Transactions in GIS, 1997, vol. 2, no. 1, p. 25

An integrated regionalization of earthquake, flood, and drought hazards in China

JINFENG WANG, STEPHEN WISE, AND ROBERT HAINING

Jinfeng Wang State Key Laboratory of Resources and Environmental Information Systems, Institute of Geography, Chinese Academy of Sciences, Beijing 100101, China. e-mail: wangjf@www.lreis.ac.cn
Stephen Wise Department of Geography, University of Sheffield, S10 2TN, United Kingdom.
Robert Haining Department of Geography, University of Sheffield, S10 2TN, United Kingdom.

Earthquake, flood, and drought data from different sources are combined in a single data set using the same data structure, projection, and scale. The intensity and frequency of each hazard is classified into severe, heavy, modest, and light, producing a classification with 64 combined states for the three kinds of hazard. These classes are then ranked according to severity. The three hazard coverages are overlaid and the polygons that are produced are coded by the classification system. A map is produced that shows the distribution of these 64 classes in regions and their areas measured from the spatial topological data file in the GIS. Spatial analysis reveals the spatial association among the three hazards and between the three hazards and human factors. There is a brief discussion of the implications of the regionalized map for hazard monitoring.

1 Introduction

Hazards can be defined as 'extreme events which may affect different places singly or in combination at different times' (Blaikie et al 1994: 21) and natural hazards as 'those elements of the physical environment harmful to Man and caused by forces extraneous to him' (Burton and Kates 1964 quoted in Smith 1992: 8). Strictly speaking, an extreme event can only be called a hazard if it has the potential to cause death or damage to humans – as Smith (1992: 9) notes: 'a severe earthquake in a remote unpopulated region is an extreme natural event, of interest to seismologists and no more' – but it is common practice to use the term hazard to refer to any extreme event which has the potential to cause damage and this is the usage which will be adopted here.

Hazards occur with varying degrees of frequency and severity and although it is often difficult to predict the occurrence of an individual hazard event, it is possible to determine the statistical likelihood of patterns of occurrence in time and space. When natural hazards occur in or near populated areas, this information can be used to calculate the risk for the population concerned of suffering damage as a consequence of a hazardous event (Cutter 1993). Risk is measured in relation to the number of people affected by and the frequency of occurrence of the hazardous event. When 'significant numbers of vulnerable people (people at risk) experience a hazard and suffer severe damage and/or disruption of their livelihood system in such a way that recovery is unlikely without external aid' (Blaikie et al 1994: 21) then a disaster is said to have occurred. The occurrence of a disaster is thus a consequence of the interaction between physical processes and human systems. Whether or not a disaster occurs depends on the severity of the natural event (the hazard) in relation to the human system that exists within the area where the event occurs.

China suffers from many disasters caused by natural hazards, especially earthquakes, floods, and droughts, losing on average 10 billion RMB (US$ 1 billion) of property annually (or 1–2 percent of total annual GNP). The death toll reached 240 000 in the Tangshan earthquake in 1976; 23 million

1361-1682/97/2-01-025

hectares are affected by drought each year, 12 million hectares of which are populated; 11 million hectares are covered by flood each year, 7 million hectares of this area being populated (People's Republic of China 1994). These events are a consequence of China's location at the junction of the huge Eurasian landmass and the Pacific Ocean: the strong thermal gradient between land and ocean induces seasonal winds that are responsible for the frequent floods and droughts; the collision between the Eurasian and Pacific tectonic plates are responsible for the frequent earthquakes. These three major hazards cause many other kinds of associated problems such as landslides. The magnitude and frequency of these hazards which impact on large numbers of the Chinese population are such that disaster relief is an important and urgent task in China.

Where hazards pose a threat to human life or property, it is important to monitor and where possible predict them, thus minimizing the risk of disaster (Alexander 1993). Three major approaches to the problem may be identified in the literature, with Information Technology (IT) and GIS playing an increasing role in all three (Gatrell and Vincent 1991; Rhind 1991): (1) predicting the occurrence of hazards; (2) general risk assessment; and (3) developing improved tools to manage a disaster should it occur.

The first approach is the prediction of the occurrence of particular events. In some cases, such as floods, a prediction may be made with some accuracy based upon monitoring of factors such as rainfall and soil moisture levels. Indeed, weather prediction generally is one area where modern computers, coupled with improved data collection from satellites, have led to enormous improvements in the range and accuracy of predictions (Tyler 1989). Some events, such as earthquakes, can only be predicted by the continuous monitoring of known precursors such as earth movements, and this has been greatly aided by the development of remote data collection via telemetry (Alexander 1991). The emphasis here is on real-time monitoring and numerical analysis, so currently GIS does not feature very much in this work.

The second approach is the more general assessment of risk, as opposed to the prediction of individual events. This is important for planning defence strategies for existing settlements and for locating new developments away from hazardous areas. It involves two elements – the mapping of the likely intensity of the hazard, followed by the combination of this with information on the location of population or economic activities, and increasingly GIS is playing a role in both these areas.

Hazard intensity may be estimated deductively, based upon a knowledge of the processes at work, or inductively, based on an analysis of the occurrence of the hazard in the region. In both cases, it is the spatial variation in intensity which is important and GIS is therefore of great potential benefit. The deductive approach is only possible where the understanding of the processes at work is sufficient, as in the case of landslide hazard (van Westen and Terlien 1996), flooding (Mejianavarro et al 1994), and groundwater contamination (Merchant 1994), and in many cases involves a combination of computer modelling and GIS. The inductive approach benefits from the ability of GIS to bring together information on a wide variety of possible explanatory factors which may then be used to derive empirical models or provide surrogates for hazard levels where measurements are sparse. This approach has also been used for the analysis of landslides, especially at regional scales where the modelling of individual landslide locations would not be feasible (McKean et al 1991; Dikau et al 1996). A key element of this approach is the use of GIS to interpolate from point measurements to provide data estimates for whole areas, as in the case of air quality and soil contamination measurements (Sengupta and Venkatachalam 1994; Szucs 1995). Finally, GIS has the ability to overlay the estimated values on the hazard intensity with information on population distribution, thus identifying the areas most at risk (Maslia et al 1994; Sengupta and Venkatachalam 1994; Hiscock et al 1995; Lowry et al 1995).

The third approach is the development of improved tools for the management of the situation when a disaster does actually occur. This normally involves the organization of whatever resources are required to deal with the disaster (emergency services, specialized equipment, etc) and the organized evacuation of the affected area. Both of these have a strong spatial element, and great interest is being shown in spatial decision support systems, based around GIS, to assist in these tasks (Alexander 1991; Gatrell and Vincent 1991; Wadge et al 1993).

The subject of this paper falls within the second of these approaches – the assessment of hazard levels at a national scale for China for three national hazards – earthquake, flood, and drought. This type of analysis has been performed for a number of countries including the United Kingdom (Perry 1981), Japan (Nakano 1974), the USSR (Gerasimov and Zvonkova 1974), and Italy (Alexander 1987), but in most cases each hazard is treated seperately, with no attempt to produce an assessment of the variation in risk from all hazards for the country. However, as Alexander (1993) points out, for anyone living in an area, the degree of risk they are subject to arises from the combination of all potential hazards. The effect is not simply additive (the more hazards the more risk) but multiplicative since some hazards are more damaging in combination than singly, for example cold temperatures are far more damaging when allied with strong winds (Gerasimov and Zvonkova 1974) and some hazards actually cause others, as in the case of earthquakes and tsunami (Alexander 1991). The focus on a single hazard was almost universal until the work of Hewitt and Burton (1971) in London, Ontario, who attempted to assess the levels of risk for the inhabitants from all potential hazards. This approach has since been applied elsewhere (Cooke 1984), but at a national level the only attempt to apply the same ideas has been the work of Gerasimov and Zvonkova (1974), although Perry (1981) makes passing reference to the importance of trying to do this. Gerasimov and Zvonkova (1974) briefly reviewed the pattern of a wide range of hazards in the USSR, and then produced a map in which the whole area was grouped into four classes, ranging from areas suffering catastrophic events causing loss of life and damage, through to areas suffering only local hazards such as frost and fog. Unfortunately, no details are provided of how these classes are defined, or of how the available data was analysed in order to produce the map, which shows the USSR divided into 29 regions on the basis of this classification.

There are two reasons why it is of interest to attempt to produce an integrated regionalization of hazards:

1. *To assess the degree to which the occurrence of hazards is spatially related:* for example, do areas which suffer severe flooding also suffer from earthquakes? In some cases there may be good reasons to expect relationships – for example, earthquakes will often lead to localized flooding because of the damage to flood defense systems. In other cases, and especially at the very broad scale, the association may be purely an accident of geography – for example the coastal region of China lies close to the tectonically active zone, but is also the area most affected by monsoon storms.
2. *To help design a sampling strategy for hazard monitoring.* Despite the advances in automated data collection, especially with the use of remote sensing (Alexander 1991), the collection of data on natural processes still relies on point sampling. This means that the location of the stations is an important issue, since it would not only be expensive to attempt to monitor natural hazards in this way over a whole country such as China, but inefficient as well, leading to oversampling in areas of low risk. It makes sense to concentrate monitoring efforts in areas which suffer high levels of risk from more than one hazard, both because of the potential economies and because of the greater potential for disaster in such areas.

China is fortunate in having a very long record of data collection on the occurrence of many natural hazards which can be analysed to produce estimates of the probability of hazards of a given magnitude occurring at any given location. This analysis has already been carried out in the case of earthquakes, floods, and drought and published in the form of a map for each hazard for the whole of China. The main aim of the work reported here has been to combine these three sets of data into an integrated regionalization, which shows the level of all three hazards which might be experienced across the country. This information is then combined with information on population and economic activity to produce a broad assessment of risk across the whole of China.

We first describe the original data which was used for the project and then the technique which was used to produce the regionalization. The reliability of the regionalization is discussed before its use in various types of analysis is described.

2 The database used in constructing the regionalization

In considering natural hazards, it is important to distinguish two aspects of the occurrence of hazard events: their magnitude and their frequency. The two are closely related, since events of large magnitude tend to occur less frequently than those of small magnitude. From data for a single recording station it is possible to construct a magnitude/frequency curve from which the following two questions can be answered:

1 How frequently will an event of size *x* occur? (Normally expressed as the recurrence interval, or number of years between occurrences of events of this magnitude.)
2 How large will an event be which has a recurrence interval of *y* years?

In undertaking an analysis of risk at a single point, the first question is often of interest – for example if it is known that existing flood defences will withstand floods up to a certain magnitude, it is important to be able to estimate the likelihood of them being breached.

In looking at the spatial variation of hazard magnitude, it is more common to select a recurrence interval (e.g. 50 years) and look at the variation in the magnitude of events of this frequency across the area.

As described above, our initial data sources were three maps each showing the pattern of occurrence of one hazard across the whole of China – two show the variation in magnitude of events at a given recurrence interval while the third combines measures of magnitude and frequency. The maps are not reproduced here, but we describe their derivation based upon personal communications from their authors (whose names are given in the acknowledgements section).

2.1 Floods

The original data on flooding were taken from 460 gauging stations, each with a 500-year record. Such a length of record means that estimating the magnitude of the 50-year flood is a process of interpolation rather than extrapolation (as would be the case in many other countries). It is well known that flood discharge is highly correlated with catchment area (Gregory and Waling 1972), and so the 50-year flood discharge was divided by the catchment area. This analysis produced a series of point estimates of flood magnitude, which were then interpolated manually to produce an isoline map of the magnitude of the 50-year flood across China. The manual interpolation allows local, expert knowledge of factors, such as topography and broad scale climatic patterns which affect flood magnitude, to be used in drawing the contours. The contour intervals (100, 500, 1000, 2000, and 4000 $m^3 s^{-1} km^{-2}$) were chosen based on a knowledge of the damage caused by floods of different magnitudes, and hence represent important breaks between minor and major flood events.

2.2 Earthquakes

The original database consisted of records from 2000 recording stations, each with a 20-year record plus historical records spanning over 4000 years. From this, the intensity of an earthquake (measured on the Richter scale) with a 10-year recurrence interval was estimated for each area of 1 degree of latitude by 1 degree of longitude.

2.3 Drought

The original data were taken from the *China Yearbook*, which records drought incidences at a provincial level. In all there were 40 years of data for each of the 31 provinces of China. Provinces were classified in terms of both frequency and severity of droughts to distinguish between areas with frequent droughts (an average of at least one a year over a 40-year record) and less frequent droughts, and between areas with extreme droughts (crop loss of more than 80 percent occurring more than 20 times in 40 years) and less extreme droughts. This meant that each province fell into one of four possible classes based on the four posssible combinations of drought frequency and severity. Provincial boundaries are not very suitable for reporting drought occurrence, since the actual areas affected will be determined by topographic and climatic factors which will span these boundaries. Therefore, local knowledge of topographic and climatic factors was used to delimit those parts of China which fell into each of the four classes, producing a map in which the areas are felt to be a more accurate reflection of drought occurrence than a map based on provincial boundaries.

2.4 The maps

Each map is therefore a representation of the spatial variation in the intensity of the natural hazard in question. In the case of flood and earthquake this has been accomplished by mapping the intensity of events at a defined recurrence interval, while the drought map shows variations in both magnitude and frequency. Note that the maps show the variation in the natural events – they do not necessarily show the full extent of the area which might be affected by an individual cover event such as a single flood. This is particularly true in the case of floods, where a flood event will cover a much greater area in the lowlands than in a mountainous region. However, this is not a serious problem given the very broad scale of the analysis and it is reasonable to assume that these maps indicate those parts of China where the intensity of each natural hazard is potentially highest.

3 Construction of the regionalization

The aim of the project was to combine these three sets of data to produce one map of 'hazard regions' (i.e. regions suffering similar levels of all three hazards) which could be used when considering the design of a sampling scheme for hazard monitoring. There were three stages to this, all of them suitable for implementation in GIS:

1. conversion of the maps to a common map projection
2. combination of the information from the three maps and construction of a regionalization
3. combination of the map of hazard regions with data at the provincial level on population and levels of economic activity.

The first stage of the analysis was relatively straightforward. All three maps were digitized and input into the ARC/INFO GIS (Environmental Systems Research Institute, Redlands, California). Two of the maps (flood and drought) were based on the Albers conic projection (lower latitude tangent of 30° N, upper latitude tangent 40° N, and a central longitude line of 105° E) while the earthquake data were reported on a latitude-longitude grid basis. This was therefore converted to Albers conic using ARC/INFO.

The combination of information on hazards measured in quite different units into a single measure for an area is a problem discussed briefly by both Perry (1981) and Hewitt and Burton (1971) who both conclude that the best approach is one based on estimating the intensity of each hazard which is likely to cause a similar amount of damage or disruption. In the case of the data for China, the authors of the original maps had chosen the isoline and class intervals based on an understanding of the effects of hazards at different levels, and it was thus possible to group the hazard intensity values on each map into four classes – Severe, High, Medium, and Light. The actual class intervals used for each map are shown in **Table 1** – in the case of flood and earthquake the intervals are based on magnitude at a set recurrence interval, in the case of drought on a combination of magnitude and frequency.

Each of the original maps can be thought of as a field (a variable which varies continuously across space) although there are differences between them. Flood magnitude is a variable which is only meaningful along the course of the rivers – at all points in between, the magnitude of the 50-year flood is a meaningless quantity. Therefore the flood map is in fact a trend surface map and the isolines cannot be taken as literal estimates of the intensity of the 50-year flood at all points (a topic which we return to later on). Earthquake and drought intensity can be conceptualized as continuous fields with a valid value at all points across the surface, but again given the scale of the analysis the maps are perhaps best thought of as trend surface maps.

The three maps between them made use of three of the six methods used to represent fields in GIS – isolines for the flood map, grid cells for earthquakes, and polygons for drought (Kemp 1993). Before combining the three maps, we used the LATTICECONTOUR command in ARC/INFO to produce an interpolated version of the earthquake data that was more comparable with the other two maps. Clearly this introduces another element of estimation into the procedure, and a different contouring algorithm would probably produce a slightly different result. However, it has the effect of removing the obvious artefacts of grid cells from the final map.

The two isoline maps were then reclassified according to the values in **Table 1**, to produce polygon maps showing the areas falling into each of the four classes, and these were overlaid with the

	Earthquake(I)	**Peak Flood(m^3/s x km)**	**Drought**
S(evere)	[7.0,12.0]	[4000.0,8999.9]	Freq>=1 time/year & Ext Dry >= 1 time/2 year
H(eavy)	[6.0,6.9]	[2000.0,3999.9]	Freq<=1 time/year & Ext Dry >= 1 time/2 year
M(oderate)	[5.0,5.9]	[500.0,1999.9]	Freq>=1 time/year & Ext Dry < 1 time/2 year
L(ight)	[0.0,4.9]	[0.0, 499.9]	Freq<=1 time/year & Ext Dry < 1 time/2 year

Notes
Freq = Frequency,
Ext Dry = Extremely dry

Table 1. Four levels of intensity of earthquake, flood, and drought.

polygon map of drought intensity.

This procedure produces 64 combinations because each hazard can be at one of four different levels. For the purposes of mapping and classification, it is useful to be able to rank these 64 combinations, from high risk (e.g. all three hazards at level S) to low risk (all three at level L).

The scheme that has been used is a hierarchical one and is shown in **Table 2**. At the top level (column 1-L) combinations are grouped according to the most severe hazard level they contain (regardless of whether this level is for flood, earthquake, or drought). This process is repeated within each of these 1-L groupings to give the classification shown in column 2-L, and repeated once more to give column 3-L, which classifies all 64 combinations into groups depending on the combination of Severe, Heavy, Modest, and Light hazards. The levels for each hazard can be distinguished in the column labelled 'Types' (or 4-L): earthquake (e); flood (f); drought (d).

The different combinations of particular hazards within each 3-L group are listed and assigned a number from 1 to 64. This is a general procedure, which could be applied if more hazards were considered, or if each were classified into a different number of severity levels. It can be used, in conjunction with the GIS, to produce maps showing where particular combinations occur and statistics on their areal extent.

4 Regionalized map

By reclassifying each of these original coverages according to the thresholds shown in **Table 1**, a new set of coverages was produced showing the areas experiencing Severe, Heavy, Moderate, and Light levels of each of the three hazards. These coverages were then overlaid, to produce the map shown in **Figure 1** (the 'Types' or 4-L column in **Table 2**).

Each polygon on the map has associated with it the levels of severity of earthquake, flood, and drought in that area derived from the three input maps.

5 Reliability of the regionalization

Prior to analysing the results of the regionalization, it is important to consider the reliability of the final map (**Figure 1**). Rejeski (1993) notes that GIS-based analyses can only make a useful contribution to the consideration of hazard assessment if their products can be believed by the public and by decision makers, and if they make clear how reliable the information they produce is. There are two forms of error associated with **Figure 1** – conceptual uncertainty and processing error.

Figure 1 divides China into a series of sharply-bounded polygons defining areas with different levels of hazard and yet we know that natural processes do not generally have sharp dividing lines: earthquake effects are centred on the epicentre, but fall away gradually with distance from this; drought areas have diffuse boundaries; and although an individual flood will generally have a well-defined spatial extent, the boundaries of areas experiencing similar levels of risk of repeated flooding will be less sharply defined. As noted above, the original maps from which **Figure 1** was derived could all be regarded as trend surface maps on which the lines (whether isolines or polygon boundaries) are cartographic devices used to indicate the trend in

1-L	2-L	3-L	Types (efd)	No	1-L	2-L	3-L	Types (efd)	No
	SSS	SSS	SSS	1				MLS	33
			SSH	2				LMS	34
		SSH	HSS	3				SLL	35
			SHS	4			SLL	LSL	36
			SSM	5				LLS	37
	SS	SSM	MSS	6		**HHH**	HHH	HHH	38
			SMS	7				HHM	39
			SSL	8			HHM	MHH	40
S		SSL	LSS	9		**HH**		HMH	41
			SLS	10				HHL	42
			SHH	11			HHL	LHH	43
		SHH	HSH	12				HLH	44
			HHS	13				HMM	45
			SHM	14	**H**		HMM	MHM	46
	S		SMH	15				MMH	47
		SHM	HSM	16				HML	48
			MSH	17				HLM	49
			HMS	18		**H**	HML	MHL	50
			MHS	19				LHM	51
			SHL	20				MLH	52
			SLH	21				LMH	53
		SHL	HSL	22				HLL	54
			LSH	23			HLL	LHL	55
			HLS	24				LLH	56
			LHS	25		**MMM**	MMM	MMM	57
			SMM	26				MML	58
		SMM	MSM	27	**M**	**MM**	MML	LMM	59
S	**S**		MMS	28				MLM	60
			SML	29				MLL	61
			SLM	30		**M**	MLL	LML	62
		SML	MSL	31				LLM	63
			LSM	32	**L**	**LLL**	LLL	LLL	64

Notes S = Severe, H = High, M = Moderate, L = Light
e = earthquake, f = flood, d = drought

Table 2. Integrated classification of intensity from S(evere) to L(ight).

the phenomenon being mapped.

It has long been recognized that current GIS use data models and algorithms which represent the world in terms of atomic units with sharp boundaries even when this may not be a good model of the phenomenon being modelled (Burrough 1992). Some progress has been made in using GIS to model phenomena which do not have sharp boundaries and this is still an active research area (Burrough and Frank 1996). Much of the early work uses fuzzy set theory (Zadeh 1965). In normal GIS operations, objects are assigned to one and only one set. For example, each point in space will be deemed to fall either in or out of the High risk area for flooding, and the cartographic representation of this is thus a sharp boundary around the High risk area. In fuzzy set theory, objects are allowed to have partial membership of a class – hence a point could be considered to belong to both the High and Moderate classes with a grade of membership for each of 0.5. This produces a representation of regions with 'core' areas where membership grades are 1 and 'edges' where membership grades are less than 1 (Leung 1987). Altman (1994) argues that in these circumstances the process of 'hardening' data into apparently well-defined, precise forms should be left as late as possible in the process of analysis, in order to minimize both the effects of the misrepresentation and the error propogation which arises when two or more 'fuzzy' regions are overlaid.

The crisp polygon boundaries in **Figure 1** are thus not an ideal representation of the edges of these hazard regions. They occur because we have treated the lines in the three original maps as sharp lines and overlaid them using a simple, geometrical operation. This operation, although a little simplistic, is defensible in this case because we do not have access to the original data from which the lines were drawn, and so we are not in a position to make any estimates of the degree of fuzziness which might be attached to the isolines.

It might be argued that it would have been preferable to return to the original data, but in fact there are distinct advantages in using the processed data:

- it saves the considerable effort which would be involved in analysing the original data
- more importantly, the maps are also based on the expertise and experience of specialists in the

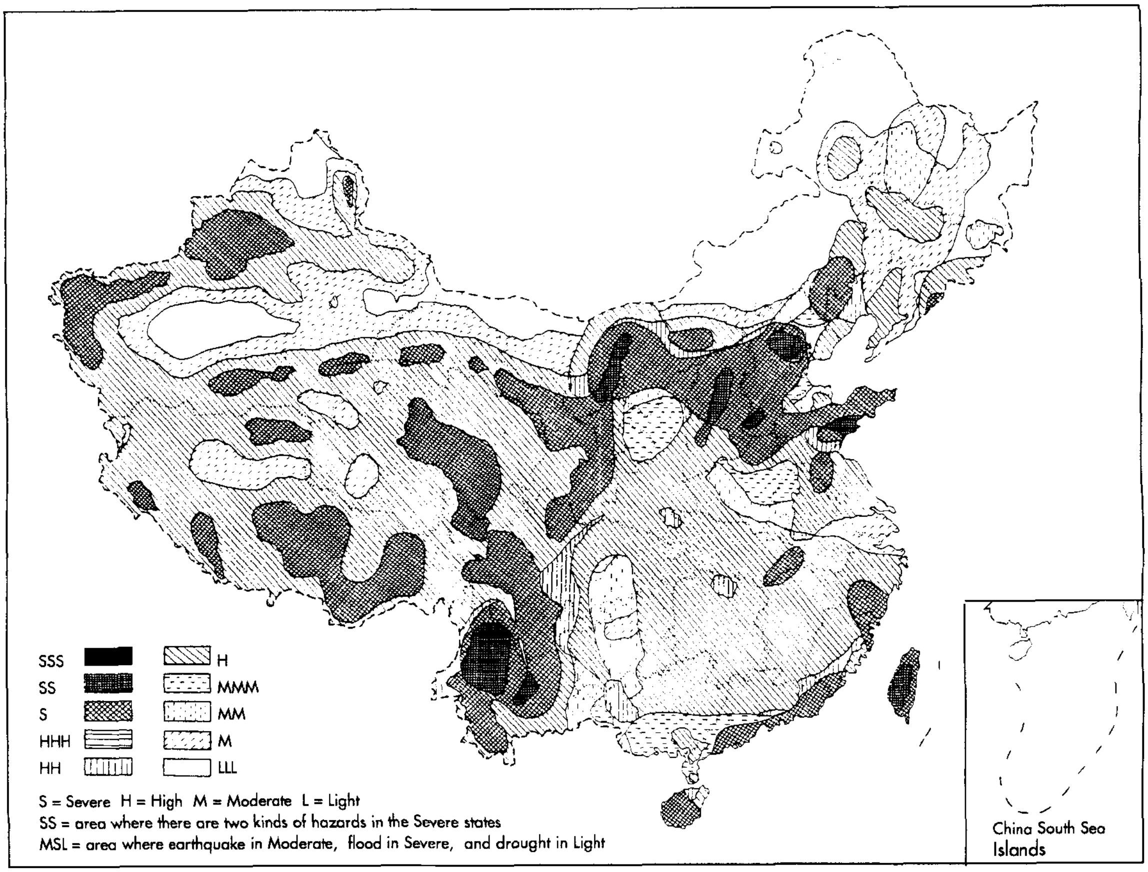

Figure 1. Map of integrated regionalization of earthquake, flood, and drought in China.

field, which could not be replicated by the automated processing of the raw data using a GIS.

Since we cannot apply fuzziness to the analysis of the data, we must apply it in the interpretation of the results. This means that the polygon boundaries on **Figure** 1 should not be treated as sharp boundaries. However, this type of map is still useful for several types of analysis:

1 It seems reasonable to suppose that Figure 1 does reflect the broad pattern of hazard occurence in China, identifying those parts of the country potentially at risk from the three hazards considered here. While the issue of fuzziness is important, it perhaps assumes rather less importance at this macro scale of analysis.

2 In general terms, the size of the polygons will give reasonable estimates of the scale of the areas under different levels of risk. Similarly, the combination of this information with population data, while not giving accurate figures on the population at risk (because of the lack of sharp boundaries, and the lack of detailed knowledge on population distribution below provincial level), will allow some estimate of those areas where the greatest numbers of people seem to be at risk.

3 The map is probably a reasonable reflection of the spatial association between the different hazards, information which may be useful in two ways. First, it is possible to assess the degree to which certain hazards occur in similar regions. Second, the spatial scale at which hazard risk varies can be used to help design sampling schemes for more detailed hazard monitoring.

The cartographic (spatial) and thematic (attribute) data in the source maps have gone through several GIS data transformations which could potentially be a further source of error in the final map. The accuracy loss (error propagation) at each step can be coarsely estimated.

The GIS processing errors as they may influence the regionalization process consist of three steps: (1) GIS processing including digitizing, scaling, and projection transformation; (2) GIS manipulation including overlay and numerical calculation according to some model or function; and (3) construction of the output map by the GIS. These three stages within the GIS and their associated errors are shown in **Table 3**. Terms in the table denote the extent to which the GIS process introduces a loss of accuracy. The term 'no' signifies no accuracy loss. The term 'precision' means that accuracy loss is purely dependent on machine precision. Error accumulation denotes the overall effect of the GIS processes.

For the purpose of primary mapping, sample data for earthquakes, floods, and droughts are rarer in West China than in East China, so we expect much more accuracy of the spatial and attribute features in Eastern China than we do of those in Western China.

In the case of the GIS processes, none of the processes change attribute values but they can produce errors in estimates of cartographic area. Digitizing causes random cartographic feature errors according to a normal probability distribution, although this can be reduced by repeating and slowing the process (Maffini et al 1989). Scaling and projecting do not affect relative spatial position and shape. Absolute locations (*x,y* co-ordinates) are only affected by the precision of the representation of features by the computer, which is not usually an important source of error (Goodchild and Gopal 1989).

GIS manipulation does not influence spatial position except perhaps through the creation of slivers formed by overlays. This effect can be estimated approximately (Veregin 1989) and

		Attributes	**Cartograph (polygon)**			
		(r)	**Absolute location**	**Relative location**	**Polygon shape**	**Area**
GIS processing	1 Digitizing	no	probably	small	small	some
	2 Scaling	no	precision	no	no	some
	3 Projection	no	precision	no	precision	some
GIS manipulation	4 Overlay	no	no	no	no	precision
	5 Functions(f)	Precision	no	no	no	no
GIS output	6 Mapping	no	precision	precision	precision	precision
Error accumulation	$\sum_{i=1}^{6}(\bullet)$	precision	probably	small	small	some
Error remedy	Reducing $\sum_{i=1}^{6}(\bullet)$	repeat 1	repeat 1	repeat 1	repeat 1	repeat 1*
Objectives			Monitor (where/what)	Statistics (area/percent)	Sampling (density)	Sampling (density)
Acceptable ?			Yes	Yes	Yes	Yes

* only digitizing can be repeated as the other activities are determined

Table 3. Accuracy losses arising from regionalization in GIS.

removed by a filter with a defined level of tolerance (Chrisman 1989). GIS manipulation of attributes is affected by the level of machine precision. For a discussion of error propagation effects arising from GIS operations on attributes see Haining and Arbia (1993). Finally, for the GIS output map, the accuracy loss depends on the precision of the output machine, and in general this is small.

In summary, the GIS operations used in our research raise two important types of problem as regards accuracy: cartographic areas and slivers. The thematic areas stored in the GIS attribute file are used instead of the cartographic ones in the GIS map when the area values are needed; slivers are removed carefully by editing in GIS when their areas are less than a threshold size. As a result, the reliability of our conclusions depends mainly upon the quality of the data sources rather than the effects of GIS operations.

6 Preliminary spatial analysis

In this section we use the regionalization map and the GIS database to explore a number of questions concerning the incidence and impact of hazards in China.

6.1 The range of the hazards in different areas

The area and the percentage of the area suffering from hazards at different levels and combinations in different administrative provinces can be measured. This is done using the attribute file of the polygons that are formed by overlaying the administrative provinces map on the regionalization map. The summary results are given in **Table** 4. The capital string BJ, TJ, etc in the first row refers to the names of provinces (see **Appendix** 1: Province acronyms). The numbers below them are the corresponding municipal codes. The numbers in the table represent the percentage and absolute area of each province under each of the hazard classes.

Table 4 is condensed into broader categories of severity levels in **Table** 5. From this table, which shows the percentage and absolute areas by province with at least one hazard at the specified severity level, we can see that both Beijing and Taiwan are covered by severe hazards, and Shanghai by moderate hazards. Most provinces are affected by three levels of severity. As a whole, 23.5 percent of China is affected by severe hazards, 46.5 percent by high hazards, 18.9 percent by modest hazards, and 11.1 percent of China is only lightly affected by hazards or free of hazards.

6.2 Spatial association of earthquakes, floods, and droughts

There are 209 valid polygons in the regionalization map and each of them combines three hazards at one of four levels of severity. To what extent do areas suffer similar levels of each hazard? Do areas that suffer from high levels of one hazard tend to have high levels of other hazards? Can this be assessed? As noted earlier, if levels of different hazards are spatially correlated then this information might be useful in deciding where to concentrate resources for monitoring.

In applying a measure of association, it must be noted that the data are at the ordinal level of measurement, and that allowance needs to be made for the variable size of the 209 regions. The Pearson correlation co-efficient assumes interval level data and makes no allowance for areas of different sizes and so is inappropriate. Tests based on contingency tables require a unit of measurement to convert the areas to counts (e.g. numbers of pixels of $100m^2$, $1000m^2$, $10000m^2$, etc.) but this will influence a measure such as chi-square reflecting the scale-dependent nature of the test. Some standardized chi-square statistics, such as Creamer's or Tschuprow's (Lewis 1977: 93–100; Gatrell 1987), have no measure of statistical significance and significant difference except at some extreme points of 0, 1, and -1.

In this paper a form of analysis similar to that introduced by Court (1970) is used (see also Norcliffe 1977: 128–38). Resemblance tables are constructed which show the proportion of China under each of the four levels of any pair of hazards (**Table 6**). Court's method requires the row and column totals of this table to be approximately equal to 0.25, using the quartile values to partition the table. This is not possible for our data given the level of attribute aggregation and so it does not seem appropriate to apply the inference theory that is available with Court's method.

However, Court's method may be an analytical technique which deserves wider adoption by GIS users, especially since it would be relatively easy to implement within GIS software. It is not clear how

%/Area	BJ	TJ	HB	SX	NM	LN	JL	HLJ	SH	JS	ZJ	AH	FJ	JX	SD	HEN	HUB	HUN	GD	GX	HAN	SC	GZ	YN	XZ	SAX	GS	QH	NX	XJ	TW	Sum
No.	11	12	13	14	15	21	22	23	31	32	33	34	35	36	37	41	42	43	44	45	46	51	52	53	54	61	62	63	64	65	71	%/A
SSS																																
SS	67	37	4.9	11	1.2										7.8							0.2		20					18		46	1.9
	11	4.1	8.8	18	14										12							1.1		80					11		17	182
S	33	50	50	54	15	9.7	0.9			12	24	5.7	23	4.9	38	14	5.9		28		69	36		34	28	10	36	36	44	16	54	21.5
	5.5	5.7	103	88	182	14	1.6			12	24	7.4	27	7.8	77	23	11		50		24	175		124	340	21	160	257	26	259	19	2043
HHH																																
HH		5.8	7.6	1.1	1	8.6				0.6		3.6	12	0.7	7.7	1.8	1.6	4.8	5.1	9.9	14	8.7	2.8	10		4.8						2.4
		0.7	14	1.7	12	12				0.6		4.7	14	1.1	12	3	2.9	10	8.9	23	4.7	48	4.8	41		9.6						228
H		1.7	11	30	14	42	42	2.2		70	57	67	65	94	23	33	92	95	50	50	18	45	47	25	63	55	35	74	17	40		43.8
		0.2	20	47	162	117	76	10		70	57	61	78	151	34	54	166	200	87	116	6.1	254	79	98	760	111	156	535	10	638		4155
MMM		5.2	8.5	*	0.2					0.2		11			2.9	23			15	30				1.2		24	4.6		1.1			2.6
		0.6	15		2.4					0.2		14			4.4	39			26	69				4.7		47	21		0.7			244
MM			7.9	0.6	1.2	2.6	7.4	16				4.8			7.3	29				8.6		8.8	39	4.8		7.6			3.4			3.7
			14	9.4	14	3.6	14	72				6.2			11	48				20		49	66	19		7.1			2			355
M			7.3		15	35	35	31	100	17	19	8.4								0.5		0.6	12		7.9		13	6.2		29		12.5
			13		182	49	63	141	6.2	17	19	11								12		3.4	20		95		58	45		461		1185
LLL			1.5		50		14	51																			1.8			11		11.1
			2.7		802		25	236																			8.1			173		1046
Sum%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100%
T-Area	17	11	180	158	1200	140	180	460	6.2	100	100	130	120	160	150	167	180	210	175	230	35	560	170	390	1200	200	450	720	60	1600	36	

Notes

S = Low H = High M = Moderate L = Light

Table 4. Hazard: percentage and absolute area of level (2-L) by administrative province (area unit: 1000km^2).

%/Area	BJ	TJ	HB	SX	NM	LN	JL	HLJ	SH	JS	ZJ	AH	FJ	JX	SD	HEN	HUB	HUN	GD	GX	HAN	SC	GZ	YN	XZ	SAX	GS	QH	NX	XJ	TW	Sum
no.	11	12	13	14	15	21	22	23	31	32	33	34	35	36	37	41	42	43	44	45	46	51	52	53	54	61	62	63	64	65	71	%/A
sumS	100	87	55	65	16	9.7	0.9			12	24	5.7	23	4.9	46	14	5.9		29		69	36		55	28	10	36	36	62	16	100	23.5
	17	9.9	112	105	196	14	1.6			12	24	7.4	27	7.8	89	23	11		50		24	176		204	340	21	160	257	37	259	36	2219
sumH		7.5	19	31	15	51	42	2.2		70	57	70	77	95	30	34	94	100	55	60	31	54	49	36	63	60	35	74	17	40		46.5
		0.8	34	49	174	129	76	10		70	57	86	92	152	46	57	169	210	96	139	11	303	84	139	760	121	156	535	10	638		4383
sumM		5.2	24	0.6	17	38	43	47	100	17	19	24			10	52			15	39		9.4	50		7.9	31	18	6.2	4.5	29		18.9
		0.6	43	9.4	199	53	77	213	6.2	17	19	31			15	87			26	90		53	85		95	54	79	45	2.7	461		1784
sumL			1.5		50		14	51																			1.8			11		11.1
			2.7		802		25	236																			8.1			173		1046
Sum%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
T-area	17	11	180	158	1200	140	180	460	6.2	100	100	130	120	160	150	167	180	210	175	230	35	560	170	390	1200	200	450	720	60	1600	36	9493

Notes

S = Low H = High M = Moderate L = Light

Table 5. Hazard: percentage and absolute area of level (1-L) by administrative province (area unit: 1000km^2).

(a) Earthquake and flood

		Flood			
Area		**S**	**H**	**M**	**L**
	S	0.0020	0.0032	0.0549	0.1027
Earthquake	**H**	0.0046	0.0290	0.0539	0.2467
	M	0.0096	0.0807	0.0523	0 1528
	L	0.0015	0.0641	0.0161	0.1261

Area ratio.
main_diag = 0.2093
opposite_diag = 0.2388

	Flood	
Area	**S + H**	**M + L**
S + H	0 0387	0.4581
M + L	0.1560	0.3472

Area ratio.
main_diag = 0.3859
opposite_diag = 0.6141

(b) Earthquake and drought

		Drought			
Area		**S**	**H**	**M**	**L**
	S	0 0160	0.0043	0.0073	0.1344
Earthquake	**H**	0.0487	0.0085	0 0342	0.2457
	M	0.0156	0.0061	0.0602	0.2122
	L	0.0000	0.0025	0.0065	0 1977

Area ratio:
main_diag = 0.2825
opposite_diag = 0.1746

	Drought	
Area	**S + H**	**M + L**
S + H	0.0775	0.4215
M + L	0.0242	0 4767

Area ratio.
main_diag = 0.5543
opposite_diag = 0.4457

(c) Flood and drought

		Drought			
Area		**S**	**H**	**M**	**L**
	S	0.0027	0.0000	0.0035	0 0188
Flood	**H**	0.0184	0.0004	0.0481	0.1832
	M	0.0312	0.0071	0.0611	0 1509
	L	0.0617	0.0229	0.0410	0.3489

Area ratio:
main_diag = 0 4131
opposite_diag = 0.1358

	Drought	
Area	**S + H**	**M + L**
S + H	0 0216	0.2537
M + L	0 1229	0.6018

Area ratio:
main_diag = 0 6234
opposite_diag = 0.3766

Table 6. The spatial association between earthquake, flood, and drought.

sensitive the inference theory is to the requirement that the row and column totals are equal. According to Norcliffe (1977), some relaxation is possible and this is an area which deserves further research.

The proportion of the map where the levels of two hazards are 'identical' is obtained from the following measure:

(sum of the diagonal values in the 4 by 4 table)

To check for 'similarity' of association collapse the 4 by 4 table to 2 by 2 (merge S & H counts, and merge M & L counts) and then repeat:

(sum of the diagonal values in the 2 by 2 table)

(Again it appears that Court's method cannot be used since the row and column totals do not sum to 0.5 as the levels cannot be manipulated to approximate the median.)

The results are listed in **Table 6**. For the 4 by 4 and 2 by 2 tables, the expected values of the proportions assuming an equal distribution are 4/16 =0.25 and 2/4=0.5 respectively. The larger the proportion of the main diagonal the stronger the tendency for spatial (positive) association between the two hazards, and the larger the proportion of the opposite diagonal the stronger the tendency for spatial separation (negative association) between the two hazards. The relative magnitudes of values within individual columns (rows) are also indicative of patterns of association at given levels of hazard.

Table 6(a) shows that there is no apparent spatial association between earthquake and flood from the 4 by 4 table, but there is evidence of some negative association from inspecting the 2 by 2 table. Severe earthquakes and floods tend not to occur in the same areas; **Table 6(b)** shows that the distribution of earthquakes tends to associate positively with drought from both the 4 by 4 and 2 by 2 tables; **Table 6(c)** shows that the distribution of flood has an association with drought. The association is dominated by the large common area with medium and light intensity for both flood and drought hazards.

6.3 Spatial correlation between human factors and hazards

The population (POP), gross national product (GNP), provincial investment (INV) and total provincial area (AREA) are available for each of the provinces of China (see **Appendix 2: Table A2**). If we assume that the three socio-economic factors (POP, GNP, and INV) are uniformly distributed within a province then the amount of the four factors that are subject to different levels of hazard in different provinces and in the whole country can be estimated. The provincial totals are given in **Table A1** (in **Appendix 2**) whilst **Table 7** shows the national figures. For a more accurate assessment in the case of the three socio-economic factors it would be necessary to use more detailed information on the intra-provincial distribution of population and economic activity.

From the provincial level data (**Appendix 2**) the Pearson correlation between population, investment, GNP, and the highest level of hazard with respect to all types of hazard (level 1 in **Table 2**) was evaluated. The hazard levels are coded, for purposes of correlation as severe (4) down to light (1). The Pearson correlation is only indicative of the association because the hazard variable is only measured at the ordinal scale. The results are given in **Table 8**. This shows that the three human factors are positively correlated. There is a negative correlation however between each of the three human factors and the severity of hazards which means that the hazard levels appear to be much

Hazard level	Area (1000 km^2)	GNP (m RMB)	INV (m RMB)	POP (10 000)	Area %	GNP %	INV %	POP %
Severe	2182.88	8370.88	3253.93	28316.1	23	26	17	24
High	4383.41	15834.65	5825.90	60939.5	46	49	49	52
Moderate	1783.57	7232.08	2705.89	24295.9	19	23	23	21
Light	1046.30	970.16	343.08	3645.7	11	3	3	3
Sum	9457.00	32070.00	11985.00	117736.0	100	100	100	100

Table 7. Human factors over risk areas (1-level) in China.

higher in remote areas with less human activity than in areas with dense human activity. This suggests that the distribution of human activity within China has, with some important exceptions such as Beijing province, adapted to the distribution of hazards at the provincial scale. However, given the simplistic assumptions made about the distribution of population and economic activity these results must be viewed with considerable caution, and more work needs to be done if a more robust estimate is to be made. It is possible, for example, that even in provinces that experience high levels of hazard, the intra-provincial distribution of human activity is located away from the hazardous areas, so that the assumptions made here would underestimate the extent to which human activity avoids areas of severe hazard.

6.4 Implications of the regionalization for hazard monitoring

We briefly consider the implications of the regionalization for hazard monitoring through the setting up of a sampling network. Sampling is unavoidable when monitoring any aspect of the natural environment since it will never be possible to monitor any natural process continuously in time and space. It is therefore important that the sampling strategy adopted is designed on the basis of some knowledge or understanding of the distribution of the natural processes at work so that the maximum information can be derived from a given number of sampling sites.

There are also distinct advantages to sampling. It reduces costs, it should facilitate the rapid collection of the most valuable data (which can be of particular importance when dealing with natural disasters) and by concentrating resources it helps to ensure that the data are of high quality. If the volume of data to be processed is reduced through sampling, then personnel of higher quality can be employed and given intensive training and this should lead to more accurate data processing.

The regionalization provides large scale information on how monitoring sites might be distributed (or how resources for monitoring might be distributed between, for example, provinces). The issue of exact siting of monitoring sites will depend on the type of hazard and an understanding of the process mechanisms giving rise to hazard events (e.g. the location of fault lines for earthquakes).

It is arguable that if resources are limited monitoring should be restricted to areas where hazard levels are severe (or at least high). The evidence of **Tables 4** and **5** are suggestive of where (at a provincial level) monitoring resources might be concentrated if the criteria are based on whether hazard levels are severe. For example, an area of north/north-east China (including the provinces of Beijing, Tianjin, Hebei, and Shaanxi) falls within this category, whereas the area further north and including the provinces of Inner Mongolia, Liaoning, Jilin, and Heilongjiang are relatively free of hazards. The provinces of Henan, Hubei, Hunan, and Guangdong in an area of Southern China also generally have low levels of hazard. Other factors will also need to be considered in relation to resource distribution for hazard monitoring. The overall population distribution will be important since there may be little purpose in intensively monitoring areas of low population. In addition, if the regionalization is spatially highly fragmented then more monitoring sites may be needed in an area than if the regionalization generates large areas (because more regions mean more strata). The more spatially autocorrelated the level of a hazard, the more information about an area any one monitoring site provides. The extent to which monitoring effort can be reduced as a function of the spatial autocorrelation in the

	Population	**GNP**	**Hazard level**
Investment	0.5556	0.9508	-0.2722
Population		0.7487	-0.2932
GNP			-0.2803

Notes Values are Pearson's product-moment correlation co-efficient

Table 8. Spatial correlation between human factors and hazards.

variable to be monitored is discussed in Ripley (1981) and Haining (1990).

7 Conclusions

This paper has presented a single, unified classification of earthquakes, floods, and droughts. The comprehensive regionalization, built using GIS, shows the large-scale spatial distribution and areas of the hazards and their intensity levels. Major hazard areas and 'islands of safety' are illustrated in the regionalization map. The reliability of the integrated regionalization depends on the reliability of the original regionalization maps of individual hazards and the effects of GIS manipulations. Whilst GIS handling does not appear to produce serious errors to the area statistics, the use of sharp boundaries rather than zones of transition overstates the extent to which clearly defined hazard regions, distinct from conditions in neighbouring areas, exist in reality. Even at this broad scale of analysis this is an area of research which justifies further consideration.

The spatial analysis on the regionalization map revealed that whilst severe earthquakes and floods do not occur in the same areas, severe earthquakes and droughts frequently happen in the same regions (although not necessarily simultaneously). There is strong spatial association between floods and droughts. Furthermore, hazards are more intense in remote areas with less human activity than in areas with dense human activity, indicating that at the provincial scale of analysis, human activities have adapted to the geographical distribution of hazards. This seems entirely reasonable, and suggests that the spatial distribution of hazards, at least at this scale, has been fairly constant over time. As noted above, more detail on the intra-provincial distribution of human activity would give a clearer picture of the extent to which adaptation has in fact taken place.

The regionalization supplies a framework for the construction of a monitoring network of the natural disasters on a nationwide scale. Other types of hazard could be integrated into the regionalization using the procedures described in the paper. The ultimate purpose of this research is to improve hazard monitoring in China, and as a contribution to this objective the work reported here provides a first indication of how resources for monitoring might be allocated. The hazard regions represent strata that justify different levels of monitoring because of the potential seriousness of the hazard events in each region. The details of any intra-provincial allocation of resources and the subsequent geographical location of monitoring sites are, however, topics requiring further research.

Acknowledgements

This work was sponsored by grants from the EU international scientific co-operation programme and the State Planning Committee of China and was undertaken whilst Jinfeng Wang held a Marie Curie Fellowship in the Department of Geography at the University of Sheffield. The following individuals provided the original data sets on which the work reported here was based: Xinlian Chen, Seismic Analysis and Prediction Center, State Seismic Bureau of China, Beijing 100080, China (earthquake data); Qichen Tang, Institute of Geography, Chinese Academy of Sciences, Beijing 100101, China (flood data); and Keran Li, Institute of Geography, Chinese Academy of Sciences, Beijing 100101, China (drought data). Jinfeng Wang would like to thank Shupeng Chen for his helpful advice. ARC/INFO was the GIS used for the work reported here. Two anonymous referees provided extremely helpful advice on the presentation of this work and the authors wish to record their appreciation.

References

Alexander D 1987 Land of disasters. *Geographical Magazine*, 59: 226–31

Alexander D 1991 Information technology in real-time for monitoring and managing natural disasters. *Progress in Physical Geography*, 15: 238–60

Alexander D 1993 *Natural Disasters*. London, University College London Press

Altman D 1994 Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems*, 8: 271–89

Blaikie P, Cannon T, Davis I, and Wisner B 1994 *At Risk*. London, Routledge

Burrough P A (1992) Are GIS data structures too simple minded? *Computers and Geosciences*, 18: 395–400

Burrough P A and Frank A U (eds) 1996 *Geographic Objects with Indeterminate Boundaries.* London, Taylor & Francis

Burton I, Kates R W, and White G F 1978 *The Environment as Hazard.* New York, Oxford University Press

Chrisman N R 1989 Modelling error in overlaid categorical maps. In Goodchild M and Gopal S (eds) *Accuracy of Spatial Databases.* London, Taylor & Francis: 21–34

Cooke R U 1984 *Geomorphological Hazards in Los Angeles.* London, Allen and Unwin

Court A 1970 Map comparisons. *Economic Geography,* 46: 435–8

Cutter S L 1993 *Living with Risk.* London, Edward Arnold

Dikau R, Cavallin A, and Jager S 1996 Databases and GIS for landslide research in Europe. *Geomorphology,* 15: 227–39

Gatrell A 1987 On putting some statistical analysis into geographical information systems: With special reference to problems of map comparison and map overlay. *Research Report No. 5 Northern Regional Research Laboratory.* Lancaster, University of Lancaster

Gatrell A and Vincent P 1991 Managing natural and technological hazards. In Masser I and Blakemore M (eds) *Handling Geographical Information.* Harlow, Longman: 148–80

Gerasimov I P and Zvonkova T B 1974 Natural hazards in the territories of the USSR: Study, control, and warning. In White G F (ed) *Natural Hazards.* New York, Oxford University Press: 243–51

Goodchild M and Gopal S (eds) 1989 *Accuracy of Spatial Databases.* London, Taylor & Francis

Gregory K J and Waling D E 1973 *Drainage Basin Form and Process.* London, Edward Arnold

Haining R P 1990 *Spatial Data Analysis in the Social and Environment Sciences.* Cambridge, Cambridge University Press

Haining R and Arbia G 1993 Error propagation through Map operations. *Technometrics,* 35: 93–305

Hewitt K and Burton I 1971 *The Hazardnous of a Place – A Regional Ecology of Damaging Events.* Toronto, University of Toronto, Department of Geography

Hiscock K M, Lovett A A, Brainerd J S, and Parfitt J P 1995 Groundwater vulnerability assessment – two case studies using GIS methodology. *Quarterly Journal of Engineering Geology,* 28: 179–94

Kemp K K 1993 Environmental modelling and GIS: Dealing with spatial continuity. In Kovar K and Nachtnebel H (eds) *Proceedings HydroGIS 93.* Wallingford, International Association of Hydrological Sciences, Publication 211: 107–15

Leung Y 1987 On the imprecision of boundaries. *Geographical Analysis,* 19: 125–51

Lewis P 1977 *Maps and Statistics.* London, Methuen

Lowry J H, Miller H J, and Hepner G F 1995 A GIS-based sensitivity analysis of community vulnerability to hazardous contaminants on the Mexico/US border. *Photogrammetric Engineering and Remote Sensing,* 61: 1347–59

Maffini G, Arno M, and Bitterlich W 1989 Observations and comments on the generation and treatment of error in digital GIS data. In Goodchild M and Gopal S (eds) *Accuracy of Spatial Databases.* London, Taylor & Francis: 55–67

Maslia M L, Aral M M, Williams R C, Susten A S, and Heitgerd J L 1994 Exposure assessment of populations using modelling, demographic analysis, and GIS. *Water Resources Bulletin,* 30: 1025–41

McKean I, Bueckel S, and Gaydos L 1991 Remote sensing and landslide hazard assessment. *Photogrammetric Engineering and Remote Sensing,* 57: 1185–93

Mejianavarro M, Wohl E E, and Oaks S D 1994 Geological hazards, vulnerability and risk assessment using GIS – model for Glenwood Springs, Colorado. *Geomorphology,* 10: 331–54

Merchant J W 1994 GIS-based groundwater pollution hazard assessment – a critical review of the DRASTIC model. *Photogrammetric Engineering and Remote Sensing,* 60: 1117–28

Nakano T 1974 Natural hazards: Report from Japan. In White G F (ed) *Natural Hazards.* New York, Oxford University Press: 231–43

Norcliffe G 1977 *Inferential statistics for Geographers.* London, Hutchinson

People's Republic of China, State Statistical Bureau 1994 *1994 Statistical Yearbook of China.*

Beijing, China Statistical Publishing House

Perry A H 1981 *Environmental Hazards in the British Isles.* London, George, Allen, and Unwin

Rejeski D 1993 GIS and risk: A three culture problem. In Goodchild M F, Parks B O, and Steyaert L T (eds) *Environmental Modelling with GIS.* New York, Oxford University Press: 318–31

Rhind D W 1991 Environmental monitoring and prediction. In Masser I and Blakemore M (eds) *Handling Geographical Information.* Harlow, Longman

Ripley B 1981 *Spatial Statistics.* New York, John Wiley & Sons, Inc.

Sengupta S and Venkatachalam P 1994 Health hazard assessment in an industrial town with the help of GRAM-GIS. *Environmental Monitoring and Assessment,* 32: 155–60

Smith K 1992 *Environmental Hazards – Assessing Risk and Reducing Disaster.* London, Routledge

Szucs M 1995 Soil trace elements availability maps of Hungary. *Computers, Environments, and Urban Systems,* 19: 117–22

Tyler C 1989 Climatology – technology's answer to the British weather. *Geographical Magazine,* 61: 44–7

van Westen C J and Terlien M T J 1996 An approach towards deterministic landslide hazard analysis in GIS – a case study from Manizales, Colombia. *Earth Surface Processes and Landforms,* 21: 853–68

Veregin H 1989 Error modelling for the map overlay operation. In Goodchild M and Gopal S (eds) *Accuracy of Spatial Databases.* London, Taylor & Francis: 3–18

Wadge G, Wislocki A, Pearson E J, and Whittow J B 1993 Mapping natural hazards with spatial modelling systems. In Mather P (ed) *Geographical Information Handling – Research and Applications.* Chichester, John Wiley & Sons, Ltd: 239–50

Zadeh L A 1965 Fuzzy sets. *Information and Control,* 8: 338–53

Appendix 1

Province acronyms

11	BJ-Beijing
12	TJ-Tianjin
13	HB-Hebei
14	SX-Shanxi
15	NM-Inner Mongolia
21	LN-Liaoning
22	JL-Jilin
23	HLJ-Heilongjiang
31	SH-Shanghai
32	JS-Jiansu
33	ZJ-Zhejiang
34	AH-Anhui
35	FJ-Fujian
36	JX-Jiangxi
37	SD-Shandong
41	HEN-Henan
42	HUB-Hubei
43	HUN-Hunan
44	GD-Guangdong
45	GX-Guangxi
46	HAN-Hainan
51	SC-Sichuan
52	GZ-Guizhou
53	YN-Yunnan
54	XZ-Tibet
61	SAX-Shaanxi
62	GS-Gansu
63	QH-Qinghai
64	NX-Ningxia
65	XJ-Xinjiang
71	TW-Taiwan

Appendix 2

pr	pop_ss	gnp_ss	inv_ss	pop_s	gnp_s	inv_s	pop_hh	gnp_hh	inv_hh	pop_h	gnp_h	inv_h
11	747.95	580.8	281.7	366.7	284.76	138.11						
12	339.99	196.4	82.68	467.28	269.95	113.63	53.38	30.84	12.98	15.6	9.01	3.79
13	310.37	76.76	22.17	3624.46	896.46	258.9	482.1	119.24	34.44	696.74	172.33	49.77
14	343.68	73.68	26.81	1691.35	362.62	131.93	33.21	7.12	2.59	913.25	195.8	71.24
15	26.78	5.84	2.48	338.52	73.8	31.34	22.32	4.87	2.07	301.32	65.69	27.9
21				392.65	175.65	68.19	346.5	154.98	60.16	3377.96	1511.1	586.6
22				23	6.05	2.23				1078.78	283.71	104.83
23										79.92	23.65	7.26
31												
32				863.91	341.56	138.53	41.8	16.53	6.7	4856	1919.88	778.66
33				1011.04	402.44	185.37				2435.89	969.58	446.6
34				336.13	55.83	16.56	212.3	35.26	10.46	2785.2	462.65	137.21
35				716.63	233.9	84.39	367.5	119.95	43.28	2058	671.72	242.36
36				194.33	34.39	9.62	27.76	4.91	1.37	3742.91	662.43	185.21
37	674.08	210.8	69.58	4447.75	1390.9	459.12	668.3	208.99	68.99	1953.09	610.76	201.61
41				1205.7	213.29	62.54	160.8	28.44	8.34	2909.76	514.75	150.92
42				332.9	76.46	22.86	90.45	20.77	6.21	5213.32	1197.42	358.02
43							303.5	57.35	15.61	6010.48	1135.63	309.12
44				1872.61	914.14	464.07	336.8	164.4	83.46	3292.17	1607.12	815.86
45							438	77.78	26.87	2238.3	397.45	137.3
46				482.82	155.07	114.79	94.54	30.37	22.48	123.43	39.64	29.35
51	22.21	3.92	1.16	3470	612.09	181.43	965.7	170.34	50.49	5036.46	888.41	263.33
52							95.45	11.44	2.87	1588.19	190.3	47.74
53	792.94	135.2	53.56	1235.23	210.56	83.44	404.4	68.94	27.32	979.22	166.92	66.14
54				65.73	10.56	5.17				146.93	23.61	11.56
61				354.63	63.29	22.16	165.3	29.5	10.33	1910.87	341.05	119.39
62				833.78	127.41	34.66				812.93	124.22	33.8
63				166.69	37.74	16.45				347.01	78.57	34.24
64	88.28	17.56	7.87	216.15	42.99	19.27				85.8	17.06	7.65
65				259.81	78.01	41.17				639.99	192.17	101.41

Table A1: Area of human factors falling in each hazard region (level 2) broken down by province.

pop_3m	gnp_3m	inv_3m	pop_mm	gnp_mm	inv_mm	pop_m	gnp_m	inv_m	pop_lll	gnp_lll	inv_lll	pr
												11
48.45	27.99	11.78										12
538.39	133.16	38.46	499.68	123.59	35.69	461	114.02	32.93	95.01	23.5	6.79	13
			180.72	38.75	14.[illegible]							14
4.46	0.97	0.41	26.78	5.84	2.48	338.5	73.8	31.34	1120	244.1	103.7	15
			105.09	47.01	18.25	1415	632.85	245.7				21
			191.63	50.4	18.62	899.9	236.68	87.45	347.8	91.46	33.8	22
			571.32	169.03	51.88	1116	330.11	101.3	1867	552.5	169.6	23
						1349	1511.6	642.4				31
13.93	5.51	2.23				1191	471.02	191				32
						806.3	320.93	147.8				33
625.99	103.98	30.84	283.06	47.02	13.94	494.4	82.13	24.36				34
												35
												36
250.62	78.37	25.87	633.75	198.18	65.42							37
2095.2	370.66	108.68	2556.09	452.18	132.58							41
												42
												43
996.71	486.56	247										44
1335.3	237.1	81.91	382.05	67.84	23.44	22.19	3.94	1.36				45
												46
			977.55	172.43	51.1[illegible]	66.62	11.75	3.48				51
			1313.47	157.38	39.48	399.1	47.82	12				52
46.62	7.95	3.15	186.28	31.75	12.58							53
						18.33	2.95	1.44				54
809.11	144.41	50.55	122.23	21.82	7.64							61
107.87	16.48	4.48				302.8	46.26	12.59	42.21	6.45	1.75	62
						28.93	6.55	2.85				63
5.44	1.08	0.49	16.83	3.35	1.5							64
						462.4	138.86	73.28	173.5	52.11	27.5	65

Appendix 2

Prov (code)	POP -10000	GNP (100m)	INV (100m)	AREA (10000 sq. km)	A_SSS	A_SS	A_S	A_HHH	A_HH	A_H	A_MMM	A_MM	A_M	LLL	Id
11	1112	864	419	16.8		11.3	5.5								11
12	928	536	226	11.3		4.1	5.7		0.7	0.2	0.6				12
13	6334	1567	452	180		8.8	103.1		13.7	19.8	15.3	14.2	13.1	2.7	13
14	3012	646	235	156		17.8	87.6		1.7	47.3		9.4			14
15	2232	487	207	1200		14.4	182.1		12.1	162.1	2.4	14.4	182	602	15
21	4042	1808	702	140			13.6		12.1	117		3.6	49.1		21
22	2555	672	248	180			1.6			76.1		13.5	63.4	24.5	22
23	3640	1077	331	460						10.1		72.2	141	236	23
31	1349	1512	642	6.2									6.2		31
32	6967	2754	1117	100			12.4		0.6	69.7	0.2		17.1		32
33	4266	1698	782	100			23.7			57.1			18.9		33
34	5897	980	291	130			7.4		4.7	61.4	13.8	6.2	10.9		34
35	3150	1028	371	120			27.3		14.1	78.4					35
36	3966	702	196	160			7.8		1.1	151.1					36
37	8642	2702	892	150		11.7	77.2		11.6	33.9	4.4	11.1			37
41	8949	1583	464	167			22.5		3.1	54.3	39.1	47.7			41
42	5653	1298	388	180			10.6		2.9	166.1					42
43	6311	1192	325	210					10.1	200.1					43
44	6607	3225	1637	175			49.6		8.9	87.2	26.4				44
45	4438	788	272	230					22.7	116.1	69.2	19.8	1.15		45
46	701	225	167	34.7			23.9		4.68	6.1					46
51	11104	1959	581	560		1.1	175		48.7	254		49.3	3.4		51
52	3409	408	102	170					4.76	79.2		65.5	19.9		52
53	3885	662	262	390		79.6	124.1		40.6	98.3	4.68	18.7			53
54	232	37	18	1200			340.1			760.1			94.8		54
61	3443	615	215	200			20.6		9.6	111.1	47	7.1			61
62	2345	358	97	450			160.1			156.1	20.7		58.1	8.1	62
63	467	106	46	720			257.1			535.1			44.6		63
64	495	98	44	60		10.7	26.2			10.4	0.7	2.1			64
65	1605	482	254	1600			259.1			638.1			461	173	65

Notes

A_SS denotes the size of area with SS severity levels in a province, etc. A blank means no record

Table A2. Spatial distribution of human factors and hazard severity (Level 2) by province (data on provincial levels of human factors from People's Republic of China 1994 *Statistical Yearbook of China*).