our extraordinary skill in perceiving the information inherent in the environment can be applied to data visualization, we will have gained a truly powerful tool.*

(Ware 2000) p.35

Geovisualisation provides a high-level view of all the data at once showing patterns and relationships, visual recognition of content similarity by spatial cues, scalability to millions of data points without any change to interface, point and click retrieval, attractive and visually simple display that encourages exploration, and various tools to explore maps in detail (Dodge 2000).

4.1 Static and Dynamic Variables

Static variables defined by Bertin (1981) can show trends and patterns and are fundamental when creating a geovisualisation. With the proliferation of computers, maps have evolved from static topics of interest into dynamic maps that promote exploration of multiple perspectives (Steiner et al. 2001). Dynamic variables in a NIR geovisualisation can provide another axis through which to project information. These variables are relatively new in geovisualisation; DiBiase et al. (1992) provided some of the first research into dynamic variations using map visualisation. DiBiases research identified three key variables when using time as a variable in cartography.

4.1.1 Duration

The duration identifies the number of units of time a scene is displayed; therefore scene duration is treated as a design variable, not applicable to static maps. This can provide emphasis to parts of a visualisation and does not have to be modelled to reflect the actual temporal collection of the data.

4.1.2 Rate of change

The rate of change is influenced by two factors, the magnitude (m) of change in position and attributes between scenes and the duration (d) of each scene. The rate or change is a proportion, m/d. Therefore m can be controlled by selecting a more or less frequent scene interval; this will be dependant on the sampling interval used to generate the scenes. The greater the magnitude of change between scenes the more abrupt the animation becomes. If d is decreased then the slower the rate of change becomes.

4.1.3 Order

The order in which the scenes are presented can be manipulated. In some instances the manipulation of the chronological order of the scenes can show different patterns and traits not usually obvious.

4.1.4 Phase

MacEachren (1994c) also discusses phase as a variable in dynamic maps. Phase implies ‘a rhythmic repetition of certain events’, allowing the length of time between repetitions to be controlled.

4.2 Patterns and Relationships

Geographic datasets contain huge amounts of patterns and relationships that only become apparent when visualised. Some patterns are subtle whilst others are obvious, but it is essential to see and understand these patterns before making applied decisions (Gahegan 2000). Also, by utilising dynamic variables, a geovisualisation highlighting areas of interest can be manipulated. This will display the information in an extent that will have added benefits to the end user. A combination of the static and dynamic variables can be combined to emphasise attributes or relationships in features (DiBiase et al. 1992). For example, each class can be shown for a particular duration, with the size of the symbols representing the magnitude of the data.

5.0 GEOVISUALISATION TECHNIQUES FOR UNCERTAINTY

Once an uncertainty measure has been extracted from the NIR data using an appropriate error model, it is useful to visualise this uncertainty in an easy and logical format. The uncertainty can be extracted using probability theory and if used in conjunction with error statistics, an effective geovisualisation of results can be achieved (Hunter and Goodchild 1995). There are many different levels of uncertainty; first, data errors can be associated to location, attributes or time (MacEachren 1994b; Davis and Keller 1997; Hunter 2002), second, aggregation will produce spatial uncertainty (MacEachren 1992; Heuvelink 1998). Uncertainty can also propagate near the boundary between classes and when classifying pixels, Goodchild (1998) proposes viewing class boundaries as a mix of shades or colours, and classifying using fading colours and the “grey area” metaphor. Recent research conducted at the International Institute for Geo-information Science and Earth Observation (ITC) by Hengl et al. (2002a) into techniques to visualise fuzziness and uncertainty in natural resource data have described novel ways to view a map and its associated uncertainty. The techniques in 5.1 and 5.2 are described in Hengl et al. (2002a). Other geovisualisation techniques for uncertainty found in the literature are also outlined.

5.1 Pixel Mixture

The pixel mixture technique originally created by De Gruijter et al. (1997) modelled soil distribution and helped to visualise the data at an appropriate level of aggregation using fuzzy membership techniques. Each pixel is divided into a grid e.g. 3x3, creating sub-pixels. Fuzzy membership values for each class are derived from a formula for each pixel. The proportion of values the membership function calculates is given to the sub-pixels. For example, if the classes were: “Fully Immunised on Time (FIT)”, “Not Immunised on Time (NIT)”, and “Not Immunised (NI)” the following could occur. If the derived fuzzy membership values for a pixel *x*, has the values FIT = 0.6, NIT = 0.2 and NI = 0.2. As Figure 1 shows, pixel *x* will produce the following sub-pixels, approximated to fit into the grid and classified. The classes are randomly distributed about the sub-pixels to avoid creating patterns.

NIT	FIT	FIT
NI	FIT	NIT
FIT	FIT	NI

Figure 1. Pixel x classified into sub-pixels proportionate to the derived fuzzy membership values.

This technique creates a map without defined boundaries, invoking a more realistic view of data.

5.2 Colour Mixture

Colour mixture uses the saturation effect of colour to visualise uncertainty. The more saturated a colour representing a particular class, the more certain the data is on that part of the map. As shown here: *http://www.itc.nl/personal/hengl/VMM/*, a colour wheel is used to define the legend with a change in hue representing a different class and the fuzziness between them. As a class projects towards the centre the saturation levels drop proportionally to the amount of uncertainty derived from a fuzzy membership function. This creates a multi-dimensional uncertainty chart, with fuzziness between classes and uncertainty in the data represented by hue and saturation respectively.

5.3 Viewing Uncertainty using Animation

Animations are becoming a more popular way to view information, adding another dimension, time to cartographic displays. Temporal information can be used in: (1) a flyby scenario where it is possible to perceive spatial change, (2) expressing the data in chronological order, or (3) manipulating the order based on a category (DiBiase et al. 1992). When the information shown in the animation is based on an uncertainty enumeration, a flyby may be a good way to change the viewpoint of the observer, helping them to see the information from a different perspective. Chronological ordering of animation is a good way to understand trends and patterns in data that occur over time. For example: *http://www.outragegis.com/animations/population-growth.htm* shows population modelling of growth over time. Category based animation provides more flexibility, in essence giving a category its own axis, therefore, showing patterns and trends in a particular category.

Fisher (1994b) explains that animation can also be used to show areas where data is uncertain, providing a map where the individual pixels constantly change. Realisations (data possibilities generated using Monte Carlo simulation) for each pixel are generated to determine the accuracy of the pixel classification. The temporal dimension in this example is used only by an uncertain pixel that changes between the different classes it could belong to. This highlights areas where data is uncertain and gives an indication towards the different classes that are present in the area, using animation.

Ehlschlaeger et al. (1997) proposes viewing uncertainty map realisations as a movie to help highlight areas where data is considered to be uncertain. This technique uses probability theory to derive uncertainty measures. The study explained that animation is a good exploratory analysis, and can assist in communicating results and clarifying important points. The changes in the realisations are easily perceived through animation, and if there is little change, then one can be fairly convinced about the extent of the uncertainty. This study showed that it is more time consuming to perceive uncertainty statically than dynamically. Interpolation is needed to smooth the transition between realisations. The non-linear interpolation method presented kept the intermediate images within the distribution from which the realisations were drawn (Ehlschlaeger et al. 1997).

5.4 Hearing Uncertainty using Sound

Human communication is primarily done through speech. Speech and sound can therefore provide an ideal medium to transfer uncertainty information. Sound can provide another dimension to a visual map, and reduce visual clutter. A study by Bly (1982) found that sound is a viable means to represent multivariate, time-varying data especially in tandem with graphic displays. Following typical cartography procedures, uncertainty figures will be represented graphically, but maps are designed to impose order and provide a representation of variables, rather than define confusion (or uncertainty) (Krygier 1994). Sound can provide a level of uncertainty at a particular point or location on a map through a variable pitch. For example, a low pitch sound could depict low uncertainty and a high pitch sound for large uncertainty (Fisher 1994a). This allows the preservation of a sharp image for the user to see, and provides access to data quality information, if and when needed (Krygier 1994).

5.5 Blinking Pixels

A number of researchers have utilised the abilities of the computer to manipulate data in the spatial display causing it to blink, hence highlighting those areas to the viewer (Fisher 1993; Monmonier and Gluck 1994;

Evans 1997). Fisher (1993) used blinking pixels to show uncertainty in soil mapping; the duration of the blink indicated the probability of class inclusion. Evans (1997) studied various methods to depict spatial data reliability and the benefits and acceptance from a users point of view. Combining data and reliability information in a static or animated map was found to be more effective than interactive "toggling" between the data and reliability information. Monmonier and Gluck (1994) had four user groups assess flashing on a map of the USA highlighting bivariate correlation. Members of all four groups found the flashing offensive, but some found that it provided rapid comprehension of the categories and symbols.

5.6 Fog Metaphor

Using the "fog" metaphor, uncertain parts of a map can become partially hidden making the areas unclear to see. The thicker the fog the more uncertainty is in that part of the map. Computer displays can make it possible to create what looks like fog passing through an "atmosphere" (MacEachren 1992).

5.7 Image Sharpness

MacEachren (1992) discusses using the clarity of a pattern to define the uncertainty of the spatial data. A sharp pattern would indicate certain information, whilst a more approximate pattern definition would indicate less certain information. Figure 3 depicts this concept using trees as the pattern and sharpness as the uncertainty measure. This could also be used for symbols that define a class, the sharper the symbol the more accurate the data at that location. MacEachren (1992) states accurate boundaries between polygons can be shown as a sharp narrow line, whilst uncertain boundaries maybe illustrated as a broad fuzzy line that can fade.

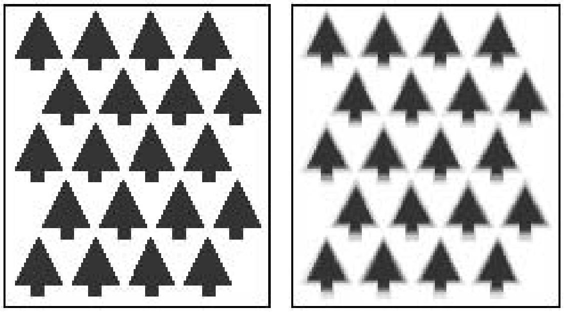

Figure 3. Certain and uncertain illustration of land cover. (MacEachren 1992)

5.8 Adjacent Maps and Texture Overlay

Adjacent maps can be used to depict reliability of data, one to show the data and another to show the reliability. A single map can also be used to show the data with an overlay of the uncertain data shown as textures. This form of visualisation has been used for health statistics data in the past (MacEachren et al. 1998), therefore, providing a foundation for more research using the NIR dataset.

The techniques described may not be an exhaustive list but form a good starting point into fresh ideas in uncertainty geovisualisation for the NIR dataset.

6.0 CONCLUSION

Currently, immunisation information storage and retrieval systems are being re-developed by the MoH after assessments of immunisation coverage returned low results. Factors contributing to previously low coverage rates include poor data collection with no central system to hold vaccine information, ethnicity of the child, parent vaccine education, geographic location of the child, and mobility of families. A central immunisation registry can help overcome some of the problems, and has been proven effective in other countries (National Health Committee. 1999). Creating geovisualisation techniques using the NIR data will help the MoH more quickly and effectively analyse the coverage in locations of interest. Incorporating herd immunity models into a geovisualisation will aid in understanding the underlying effect of vaccine programmes on a disease, and determine if outbreaks are imminent. Geovisualisation helps combine multiple records spatially by topic and then display the information in a format that is easy to understand and make informed decisions from. Aggregating records into a generalised dataset inherently produces spatial uncertainty. Data collection techniques for location, attribute and temporal data can also produce data uncertainty. Being able to see or understand the spatial error associated with any individual geovisualisation is also helpful. The main focus of this research is to find uncertainty sources in the NIR dataset, generate uncertainty figures from the uncertainty sources using a spatial data uncertainty model, and create a new geovisualisation technique to represent uncertainty in the NIR. It is hoped that there will be differences in the coverage rate and herd immunity geovisualisations, emphasising that herd immunity is a more viable way to understand immunisation data. This will aid in increasing immunisation coverage and herd immunity by highlighting areas of need. An effective geovisualisation of the uncertainty data will benefit the end user, enabling them to make accurate decisions, therefore decreasing the spread of vaccine preventable disease.

ACKNOWLEDGEMENTS

Part of this work is not possible without help from Chris Skelly and the Ministry of Health, and the School of Business at the University of Otago.

REFERENCES

Anderson, R. M. and R. M. May (1991). Infectious diseases of humans : dynamics and control. Oxford ; New York, Oxford University Press.

Andrienko, N., G. Andrienko and P. Gatalsky (2000). Mapping Spatio-temporal Data for Exploratory Analysis, German National Research Center for Information Technology, Institute for Autonomous Intelligent Systems, Knowledge Discovery Research Group. **2002**.

Bertin, J. (1981). Graphics and graphical information processing. Berlin, Walter de Gruyter.

Bly, S. (1982). Sound and Computer Information Presentation. Unpublished PhD. Davis, University of California.

Burrough, P. A. and R. McDonnell (1998). Principles of geographical information systems. Oxford ; New York, Oxford University Press.

Davis, T. J. and C. P. Keller (1997). “Modelling and Visualizing Multiple Spatial Uncertainties.” Computers and Geosciences **23**(4): 397-408.

De Gruijter, J. J., D. J. J. Walvoort and P. F. M. van Gaans (1997). “Continuous soil maps -- a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models.” Geoderma **77**(2-4): 169-195.

DiBiase, D., A. M. MacEachren, J. B. Krygier and C. Reeves (1992). “Animation and the role of map design in scientific visualization.” Cartography and Geographic Information Systems **19**(4): 201-214.

Dodge, M. (2000). NewsMaps: Topographic Mapping of Information, Mappa Mundi Magazine. **2002**.

Edwards, J. (2002). Privacy Impact Assessment: In respect of a Proposal to Develop a National Immunisation Register. Wellington, Minisrty of Health**:** 31.

Egan, D. (1988). Individual Differences in Human-Computer Interaction. Handbook of Human-Computer Interaction. M. H. (Ed.). North-Holland, Elsevier.

Ehlschlaeger, C. R. (2002). Can we construct error buttons for every application? Proceedings : International Symposium on Spatial Accuracy in Natural Resources and Environmental Science10-12 July 2002, Melbourne, Australia, University of Melbourne and RMIT University.

Ehlschlaeger, C. R., A.M. Shortridge and M. F. Goodchild (1997). "Visualizing spatial data uncertainty using animation." Computers and Geosciences **23**(4): 387-395.

ESR:Communicable Disease Centre., New Zealand. Public Health Commission. and New Zealand. Ministry of Health. (1998). Immunisation coverage and risk factors for immunisation failure in Auckland and Northland, The New Zealand public health report. [Wellington, N.Z.], ESR:Health Communicable Disease Centre for the Public Health Commission and Ministry of Health. **5:** 49-51.

ESR:Communicable Disease Centre., New Zealand. Public Health Commission. and New Zealand. Ministry of Health. (2000). Improving immunisation coverage: what needs to be done, The New Zealand public health report. [Wellington, N.Z.], ESR:Health Communicable Disease Centre for the Public Health Commission and Ministry of Health. **7:** 11-14.

Evans, B. J. (1997). "Dynamic display of spatial data reliability: does it benefit the Map user?" Computers and Geosciences **23**(4): p. 409-422.

Fabrikant, S. I. (2000). "Spatialized Browsing in Large Data Archives." Transactions in GIS **4**(1): 65-78(14).

Fisher, P. (1993). "Visualizing uncertainty in soil maps by animation." Cartographica **30**(2-3): p. 20-27.

Fisher, P. (1994a). "Hearing the Reliability in Classified Remotely Sensed Images." Cartography and Geographical Information Systems **21**(1): pp. 31-36.

Fisher, P. (1994b). Animation and Sound for the Visualization of Uncertain Spatial Information. Visualization in geographical information systems. H. M. Hearnshaw, D. J. Unwin and Association for Geographic Information. Chichester, West Sussex, England ; New York, Wiley & Sons**:** xv, 243.

Fisher, P. and K. C. McGwire (2001). Spatially Variable Thematic Accuracy: Beyond the Confusion Matrix. Spatial uncertainty in ecology : implications for remote sensing and GIS applications. C. T. Hunsaker. New York, N.Y., Springer**:** xiii, 402.

Gahegan, M. (2000). Visual exploration in geography: analysis with light. Geographic knowledge discovery and spatial data mining. H. J. Miller and J. Han. London, Taylor & Francis.

Goodchild, M. F. (1991). Issues of quality and uncertainty. Advances in Cartography. J. C. Muller. New York, Elsevier**:** 113-39.

Goodchild, M. F. (1994b). GIS error models and visualization techniques for spatial variability in soils. Proceedings 15th International Congress of Soil Science, Acapulco, July 10-16, 1994.

Goodchild, M. F. (1998). "Uncertainty: the Achilles heel of GIS?" Geo Info Systems **November**: 50–52.

Hanna, J., J. Wakefield and C. Doolan (1994). "Childhood immunisation: factors associated with failure to complete the recommended schedule by two years of age." Australian Journal of Public Health **18**: 15-21.

Hengl, T., D. J. J. Walvoort and A. Brown (2002a). Pixel and Colour Mixture: GIS techniques for visualisation of fuzzyness and uncertainty of natural resource inventories. Proceedings :

International Symposium on Spatial Accuracy in Natural Resources and Environmental Science10-12 July 2002, Melbourne, Australia, University of Melbourne and RMIT University.

Herbert, R. (2002). National Immunisation Register Project Workplan. Wellington, Health By Design**:** pp 92.

Heuvelink, G. B. M. (1998). Error propagation in environmental modelling with GIS. London ; Bristol, PA, Taylor & Francis.

Hook, K., M. Sjölinder and N. Dahlback (1996). Individual Differences and Navigation in Hypermedia. In the Proceedings of the Eighth European Conference on Cognitive Ergonomics (ECCE-8), Spain.

Hunter, G. J. (2002). Re: Immunisation Visualisation. J. Kardos. Melbourne, The University of Melbourne. "Personal Communication".

Hunter, G. J. and M. F. Goodchild (1995). “Dealing with error in spatial databases - a simple case study.” Photogrammetric Engineering and Remote Sensing **61**(5): 529-537.

Iowa Department of Public Health. (1997). "Immunization and infectious diseases" in Healthy Iowans 2010. Des Moines, Iowa Department of Public Health**:** pp1-12.

Krygier, J. B. (1994). Sound and Geographic Visualization. Visualization in the Modern Cartography. A. MacEachren and D. R. F. Taylor. New York, Pergamon**:** pp. 149-166.

Lilburne, L. and H. North (2002). Quantifying map confusion due to uncertainty. Proceedings : International Symposium on Spatial Accuracy in Natural Resources and Environmental Science10-12 July 2002, Melbourne, Australia, University of Melbourne and RMIT University.

Longley, P., M. F. Goodchild and D. J. Maguire (2001). Geographic information systems and science. Chichester ; New York, Wiley.

MacEachren, A., C. A. Brewer and L. W. Pickle (1998). “Visualizing georeferenced data: representing reliability of health statistics.” Environment and Planning A **30**: p1547-1561.

MacEachren, A. M. (1992). “Visualizing Uncertain Information.” Cartographic Perspective **13**(Fall): pp. 10-19.

MacEachren, A. M. (1994a). Introducing Geographic Visualization (GVIS). Visualization in modern cartography: Modern cartography ; v. 2. A. M. MacEachren and D. R. F. Taylor. Oxford, U.K. ; New York, Pergammon**:** xviii, 345.

MacEachren, A. M. (1994b). Some truth with maps : a primer on symbolization and design. Washington, D.C., Association of American Geographers.

MacEachren, A. M. (1995a). Approaches to Truth in Geographic Visualization. Proceedings, Auto Carto 12, Charlotte, North Carolina, ACM/ASPRS.

MacEachren, A. M. (1995b). How maps work : representation, visualization, and design. New York, Guilford Press.

Ministry of Health. (2002). National Immunisation Register Report for the Privacy Commission (DRAFT). Wellington, Ministry of Health**:** pp. 30.

Monmonier, M. and M. Gluck (1994). “Focus groups for design improvement in dynamic cartography.” Cartography and Geographical Information Systems **21**(1): p. 37-47.

National Health Committee. (1999). Review of the wisdom and fairness of the Health Funding Authority strategy for immunisation of 'hard to reach' children. Wellington, National Health Committee**:** 78.

Northern Regional Health Authority. (1997). Immunisation coverage in North Health : comparative results from North Health's 1996 immunisation coverage survey. [Auckland, N.Z.], North Health.

Pagh, S. (1999). SPIRE-Spatial Paradigm for Information Retrieval and Exploration, Pacific Northwest National Laboratory. **2002**.

Prislin, R., J. Dyer and C. Blakely (1998). "Immunization status and sociodemographic characteristics: the mediating role of beliefs, attitudes and perceived control." American Journal of Public Health **88**: 1821-6.

Snowdon, S. (2000). Herd Immunity, Vaxchoice. **2002**.

Statistics New Zealand. (2002a). Your complete guide to the INFOS Database. Wellington, Statistics New Zealand**:** 165.

Statistics New Zealand. (2002b). Information about Births. Wellington, Statistics New Zealand.**:** 15.

Steiner, E. B., A. M. MacEachren and D. Guo (2001). Developing and assessing light-weight data-driven exploratory geovisualization tools for the web. International Cartographic Association: Visualization and Virtual Environments, Taupo, New Zealand.

The National Vaccine Advisory Committee. (1999). Strategies to sustain success in childhood immunisations. JAMA, The National Vaccine Advisory Committee. **282:** 363-370.

Vicente, K. and R. Willeges (1988). "Accommodating individual differences in searching a hierarchical file system." International Journal of Man-Machine studies **29**: 647-668.

Ware, C. (2000). Information visualization : perception for design. San Francisco, Morgan Kaufman.

Zadeh, L. (1965). "Fuzzy sets." Information and Control **8**: pp. 338-353.