Effect of topography on the risk of malaria in the Usambara Mountains, Tanzania

Balls, Mandy Jayne

Effect of topography on the risk of malaria in the Usambara Mountains, Tanzania

Mandy Balls

Abstract

There has been a progressive rise in malaria in parts of the African highlands over the last 50 years. In this area of unstable malaria, devastating epidemics are experienced at irregular intervals. Altitude plays a very important role in determining malaria transmission and infection. However, other landscape features may also influence this relationship. This research investigates whether the risk of malaria is related to the shape of the surrounding land, at various altitudes. We hypothesized that households situated close to flat areas where water is expected to accumulate, and are thus potential mosquitoes breeding sites, are at greater risk from malaria than those further away.

Cross-sectional clinical surveys were carried out in seven villages along an altitudinal transect rising from 300 m to 1650 m in the western Usambara Mountains, Tanzania. Each village was mapped and incorporated within a geographical information system (GIS). Univariate analysis showed that the risk of an enlarged spleen was positively correlated with decreasing altitude. Other influential topographic variables identified were: water accumulation, flatness and swampiness. Logistic regression analysis produced two models and their equations were used in the GIS to map the risk of malaria infection within each village area. Model 1included only altitude and correctly predicted the malaria status of 73% of households, whereas Model 2 incorporated altitude and the amount of swampiness within 400 m radius of each household to predict with 76% accuracy whether households were positive or not.

We have identified that between 750 m and 1200 m, characteristics of the landscape play an important role in governing malaria risk. At these elevations malaria is highly unstable, and favourable meteorological conditions can cause malaria epidemics. This novel approach of exploring how topography affects the risk of malaria could be used to identify epidemic-prone areas in other African highland regions and help to improve the targeting of control activities in high-risk areas.

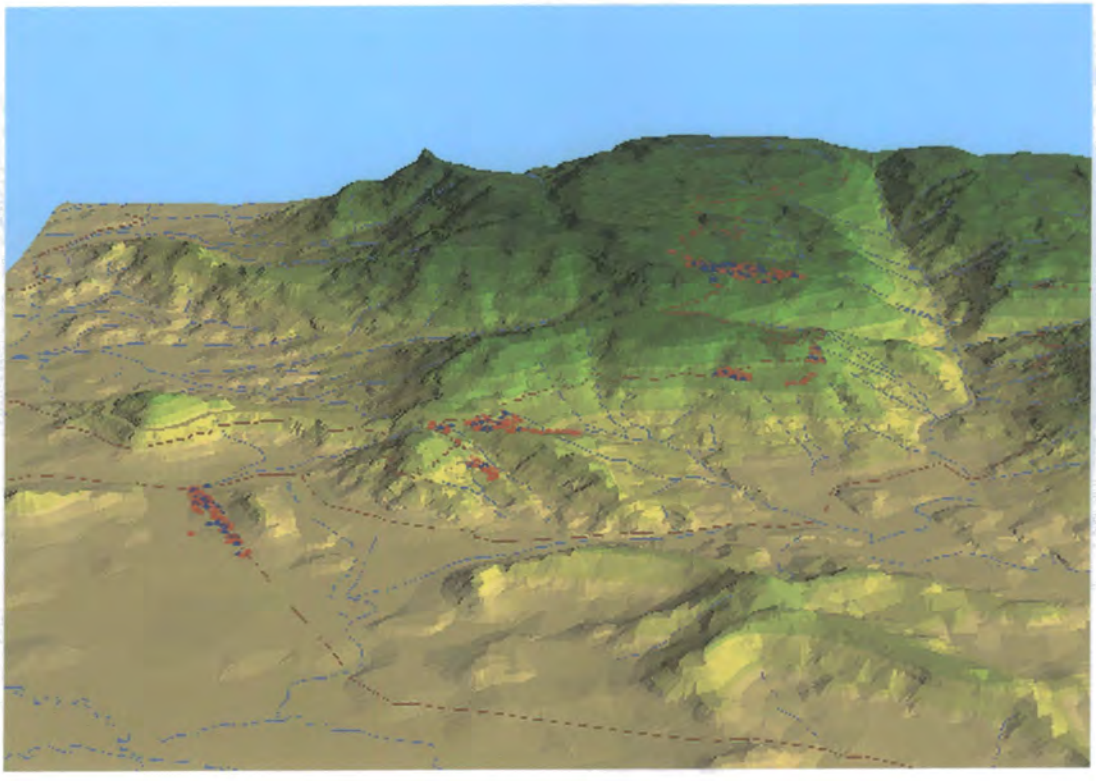
Effect of topography on the risk of malaria in the Usambara Mountains, Tanzania

Mandy Jayne Balls

Presented for the qualification:
Degree of Master of Science (By Thesis)
in accordance with the regulations of the University of Durham

Department of Biological Sciences
Science Laboratories
South Road
Durham DH1 3LE

2001



Surface representation showing the surrounding terrain for the villages situated at 300 m, 600 m, 750 m and 1000 m (malaria positive houses shown in red, negative houses in blue)

Acknowledgements

I would like to thank the following people who have been a great help to me during the course of this research:

Firstly, I am very grateful to my partner Simon Jones for his support and encouragement.

I would like to thank Dr Steve Lindsay from the University of Durham, who originally presented me with the opportunity to carry out this research. Dr Lindsay has been an excellent tutor to me regarding the complexities of malaria epidemiology and has inspired me throughout this work. I really appreciate all the time, advice and support he gave. I also wish to thank Dr Chris Thomas from the University of Durham for his help, encouragement and guidance regarding the GIS aspects of this project.

I am very grateful to Dr René Bødker from the Danish Bilharziais Laboratory (DBL), who allowed me to use clinical and census data collected for his PhD Thesis. I would also like to thank him for everything he organised for me regarding my fieldwork in Tanzania.

I am grateful to Mr W. Kisinza from the National Institute for Medical Research in Tanzania (NIMR), whose sketch maps and knowledge of the villages were a big help to me. I would also like to thank Prof. W. Kilama, Dr J. B. Rugemalila and Dr C. Njumwa from NIMR for allowing this project to go ahead. I am grateful to Dr H. Msangeni from NIMR and D. Shayo from DBL for planning and organising the parasitological surveys and L. Malle, Z. Savael, W. Chambo, G. Kilua and J. Raphael for collecting and reading the slides.

Towards the end of my fieldwork I met Miss J. Lucas who translated for me – I wish I had made her acquaintance at the beginning. Mapping the villages could not have been done without the help of the village assistants: Mr Y. Gumbo, Mr Hussein, Mr Chambou, Mr R. Billa, Mr S. Mahanyo, Mr S. Kipingu and Mr H. Bakari. I am extremely grateful to everyone from the seven villages who took part in the surveys.

Finally, I would like to thank staff from the Environmental Science Department of the University of Durham for their support, especially Prof. Peter Evans and Mr Mike Alexander.

Contents

List of Tables

List of Figures

List of Abbreviations and glossary of terms used

AVHRR	Advanced Very High Resolution Radiometer
Basic reproduction rate	The estimated number of secondary malaria infections potentially transmitted within a susceptible population from a single non-immune individual
Coverage	Map layer, i.e. stream coverage, contour coverage, road coverage, etc
DEM	Digital elevation model
DTM	Digital terrain model
EIR	Entomological Inoculation Rate: The number of infectious mosquito bites a person is exposed to in a certain time period, typically a year
Endemic malaria	Measurable transmission and incidence every year (stable malaria)
ENSO	El Niño-Southern Oscillation
Epidemic malaria	Occasional malaria outbreaks in normally malaria-free regions; a particularly severe malaria season in a normally low-risk area (unstable malaria)
GCM	General Circulation Models
GIS	Geographical Information System(s): Computer programmes that combine spatial and descriptive (attribute) data for mapping and spatial analysis
GPS	Global Positioning System
HIMAL	Epidemiology of highland malaria project
Holoendemic	Areas of highest endemicity. Spleen rates in children (2-9 years) constantly over 75%, but low in adults
Hyperendemic	Spleen rates in children (2-9 years) are constantly over 50% and also high in adults (over 25%)
Hypoendemic	Low malaria endemicity. Spleen rate in children (2-9 years) not exceeding 10%
IDRC	International Research Centre of Canada

MARA/ARMA	Mapping Malaria Risk in Africa/Atlas du Risque de la Malaria en Afrique
MBR/HBR	Man Biting Rate/Human Biting Rate: Bites per person per night by vector population
Mesoendemic	Spleen rates in children (2-9 years) between 11-50%. Development of immunity is delayed and so adults may also display symptoms
MIASMA	Modelling framework for the health Impact Assessment of Man-induced Atmospheric changes
NDVI	Normalized Difference Vegetation Index: A satellite derived index which gives a measure of the "greeness" of plant growth - measures chlorophyll activity in plants
NOAA	National Oceanographic and Atmospheric Administration
Prevalence	Percentage of survey population testing positive for malaria
Raster format	Spatial objects are represented as cells within a grid
RS	Remote Sensing
Sporogonic cycle	Sexual development of / incubation period of malaria parasite in mosquito; time required for a mosquito to become infective after feeding on an infected person
TIN	Triangulated irregular network
Transmission	Spread of malaria by completion of a full transmission cycle (man - mosquito - man)
UNDP	United Nations Development Programme
Vector format	Spatial data representing points, lines or area information
Vectorial capacity	The potential number of inoculations originating from one infective person each day
VOC	Volatile organic compound
WHO	World Health Organisation

Declaration

I confirm that the thesis conforms with the prescribed word length for the degree for which I am submitting it for examination.

I confirm that no part of the material offered has previously been submitted by me for a degree in this or in any other University. If material has been generated through joint work, my independent contribution has been clearly indicated. In all other cases material from the work of others has been acknowledged and quotations and paraphrases suitably indicated.

1 Introduction

1.1 Background and Rationale

Malaria remains one of the most serious threats to human health in the developing world, with 40% of the world's population at risk from this disease (WHO, 2001). About 90% of the clinical cases occur in sub-Saharan Africa, where 1-2 million deaths occur each year (WHO, 1997). In recent years, there has been a growing recognition of the importance of highland malaria, and the need for greater scientific understanding of its epidemiology. Malaria in the African highlands is on the increase. Historically the highlands have been a refuge from the malarial lowlands, malaria being uncommon or absent at higher elevations (Lindsay & Martens, 1998). Many reasons have been put forward for this increase, including: climate change, population growth and migration promoting the extension of parasites and their resistance to antimalarials, development of land for agriculture and subsequent irrigation projects increasing the number of potential mosquito breeding sites, the transportation of mosquitoes from the endemic lowlands via newly built roads and railways, and declining or non existent health systems.

In the highlands, transmission is probably much more focal in its distribution than in many lowland areas; breeding sites are more common on the valley floor than the steep valley slopes and the colder nights at higher altitudes restrict the dispersal of adult mosquitoes from these sites. There is evidence that *Anopheles gambiae s.l.*, the principal vector of malaria in Africa, may fly as little as several 100 m in altitude up the hill (Manga *et al.*, 1993), although lateral movement along upland valley floors may be substantially greater. Altitude is a major factor governing the distribution of malaria in the highlands, as altitude increases temperature declines by approximately 0.6°C per 100 m (Adams *et al.*, 1996). Highland populations are especially at risk from malaria epidemics due to their lack of acquired immunity, and deaths can occur in all age groups. Identifying epidemic-prone villages would support malaria control activities, allowing interventions to be directed at these high risk communities.

The Danish Bilharziasis Laboratory, University of Durham and the National Institute of Medical Research, Tanzania have been researching the epidemiology of malaria in the Usambara Mountains in Tanzania. The first stage of this research was to describe the

microepidemiology of malaria along an altitudinal gradient. In 1996 cross-sectional clinical surveys and longitudinal studies of the vectors were carried out in seven villages along a transect between 300 m and 1650 m above sea level. One of these villages, at 1200 m, experienced an epidemic at the end of 1996, where 60% of the population were infected with malaria parasites (Bødker, 2000).

In this study, using the clinical data collected in 1996, we examined how local topography affects the risk of infection with malaria parasites within a village. As long ago as 1883, even before the mechanisms of malaria transmission had been discovered, Hirsch (1883) recognised that highland epidemics were persistently located in *'a valley with a small declivity'* or a *'basin-like depression in a plateau'*; places where water collects and malaria vectors breed. Since then, this interesting observation has not been pursued, until now. The focus of this study is therefore to identify areas where water may collect in this mountainous area, and relate this to the risk of malaria between villages using geographical information systems (GIS).

Recent technological advances have supplied new tools for mapping malaria and identifying ecological determinants of disease distribution. There is now an international initiative, which is using historical data sets to map the distribution of malaria at a continental scale for Africa (MARA/ARMA, 1998; Snow *et al.*, 1999). At a local level remote sensing techniques and GIS have been used to locate mosquito-breeding sites (Wood *et al.*, 1992; Thomson *et al.*, 1996; Hay *et al.*, 1998). One area where there has been little research concerns the influence that topography plays on the distribution of malaria in highland regions. GIS can provide the means to model topographical phenomenon. Digital Elevation Models (DEMs) have been used regularly in the ecological and environmental sciences for estimating topographic variables such as: elevation, slope, aspect and ruggedness (Petrie & Kennie, 1990). These data are used frequently to build surface hydrological models from which parameters such as flow channels, direction of flow, flow rates, flow accumulation, sink and pour points can be derived. Thus, DEMs are an effective tool for quantifying many landscape features of mountainous regions that are likely to affect biological systems, such as malaria. Models of infection risk based upon these parameters could provide a fast and accurate method for identifying malaria 'hot spots' in remote highland areas. Such models may have wide applicability and be relevant to other highlands in East Africa.

1.2 General Goal

To investigate the role of topography to the risk of malaria in the western Usambara's, Tanzania.

1.3 General objectives

1. Produce digital elevation models (DEMs) of each study area to include the location of each survey village and their houses.

2. Generate hydrological and terrain models that identify flat areas and water inundation.

3. Identify topographic variables that influence malaria prevalence and describe the relationship between topography and malaria distribution.

4. Produce models that map the risk of malaria for the areas surrounding and including each of the survey villages.

5. Investigate spatial disease clustering of malaria for each village in relation to the topographic variables.

1.4 General Hypothesis

To test the hypothesis that elevation and terrain are strong predictors of malaria prevalence in the Usambara Mountains, in north-eastern Tanzania.

Are individuals living close to flat areas at greater risk of malaria than those living further away?

1.5 Strategy of study

Chapter 2 introduces the main aspects of malaria epidemiology, identifying the main geographical areas at risk and the populations involved. It also describes the major historical breakthroughs in the discovery of the life-cycle of the malaria parasite, its mode of transmission, and measures to combat both the mosquito vector and parasite.

The factors that affect the distribution are discussed in relation to the parasite, vector and human host. The consequences of global warming on the distribution and transmission intensity of malaria are also examined. Highland malaria and the reasons for its increase are considered, and the need for reliable models to map aspects of malaria epidemiology is supported and examples of such models are introduced. Chapter 3 provides some background information regarding the use of GIS in disease mapping, including the use of satellite remote sensing. Some of the theory relating to the production of DEMs and hydrological models is explained, together with a critique of their value with regard to other terrain models that have been developed in other research. In Chapter 4 the research methods are described, from collecting the house locations of the survey villages in the Usambara Mountains using a Global Positioning System (GPS), to the development of the GIS, followed by the statistical analysis carried out to identify the topographical variables that influence malaria prevalence to producing the predictive risk models. The results of this research are presented in Chapter 5. Finally, in Chapter 6 the research findings are discussed with an examination of the influence the topographic variables have on prevalence with regard to the predictive models produced. The models limitations are presented, considered, and put into context alongside other malaria mapping and modelling research. The relevance of this study is argued, and suggestions for future work are put forward.

2 Malaria epidemiology

2.1 Who is at risk?

Malaria is the most prevalent vector-borne disease in the world, affecting people in some 90 countries. It is estimated that about 40% of the world's population is at risk, and there are around 300 to 500 million clinical cases of malaria each year (WHO, 2001). Malaria is responsible for between 1.5 and 2.7 million deaths annually (Knell, 1991; Butler, 1997; Wellcome Trust, 1999). Most deaths are in sub-Saharan Africa, where some 90% of the clinical cases occur (Dobson *et al.*, 1999); two-thirds of the remaining cases are concentrated in just six countries. These are in order of decreasing prevalence, India, Brazil, Sri Lanka, Vietnam, Colombia and the Solomon Islands (Butler, 1997). Warm, humid, tropical areas are most at risk from malaria, the climate experienced in equatorial countries. Regions free from indigenous malaria are those located above latitudes of 60° N and 30° S, and dry desert regions. Altitude also constrains malaria transmission, due to its effect on temperature; it cannot survive above 3000 m anywhere in the world (Knell, 1991). This altitude limit decreases rapidly with distance from the equator.

The risk of severe malaria is almost exclusively limited to individuals that have no acquired immunity. In highly endemic areas the most vulnerable are children older than 3 – 6 months, who have lost the immunity transferred from their mother, up to the age of 5 years, when surviving children have developed their own immunity (Najera, 1992). Pregnant women are also susceptible, especially in parts of the tropics where malarial transmission is unstable and they have a poor degree of acquired immunity. Malaria is a major cause of maternal death, abortion, stillbirth, premature delivery and low birth weight, particularly for those women undergoing their first pregnancy (Gilles & Warrell, 1993). In areas of low and infrequent transmission, individuals will not have acquired this protective immunity and as such will be vulnerable to malaria epidemics when favourable conditions arise to intensify transmission. One group at high risk from malaria are those communities in the African highlands, where there has been a gradual rise in malaria over the last fifty years (Lindsay & Martens, 1998). In this area of fringe transmission malaria is inherently unstable, and epidemics may result from relatively subtle changes in climate. These outbreaks can have devastating consequences, since the highland communities often have little, or no immunity to malaria. Other non-

immunes also at risk are immigrants and travellers from non-endemic areas who may relocate to or visit regions where malaria is endemic.

In Africa especially, mechanisms to overcome malaria are hindered by many obstacles. In endemic areas individuals may receive up to 1000 infectious bites per year (Butler, 1997), gradually acquiring a natural immunity that allows them to tolerate a parasite load without developing severe symptoms. However, if they were to travel to an area where the risk of malaria is low they would lose this protective immunity after a few months and would then be at risk from developing clinical malaria if they return. Resistance to chloroquine is spreading, reducing the impact of a drug which in the past had helped to limit the damage caused by the disease. Resistance to Fansidar, the only affordable drug for extensive use in Africa is also widespread (Knell, 1991). African Ministries of Health have very limited budgets to tackle such disease burdens and the decision of where to target such limited resources is not an easy one. Moreover, health and malaria in particular is low on the list of priorities for government spending in most African countries. There is a great need for political commitment to reverse widespread pessimism about malaria in Africa (Butler, 1997).

2.2 Brief history of malaria discovery and control activities

Since Prehistoric man, malaria has been a feature of humankind's existence. However, due to the complicated nature of its transmission and life cycle very little was known of its range, cause or treatment for many thousands of years (Gilles & Warrell, 1993). The most important events in the history of malaria took place at the end of the nineteenth century. Before then the mode of transmission from person to person was still unknown, although the relationship between marshes, fevers and mosquitoes were considered. In 1878, Patrick Manson, a Scotsman practising medicine in China showed that a mosquito can act as a vector of human filaria, and shortly afterwards, in 1880 the malaria parasites were first seen in human red blood cells and described by Laveran, a French army surgeon in Algeria. Improvements to the microscope, and advances in developing new methods of staining the malaria parasites in the blood were soon to follow and helped further the study of the *Plasmodia* parasite. The actual mode of transmission became clear when Ronald Ross working in India in 1897 found a developing form of the malaria parasite in the body of a mosquito that had previously fed on a patient with the *Plasmodia* in his blood. By 1900, in Italy Grassi, Bignami and

Bastianelli had described the cycle of human malaria parasites in *Anopheles* mosquitoes, and Manson had confirmed the mosquito-malaria transmission theory by experiments on human volunteers near Rome and in London.

During the twentieth century much research has been devoted to malaria control. Larvicides such as oil and Paris green were first used to prevent the breeding of mosquitoes in various types of waters. Such measures proved highly successful in controlling malaria in Cuba and the Panama Canal Zone and other control initiatives followed. The search for synthetic antimalarial drugs began in earnest after the severe effect of malaria experienced during the First World War. Pamaquine was discovered first in 1924 followed by mepacrine in 1930, which played a critical role during the Second World War. The development of other valuable synthetic drugs then followed. Another major breakthrough was the discovery in 1942 of the highly insecticidal action of DDT. The high potency against the mosquitoes, low toxicity to humans, ease of application in rural areas and fairly low cost of DDT, persuaded many health workers to press for malaria eradication. This concept was adopted by the World Heath Organisation (then the World Health Assembly), and in 1955 it launched a global eradication campaign (Dobson *et al.*, 1999). Over the next 15 years excellent results were achieved in Europe, North America, parts of Asia, the former USSR and Australia, but in tropical countries and Africa in particular the outcome was much less impressive. In 1969 the WHO revised the policy for malaria eradication by underlining the need for greater involvement of general health services and for extension of research on new insecticides, improved surveillance, development of new antimalarial drugs and alternative methods of malaria control (Gilles & Warrell, 1993). However, today the malaria situation is depressing since it continues to be a drain on human resources in the tropics. Most importantly, the malaria parasite has proved to be extremely resilient and adaptable, fast becoming resistant to antimalarial drugs such as chloroquine. There does not seem to be an easy solution, but the WHO is determined to reduce mortality by 50% by 2010 in their new campaign to 'Roll Back Malaria' (Dobson *et al.*, 1999).

2.3 Malaria ecology

2.3.1 The vectors and parasites responsible

Malaria is a complex disease and is influenced by a variety of environmental and socio-economic factors. The incidence of malaria is governed by the abundance of anopheline

mosquito species, human behaviour, and the presence of malaria parasites. Humans are the vertebrate host of the human *Plasmodia*; the female *Anopheles* mosquito is the invertebrate host and is the agent of transmission; the malaria parasite is the agent of infection. The distribution and population dynamics of malaria are most influenced by abiotic rather than biotic factors (Martens *et al.*, 1995). The life cycle of the malaria parasite involves many stages that rely on the mosquito and human host. This makes targeting and destroying the malaria parasite a very complicated task. To understand the disease we need to know the factors that determine its spatial distribution and that of the mosquito host. However, this relationship is further complicated due to the tendency of the different species of parasite and mosquito to specialise in different environments. For instance, malaria is caused by four species of the genus *Plasmodium. Plasmodium vivax* has the widest geographical range in temperate and tropical zones, but is rarely fatal; *P. falciparum* predominates in Africa, south-east Asia and Central America, and is the most dangerous clinically; *P. malariae* has a similar range to *P. falciparum*, but is much less frequent; and *P. ovale* is similar to vivax and supersedes it in West Africa (Knell, 1991; Martens *et al.*, 1999).

The vector responsible for malaria transmission is the mosquito of the genus *Anopheles*. There are around 400 known species and their distribution is worldwide, however, only about 70 species are vectors of malaria under natural circumstances (Gilles & Warrell, 1993). These species vary considerably in their ability to transmit malaria. The most important vector species in Africa, the *Anopheles gambiae* complex, are the most efficient vectors in the world and are a main reason for the severe burden of disease in this region (Martens *et al.*, 1999). The female mosquitoes obtain the infection from a human individual with parasites in the blood. To be a successful vector a species must be present in sufficient numbers in or near human habitations. A species with a clear preference for human rather than animal blood is a better vector.

2.3.2 The effect of temperature and rainfall

The longevity of the mosquito is an overriding factor in malaria transmission. This varies in different places in relation to temperature and humidity. The development of *Plasmodia* in the *Anopheles* depends on a minimum temperature below which it does not occur. Malaria parasites cease to develop in the mosquito when temperatures are below 15°C (Gilles & Warrell, 1993). Above this threshold, the amount of transmission is dependent on temperature and other environmental variables. The best conditions for

the development of *Plasmodia* in the *Anopheles* and the transmission of infection are when the mean temperature is within the range of 20-30°C, while the mean relative humidity is at least 60% (Gilles & Warrell, 1993). Table 2.1 shows how the length of the sporogonic cycle, which is the sexual phase of development within the *Anopheles,* shortens with increasing temperature. Very high temperatures of over 35°C are lethal to the malaria parasite.

Table 2.1. Length of sporongic cycle in *Plasmodiam vivax* and *P. falciparum* under different temperatures (Gilles & Warrell, 1993)

Malaria parasite	Temperature	Length of sporogonic cycle
P. vivax	below 15°C	completion of sporogonic cycle unlikely
	20°C	16 days
	28°C	8-10 days
P. falciparum	below 16°C	completion of sporogonic cycle unlikely
	20°C	22 days
	23°C	15-17 days
	28°C	9-10 days

It is only the female *Anopheles* that transmit malaria. The female mosquito requires a blood meal before the first batch of eggs can develop, this is termed the gonotrophic cycle. Females usually lay their first batch of eggs within 3-6 days after emergence (Gilles & Warrell, 1993). In subsequent cycles a batch of eggs produced by the ovaries develops after each blood meal. They generally bite at regular intervals of 2-5 days. A small rise in temperature can make a big difference illustrated by *An. gambiae* and *An. funestus* that will feed every 2 days at around 25°C, whereas they only feed every 3 days at lower temperatures (Gilles, 1953). Therefore, in warm conditions their reproductive potential is enormous and is a major factor in the success of the malaria parasite (Knell, 1991). Eggs are always deposited in or near water, and the number per batch varies between 70 and 200 (Najera, 1992). The duration of the cycle from egg to adult *Anopheles* may vary between 7 days at 31°C and 20 days at 20°C. Each species has its own optimum temperature range. The length of life of adult *Anopheles* varies between species also, but even more so in relation to external factors, particularly temperature, humidity and the presence of natural enemies. When the mean temperature is greater than 35°C or humidity is below 50%, the longevity of *Anopheles* is drastically reduced (Gilles & Warrell, 1993). Under favourable climatic conditions the female *Anopheles* often lives between 10-14 days, but occasionally up to a month and longer (Knell, 1991). A high relative humidity lengthens the life of the mosquito and enables it to live long enough to transmit the infection to several individuals.

Temperature and humidity therefore play a very important role in determining the range and transmission of malaria. Higher temperatures and humidity generally intensify the rate of development and activity of these organisms, therefore helping to increase the transmission of the disease. In areas where malaria is endemic, a small rise in temperature will not have a dramatic effect, but in areas where temperature is lower and there is no malaria, or where malaria is unstable, a small rise in temperature may increase the risk of malaria dramatically (Lindsay & Birley, 1996). Predictions that global warming will lead to a rise in mean global temperature of between 1.0 – 3.5°C by 2100 (Githeko *et al.*, 2000) will have significant impacts on the transmission of malaria. Areas most at risk will be those at higher altitudes and latitudes where malaria is so far presently constrained by lower temperatures (Martens, 1999; Rogers & Randolph, 2000).

The main factor limiting the number of mosquitoes in any area is generally the availability of mosquito breeding sites (Lindsay & Birley, 1996). Mosquitoes normally lay their eggs in water. Most *Anopheles* larvae favour clean, fresh water, although some prefer brackish or salt water. These habitats can be temporary or permanent, natural or man-made. There is great diversity in the types of habitats used by the various species that depend on many factors. Examples of breeding habitats include: large open marshes, open pools in fields, cattle hoof prints, running fresh water and tree holes (Gilles & Warrell, 1993). Rainfall distribution plays a major influence on the seasonality of suitable breeding places and hence is largely responsible for the marked seasonal cycles in mosquito densities and malaria transmission (Najera, 1992). When the rainy season starts many breeding places are created and the anopheline population increases dramatically as a result. However, the amount and duration of the rainfall influences the availability of breeding sites in different ways, depending on the environment. Excessive rainfall may transform small streams into rapid torrents, flushing out and stranding anopheline larvae. Too little rain on the other hand may change many rivers into pools in which certain *Anopheles* would breed in profusion, thereby intensifying transmission (Loevinsohn, 1994). Craig *et al* (1999) demonstrate that for stable malaria to occur in sub-Saharan Africa, 80 mm of rainfall is required per month for at least five months as long as temperatures are suitable. However, it is difficult to make a direct link between the amount of rainfall and the abundance of the mosquito vector. For example, *An. gambiae s.l.* reproduce rapidly in small temporary

pools formed after rainfall, whereas in permanent water bodies predation becomes important. Conversely, *An. funestus* favour more permanent water bodies. Rainfall also increases relative humidity and thus, the longevity of female *Anopheles.*

2.3.3 Human influence

Environmental development will also influence the transmission of malaria, in many ways. By altering their surroundings, humans can inadvertently create favourable habitats for the mosquito to live and breed. In tropical Africa, deforestation may promote malaria transmission, by producing large open areas, often pitted with depressions in the treeless landscape. Rainfall may collect in these 'sinks', creating unshaded freshwater pools, ideal breeding habitats for the efficient vector, *An. gambiae* (Gilles & deMeillon, 1968). In addition, according to Hamilton & Macfadyen (1989), deforestation can also lead to an increase in temperature of 3-4°C. Any land use change that can potentially increase the number of breeding sites for the mosquito vector and can affect the intensity of malaria transmission. Most developing countries in Africa and elsewhere depend on agricultural schemes for their economic development. Unfortunately, well-intentioned agricultural projects requiring irrigation canals, have often resulted in the creation of new breeding sites for mosquitoes. In Tigray region in northern Ethiopia, many mirodams have been built and more are planned to minimise the dependence on rain fed agriculture, and improve agricultural productivity in response to severe famines. However, together with the advantages, children in villages near recently constructed microdams had a significantly increased risk of malaria (Ghebreyesus *et al.*, 1999). Other agricultural developments, such as inland valley swamp development for the cultivation of rice in South Central Sierra Leone was associated with an increase in the prevalence of malaria infection (Gbakima, 1994).

Vegetation also plays an important role. Although mosquitoes usually prefer sunlit areas for breeding, they favour shaded dark areas to rest. In Sri Lanka individuals whose houses are near the forest fringe are more at risk from malaria (Gamage-Mendis *et al.*, 1991). Human dwellings and animal shelters are also good resting places where mosquitoes can digest the blood they have consumed, while their eggs mature (Najera, 1992), and Gamage-Mendes *et al* (1991) have shown that there is a link between the quality and type of housing construction and malaria incidence and transmission. Poverty is another important factor as summarised by Asenso-Okyere (1994). *"Malaria is linked to poverty in a vicious cycle: people become sick because they are poor, they*

become poorer because they are sick, and they become even more ill as their poverty increases." Poor people tend to live in more crowded poor quality housing, where transmission becomes more likely. Often they cannot afford preventative measures, such as antimalarials, pesticides and bednets. In poor remote rural communities, local health centres will often be out of reach, and so medical treatment unavailable. Faced with a heavy burden of disease, not just from malaria, poor communities struggle to survive. Malaria is less of a problem in urban areas because transmission is lower and individuals have greater access to medical facilities and pharmacies. There are generally fewer potential breeding sites in larger towns and cities, due to the presence of better sanitation and housing. Targeting and controlling the mosquito vector is also more feasible in an urban environment. However, in slum areas and shantytowns that are a feature of many cities in developing countries, the risk of malaria is still a problem. The world's urban population is growing at a phenomenal rate, mainly because of migration from the countryside. Rapid, unregulated urbanisation can bring about an increase in or re-establishment of malaria transmission. (Martens & Hall, 2000).

Human population movement has been shown to have a great impact on malaria transmission patterns (Martens & Hall, 2000). People moving from areas where malaria is endemic to areas where the disease has been eliminated have reintroduced the disease. The malaria parasite will be carried in their blood and so transmittable to areas where environmental conditions are favourable and the malaria vector is present. Conversely, if these migrants have a low level of immunity or are nonimmune and migrate to disease endemic areas they will have an increased risk for malaria. These people may be agricultural labourers and colonisers of new territories seeking employment, or refugees escaping natural disasters or conflict. Natural disasters that include droughts, floods, and earthquakes have created approximately 25 million environmental refugees (Martens & Hall, 2000). In large refugee camps, devastating epidemics can arise. Infectious mosquitoes can also be unintentionally transported to malaria-free areas, reintroducing disease. Air transport has enabled people and mosquitoes to travel vast distances taking the threat of malaria with them. There have been cases of so called 'airport malaria' in and around Roissy-Charles-de-Gaulle Airport in France in 1994 (Martens & Hall, 2000) and Gatwick Airport in the UK in 1983 (Marchant *et al.*, 1998), where hot humid weather may have assisted the survival of an imported mosquito.

Population movement is also increasingly held responsible in the spread of drug resistance in malaria.

2.4 Climate change and the implications for malaria transmission

It is generally recognised that human activities since the industrial revolution have begun to have a detectable influence on the world's climate, causing it to warm. Climate plays a major role in the geographical distribution and seasonal abundance of the mosquito vector. Changes in rainfall, temperature and humidity will directly affect their reproduction, development and longevity (Martens *et al.*, 1999). Within the next century, global temperatures are estimated to rise by 1.0-3.5°C (Githeko *et al.*, 2000). The greatest effects on transmission are likely to be observed at the extremes of the range of temperatures at which transmission occurs. For malaria, the lower temperature threshold is around 16°C, with the upper end above 35°C. Global warming at the lower threshold will have a significant and non-linear impact on the vectorial capacity, which is the potential number of inoculations originating from one infective person each day (Lindsay & Birley, 1996). However, warming at the upper temperature limit will mean transmission may cease. Global warming may extend the season of malaria transmission in some regions as well as increasing the altitude and latitude range of the mosquito vector. The most vulnerable areas will therefore be those currently at the fringe of transmission. Tropical highlands, especially those in sub-Saharan Africa, with their lower temperatures are currently a refuge from malaria for large populations. The elevation at which transmission ceases varies by region, but a small temperature rise will allow transmission to occur. Highland communities have very low or no immunity to malaria and so they may become prone to devastating epidemics. Malaria incidence associated with local climate change has been observed in Rwanda, when in 1987, expansion of malaria into higher altitudes occurred after record high temperatures (Loevinsohn, 1994). Malaria could also advance into temperate regions; its latitudinal range generally extends to the 16°C winter isotherm (Patz *et al.*, 1996). This isotherm would shift further north and south as a consequence of global warming. Areas such as the former republic of the Soviet Union where malaria has been eradicated in the past and where health services are poor will be at risk from severe epidemics (Patz & Lindsay, 1999). Many wealthier countries have eradicated malaria, using insecticides, antimalarials and other control measures, however, the vector responsible for its transmission is still present. Malaria may be imported through air travel, leading to the

potential risk that a local epidemic may occur (Martens, 1999). Regions at risk from small, local epidemics due to a warmer more humid climate will be those in temperate zones. In Europe the potential risk is greatest in countries that surround the Mediterranean (Martens, 1999; Rogers & Randolph, 2000).

Global warming may lead to the melting of polar icecaps and alpine glaciers and consequently sea levels are predicted to rise by around 50 m by 2100 (Houghton *et al.*, 1996). Many low lying coastal areas may be flooded and thus lead to an increase in vectors that breed in brackish waters, such as *An. sundaicus* in South-East Asia. However, salt water contamination of inland water may lead to a reduction in transmission since the salt water tolerant species *An. melas* and *An. merus* are less efficient vectors than the *An. gambiae* complex (Lindsay & Birley, 1996).

The effect on rainfall patterns due to global warming is less predictable. Some parts of the world may have more rainfall than present and others less. In areas that receive less rainfall where malaria is currently present mosquito breeding sites may be reduced and eventually disappear. This would lead to a reduction in transmission. However, if drier conditions lead to drought, people may be forced to migrate to wetter areas where there are more vectors and are thus exposed to the threat of malaria. An increase in rainfall will often lead to increased breeding sites and thus transmission intensity. The El Niño-Southern Oscillation (ENSO) is a natural phenomenon that occurs every 2-7 years (Kovats, 2000) and tends to exaggerate the extremes of climate in certain regions, causing floods and drought. Many parts of South America are affected by El Niño and in 1983 epidemics of malaria occurred in Bolivia, Ecuador and Peru due to heavy rainfall linked to a strong El Niño event (Kovats *et al.*, 1999). However, in Tanzania there was a reduction in malaria after one of the largest El Niño events on record occurred in 1997-8. Even though there was a 2.4 increase in rainfall, the reduction in malaria was perhaps due to mosquitoes being washed out of their breeding sites (Lindsay *et al.*, 2000). Malaria can also be associated with droughts linked to ENSO phenomenon, where rivers dry up, leaving pools in river beds which become ideal breeding sites for the mosquito vector, as witnessed in Sri Lanka (Patz *et al.*, 1996).

Climate change will inevitably affect humans in many ways. Agricultural practices may alter if areas become drier due to increased temperatures and less rainfall leading to the development of irrigation schemes that become breeding sites for the mosquito vector.

If areas suffer from drought, people may have to migrate to wetter regions where there is more malaria. Poor rural communities may be lured to urban areas where the chance of work may be greater, but housing and sanitation very poor and conducive to malaria transmission. Floods will displace thousands of people to crowded refugee camps where disease will proliferate. Although we may attempt to predict how globally climate patterns will change with increased warming, trying to estimate the effect on a regional and local level is a much more difficult task and consequently how humans will respond to such changes can only be guessed at.

2.5 Highland malaria in sub-Saharan Africa

Historically, the African highlands have always been a refuge from the heat and humidity of the lowlands. The highlands also offered protection from the many diseases found in the lowlands, especially malaria. In fact, the highlands have always been regarded as areas of little or no malaria transmission, mainly because of low temperatures. However, this appears to be changing and in the past 50 years the highlands have seen a progressive rise in malaria transmission (Lindsay & Martens, 1998). Generally, elevations greater than 1500 m were considered malaria-free zones, but this does vary by region, and seems to be changing. In Ethiopia, Tanzania, and Kenya the upper limit is around 2000 m (Garnham, 1948; Lindsay & Martens, 1998; Bødker, 2000), whereas for Uganda it is about 1900 m (Heisch & Harper, 1949; Lindblade *et al.*, 2000), and for Zimbabwe it is 1200 m (Martens, 1999). However, malaria has occurred in Kenya at over 2500 m (Bhatnagar, 1954).

In most of the highlands, local communities have little or no immunity against the malaria parasite. Individuals of all ages are at great risk if an epidemic occurs, with high morbidity and mortality characterising such epidemics. Communities in the lowlands however, where malaria is endemic have an acquired immunity and therefore mortality is confined mainly to young children. Areas of highest endemicity, where the spleen rate in children (2-9 years old) is constantly over 75%, but low in adults are classified as holoendemic (Gilles & Warrell, 1993). These areas are often characterised by extensive forest or savannah up to 1000 m high with over 2000 mm of rainfall annually (Najera, 1992). Hyperendemic areas are often between 1000 and 1500 m high, or lowland areas that receive between 1000 and 2000 mm of rainfall (Najera, 1992), where spleen rates in children are constantly over 50% and also high in adults. As

altitude increases above 1500 m or rainfall decreases below 1000 mm a year, malaria becomes less endemic and more seasonal. Mesoendemic malaria is thus, usually found in areas where transmission of the disease is for a short period, following on from the rainy season. In these areas spleen rate in children varies between 11 and 50% (Gilles & Warrell, 1993), however, development of immunity is delayed so that not only children, but adults can display clinical symptoms. These are areas on the extreme limits of malaria distribution in Africa, and seasonal epidemics may occur following exceptional meteorological events, such as heavy rainfall after drought. Low malaria endemicity, or hypoendemic malaria is usually found in areas where malaria transmission is absent but in which it may occur following remarkably favourable meteorological variations (Gilles & Warrell, 1993). The immunity of the population in such areas is virtually non-existent, and when transmission begins it may quickly reach epidemic status, with high mortality among all the different age groups of the population. In the hypoendemic highlands of western Ethiopia, a devastating epidemic in 1958 caused the death of some 150,000 people (Fontaine *et al.*, 1961). This area is normally free from malaria and so the population were nonimmune. This devastating epidemic was associated with abnormally high rainfall coupled with elevated temperatures and relative humidity. The high number of fatalities was largely due to limited or lack of access to health services and antimalarial drugs, with poor nutrition worsening the situation (Alles *et al.*, 1998).

As previously discussed, malaria is very sensitive to temperature; both the mosquito and parasite survival and development rates are greatly reliant on ambient temperature. Altitude is a very good proxy for temperature; the average temperature decrease with height is approximately 0.6°C per 100 m (Adams *et al.*, 1996). A minimum constant temperature of 18°C is necessary for the development of *P. falciparum* (Gilles & Warrell, 1993). In most highland areas such a temperature is never reached. However, the mosquito vector spends much of its life inside human dwellings that preferably have little ventilation, a constant fire and are overcrowded with people and domestic animals. At high altitudes an occupied hut can be at least 3°C warmer than outside or in an unoccupied hut. Moreover, the relative humidity is much more stable and suitable in an occupied hut, creating a more suitable environment for the survival of the mosquito (Garnham, 1948).

Several theories to explain the increase in highland malaria transmission have been presented, the most recent of which is climate change. Global warming could shift malaria up the slopes, increasing highland populations at risk. Recent mathematical models indicate that the threshold temperature necessary to trigger an epidemic lies above 20°C (Lindsay & Martens, 1998). The altitude at which these temperatures occur will vary in relation to latitude, but generally, areas between 1000 m and 1500 m are thought to be susceptible (Patz & Lindsay, 1999). Evidence for climate change initiating epidemics can be found in Rwanda (Loevinsohn, 1994). Here temperature has increased progressively since 1960-1990, with 1987 being the peak for both temperature and rainfall, and malaria incidence increased by 337%. The increase was greatest in communities with little acquired immunity, especially those living in high altitude areas. The average minimum temperature appeared to be most important at higher altitudes, with increased rainfall effecting incidence in the lower altitude zone which borders broad valleys. Malaria also reappeared in the central highland plateaux of Madagascar. In 1949, indoor spraying with DDT and mass drug administration resulted in the disappearance of the principal vector, *An. funestus*, and the elimination of malaria. In 1987, an unusual increase in mortality revealed the epidemic re-emergence of *P. falciparum* malaria in the plateaux area. Since then malaria seems to be highly endemic and transmitted by both *An. funestus* and *An. gambiae* (Lepers *et al.*, 1990).

Although, it is apparent that climate change will affect the risk of malaria in highland Africa, development of the highlands, due to rapid population growth is thought to be the most important factor contributing to increased malaria transmission. Garnham (1948), noted the increased incidence of malaria in Kenya after 1918. He blamed this increase on the gradual development of the country. Much of this development created more breeding sites for the mosquito vector, such as the introduction of dams, the construction of rough roads for the ox wagon, where their wheel ruts developed into pools, deforestation, increasing the facilities for mosquito breeding in sunlit pools, and the opening of big estates resulting in changes to land use, and hence the creation of more breeding places. He also blamed the transportation of mosquitoes from the endemic lowlands via newly built railways and roads, and the importation of infected labour that spread malaria to the estates, road gangs and railway construction works, etc, and from there to nearby local populations. The colonisation of the highlands in Tanzania and the subsequent intensification of agriculture with terracing and levelling of land in and around human settlements have increased the number of potential

anopheline breeding sites, and consequently increased malaria transmission (Matola *et al.*, 1987). In the south-western highlands of Uganda, higher malaria prevalence was attributed to villages located near cultivated swamps compared to villages situated near natural swamps (Lindblade *et al.*, 2000). In the Ethiopian highlands, the construction of microdams lead to a substantial increase in the incidence of malaria among children living nearby (Ghebreyesus *et al.*, 1999).

Increased malaria transmission has also been linked to human movement. As farming increased in the highlands, many immigrants attracted to the area would have brought their parasites with them. As the highland populations grew, increased pressure on land meant people living in the highlands began farming in the lowlands where transmission is high. On return to their highland villages they may be carrying an infection that could potentially spread to their local community if conditions are favourable (Lindsay & Martens, 1998). Poor access and the decline and lack of basic health care and the growing resistance to antimalarials, especially chloroquine pose a great threat to the highlands where populations have little or no immunity to the disease (Bødker *et al.*, 2000); if epidemics arise, there will be few resources to combat transmission. The increase of malaria in many parts of the highlands can also be blamed on inadequate control measures. The cessation of vector control measures can also be held responsible, as previously described in Madagascar (Lepers *et al.*, 1991).

The highlands are not a uniform environment, and topography may play a key role regarding which areas are more susceptible to malaria outbreaks than others. Over a century ago, Hirsch (1883) recognised the protective role the highlands played on people's health. He also observed that flat valley bottoms were more prone to outbreaks of malaria *"...that wherever malaria fever is endemic at more or less considerable elevations, the seat of the disease is always a valley with a small declivity, or basin-like depression in a plateau."* Garnham (1948), commented on the inability of mosquitoes to penetrate high valleys exposed to wind, whereas on the opposite side of the range where calmer conditions exist they rise to much higher altitudes, he also states that steeper slopes are less likely to provide suitable breeding sites for the anopheline vector. The effect of the prevailing wind has been studied by Gilles (1961), who concludes that in a highly populated hilly region, wind direction was a minor factor in contributing to the dispersion of mosquitoes. Manga *et al* (1993) also noted that in Cameroon the hilly slope limits the dispersion of *An. gambiae*, and that the man-biting rate and inoculation

rate decreased rapidly from the swampy valley bottoms to the hilltops. Assessing the role of topography in relation to malaria transmission is one area where there has been little research.

It is difficult to conclude that any one factor is more responsible for the increase of malaria in the African highlands than any other. Environmental as well as socio-economic factors can influence malaria transmission. The highlands are areas of unstable transmission and subtle changes in temperature can have devastating consequences if land-use modifications have altered the environment, thereby increasing the number of potential mosquito breeding sites. Climatic change can therefore have a substantial impact in areas near the limits of malaria's distribution. Highland communities have little or no acquired immunity to malaria and so are extremely vulnerable to severe outbreaks. The situation is exacerbated by the lack of health services and the decline in control initiatives. Since the highlands are home to millions of susceptible individuals, it is important that more is understood about this increasing phenomenon. An international workshop held in Addis Ababa in 1996, organised by the UNDP (United Nations Development Programme), World Bank, WHO Special Centre for Research and Training in Tropical Diseases (TDR), and the International Research Centre of Canada (IDRC) established the project, 'Epidemiology of Highland Malaria in Africa' (HIMAL). It was agreed at this meeting that malaria in the African highlands should be acknowledged as a distinct entity, and a regional programme should be launched to: define epidemic prone areas, identify the reasons for increased malaria, and help develop solutions to protect these vulnerable communities from this growing problem (Lindsay & Martens, 1998).

2.6 Mapping and modelling malaria

Many attempts have been made to model and map malaria distribution throughout the world. This is an essential task since the disease affects a great number of the world's population. However, so many factors affect the transmission of malaria that it is impossible to account for, and include them all in a comprehensive model to predict its distribution. On a global scale it is important to know where malaria is present and how prevalent it is, so that reliable estimates can be made of the number of people at risk, allowing governments and international health organisations to plan and budget for its control. More importantly, with the advent of global warming, modelling and

predicting the distribution of malaria and new populations at risk in future years is crucial so that preparations can be made to guard against this threat in areas that are currently free from the disease. Predicting future populations at risk due to climate change relies on the use of General Circulation Models (GCM) that attempt to forecast future climates due to global warming. The MIASMA (Modelling framework for the health Impact Assessment of Man-induced Atmospheric changes) model uses the future climate scenarios produced by the HadCM2 and HadCM3, which are global climate models developed by the UK Hadley Centre. The MISAMA model incorporates knowledge about the current distributions and transmission potential (based on the basic reproduction rate and vectorial capacity) of the main mosquito vectors, and data on monthly mean temperature and precipitation predicted by the climate change scenarios. The greatest changes in potential transmission are predicted to arise in temperate zones, in areas where vectors are present but where temperatures are currently too cold for transmission. These include such areas as northern Europe, North America and Asia. Worldwide, between 360 and 520 million additional people are thought to be at risk of *P. falciparum* and *P. vivax* malaria in 2080 due to climate change (Martens *et al.*, 1999). However, predictions of reduced rainfall imply that some areas that presently experience year round malaria may experience only seasonal transmission in the future. In a different model, GCMs predicting future temperatures and precipitation patterns were employed, together with a biological function known as the 'epidemic potential'. The model therefore incorporated factors describing climate, vectors, parasites, hosts and health impact, and revealed that a further 50-80 million cases may occur each year as a direct result of a 3°C rise in temperature by 2100. This predicted increase is most noticeable at the boundaries of endemic malaria areas and at higher altitudes within malarial areas (Martens *et al.*, 1995). However, Rogers and Randolph (2000) are critical of such predictions that are based on biological transmission models influenced mainly by temperature. They propose an alternative statistical approach, where the documented, current global distribution of *P. falciparum* malaria was used to determine the present multivariate climatic limitations. These results were then included in future climate scenarios to predict future distributions, which displayed, only a relatively small increase and expansion as compared to the situation today: northward into the southern United States, and into Turkey, Turkmenistan and Usbekistan; southward in Brazil; and westward in China.

The need for up-to-date, reliable distribution maps of malaria is essential for Africa, where 90% of mortality occurs (Dobson *et al.*, 1999). Africa experiences a variety of malaria epidemiology, ranging from intense perennial transmission to unstable, epidemic-prone areas. A thorough understanding of the transmission intensity, distribution, seasonality, environmental influences and populations at risk are required to aid the targeting of resources for its control on a continental, national and local level. Such information is generally not available in an accessible and standardised form for many countries (Snow *et al.*, 1996). The Mapping Malaria in Africa/Atlas du Risque de la Malaria en Afrique (MARA/ARMA) collaboration is an initiative that has been set up to address this problem and provide an atlas of malaria risk for Africa. A continental database of the spatial epidemiology of malaria has been produced using a combination of empirical data collection and modelling, that will include information on population, topography, climate and administrative boundaries (MARA/ARMA, 1998). The type of empirical data collected is parasite rates in children under 10 years old, sourced from an extensive number of published and unpublished survey data. Data on the distribution of vectors and epidemic highland malaria were also collected. To supplement the empirical data collection in areas where there was no data, predictive modelling, based on climate and other environmental factors was employed (Craig *et al.*, 1999).

The modelling and mapping of malaria was approached on four levels: continental, sub-continental, regional and national, and at a scale of 30 km^2 and below. On a continental level, the risk of malaria transmission was calculated for Africa on a scale of 0 to 1, based on climate conditions in an average year. From the model, it was concluded that in sub-Saharan Africa malaria is responsible for about 1 million deaths and just over 200 million episodes of clinical disease (Snow *et al.*, 1999a). The seasonality of malaria transmission was also modelled on a continental level and divided into two zones: seasonal and stable. At the sub-continental level, the distribution of malaria at the margins was refined to take into account inter-annual changes in climate and differences between major ecological zones. The model produced indicates where epidemics are likely to occur. Regional models were produced for Kenya (predicted levels of transmission intensity) (Omumbo *et al.*, 1998) and Mali (predicted prevalence rates) that were useful, and reflected the situation in these countries well. The regional models involved related parasite rates to climate and other environmental data and defined the transmission intensity within a region of transmission ecology (perennial, seasonal or bi-seasonal transmission). At a scale of 30 km^2 and below investigates

variation in transmission on a localised level (MARA/ARMA, 1998). Within the HIMAL project, MARA/ARMA have also mapped potential highland areas at risk by using climate and altitude data plus historical information on the distribution of past epidemics to produce epidemic risk maps for individual countries. The distribution of malaria vector mosquitoes of the *An. gambiae* complex has also been mapped (Coatzee *et al.*, 2000) and compared to a previous model that predicts the species ranges and relative abundance of *An. gambiae* and *An. arabiensis* (Lindsay *et al.*, 1998).

At a local level there have been many studies investigating the relationship between the risk of malaria and the environment. Often the research has concentrated on identifying and mapping potential breeding sites for the mosquito vector using satellite remote sensing. In The Gambia, Thomas & Lindsay (2000) showed that proximity to the river was a good predictor for the number of mosquitoes found in a village. However, the most prevalent areas were those with fewer mosquitoes further away from the river. This paradox was also found in a previous study in The Gambia (Thomson *et al.*, 1996) and was attributed to the tendency of people to use bednets in villages where mosquito numbers were high. Other examples include: Hay *et al.* (1998), where malaria seasons were predicted in Kenya using multitemporal meteorological satellite sensor data. Beck *et al.* (1994) produced a computer-generated map of landscape elements from two Landsat Thematic Mapper scenes for southern Chiapas, Mexico. The relationship between vector abundance and landscape element were investigated and they were able to predict villages at risk for malaria transmission with an overall accuracy of 90%. Environmental proxies such as the normalised difference vegetation index (NDVI) were shown to be associated with a suitable environment for the survival of infective mosquitoes in The Gambia, with high NDVI values being indicative of a moist environment supporting both mosquito breeding and adult survival (Thomson *et al.*, 1999). The use of satellite remote sensing is becoming increasingly widespread in mapping disease epidemiology. One of the biggest advantages of meteorological and environmental satellites is that they are able to observe large areas of land on a regular basis. This is particularly helpful for developing countries, as it provides them with an affordable and effective way to collect large amounts of detailed ecological and climatic information to assist them in tackling not only malaria, but many other issues.

On a much smaller scale, Smith *et al.* (1995) mapped the densities of malaria vectors within a single village in Tanzania using Bayesian techniques. They showed that areas

of high mosquito density were predominantly those nearest the low lying areas in which rice was grown, including the fringes of the streams running through the main part of the village. Another study at the village scale looked at the variation in exposure of Gambian children to *Anopheles gambiae* in an irrigated rice production area (Lindsay *et al.*, 1995). They found that the spatial distribution of mosquitoes varied between rainy and dry seasons and was associated with the prevailing wind direction at night, indicating that wind aided the dispersal of mosquitoes from their breeding sites. They also concluded that increased exposure to malaria vectors for individual children depended on a variety of factors (adjacency to mosquito breeding sites, being resident in larger compounds, having open eaves in a house, the absence of a ceiling in the bedroom, the absence of wood smoke indoors, leaving the bednet untucked at night, using insecticide) that varied between the dry and wet season. Understanding spatial and temporal variation in vector densities within villages would have immediate implications for inhabitants who wish to avoid anthropophilic mosquitoes.

Maps are a very useful way of displaying complicated information simply, in a way that is understandable and instructive. Maps and models allow a means of understanding malaria epidemiology and provide essential help in targeting control measures where they are most needed. They may also be used to identify epidemic-prone areas and integrated into adequate early warning systems. However, a model cannot include every feature of the problem. Malaria is a very complex disease that involves many factors: environmental, biological and socio-economic. Successful models should balance over-simplification and over-complexity and not leave out the crucial features. Understanding the effects of major issues such as chloroquine resistance, urbanisation and variations in climate will be difficult in the absence of high-quality maps of malaria distribution on a global, continental, regional and local scale. McKenzie (2000) assesses why we should model malaria and sums up the usefulness of models in the following quote: *"A model is useful if it sharpens questions, points to what is missing in our data or in our conceptual grasp and contributes to a larger process of discovery."* In other words, models are useful products that lead us on to a greater understanding of the malaria situation.

3 Geographical Information Systems and their use in epidemiological research

3.1 What are Geographical Information Systems?

GIS can be simply defined as *"...a computer system that is able to store and utilise data describing places on the earth's surface"* (ESRI, 1991). GIS has emerged due to the increasing availability and advances in computer technology, along with the copious amount of geographical data that continues to be collected, and the need to be able to analyse and use this information, enabling an understanding of the world in which we live (Brown *et al.*, 1991).

A GIS is able to capture, store, handle and analyse spatial data. It is not just a computer system for producing maps, but also allows spatial operations on the data, and is able to identify the spatial relationships between features (Hicken *et al.*, 1991). This leads to a wide range of information processing and display operations, providing a powerful tool for engineers, geographers, economists, planners, cartographers, military organisations, local governments, health authorities and epidemiologists, to name a few.

GIS is particularly appropriate for use in epidemiological research. For instance, location queries can be answered to discover the type of features that exist at a specific location, such as the type of vegetation that may help to identify mosquito-breeding sites (Hugh-Jones, 1989; Rejmankova *et al.*, 1995; Wolock & McCabe, 2000). Conditional questions, for instance pinpointing the locations of health care services and the area they cover can be easily undertaken (Bamford *et al.*, 1999; Tanser & Wilkinson, 1999). Trend analysis, may be carried out to identify if disease patterns have changed over a number of years in an area or vary on a seasonal basis (Snow *et al.*, 1993). GIS can help determine where it is most efficient and effective to target disease control resources (Gabinaud, 1987; Sharma & Aruna, 1997). Patterns may be determined such as the incidence of childhood leukaemia near nuclear power stations (Knox, 1964). It is also a powerful modelling tool enabling the user to determine the possible outcome of certain events, from identifying highland villages in Africa at risk from malaria, to predicting the global consequences to human health from climate change (Martens *et al.*, 1995; Martens, 1999).

Most geographic information may be represented in two dimensions, conveying spatial information well, such as the location of drainage networks, political boundaries, land use and population densities within an area. However, it is also possible to convert the conventional 2-D image into a 3-D representation. This is referred to as a 'digital elevation model' (DEM) or 'digital terrain model' (DTM) and they are typically used to represent terrain relief (Price & Heywood, 1994). The ability to produce a 3-D image of an area allows complex analysis, such as the computation of the amount of water flowing across a landscape. Perhaps a more important function of a DEM, however, is that modelling displayed graphically is a powerful medium that may directly support decision making without making quantitative analysis. Visualisation thus plays a vital role in a DEM system (Weibel, 1991).

In a GIS 2-D spatial data may be represented in vector format, representing points, lines or area information, or a raster format where the spatial objects are represented as cells within a grid. The cells are allocated different values depending on the features being mapped. Vector models therefore convey detail of specific features such as displayed on a conventional cartographic map, whereas raster models are a much coarser representation, providing a powerful modelling environment for analysing geographic features, i.e. for storing and representing continuous spatial surfaces and points, lines and area information in the same format (ESRI, 1991). Both formats represent spatial objects or regions by measurement of x and y co-ordinates. Three-dimensional data, however, involves the spatial object or domain extending through 3-D space defined by x, y and z co-ordinates.

3.2 Mountains environments and terrain modelling

Mountains are characterised by high altitude, steep slopes, and sharp climatic and ecological gradients. In general, daytime temperatures decrease with increasing altitude by approximately 6°C per 1000 m (Adams *et al.*, 1996), although this does vary with differences in spatial and temporal atmospheric humidity. Rainfall is also influenced by altitude. Tropical mountains generally exhibit highest precipitation levels at 1500 m, above which levels can rapidly decrease (Adams *et al.*, 1996). The altitude-driven differences in climatic conditions create different environments at different elevations. Together with this altitudinal zonation, are variations that result from the aspect, slope

and topography of a particular mountain or region, influencing local and micro-climates in valleys and on upper slopes. Aspect influences insolation levels received, rates of evaporation, soil chemistry and stability, and the availability of moisture. Shaded slopes will experience lower than ambient temperatures with the effect being most evident in the valleys within which cool air has accumulated at night. Consequently, the typical decrease of temperature with increasing elevation does not apply, with the upper slopes being far warmer than valley bottoms and lower slopes. However, upper slopes and summits are more exposed and tend to experience harsher conditions, characterised by strong winds, low temperatures and moisture deficits. Aspect also affects rainfall levels. As well as the physical and biological zonation, variations in local and micro-climates influence patterns of settlement and agriculture with valley floors being particularly desirable for these purposes. In Africa mountainous regions tend to be highly populated, with communities relying on subsistence agriculture (Berry, 1971).

One necessity in the development of a GIS for a mountain area is the ability to model its three-dimensional complexity. The most powerful method of representing relief is to construct a mathematical model of the earth's surface, commonly termed a DEM or DTM. These models can be used to calculate information on height, aspect, slope angle, watersheds, insolation, and hill shadows. In regions where no elevation data exists, interpolation is carried out through the modelling process to estimate missing elevations.

DEMs may be derived from three principle sources: ground surveys, photogrammetric data capture, or from digitised cartographic data sources (Weibel, 1991). Global positioning systems (GPS) are increasingly becoming a more common tool used to supply supplementary height information (Price & Heywood, 1994). For this research the third source type was used and the contours of the 1:50,000 maps of the region were digitised at 40 m intervals. Digitising contours is one of the major sources for constructing a DEM. Contour maps are generally widely available and for many projects may be the only affordable data source of elevation data from which to create DEMs. However, they have the disadvantage that they over sample elevation at certain heights, with no estimates of elevation between these heights. This means that in areas of low relief, there may be very little information about the terrain height on which to base an interpolation (Wise, 2000).

The majority of DEMs conform to one or two data structures: rectangular grid (or lattice) or TIN (Triangulated Irregular Network). TIN structures are based on a network of triangular surfaces of irregular size, shape and orientation. They work on randomly located height (z) values; the vertices of the triangle are formed from each observed height value (Price & Heywood, 1994). TINs are especially useful for surface representation and surface modelling and display (e.g. contouring, visibility, 3-D display profiles). A lattice represents a matrix structure that records topological relations between data points, therefore this grid model can be stored as a 2-D array of elevations (Weibel, 1991). Grid-based terrain modelling algorithms tend to be relatively straightforward and make fewer demands upon computer processing resources compared to TIN models. However, as with 2-D grids, the resolution is coarser and so the complexity of relief may not be depicted as well as with a TIN. Unlike the grid, the TIN allows dense information in complex areas, and sparse information in simpler or more homogenous areas, according to the variation in the relief. From the TIN model the slopes, aspect and surface changes for any type of surface can be calculated (DeMers, 1997). Both data structures are useful for elevation analysis, depending on the nature of the project. However, Wise (2000) points out that in the analysis of DEMs there will be two main sources of error in the final results: error in the DEM itself, and errors caused by the algorithms used in the analysis. Even if the elevations in a DEM are totally error free, the fact that the DEM is a sample from a continuous surface means that measures such as slope and aspect will only ever be estimates of the true values, and as such will contain an element of error. The accuracy of a DEM also depends on the spatial density and accuracy of the sampled data points. DEMs derived from contour maps typically have a standard error that is usually half the contour interval (Price & Heywood, 1994). Despite a small loss of precision, DEMs are still an extremely valuable method of modelling terrain.

Topography is an important land-surface characteristic that affects most aspects of the water balance in a watershed, including the flow direction followed by water as it moves down and through hill slopes and the rate of water movement (Wolock & McCabe, 2000). DEMs are now a standard tool in hydrological analysis and many models have been developed to calculate hydrological phenomenon such as overland flow, watershed delineation, soil water accumulation, surface saturation, and stream channel networks (Anderson & Kneale, 1982; Burt & Butcher, 1985; Moore *et al.*,

1988; Haines-Young *et al.*, 1993). Burt & Butcher (1985) identified hill slope hollows and low gradient slopes as areas within a catchment where water table levels were likely to be. They noted that where the hill slope profile and plan were concave the soil tended to be very wet. Moore *et al* (1988) produced a topographic model for determining slope, aspect and upslope contributing area across three-dimensional terrain. They present and discuss three example applications: the prediction of zones of surface saturation, the prediction of the distribution of potential daily solar radiation and the prediction of zones of erosion and deposition in a catchment. Variations in wetness within a catchment are explained in terms of the local topography (slope and degree of convergence of the hill slope) and the hydraulic properties of the soil profile. The effect of topography on the amount of radiation received at the land surface is particularly important in understanding the ecology and the hydrology of catchments, and they show that biomass production is related to aspect, radiation and wetness. Haines-Young *et al* (1993) used digital elevation data alongside hydrological simulation modelling within a GIS to assess the time-varying distribution of temporary water on a farm scale to estimate the possible patterns of redistribution that might arise from drainage or consolidation. This is a major management concern in western Canada, due to the desire of farmers to drain or combine small bodies of non-permanent surface water.

In this research DEMs of each village area were produced so that the topographical characteristics of altitude and slope, and hydrological events of flow direction and flow accumulation could be derived. These features were further manipulated to define potential mosquito breeding sites.

3.3 Terrain and hydrological modelling using ARC/Info

This project involved the use of both TIN and lattice data structures. The former is used to display the study areas as a 3-D landscape and is incorporated into producing the elevation grid. The elevation grid is then manipulated through ARC/Info to derive slope values across the study areas producing a grid depicting slope values across a surface. Slope is defined by a plane tangent to the DEM surface at any given point, and comprises two components: gradient (maximum rate of change in altitude) and aspect (the compass direction of this maximum) (Chrisman, 1991). The hydrologic modelling tools in ARC/Info were used to calculate flow direction and flow accumulation across the study areas. The remaining part of this section briefly explains some of the

principles involved to produce the hydrological models. This information is described in more detail in ESRI (1991), 'Hydrologic; modelling tools'.

The source for modelling hydrological characteristics of a catchment area in ARC/Info is the digital elevation grid or lattice. The tools within ARC/Info help model the movement of water across a surface, i.e. surface run-off, and so the physical characteristics of a surface determine the characteristics of flow across it. For the purpose of this study the main considerations were to determine the direction of flow across the surface of the study areas, where the flow accumulates, and how much is accumulating.

3.3.1 Determining flow direction

Flow direction is determined by the slope direction and the maximum rate of change of elevation from each cell. Therefore, the steeper the slope the greater likelihood that water will flow in that direction. Flow direction is determined by the 'flowdirection' function. This function takes the digital elevation grid as input, resulting in a grid showing the direction of flow out of each cell. There are eight valid output directions relating to the eight adjacent cells into which flow could travel as shown in Figure 3.1.

Figure 3.1. Grid cells denoting flow direction value

32	**64**	**128**
16		**1**
8	**4**	**2**

The direction of flow is determined by finding the direction of steepest descent from each cell as follows:

maximum drop = change in z value / distance

When the direction of steepest decent is found, the output cell is coded with the value representing that direction.

At this stage it is important to identify any sinks or depressions in the DEM. A sink is an area surrounded by higher elevation values and is usually an imperfection in the DEM. Sinks, being areas of internal drainage may cause undesirable results when calculating flow direction and therefore should be removed. The sinks were filled using the 'fill' function in ARC/Info. Fixing sinks can sometimes cause sinks to occur in neighbouring cells, so the procedure often needs to be repeated until the DEM is free of all depressions (Pullar & Springer, 2000).

3.3.2 Determining flow accumulation

The 'flowaccumulation' function takes the flow direction grid as input resulting in a grid depicting flow accumulation. Accumulated flow is calculated by the accumulated weight of all cells flowing into each down slope cell in the output grid. The weighting of one is applied to each cell and the value of the cells in the output grid will be the number of cells that flow into each cell. In this way cells with a high flow accumulation are areas of concentrated flow and may be used to identify stream channels.

In Appendix A, Figures A9, A10, A11 and A12 display the surface model of each study area represented by the TIN overlaid with the 'flowdirection' grid. The 'flowaccumulation' grids are combined with the grid identifying all flat areas within each study region ('slope' grid reclassified to 'flat' grid) to produce the 'swamp' grids, which locate areas where water is likely to accumulate and hence could be potential mosquito breeding sites. These 'swamp' grids are shown in Figures A25, A26, A27 and A28 in Appendix A.

3.4 Using GIS and remote sensing to map disease

3.4.1 What is remote sensing?

Although remote sensing (RS) has not been used in this research, it is important to recognise the valuable contribution that it makes towards greater knowledge of disease patterns, especially our understanding of malaria transmission. This section therefore gives a brief outline of the basic principles of RS and considers their role, together with GIS, in the mapping, monitoring and control of disease.

RS is a term used to describe any method where information about the earth, or objects on the earth, is obtained without any physical contact between the subject of study and the sensor. Normally, the term is applied to the collection of data by instruments placed on satellites or aircraft. Every object, area, or phenomenon reflects and emits energy at specific and distinctive wavelengths of the electromagnetic spectrum. Hence, there are many types of RS device that record different forms of reflected electromagnetic energy from different parts of the electromagnetic spectrum. The commonest are:

- Cameras: sensors that record reflected energy in the part of the electromagnetic spectrum that is detected by the human eye, i.e. visible light.
- Infrared sensors: these record infrared radiation and create images showing temperature variation in an area (e.g. sea surface temperatures)
- Microwave sensors (e.g. radar): these transmit very long wavelength electromagnetic energy towards objects and record the resulting reflections. Such energy is capable of penetrating cloud cover and microwave sensors have important implications in geological mapping, the mapping of variations in soil moisture content and the extent and movement of sea ice.
- Multispectral scanners (e.g. the Thematic Mapper deployed from the Landsat satellite): such scanners produce images spanning both the visible and infrared parts of the electromagnetic spectrum (Jones *et al.*, 2000).

All satellite RS data is digital. Satellite scanners numerically record radiation from the earth's surface and transmit this to a receiving station as numbers. These numbers are then used by image processing computers to generate the image. The basic unit created by the numbers is a picture element or pixel. The area on the ground covered by one pixel gives the spatial resolution. The single numerical value of each pixel is the combined reading for the whole area that the pixel covers on the ground (Hilton, 1991). The spatial resolution varies according to the type of system used. The Landsat Thematic Mapper, for example permits the observation of objects with an area of 30 x 30 m, whereas a sensor such as the NOAA-AVHRR (National Oceanographic and Atmospheric Administration – Advanced Very High Resolution Radiometer) has a lower spatial resolution of 1.1 km (Connor *et al.*, 1998). The area on the ground a satellite images covers will depend on the field of view of the sensor (swath width) and the satellites altitude. The AVHRR sensor on the NOAA satellites has a swath width of 1500 km, which enables it to complete global coverage every 12 hours. In contrast,

earth observation satellites such as Landsat have a swath width of 185 km, and thus sense smaller strips of the earth as they travel from pole to pole. Consequently, they take 16 days to revisit any particular area (Hilton, 1991).

Since satellites measure only radiation, this measurement needs to be interpreted as to what has produced the radiation. Raw digital data from a satellite therefore needs processing and enhancing to a form where it can be understood. The basic reflectance and emissivity characteristics of familiar earth surface features are generally well known and can be measured on the ground. Using these known characteristics, methods can be designed to detect and identify various features and extract information about them. Common products derived from satellite data to indicate ecological conditions on the ground include vegetation indices, land and sea surface temperatures and cloud temperature (Connor *et al.*, 1998). The Normalised Difference Vegetation Index (NDVI) is the most widely used vegetation index and measures chlorophyll activity in plants. It not only provides information about vegetation status, but can also be used to monitor soil dryness, land use changes and surface moisture. Land surface temperature provides information about local climate conditions such as surface temperature, moisture content and vegetation status. Rainfall estimates can be made using cold cloud duration images, which are a good indicator of the relative distribution of rainfall (Connor *et al.*, 1998).

There are many advantages in using satellite systems to map the environmental characteristics of the earth. They cover large and remote areas instantaneously, gathering data from a wide spectral range, at consistent and frequent intervals; such information gathering would be unrealistic using regular ground survey methods. Satellite RS can thus be applied to the mapping of isolated and inaccessible areas, mineral exploration, and the monitoring of phenomenon such as pollution, forest fires, weather, sea temperatures polar ice-caps, deforestation and other land use changes. It is a relatively cheap way of mapping and monitoring large areas, and the data produced is digital and so accessible in a form that can be used via a computer. RS data can be linked as an information layer in a GIS. Combing data in this way is possible because both satellite images and maps can be connected to the same map reference system. Linking satellite imagery with other digital data sets allows further analysis and understanding of the geographical area and features of interest. Combined with local

information within GIS they enhance local knowledge and help to improve mapping, monitoring and resource management.

3.4.2 Application of remote sensing data in disease surveillance and control

There has been much interest in recent years in applying RS and GIS techniques to the study of disease transmission and vector ecology (Hugh-Jones, 1989), especially with regard to malaria. These tools greatly improve our capacity to investigate landscape level relationships of vectors and diseases. However, their successful use can only be achieved when a thorough understanding of the ecological and epidemiological processes of disease transmission are known.

The information that can be gathered via RS such as vegetation growth, land and cloud temperatures are extremely valuable indicators of ecological activity that influence disease patterns. Processes that operate at a variety of landscape scales influence the mechanisms of vector-borne disease at any point. Malaria for example can be regarded as a local as well as a global problem. RS imagery using high-resolution aerial photography or coarse-resolution satellite imagery, when linked with GIS spatial analyses tools can play a key role in current vector surveillance and control programmes at local and regional scales (Washino & Wood, 1994). It is important to consider the appropriate spatial, spectral and temporal resolution required when selecting RS data. When less precision is required and the areas of interest are large, AVHRR data from the NOAA satellites can be used. The AVHRR data can map vegetation with accuracy similar to the multispectral scanner Landsat data. Landsat imagery is ideal for accurately mapping land coverage because of the 30 m resolution, the bands used, and the different reflectance signatures of land use features. For the analysis of smaller sites, the SPOT (Systeme Pour l'Observation de la Terre) satellite with its 10 m and 20 m resolution could be used. However, if the area of interest is small and a higher resolution is required it can be cheaper and quicker to rent an aeroplane and photograph ground sites using colour infrared film (Hugh-Jones, 1989).

A good deal of the initial work in using RS has been concentrated on the remote identification of vectors habitats. For instance, adult mosquito abundance was predicted in Belize by Rejmankova *et al* (1995) based on distances to remotely sensed larval habitats. They found that the malaria vector *Anopheles albimanus* was positively

associated with floating mats of blue green algae and river margin vegetation. RS was used to identify these habitats and predictions were made based on the hypothesis that houses situated close to these potential productive larval sites would have higher densities of adult mosquitoes. In another study, Kitron *et al* (1996) found that the Landsat Thematic Mapper band 7, which is associated with the moisture content of soil and vegetation, emerged as being consistently highly correlated with the density of tsetse flies in the Lambwe Valley, Kenya. Tsetse flies are vectors of animal and human trypanosomiasis, a serious disease of cattle and people in Africa. Control of the disease mainly depends on the management of tsetse fly populations. Therefore, incorporating RS imagery in a GIS with ground data on fly density and environmental conditions to predict favourable fly habitats in inaccessible sites, will help to determine the number and location of fly suppressing traps in a local control programme.

A disease vector can often be attacked at the larval stage and much research has been carried out to identify potential larval habitats using RS. Aerial colour infrared photography has been used as a survey technique for identifying potential *Psorophora columbiae* egg-laying habitats in Texas rice fields, allowing more effective and efficient control of the mosquito population (Welch *et al.*, 1989). Wood *et al* (1992) combined RS reflectance measurements of early season canopy development and GIS measurements of distances between rice fields and pastures with livestock to distinguish between high and low mosquito producing rice fields in California. They were able to identify high mosquito producing fields with 85% accuracy nearly two months before anopheline larval populations peaked, thus allowing effective larval control. Thomas and Lindsay (2000) used 20 m spatial resolution multispectral SPOT satellite imagery to determine potential mosquito-breeding habitat in The Gambia to assess the local-scale variation in malaria infection in children. A widely used indicator of vector distribution is the NDVI. It has been applied to the detection of Rift Valley Fever virus vector mosquitoes (*Aedes* and *Culex* spp.) in Kenya (Linthicum *et al.*, 1990). In this case the NDVI was used to derive ground moisture and rainfall patterns to quantify the potential for flooding of mosquito larval habitats.

As well as identifying vector habitats, RS has also been applied to identifying communities at risk, especially to malaria. NDVI has been used in this context by Thomson *et al* (1999) who were able to show a significant association between age related malaria infection in Gambian children and the amount of environmental

greenness as measured using NDVI. In another study, Beck *et al* (1994) processed and classified RS data to generate a map of landscape elements in southern Chiapas, Mexico. In a GIS the proportion of mapped landscape elements surrounding 40 villages where *An. albimanus* abundance data had been collected was determined. They found that the most important landscape elements in terms of explaining vector abundance were transitional swamp and unmanaged pasture. Including these two elements in discriminant analysis they were further able to distinguish between villages with high and low vector abundance with an overall accuracy of 90%. Many diseases display seasonal patterns. Seasonality is a widely recognized aspect of clinical malaria in Africa. However, there are very few examples of reliable maps showing the relationship between malaria seasonality and the clinical outcomes of infection. Hay *et al* (1998) studied the seasonal variations in clinical malaria in relation to a variety of proxy meteorological and vegetation variables recorded on board polar-orbiting and geostationary satellites. They found that there was a high and consistent correlation between temporal changes in NDVI and the temporal changes in malaria cases across the Kenyan communities studied. Thus, they were able produce maps predicting the number of months for which malaria transmission was possible in Kenya. Such maps, coupled with other models of disease transmission intensity, demographic and health service information could be used to guide the timing and selection of interventions. Programme managers often have few resources and this information would provide a more cost-effective and objective basis for malaria control.

These few examples illustrate the value of RS in our understanding of disease. All these examples have integrated RS data with GIS processes to analyse disease patterns, whether related to vector abundance or disease prevalence and risk. In a study by Patz *et al* (1998), a comparison was made between the use of NDVI data and hydrological modelling of soil moisture in predicting the human biting rates (HBR) and entomological inoculation rates (EIR) in Kenya. It was found that soil moisture modelling and satellite NDVI nearly equally predicted HBR. However, there were several benefits recognised in using the hydrological model: Daily and weekly modelled soil moisture can be calculated, whereas satellite NDVI cannot capture the weekly variation in biting rates. The most crucial time to measure surface water for mosquito breeding sites is during the rainy season and acquiring a cloud-free satellite image during this time is difficult, whereas a soil moisture model does not have this drawback. Hydrological modelling is relatively inexpensive and may be more practical for use by

resident public health scientists who have access to local stream flow and weather data in areas with disease risk.

3.4.3 The role of geographical information systems

Understanding the complex spatial and temporal interactions between the environment and disease, and identifying exposure to environmental and disease risk in vulnerable populations are essential elements of an effective environmental and public health management programme. GIS provide cost-effective tools for evaluating interventions and policies potentially affecting health outcomes. In environmental health research GIS provide a way of combining mortality, morbidity, incidence and prevalence data with environmental, socio-economic, and demographic data to help present a clear picture of the role of environmental variables and lifestyle in determining health status (Tim, 1995). A GIS allows an extensive range of appropriate non-spatial data to be incorporated into a reliable framework to improve analysis of the geographical distribution of disease. GIS therefore have a vital role to play in mapping disease. This section considers some of these applications at a variety of spatial scales

local and regional scale

All of the examples so far have been related to vector-borne diseases. However, air quality has also been linked with ill health and research has been carried out using GIS to see if there is a causal link (Dunn *et al.*, 1995; Dunn & Kingham, 1996). In a town in northern England, factory emissions were thought to be the cause of respiratory health problems in the local population. Dunn *et al* (1995), using GIS found that there was an increased rate of asthma prevalence in middle aged and elderly adults living in the area between 0.5 and 1 km to the north east of the factory. This was the area where airborne emissions from the factory were expected to be carried by the prevailing south westerly winds. In another study, Dunn and Kingham (1996) used GIS to predict and model a surface of isolines representing different levels of volatile organic compound (VOCs) emitted to the air from a point source. The interpolated VOC levels then formed the basis of an analysis to test whether spatial variations in air quality were associated with respiratory symptoms reported by respondents of a questionnaire survey. However, they did not find a statistical association between modelled VOC levels with reported long-term illness.

Planning the location of health services can be aided greatly by the use of GIS. In remote rural areas many communities are located far from health services. This is certainly true in South Australia where GIS has been used together with the Accessibility/Remoteness Index for Australia (ARIA) to define distance from populated locality to health services. The resulting maps show populated localities and the location of health services and provide a good visual guide to help contribute to health service planning (Bamford *et al.*, 1999). In another study, Tanser and Wilkinson (1999) used GIS and GPS to document access to community-based treatment for tuberculosis in Hlabisa, South Africa. By identifying areas where treatment was lacking, GIS can help guide the location of new primary care clinics and hence the treatment of tuberculosis. Although not specifically related to epidemiology, identifying and mapping community vulnerability is another important role for GIS. Natural disasters generally cause the greatest harm to communities that are socially and economically poor. In the developing world this may include the majority of the population, however in developed countries such vulnerable communities tend to live in specific neighbourhoods. Knowledge of where these groups are concentrated within communities, and the general nature of their circumstances is an important step towards effective emergency planning (Morrow, 1999).

GPS are increasingly becoming common mapping tools that are easily incorporated in a GIS. Hightower *et al* (1998) accurately mapped geographical features of interest associated with a longitudinal study of malaria in 15 villages in western Kenya. They used distance from households to the nearest major potential larval habitat and observed distinct patterns of abundance by household and village for each mosquito species that change between rainy and dry seasons.

National scale

Understanding the level of endemicity of malaria infection would assist managers of control programmes to use interventions rationally, where they are most likely to succeed. Omumbo *et al* (1998) were able to map the estimated risk of infection with *P. falciparum* among Kenyan children using parasitological, geographical, demographic and climatic data in a GIS platform. Their research concluded that the risk is low for 2.9 million, stable but low for another 1.3 million, moderate for 3.0 million and high for 0.8 m. In Israel/Palestine, malaria has been successfully eradicated. However, since the early 1980's several important malaria vectors have expanded their geographic range.

At the same time the number of malaria carriers increased significantly as a result of immigration from Ethiopia, with more than 2000 cases of imported malaria diagnosed from 1981 to 1990. As a result, concern about the infection of local anopheline mosquitoes and the possible resurgence of malaria transmission has increased. Kitron *et al* (1994) have used GIS to develop a surveillance system of breeding sites of the *Anopheles* mosquitoes and imported malaria cases. Distances between breeding sites and population centres were calculated. Risk of malaria transmission was assessed with consideration of vectorial capacity and flight range of each *Anopheles* species. If a localized malaria outbreak does occur, it will be associated rapidly with a likely breeding site, a specific *Anopheles* vector and a probable human source, so that prompt control measures can be most efficiently targeted. Dengue and malaria are important mosquito-borne diseases affecting economic development in Southeast Asia. Indaratna *et al* (1998) employed GIS to examine the distribution of both diseases in relation to economic resources in Thailand. The two diseases vary greatly in overall seasonal patterns and in relation to provincial economic status, and create differing demands on resource utilization. Malaria, and to a lesser extent dengue tend to be concentrated along international border areas and exposes the need for multi-country co-ordination of disease management and control programmes. By illustrating the varying geographic patterns of malaria and dengue this investigation may help contribute to the rationalizing of control programmes in terms of material, economic and social resources with respect to time, season and provincial demands.

Continental scale

There has been much work recently concentrated on mapping disease on a continental scale. The MARA/ARMA collaboration has made much progress in mapping various aspects of malaria in Africa (MARA/ARMA, 1998; Craig *et al.*, 1999; Snow *et al.*, 1999a; Snow *et al.*, 1999b). GIS was used extensively for these projects. GIS was also used by Lindsay and Thomas (2000) to the map risk of lymphatic filariasis infection across Africa. They hypothesized that the distribution of lymphatic filariasis is governed by climate. Using computerized climate surfaces they were able to distinguish the climate at sites in Africa where surveys for lymphatic filariasis had taken place. Logistic regression analysis of the climate variables predicted with 76% accuracy whether sites had microfilaraemic patients or not. The logistic regression equation was then entered into the GIS to map the risk of lymphatic filariasis infection across Africa.

These examples illustrate that GIS are now playing a major role in our understanding of disease patterns and health status. The growing availability of demographic, environmental, socio-economic and health data with the increasing accessibility of GIS software packages, provides the means to examine the role of disease factors in determining health outcomes. GIS together with RS data empowers us to view the landscape in a very new spatial dimension. GIS is one of the few tools meeting the need of monitoring the distribution of a disease in space and time, GIS can therefore form an integrated part of surveillance systems. This facility is now within the reach of many health scientists and researchers and is likely to become a fundamental tool in monitoring, forecasting, understanding and resource allocation planning of health and disease.

4 Materials and methods

4.1 Study area

The research was carried out in the West Usambara Mountains in north-east Tanzania. The Usambara's are situated near the Indian Ocean and the Kenyan border and cover an approximate area of 2500 km^2. The highest peak reaches 2314 m. To the east and separated by the deep Lwengera Valley are the East Usambara's, a smaller and lower highland block, covering an area of roughly 330 km^2. The mountain ranges climb sharply from the surrounding lowland plains, and are characterised by steep undulating slopes (Bødker, 2000). Figure 4.1 displays the location of the Usambara Mountains in Tanzania.

The seven villages selected for this study were situated along an altitudinal transect rising from 300 m to 1650 m (Table 4.1).

Table 4.1 Characteristics of study villages

Village Name	Referred to as	Altitude (m)	Location	No. Houses	No. Individuals
Kwameta	Village1	300	5°06' S, 38°29' E	155	760
Magundi	Village2	600	5°05' S, 38°29' E	140	702
Kwamhanya	Village3	750	5°03'S, 38°28' E	77	391
Bagamoyo	Village4	1000	5°03'S, 38°26' E	128	611
Ubiri	Village5	1200	4°50' S, 38°19' E	141	788
Balangai	Village6	1300	4°56' S, 38°28' E	172	832
Milungui	Village7	1650	4°46' S, 38°21' E	144	767

In this region, the climate is monsoonal, with a long rainy season from March to May, and short rains commencing October through to December. The spatial distribution of rainfall is mainly associated with altitude and aspect. In general, the higher areas receive more rain than the lower areas with a similar aspect. Areas on the south east of

Figure 4.1 Map of Tanzania showing location of Usambara Mountains

the mountain range tend to receive more rainfall than areas of a similar altitude on the northern or western slopes. All villages received between 1000-1300 mm of rainfall in 1996, except for the village at 1000 m, which received considerably more rainfall than the other villages throughout the year with 1920 mm. The difference in mean outdoor temperature between the top and bottom village is about 10°C, ranging from about 25°C at 300 m to 16°C at 1650 m (Bødker, 2000). Relative humidity is high in the lowlands for most of the year.

Most areas of the Usambara's were once covered with submontane tropical rainforest. Nowadays the only forests found in the West Usambara's are confined to reserves and plantations. Most deforestation has occurred during the last century, mainly due to population pressure. The land has been cleared for agriculture and the timber has been used for firewood and other local use. The introduction of tea and coffee plantations, and commercial logging activities has also contributed to the widespread deforestation (Hamilton & Bensted-Smith, 1989). Since the 1900's, the population of the Usambara's has grown dramatically. The current population averages over 300 people/km^2, representing one of the highest densities in the world for agriculture-based economies (Adams *et al.*, 1996). Deforestation has been much slower and less extensive in the East Usambara's and concerted efforts are being made to protect forest reserves. The West Usambara Mountains are one of the few areas of Tanzania where tightly clustered villages are found (Berry, 1971). Most highland villages are situated on ridges. The agricultural system is essentially individualistic. Each household owns a number of plots scattered in the valleys and on the hillsides surrounding the village. The household farms these plots with some help from neighbours if needed. Bananas were formerly the staple food, but they have been largely superseded by maize and cassava. Coffee, tea, and more recently market vegetables are grown as cash crop in parts of the area.

4.2 Prevalence Data

Clinical surveys were carried out at the end of the transmission season, between November and December 1996. All households included in the study villages were numbered and household members enumerated. All subjects had their spleen palpated and it was noted if their spleen was enlarged or not. Subjects were also finger pricked and a thin and thick blood smear prepared on a glass slide. Slides were stained with

Geimsa, and 100 fields examined for asexual malaria parasites at 1000x magnification. Positive slides were counted against 200 leucocytes, and counts converted to asexual parasites per µl blood, assuming an average concentration of 8000 leucocytes per µl blood. The quality of the slide was recorded and only thick films judged as "good" were used in analysis. Parasite densities were determined for 100 fields analysed. Thus, a slide may be positive in the 100 fields examined, but negative in the counting against 200 leucocytes, since the latter examination has a lower sensitivity (Bødker, 2000).

For this research the following data were used:

House number
Person identification number
age at survey
spleen (enlarged or not)
parasites present in blood (yes or no)

4.3 Mapping - data acquisition and field work

A sketch map was made of each village showing the position of all households included in the survey, together with local stream and road networks. Using a Global Positioning System (GPS; Trimble Geoexplorer II), the longitude and latitude was measured for 30-60 houses in each village. The number of households located in this way varied according to the size and extent of each village, so that fixed GPS points were evenly spread throughout all areas of the village. To increase the precision of each house measurement, 300 fixes were taken. These fixes were later averaged using the Trimble Navigation Pathfinder software, thus locating each house position to within 25 m (August *et al.*, 1994).

4.4 GIS Construction and data extraction

4.4.1 Data sources

The following data sources were used to construct the GIS and for statistical analysis:

- Ordnance Survey for the Government of the United Republic of Tanzania – 1:50,000 scale map of Korogwe District, Series Y742 (D.O.S.422), Sheet 129/2 Edition 3-TSD/OSD 1988

- Ordnance Survey for the Government of the United Republic of Tanzania – 1:50,000 scale map of Lushoto District, Series Y742 (D.O.S.422), sheet 109/4, Edition 3-TSD/OSD 1988
- Ordnance Survey for the Government of the United Republic of Tanzania – 1:50,000 scale map of Mlalo District, Series Y742 (D.O.S.422), Sheet 109/2, Edition 3-TSD/OSD 1988
- Sketch maps of each village
- GPS data collected for each village, using the Trimble Geoexplorer II and converted via the Trimble Navigation Pathfinder software
- Clinical data was acquired from Bødker, R. (1996) – The data were contained in 2 files: *per.rec* - details about each person in the survey (house identification number, gender, age at survey)
  mal.rec – malaria status (spleen classification, parasite load in blood sample)

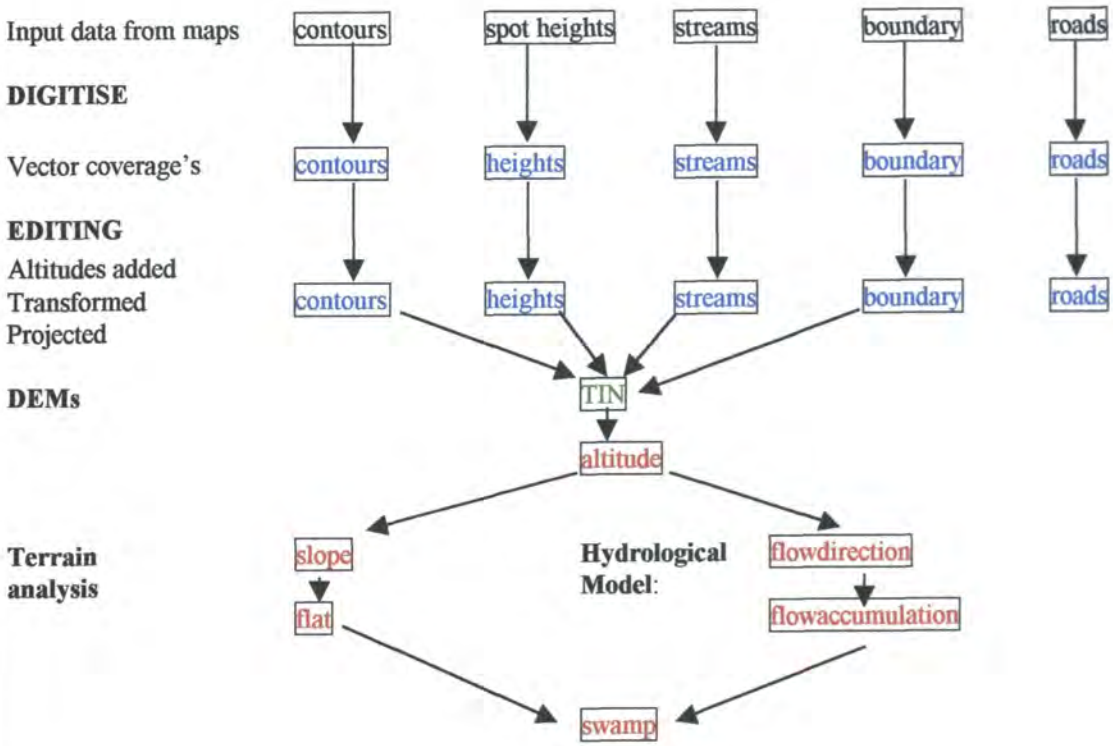
4.4.2 Data input:

The GIS software system used was ARC/Info Versions 7 and 8.

First, the relevant map data was digitised. A digitiser is a device for capturing maps in a format suitable for storage and manipulation in a computer (Mitchell, 1991). The digitiser is essentially a 'tablet' of electronic graph paper to which the baseline map for digitising is secured. Map features were then traced with a flat cursor that detects the signal of any intersection of grid wires and the signal is coded in the computer into x and y co-ordinates. The different features were digitised and stored within the GIS as individual layers. In ARC/Info, each layer is termed a coverage.

The first stage in the digitising process is the entry of four or more 'tic' points, which are reference markers identified by known co-ordinates on the base map. These markers are used to align the coverage's and pinpoint their position in relation to the 'real-world'. Figure 4.2 outlines the steps taken in constructing the GIS.

Figure 4.2: GIS Stages involved in entering and manipulating the topographic data



Data digitised and coverage's produced for each village area:

- *Boundary*

  The boundary coverage encloses all the required village features and was used to associate all digitised coverages to their 'real-world' coordinates. The four corners of this coverage were digitised as 'tic' points.

  Villages at 300 m, 600 m, 750 m and 1000 m were contained in one boundary coverage of 14 km x 13 km, due to their proximity to each other. Villages located at 1200 m, 1300 m and 1650 m were situated in separate boundary coverages 8 km^2.

- *Contours*

  Contour lines at 40 m intervals
- *Spot Heights*
- *Streams*
- *Roads*
- *Village*

  House positions taken from each village sketch map. The GPS located houses were digitised first as 'tic' points. Then all houses included in the survey were entered as point features, including the GPS located houses. Figure 4.3 summarizes the stages followed in entering and manipulating the house data.

4.4.3 Editing the topographic coverage's

Once the relevant topographic map features had been digitised, they were examined for errors, and edited. The altitude of each contour was also entered in the relevant contour coverage attribute file (*contour.aat*). All digitised coverages were transformed from digitiser units to their real-world co-ordinates. This was done using the four 'tic' points of the *boundary* coverage, where the real world co-ordinates were known. Each coverage was then associated with its correct co-ordinate system (UTM, Units: meters, Zone: 37, Spheroid Clarke1880).

4.4.4 Creating Digital Elevation Models (DEMs)

At this stage the study areas are represented as 2-D images, in vector (line) format on the PC screen, just as they appear on hard copy maps. Although relief may be represented as a 2-D image in the form of a contour map within a GIS, such an image will not clearly convey the surface shape and characteristics relating to its hydrology, such as slope and flow. Consequently, a 3-D representation of each area is required. This was achieved by constructing a DEM, where digitised contour maps were interpolated, allowing the estimation of elevations in regions where no data exist. . Each square in a grid is called a 'cell'. In this study, a cell resolution of 50 m was used. The value of each cell in the elevation grid represents different altitudes across a continuous surface. A grid is therefore a 2-D array of elevations (Weibel, 1991). However, with a 2-D grid the resolution is coarser, and so the complexity of relief may not be depicted as well as with a TIN. Both data structures are useful for elevation and terrain analysis, depending on the nature of the project. We used both TIN and lattice data structures. The former was used to display each area as a 3-D landscape and was

incorporated in generating the elevation or *altitude* grid. The *altitude* grid was later used to determine accumulated flow and flat areas within each village site.

4.4.5 Determining accumulated flow and identifying flat areas:

Before calculating the amount of water collecting in each village landscape, the direction of flow was established using the *altitude* grid. The flow direction information, held in the *flowdirection* grid, was then used to compute accumulated flow for every cell within each village area. The cell values in the *flowaccumulation* grid produced have no units. The values simply represent the accumulated weight of all cells flowing into each down slope cell. A weighting of one is applied to each cell and the value of the cells in the output grid will be the number of cells that flow into each cell (ESRI, 1991).

The *altitude* grids were used to determine the gradient or slope value of each cell. To distinguish flat zones within each village, cells with a gradient between 0-5° were selected and the resulting grid named *flat*.

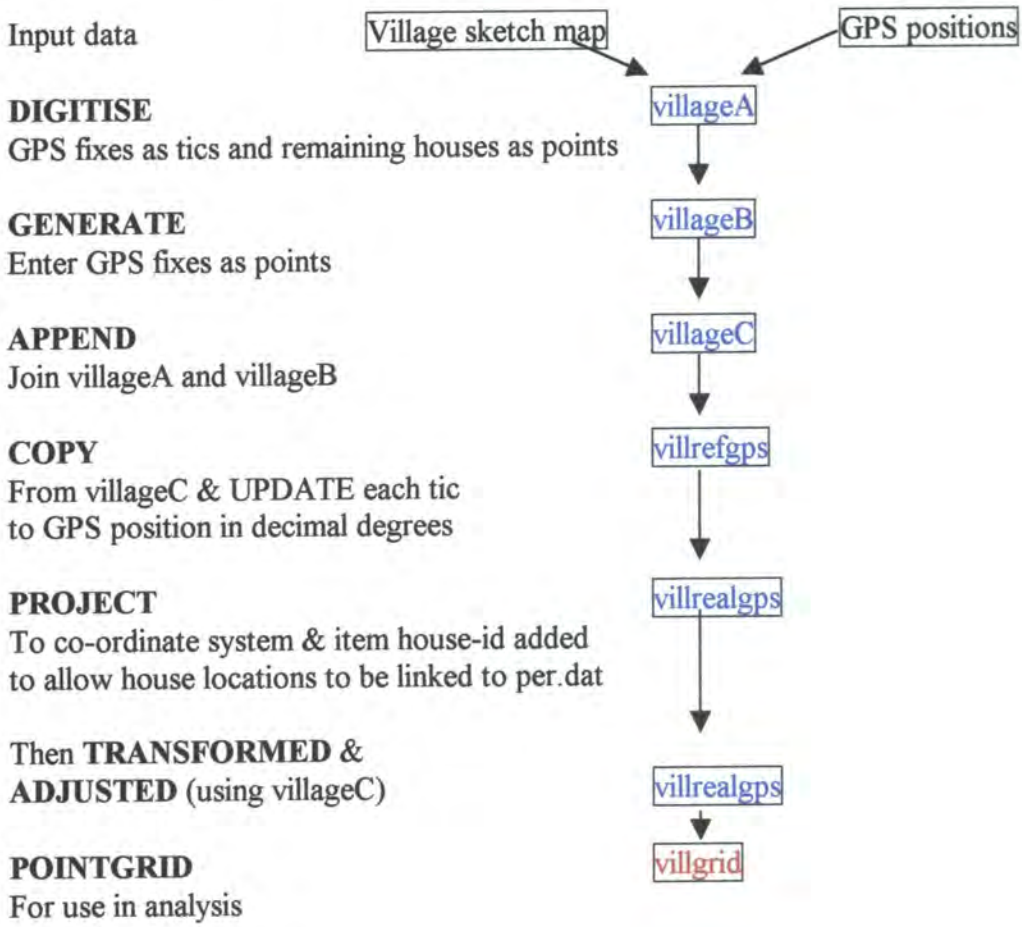
The *flat* grids and their corresponding *flowaccumulation* grids were then multiplied together to produce a set of grids that pulled out only flat areas where flow was accumulating. This was done by multiplying the cells in the *flat* grid with a value of 1 by the flow value in the *flowaccumulation* grid. The resulting set of grids were named *swamp*. The term swamp is for naming purposes only and should not be taken as a suggestion that the areas represented in this grid are swampy, although they may have the potential to be swampy.

4.4.6 Houses and GPS data:

The house positions for each village, in the coverage called *villageA*, are represented as: 'tics' (houses located with the GPS), and points (remaining houses in the survey). Figure 4.3 outlines the steps taken to incorporate the house position information within the GIS. Next, using the generate command a coverage called *villageB* was produced. The generate command copied the 'tics' from *villageA* and then allowed the user to manually enter a series of x and y co-ordinates as points. In this way, the GPS positions were entered as points (the x and y co-ordinates were extracted from the 'tic' locations in the 'tic' file of the *villageA* coverage, *villageA.tic*, at this stage still in digitiser units).

VillageA and *villageB* were subsequently joined using the append command, producing a third coverage called *villageC* that contained all households in the survey as ordinary points.

Figure 4.3: GIS stages involved in entering and manipulating house location data



Another coverage was then created, *villrefgps,* by copying the tics and house positions from *villageC.* In this new coverage each tic position was selected and updated with its GPS position in decimal degrees. *Villrefgps* was then projected to associate it with a co-ordinate system, and the resulting coverage was named *villrealgps.* The digitised co-ordinates for *villageC* were then transformed to their real-world co-ordinates using the projected positions in *villrealgps.* At this stage the house positions were not accurately placed, except for those houses located with the GPS. To correct each house position to its exact position, links were created between the tics in *villageC* and their exact positions in *villrealgps.* The links 'nail' the *villageC* tics to their proper positions, and

the remaining house points are moved in relation to them. This adjustment of the map is called 'rubber-sheeting', and it is a way of warping a map, using known positions as reference markers, to move the points where exact positions are unknown, to their correct location.

4.4.7 Incorporating the clinical data in the GIS

The information about each person included in the survey was contained within files in Epi Info (Dean, 1995) format, a statistical package that analyses epidemiological data. Data designed for this research were extracted and edited to ASCII format. Two files were used; one called *'per.rec'* containing information relating to each person, such as house-id number, person identification number, date of birth and sex; and another that included the malaria status data for each person named, *'mal.rec'*. In ARC/Info tables were defined, allocating heading names, column widths and data type. The Epi Info data were downloaded to the relevant tables created in ARC/Info (*per.dat* and *mal.dat*); both tables were then joined using the 'person identification number' contained in each table as the link. The combined table was called *per.dat*.

To enable the association of each person contained in the *per.dat* tables to their appropriate house in the coverage *villrealgps*, the item 'House-id' was added to the house position attribute table (*villrealgps.pat*). The house identification numbers were then entered in this column alongside their x and y co-ordinates. All individuals were related to one household as specified by the column 'House-id' in each *per.dat*. Using 'House-id' as the key, the attribute information for individuals can now be linked to their house location in the GIS.

4.4.8 Extracting environmental data for each house

The following data was extracted for the location of each house:

- altitude in meters
- gradient in degrees
- accumulated flow (no units)
- the sum of accumulated flow within 50 m, 100 m, 250 m, and 500 m radii of each house
- the sum of flat cells (between 0-5°) within 50 m, 100 m, 250 m, and 500 m radii

- the sum of swamp cells within 100 m, 200 m, 300 m, 400 m and 500 m radii

Figures in Appendix A illustrate a selection of coverages, grids and surface models produced during this data exploration.

4.5 Statistical methods

Tables were created in SPSS (Statistical Package for Social Sciences Release 9.0.0, Copyright © SPSS Inc., 1989-1999) for the entry of the GIS data extracted for each household mentioned previously and the information regarding each person included in the survey. The main database contained the following information for each person:

Person identification number

House-id

Village number (1,2,3,4,5,6 or 7)

x co-ordinate

y co-ordinate

Gender

Age at survey

DEPENDENT VARIABLES:

Spleen (0 = not enlarged / 1 = enlarged)

Pfcount (parasite count in blood – parasites present = 1 / no parasites = 0)

INDEPENDENT VARIABLES:

Altitude (meters)

Slope value (degrees)

Flowacc – accumulated flow (no units)

Flow50/100/250/500 – total accumulated flow within 50, 100, 250, 500 m radii of each house

Flat 50/100/250/500 – number of flat cells within 50, 100, 250, 500 m radii

Swamp100/200/300/400/500 - swamp value within 100, 200, 300, 400, 500 m radii

4.5.1 Preliminary data exploration

Using SPSS, exploratory analysis was carried out to see if the data were normally distributed and if there were any outliers or extreme values in the data set. This was done using 'box plots' by comparing the controls (no malaria) and cases (positive for malaria) for each independent variable.
Cases and controls were determined by the dependent variables:

pfcount (pfcount of 0 = no malaria, pfcount > 0 = malaria) and
spleen (no enlargement = no malaria, enlargement = malaria).

The analysis was performed for households, over three age categories (all ages, under 10 years, and under 5 years), for each village, and for the villages collectively. When comparing cases and controls for households, it should be noted that a house is positive for malaria if there is just one positive individual in a particular age category living in that household. It became clear after this exploratory analysis that these data were not normally distributed and so non-parametric tests were then used to compare topographic values between the cases and controls.

4.5.2 Mann-Whitney *U*-test

The Mann-Whitney *U*-test is a non-parametric technique for comparing the medians of two unmatched samples. The test is distribution-free and is suitable for data that are not normally distributed (Fowler *et al.*, 1998). This test was used to determine whether there were significant differences between the cases and controls for the various geographical independent variables. The cases and controls (for spleen and pfcount) were compared for households, over the three age categories and for each village and the villages collectively. The median values and interquartile range were also recorded for the criteria mentioned above.

4.5.3 Logistic regression analysis

Logistic regression analysis is used for predicting whether an event will or will not occur, as well as identifying the variables useful in making the prediction (Norusis, 1994). Thus, logistic regression analysis can be used to predict a binary dependent variable (an event occurring or not occurring), allowing both the inclusion of a set of

continuous explanatory variables and the assessment of interaction (Kirkwood, 1988). In this case, we are interested in which topographical variables are important in determining whether malaria will be present or not.

Forward stepwise logistic regression was performed, incorporating the most significant geographic variables identified by the univariate analysis using the Man-Whitney *U*-test, to find the best predictive model. This was done for households, examining prevalence indicated by spleen or parasite count in the three age categories for the each village and for the villages collectively. This technique should identify the topographic variables that are most important in determining the risk of malaria in this highland region and generate a model predicting which areas are at risk. Standard diagnostics were carried out to assess the quality of the models.

4.5.4 Spatial Analysis

Spatial autocorrelation is said to exist when an observed value of a variable at one locality is dependent on the values at neighbouring localities (Sokal & Oden, 1978). This occurrence is reiterated in the statement which (Tobler, 1970) referred to as the "*first law of geography: everything is related to everything else, but near things are more related than distant things.*" Positive spatial autocorrelation is said to exist when variables that are similar in location also tend to display similar characteristics. Alternatively, negative spatial autocorrelation is present when variables that are close together in space tend to be less similar in attribute than those which are further apart (Goodchild, 1986). One of the assumptions in regression analysis is that observations are independent (Smith, 1994; Cliff & Ord, 1981), but this is not true for spatial data. For example, elevation is a continuous variable that generally increases, decreases or remains the same depending on neighbouring regions.

To examine whether spatial autocorrelation exists for the variables used in the logistic regression analysis carried out, Rook's Case Spatial Autocorrelation V.0.9. (Sawada, 1999), a computer program that can be incorporated with Microsoft ® Excel to test for spatial autocorrelation was used. The Moran's I index was calculated for each village over distances at 50 m intervals, from 0 – 1600 m, for *Altitude*, *Swamp400*, and malaria prevalence in each household determined by spleen rate. The Moran's I index varies from + 1 to –1. It is positive when nearby areas tend to be similar in attribute, negative

when they tend to be more dissimilar than might be expected, and approximately zero when attribute values are arranged randomly and independently in space (Goodchild, 1986). The standard z-statistic under the randomisation hypothesis (z rand I) is also calculated, this statistic determines whether the Moran's I index is significant or not. Using the Moran's I index, correlograms were produced to summarise the distribution of the variable over the space studied. The regression residuals from Models 1 and 2 were also analysed for spatial autocorrelation and correlograms produced. If the z rand I is < 1.96 then significant spatial autocorrelation is not present and therefore the models will be robust and reliable (Cliff & Ord, 1981).

4.5.5 Cluster Analysis

GIS is an extremely good visual tool for identifying many environmental and socio-economic characteristics within a defined geographic area. The GIS developed for this research can be viewed to see whether malaria positive houses appear clustered. Spatial clustering was examined for using SatScan software (Kulldorff *et al.*, 1998).

The Bernoulli model in SaTScan was used to calculate the spatial scan statistic for each survey village. The Bernoulli model deals with 0/1 event data such as cases and controls. The spatial scan statistic imposes a circular window on the village area. The window in turn is centred on each of the houses throughout the study region. For each household the radius of the window varies continuously in size from 0 to some upper limit. In this way the method creates an infinite number of distinct geographical circles, with different sets of neighbouring census areas within them, each being a possible candidate for a cluster. Each household used in the analysis is defined in the co-ordinates file which provides the geographic co-ordinates of each house in the village. Each line in the file represents one geographical location. Other files used in the analysis are the control and the case files. These files are defined by entering the number of individuals with malaria for the case file and the number of individuals without malaria in the control file for each house. The cases and controls were determined by spleen rate. Therefore, to calculate the scan statistic for each village, three ASCII files are required. For each village we used SaTScan to carry out 999 Monte Carlo replications to produce the spatial scan statistic and identify any malaria clusters.

4.6 Ethics

All subjects with malaria and or anaemia were offered suitable treatment at no cost. Subjects were also offered treatment for minor ailments or presented to a health service without charge. Ethical approval for this study was given by the Medical Co-ordinating Committee of the National Institute for Medical Research in Tanzania and registered at the Tanzania Commission for Science and Technology (No. 95-217-CC, 96-264-ER-95-121 and 98-00ER-ER-95-21).

5 Results

5.1 Prevalence

Malaria prevalence was measured in each village by the proportion of enlarged spleens and the proportion of blood films showing malaria parasites in the survey population. Table 5.1, and Figures 5.1 and 5.2 display the prevalence rates for households in each village across three age categories. In this instance, a household is positive for malaria if one person of a particular age group in that house has an enlarged spleen and/or has parasites present in their blood.

Examining spleen rates first (Figure 5.1), we see an initial rise followed by a progressive decrease with increasing altitude. Spleen rates were highest over all age categories at 600 m, with 81%, 85% and 79% prevalence for all ages, under 10 years, and under 5 years respectively, compared to 8%, 6%, and 6% for the same respective age groups in the highest village at 1650 m. Generally, spleen rates were highest in the all ages category, except at 300 m and 600 m. In the lowest village spleen rates were greater in the under 5's, whereas at 600 m it is in the under 10 year age group where spleen rates were higher.

Considering the proportion of households that have at least one person with parasites in their blood (Figure 5.2), the pattern is quite different. The highest prevalence was at 1200 m, with 87% of households with at least one person of any age positive for malaria. This village experienced a malaria epidemic just before it was surveyed. Parasite rates were also high in this village for children under 10 years, with 80% of households positive for malaria. Ignoring this village, and only taking into account children under 10 and under 5 years, there was again an initial increase in prevalence, followed by a decrease with rising elevation, although this trend is not as pronounced as with spleen rates. Considering the all ages category, parasite rates increased from 300 m to 750 m and then declined, however, there is less disparity between villages, parasite rates being high (between 87% and 74%) in all villages apart from the highest village where the proportion of malaria positive houses is 44%.

Table 5.1. Malaria prevalence rates in households for the seven survey villages. Confidence intervals (95%) in parenthesis

	300 m	600 m	750 m	1000 m	1200 m	1300 m	1650 m
Spleen rate							
All ages	67	81	78	63	52	19	8
	(60-75)	(75-88)	(68-87)	(54-71)	(44-61)	(13-25)	(4-13)
< 10 years	68	85	60	49	46	13	6
	(58-77)	(78-93)	(47-74)	(39-59)	(37-56)	(7-19)	(2-11)
< 5 years	74	79	65	51	32	9	6
	(63-84)	(69-88)	(49-81)	(38-64)	(22-42)	(3-15)	(1-12)
Parasite rate							
All ages	74	81	86	84	87	77	44
	(67-81)	(74-87)	(78-94)	(77-90)	(81-92)	(71-84)	(36-52)
< 10 years	72	84	83	68	80	62	25
	(63-81)	(77-92)	(73-93)	(58-77)	(72-87)	(53-70)	(16-33)
< 5 years	77	86	78	67	67	46	19
	(67-87)	(78-94)	(64-92)	(55-79)	(57-77)	(36-55)	(10-28)

Figure 5.1. Prevalence for households according to spleen rates for the seven survey villages.

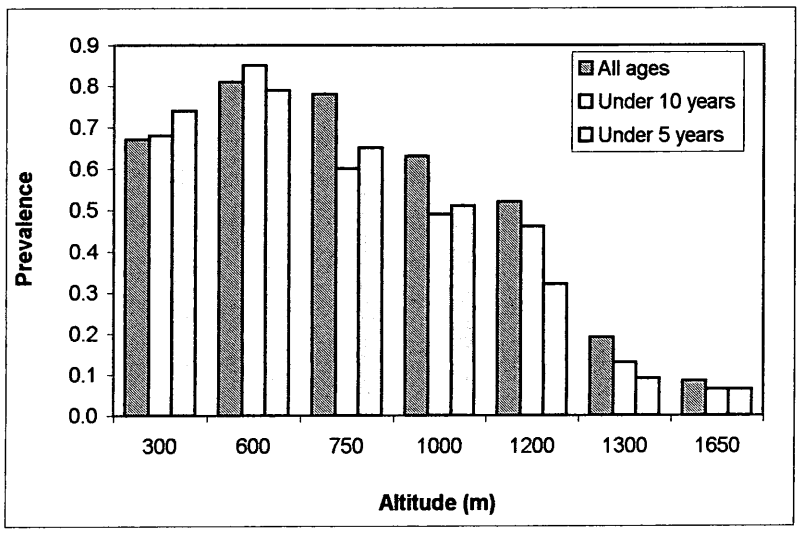


Figure 5.2. Prevalence for households according to parasite rates for the seven survey villages.

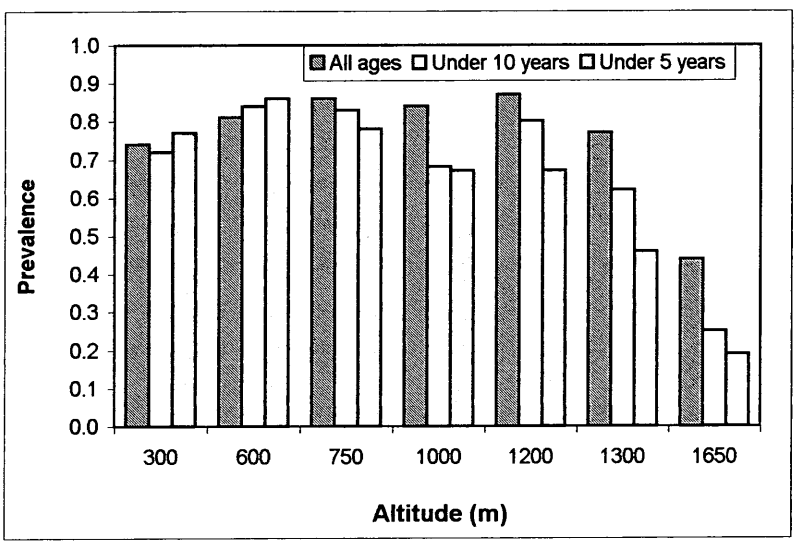


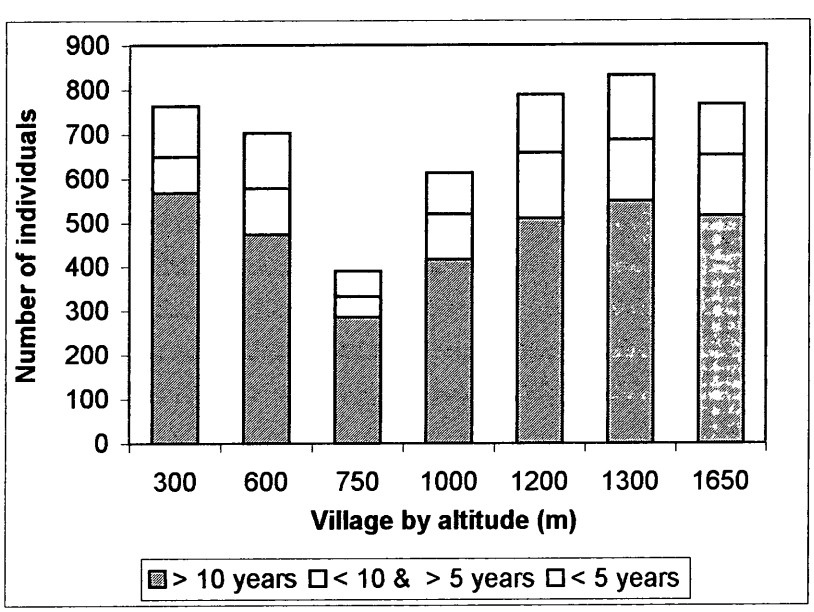
Table 5.2 displays the mean number of people in each household for each village, the standard deviation, and the range. In every village the smallest household size is one person. The number of people living in one house can be quite high, with the largest household size being 22 in the village situated at 600 m. However, in each village the mean household size is 5, the exception being the village at 1200 m where the mean number of people per household is 6. Yet, there is no significant variation in household size between the seven villages ($F_{6,950} = 1.26$, NS)

Table 5.2. Mean and range of household size for each village

Village	**Mean number of people in household**	*s.d*	*Range*
300 m	5	3	1 – 14
600 m	5	3.12	1 – 22
750 m	5	2.58	1 – 13
1000 m	5	2.57	1 – 12
1200 m	6	3.11	1 – 17
1300 m	5	3.09	1 – 18
1650 m	5	3.54	1 – 20

The population structure of each village is shown in Figure 5.3. The survey populations range between 391 and 832 individuals. The percentage of children under 5 in each village is very similar, between 15-18% of the survey population (χ^2 with 6 df = 1.56, NS). The proportion of children under 10 years old is also comparable, between 26-35% of the survey population (χ^2 with 6 df = 2.34, NS). From Table 5.2 and Figure 5.3 it is apparent that all the villages included in the survey have a similar age structure (considering the three named age categories only) and household size.

Figure 5.3. Breakdown of survey population by three age categories for each village



5.2 Man-Whitney *U*-test

Tables B1 to B6 in Appendix B display the results from the non-parametric analysis carried out to compare all the topographic independent variables for the cases and controls. The comparison was done for households over three age categories. Where

there was a significant difference between topographic values for the case and control houses, these data have been highlighted.

5.2.1 Individual villages

Spleen rates

The relationship between topographic measurements and the presence of an enlarged spleen is shown in Appendix B, Tables B1, B3 and B5. The risk of an enlarged spleen in different age groups was significantly related to altitude, flow accumulation between 100 m and 500 m of each household, and flat areas between 50 m to 500 m of each household for villages situated at 1000 m and 1200 m. The epidemic village situated at 1200 m demonstrated the most significant relationship between these variables and an enlarged spleen over all age groups compared to the village at 1000 m. In addition, the amount of swampland surrounding the epidemic village was also significant. None of the other villages displayed such significant relationships. Villages located at 600 m and 1300 m showed no significant differences in any of the topographic variables over all age groups. In the lowest village, the only significant variable was the gradient at the location of every house for all ages. At 750 m, altitude was significant in the under 10 and under 5 age groups, with gradient and the amount of accumulated flow within 500 m radius of each house being significant in the under 5's only. In the highest village, whether a house was situated on a swamp cell or not was the only significant independent variable in the under 10 households only.

Parasite rates

The relationship between the topographic variables and the presence of parasites in the blood is shown in Tables B2, B4 and B6. The risk of parasites in the blood was only significantly related to a few topographic measurements for children under 5 and 10 years old in villages located at 750 m, 1000 m and 1300 m. At 750 m, the only significant variable was the gradient each household with under 5's was located. The village situated at 1000 m displayed the most significant relationships, there being significant differences between altitude, the amount of accumulated flow between 250 m and 500 m radius of each house, the number of flat cells within 250 m and 500 m of each household, and the amount of swampland within 300 m and 500 m. Lastly, at 1300 m, gradient and the number of flat cells within 50 m were significant in the under 10 households only.

To summarise the village results, only villages at 300 m, 750 m, 1000 m, 1200 m and 1300 m show any significant differences in their topographic values for cases and controls. The village at 1200 m exhibits significant differences only for enlarged spleen over all age groups whereas villages at 750 m and 1000 m tend to show significant differences for both spleen and parasite rates. However, more significant differences were demonstrated by spleen rates. It should be noted that in and surrounding the village at 750 m there are very few flat cells, and hence there are no values for any of the swamp variables or the flat cell values, except within radii of 500 m. This village is located on very steep terrain, illustrated by the median slope values for both cases and controls. Villages at 600 m, and 1650 m display no significant differences across any of the criteria.

5.2.2 Villages combined

Tables B1 – B6 also display results from collating the seven villages. These results are summarised in Table 5.3 below. The most significant independent variables calculated across all categories were: *Altitude*, *Flow100*, *Flow250*, *Flow500*, *Flat500*, *Swamp200*, *Swamp300* and *Swamp400*.

Table 5.3. Significant differences between topographic variables for cases and controls for all villages combined (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$)

Independent variable	SPLEEN RATE			PARASITE RATE		
	All ages	Under 10 years	Under 5 years	All ages	Under 10 years	Under 5 years
Altitude	***	***	***	***	***	***
Gradient		*	**			
Flowacc	***	**	*			
Flow50	***	***	***	*	*	*
Flow100	***	***	***	**	*	**
Flow250	***	***	***	***	***	***
Flow500	*			*	*	
Flat50						
Flat100						
Flat250						
Flat500		*	*			
Swamp	**	**	**			
Swamp100	*		*			
Swamp200	***	***	***			
Swamp300	***	***	**			*
Swamp400	**	**	*			*
Swamp500	*					

Figures 5.4 – 5.9 below depict the significant differences for these independent variables.

Spleen rates

From Table 5.3 it can be seen that there are more significant differences between the topographic variables for spleen rates compared to parasite rates. Figures 5.4, 5.6 and 5.8 illustrate the significant differences for spleen rates in more detail. Altitude is very highly significant ($P < 0.001$) across all age groups. Households with malaria are situated at lower elevation than those where there is no malaria. In the under 5 age group the difference in altitude is 682 m (Figure 5.8). Significant differences were also demonstrated by the accumulated flow values, with cases encountering more accumulated flow than controls. The number of flat cells within different radii of each household demonstrates very little significance, except at 500 m radius. There are however, significant differences indicated by 'swampiness', especially within 200-400 m of each household. For example, in the under 10 age group the swamp value within 400 m radius of each household is 448 for cases compared to 133 for control houses (Figure 5.6).

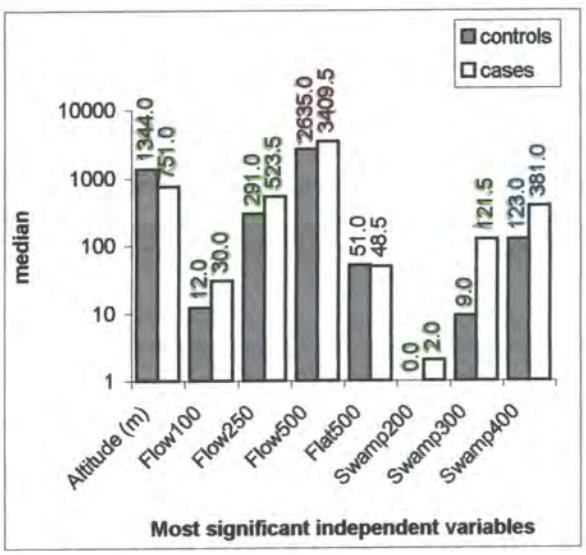


Figure 5.4. Comparing the topographic variables for cases and controls classified by enlarged spleen for all ages and for all villages combined. Coloured values refer to different levels of significance (< 0.05, < 0.01, < 0.001)

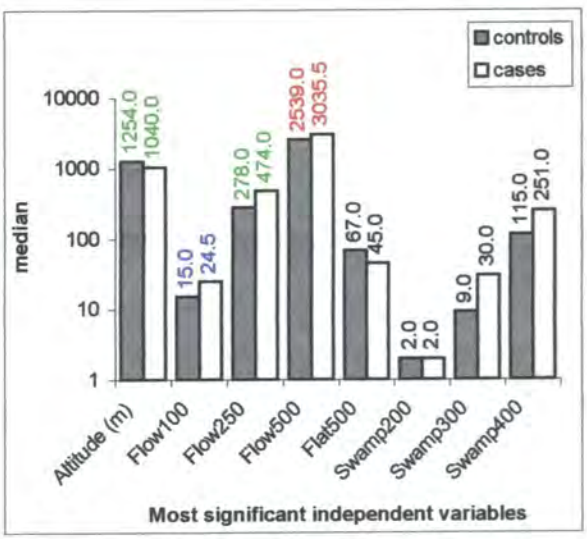


Figure 5.5 Comparing the topographic variables for cases and controls classified by parasites present in the blood for all ages and for all villages combined. Coloured values refer to different levels of significance (< 0.05, < 0.01, < 0.001)

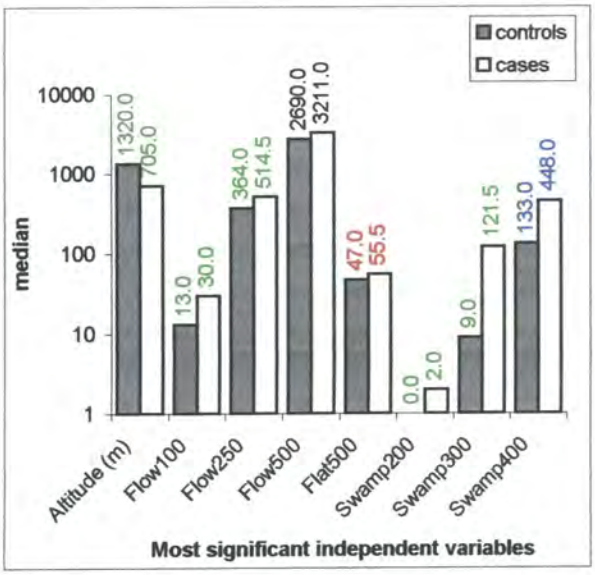


Figure 5.6. Comparing the topographic variables for cases and controls classified by enlarged spleen for children under 10 years and for all villages combined. Coloured values refer to different levels of significance (< 0.05. < 0.01. < 0.001)

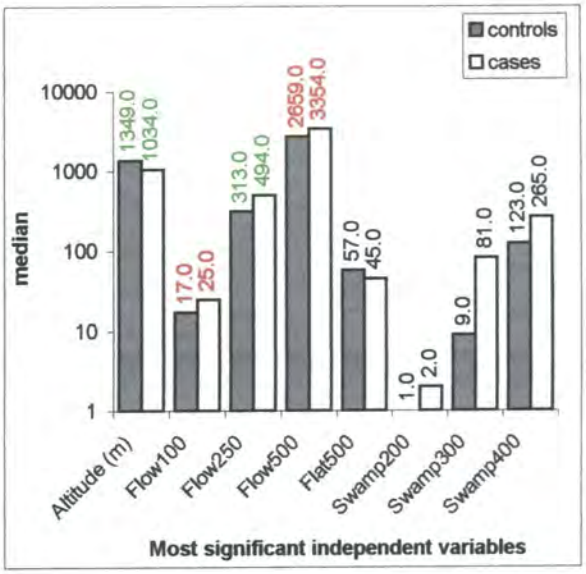


Figure 5.7. Comparing the topographic variables for cases and controls classified by parasites present in the blood for children under 10 and for all villages combined. Coloured values refer to different levels of significance (< 0.05. < 0.01. < 0.001)

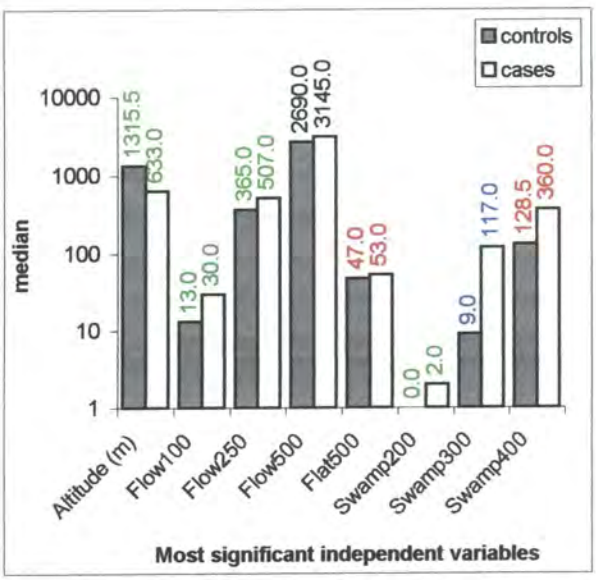


Figure 5.8. Comparing the topographic variables for cases and controls classified by enlarged spleen for children under 5 years and for all villages combined. Coloured values refer to different levels of significance (< 0.05. < 0.01. < 0.001)

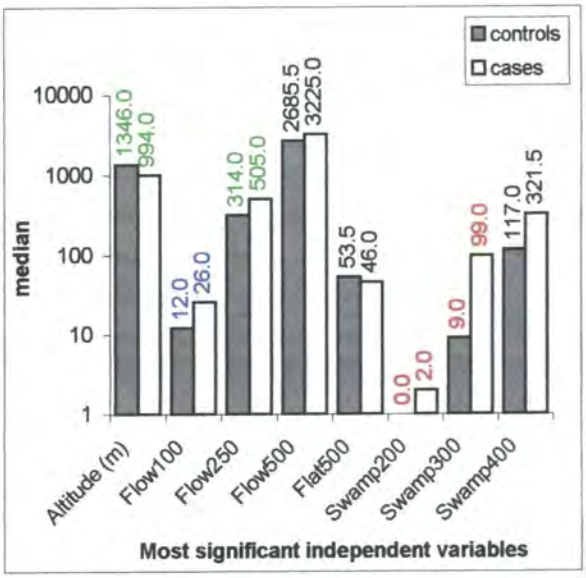


Figure 5.9. Comparing the topographic variables for cases and controls classified by parasites present in the blood for children under 5 and for all villages combined. Coloured values refer to different levels of significance (< 0.05. < 0.01. < 0.001)

Parasite rates

Altitude is again very highly significant, illustrated in Figures 5.5, 5.7, and 5.9. However, there is less disparity between cases and controls compared to spleen rates. For example the greatest difference in altitude between cases and controls for parasite rates is in the under 5 households, controls are situated 352 m higher than cases. Accumulated flow between 50-500 m of each household also exhibits significant differences between cases and controls, especially at 250 m where differences are very highly significant ($P < 0.001$) across all age groups. The number of flat cells and swamp cells at the site of each house and up to 500 m radius of each household were not significant.

To summarise the results presented in Figures 5.5 – 5.10, the most significant topographic variable is altitude, in each Figure the values for altitude are very highly significant ($P < 0.001$). The cases and control houses determined by spleen rate show more significant differences in their topographic values compared to parasite rates across all age categories. The accumulated flow value at 250 m radius is also very highly significant ($P < 0.001$) across all criteria followed by accumulated flow at 100 m. The variable for accumulated flow, flat cells and swamp cells at 500 m although showing some significance, do not appear to be as significant as the same variables at smaller radii. These eight topographic variables will now be investigated further to see if they can be useful in producing a model that can predict malaria prevalence in the survey villages.

5.3 Predictive models

The geographic variables identified by the univariate comparisons between houses with and without malaria were: *Altitude, Flow100, Flow250, Flow500, Flat500, Swamp200, Swamp300* and *Swamp400.* Many of these variables are highly correlated and therefore cannot be used together when carrying out the logistic regression analysis. The accumulated flow grids cannot be combined because they all represent the same variable, but at different radii. This is also the case for the swamp grids. The *Flat500* grid and any of the flow accumulation grids cannot be combined with the swamp grids because the swamp grids were produced by combining accumulated flow with flat cells in those respective grids. This reduces the number of topographic permutations that can be used in the logistic regression equation. Table 5.4 summarises the Spearman Rank

Correlation Coefficient for the topographic variables. There is a weak negative correlation between altitude and the swamp, flow and *Flat500* grids, as altitude increases these variables decrease. Although the swamp grids are produced from the accumulated flow and flat grids, there is only a weak correlation between the swamp and flow grids, whereas the correlation is moderate between the swamp and *Flat500* grid. The accumulated flow grids and *Flat500* demonstrate a very weak positive correlation.

Table 5.4. Topographic grids and their Spearman Rank Correlation Coefficient

GRIDS	Altitude	Flow	Flat500	Swamp
Altitude	X	-0.241	-0.226	-0.283
Flow		X	0.182	0.270
Flat500			X	0.467
Swamp				X

The univariate analysis indicated that spleen rates were the most useful malaria indicator, more significant differences were found between the case and control houses determined by spleen rate compared with results for parasite rate. However, both criteria were used in the logistic regression analysis, and compared to see which produced the best model. The logistic regression analysis was also carried out for households using the three ages categories (all ages, children under 10 years, and children under 5 years) to see which one supports the best model.

The aim in this stage of the analysis is to produce a good model that predicts the risk of malaria between these survey villages, that may be later used as a predictive model for the Usambara Mountains as a whole. However, logistic regression analysis was also be used to explore and produce a model to predict malaria within each village.

5.3.1 Predictive models - between villages

Results of the four best logistic regression models predicting the risk of malaria between villages are displayed in Tables 5.5 and 5.6. Table 5.5 details the predictive capability of each model and Table 5.6 lists the variables and their coefficient for the logistic regression equation. All models used spleen rate as the malaria indicator; parasite rates did not produce consistent models.

Table 5.5. Predictive capability of each model

Model	No. houses in model	Correctly identified		Overall prediction
		cases	controls	
1	811	278/402 (69%)	315/409 (77%)	593/811 (73%)
2	797	289/372 (78%)	319/425 (75%)	608/797 (76%)
3	418	132/177 (75%)	194/241 (81%)	326/418 (78%)
4	346	100/124 (81%)	183/222 (82%)	283/346 (82%)

Table 5.6. Results from the logistic regression models (between villages)

Variable	β	95% C. I.	*P*
Model 1			
Altitude	- 0.0026	-0.0030 - -0.0022	< 0.0001
Constant	2.5093		< 0.0001
Model 2			
Altitude	- 0.0041	-0.0047 - -0.0035	< 0.0001
Swamp400	0.0004	0.0002 - 0.0006	< 0.0001
Constant	4.2186		< 0.0001
Model 3			
Altitude	- 0.0033	-0.0040 - -0.0027	< 0.0001
Constant	2.8985		< 0.0001
Model 4			
Altitude	- 0.0049	-0.0058 - -0.0039	< 0.0001
Flat500	0.0185	0.0067 - 0.0303	0.0022
Constant	4.0230		< 0.0001

All models use spleen rates as the malaria indicator.
Age groups and villages included in models:
Model 1: all ages, villages 1,2,3,4,6,7; Model 2: all ages, villages 2-7;
Model 3: under 5's, villages 1,2,3,4,6,7; Model 4: under 5's, villages 2,3,4,6,7

Altitude has a very strong influence and is therefore incorporated in all the models. Models 1 and 3 are the same, using only altitude as the topographic variable, except Model 1 was performed on all ages, whereas Model 3 was carried out on the under 5 age group. Both models include survey populations from all villages except for the village situated at 1200 m. Model 1 correctly identified 69% of the positive houses and 77% control houses; the overall number of sites correctly classified were 73%. Model 3

produced an improved model, its overall predictive ability was 78%, but the survey population was nearly half that of Model 1.

The general formula for a logistic regression equation is:

$$\text{Probability} = 1 / (1 + e^{-z})$$

For Model 1: $z = 2.5093 - (0.0026 \times \textit{Altitude})$.

We used this equation to map the risk of malaria in the GIS, producing four grids covering the survey villages. These results are displayed in Figure 5.10.

Model 2 included *Altitude* and *Swamp400* (the amount of 'swampiness' within 400 m of each household) in the logistic regression analysis. Households of all ages from villages at 600 m to 1650 m were included in the model. The models overall predictive capability was 76%; correctly classifying 78% of the cases and 75% of the controls. The logistic regression equation was used to map the risk of malaria in the GIS, to produce four more grids, but this time including the contribution of the number of swamp cells within 400 m of each household together with altitude to produce the risk maps (Figure 5.11).

For Model 2: $z = 4.2186 - (0.0041 \times \textit{Altitude}) + (0.0004 \times \textit{Swamp400})$

In Model 4 logistic regression analysis was performed on *Altitude* and *Flat500* (the number of flat cells within 500 m radius of each household). This model included only children under 5 years of age and villages located at 600 m, 750 m, 1000 m, 1300 m and 1650 m. Model 4 correctly identified 81 % of the positive houses and 82% of control houses, therefore, correctly classifying 82% of houses.

The two sets of risk maps produced from Models 1 and 2 (Figures 5.10 and 5.11 respectively) demonstrate how the risk of an enlarged spleen changes with altitude and terrain throughout the seven survey villages. Model 1 illustrates how risk decreases with increasing altitude. In the lowlands, at 300 m the median risk is 85% compared to only 14% in the highest village. Figure 5.12 illustrates the linear relationship between altitude and predicted risk of malaria.

Figure 5.10. Model 1: Risk map of malaria prevalence incorporating altitude

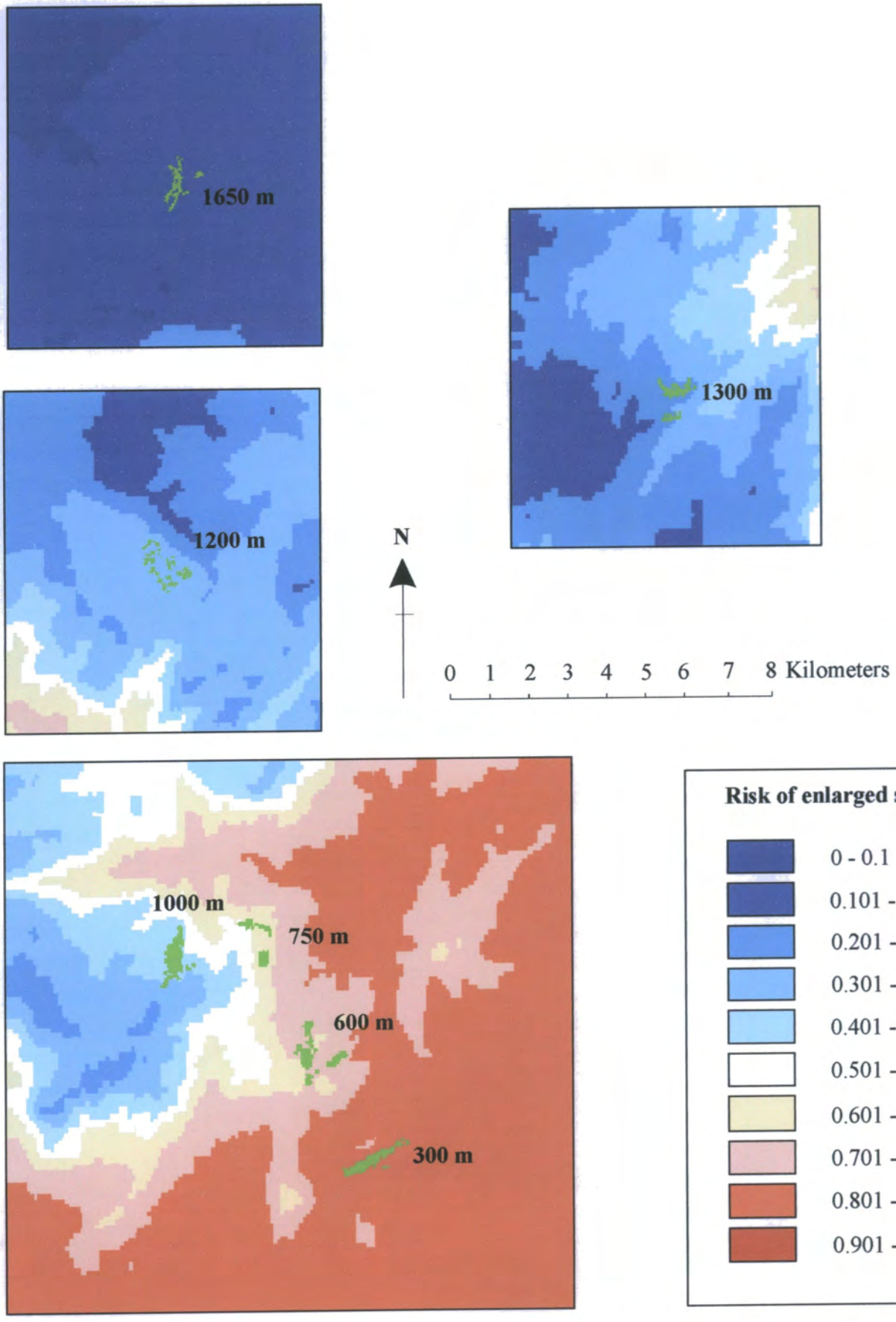
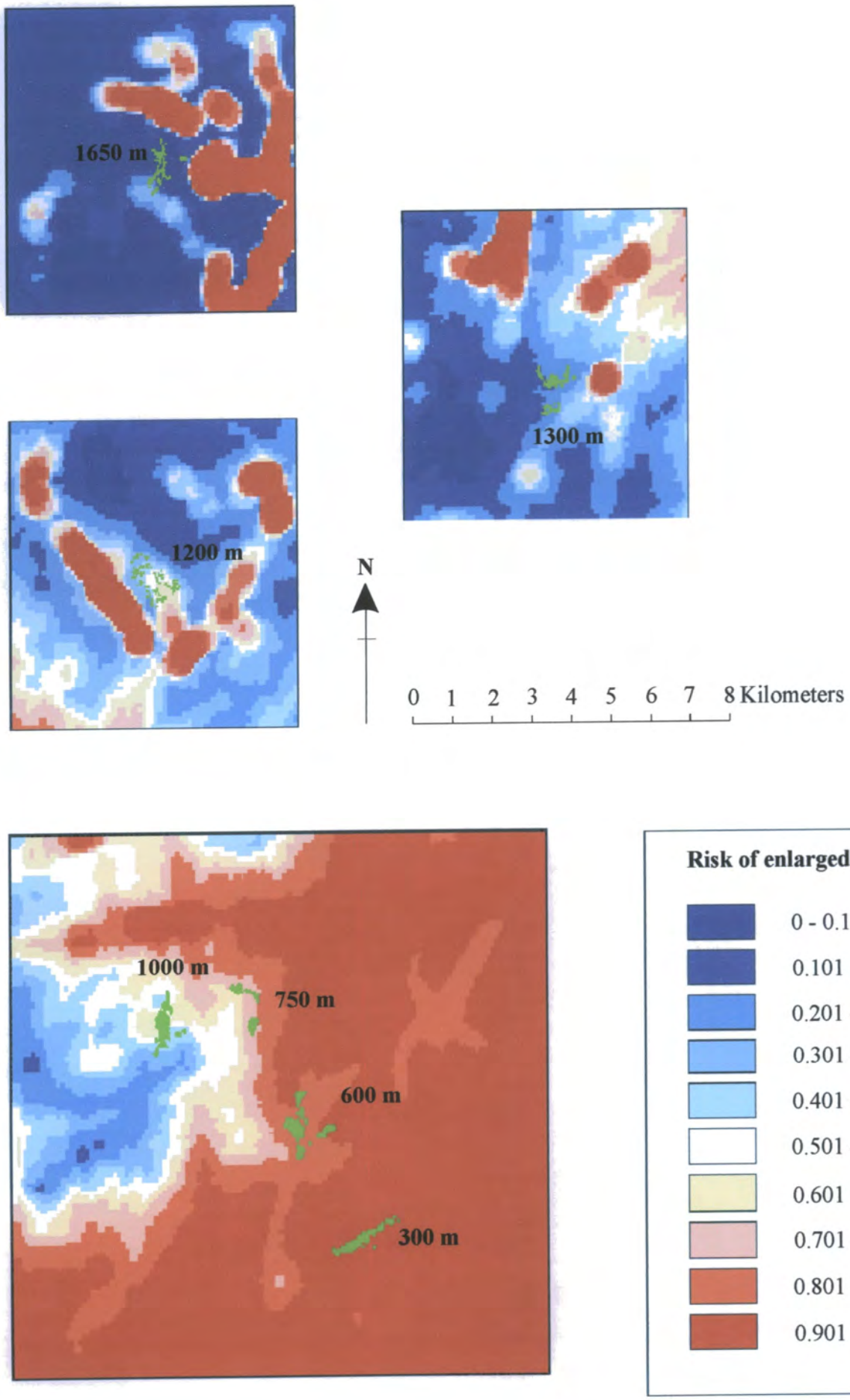


Figure 5.11. Model 2: Risk map of malaria prevalence incorporating altitude and the amount of 'swampiness' within 400 m of each household.



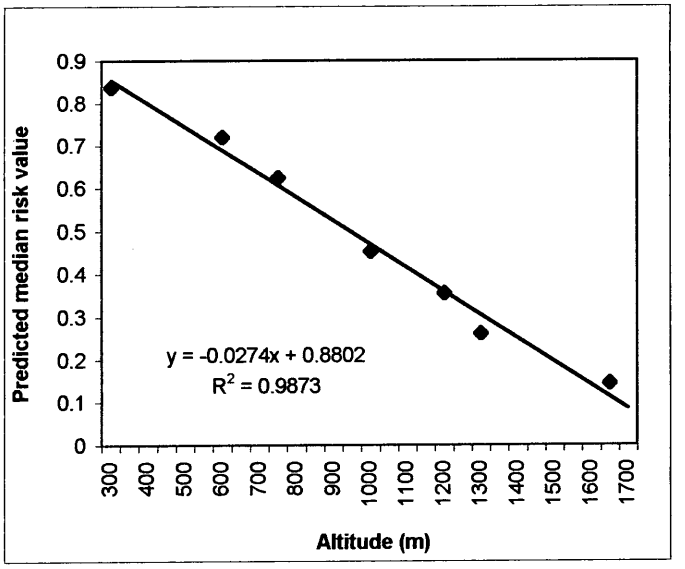


Figure 5.12. Median predicted values for Risk Model 1.

The risk map produced for Model 2 illustrates how the combination of altitude and the amount of 'swampiness' within 400 m of houses affects malaria risk. Again, a clear pattern can be discerned relating decreasing risk with increasing altitude, but within different altitude regions there are pockets of high risk where the terrain becomes flatter and more susceptible to water inundation. Compared to Model 1, results in Model 2 show that median risk is greater in all villages apart from the two at highest elevation. This is most evident at 1200 m where median risk increases from 36% in Model 1 to 51%, illustrated well in Figures 5.10 and 5.11 where the risk shades change from blue (low risk) to white and pink (increasing risk). In the lowest village, risk is highest with a median risk of 99%, even higher than found in Model 1, and lowest at 1650 m, with a median risk of 8%.

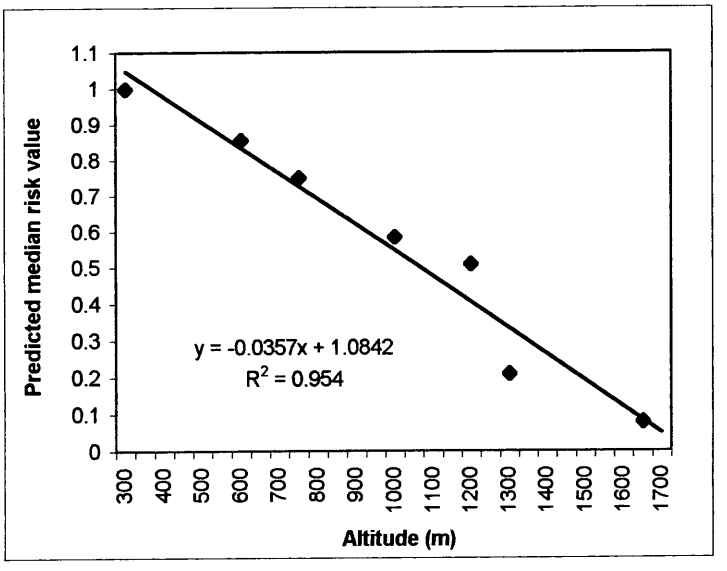


Figure 5.13. Median predicted values for Risk Model 2.

The relationship of risk with altitude is again almost linear as illustrated in Figure 5.13. However, predicted values for villages at 1200 m and 1300 m show a slight departure from this relationship. The regression line also implies that below 300 m predicted risk rises to 1.0842, which cannot be the case.

Univariate comparisons of the predicted median values of Models 1 and 2 were made between the positive and control houses. These results are displayed in Figures 5.14 and 5.15. For Model 1 (Figure 5.14), the predicted median risk values are higher for the cases in all villages apart from the two highest villages at 1300 m and 1650 m, where the predicted median risk is the same for cases and controls. The difference between cases and controls is only significant in the village at 1200 m ($P < 0.01$). Comparing the predicted risk for cases and controls in Model 2 (Figure 5.15), again the predicted median risk values are higher in villages from 300 m to 1200 m, but only significant for the village at 600 m ($P < 0.05$) and again at 1200 m where the difference is more pronounced ($P < 0.01$).

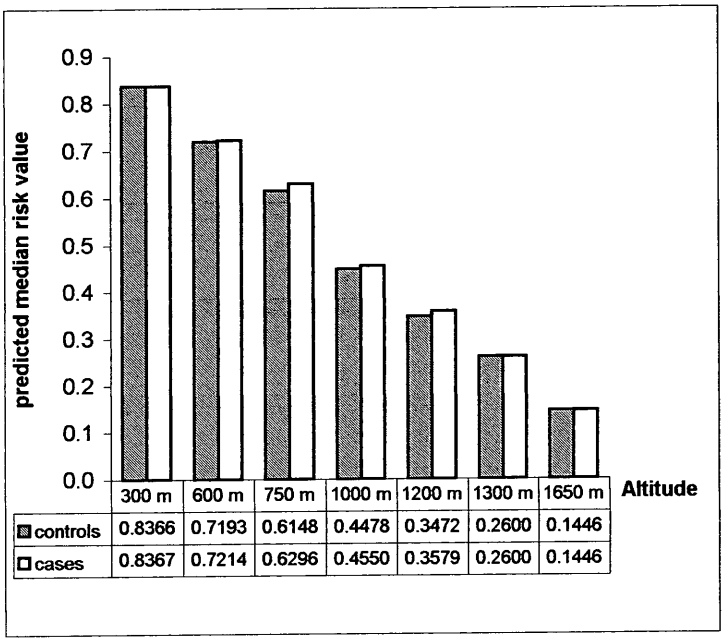
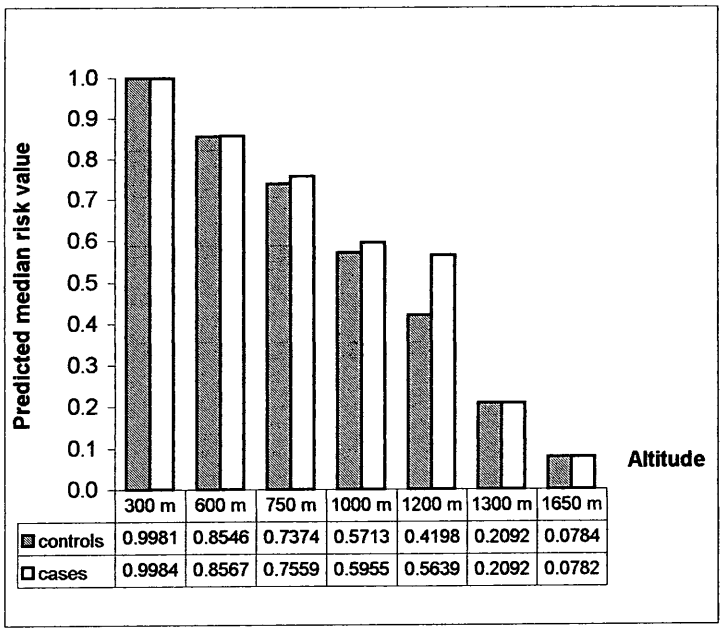


Figure 5.14. Comparing the predicted risk for the cases and controls using the Mann-Whitney U test for Model 1



	300 m	600 m	750 m	1000 m	1200 m	1300 m	1650 m
controls	0.9981	0.8546	0.7374	0.5713	0.4198	0.2092	0.0784
cases	0.9984	0.8567	0.7559	0.5955	0.5639	0.2092	0.0782

Figure 5.15. Comparing the predicted risk for the cases and controls using the Mann-Whitney U test for Model 2

5.3.2 Predictive models - within villages

Logistic regression analysis was also carried out on each village individually. Villages at 750 m, 1000 m and 1200 m were the only ones that produced reasonable models. The results from the best models for these villages are displayed in Tables 5.7 and 5.8. Table 5.7 details the predictive capability of each model and 5. 8 lists the variables and their coefficient for the logistic regression equation. All models used spleen rate as the malaria indicator. *Altitude* featured in each model; in Model 5 it was the only independent variable used. The number of survey households was much reduced compared to the analysis carried out between villages. Model 5 included just households with children under 5 years old and it predicted 68% of the cases and controls correctly. Model 6 was the best model for individual villages. The analysis was carried out again on households with children under 5 years, using *Altitude* and *Flat250* (the number of flat cells within 250 m radius of each house) as the predictive variables. The overall prediction ability of this model was 78%; with 86% of the cases correctly identified and 70% of controls. The village at 1200 m was incorporated in Model 7. This time the analysis was carried out on households with children under 10 years of age, using *Altitude* and *Flat500* as the independent variables. This model was the least reliable predicting only 61% of the cases and controls correctly.

Table 5.7. Predictive capability of each model for individual villages

Model	No. houses in model	Correctly identified		Overall prediction
		cases	controls	
5	37	18/24 (75%)	7/13 (54%)	68%
6	59	25/29 (86%)	21/30 (70%)	78%
7	114	32/53 (60%)	38/61 (62%)	61%

Table 5.8. Results from the logistic regression models (within villages)

Variable	β	95% C. I.	*P*
Model 5			
Altitude	-0.0228	-0.0426 - -0.0029	P < 0.05
Constant	18.3003		
Model 6			
Altitude	-0.0461	-0.0739 - -0.0185	P < 0.01
Flat250	0.0713	0.0713 - 0.1423	P < 0.05
Constant	46.7866		
Model 7			
Altitude	-0.0207	-0.0397 - -0.0016	P < 0.05
Flat500	0.0287	0.0288 - 0.0556	P < 0.05
Constant	22.1423		

All models use spleen rates as the malaria indicator.
Age groups and villages included in models:
Model 5: under 5's, village 3 (750 m), Model 6: under 5's, villages 4 (1000 m)
Model 7: under 10's, villages 5 (1200 m)

5.4 Spatial analysis

Figures 5.16, 5.17, 5.18, 5.19 and 5.20 are correlograms, measuring spatial autocorrelation by distance for *Altitude*, *Swamp400*, malaria prevalence and the residuals in Models 1 and 2 respectively.

Figure 5.16. Correlogram of Moran's I for altitude

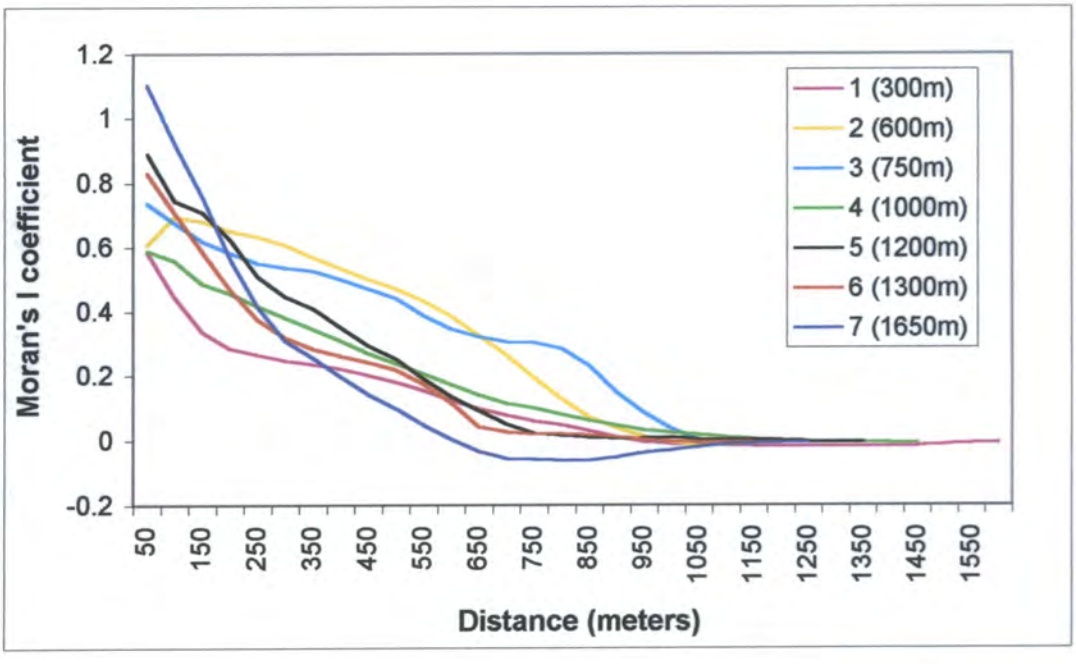


Figure 5.17. Correlogram of Moran's I for the amount of 'swampiness 'within 400 m of each house

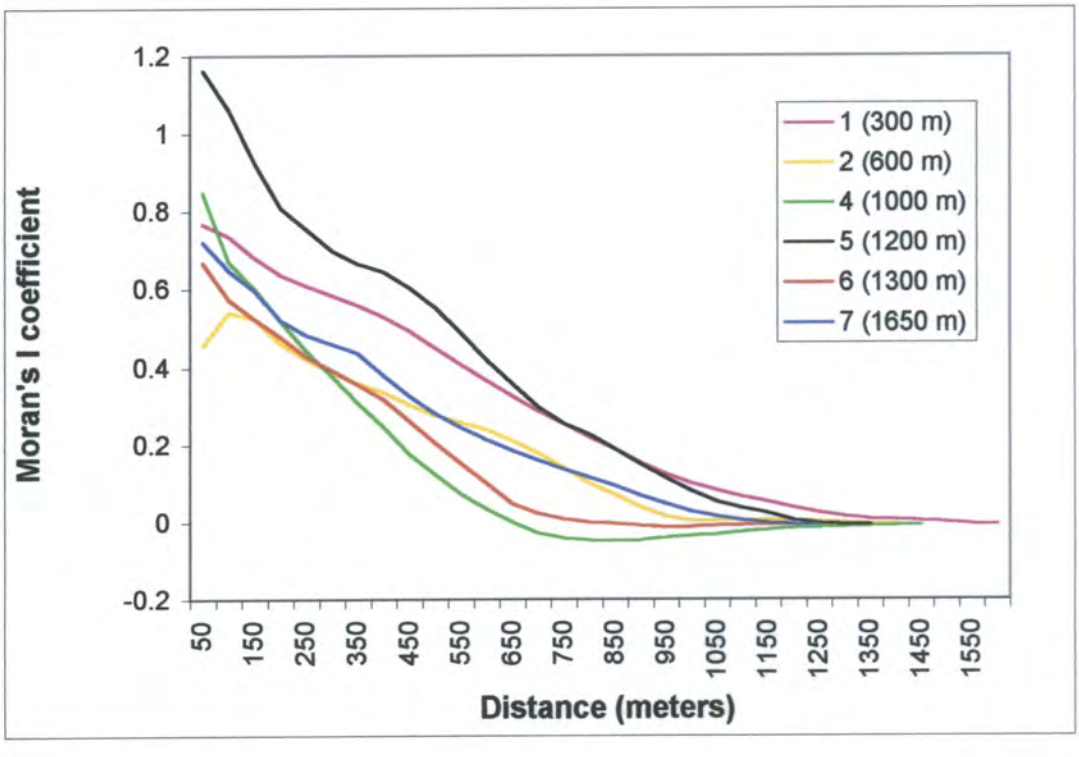


Figure 5.18. Correlogram of Moran's I for malaria prevalence (by enlarged spleen)

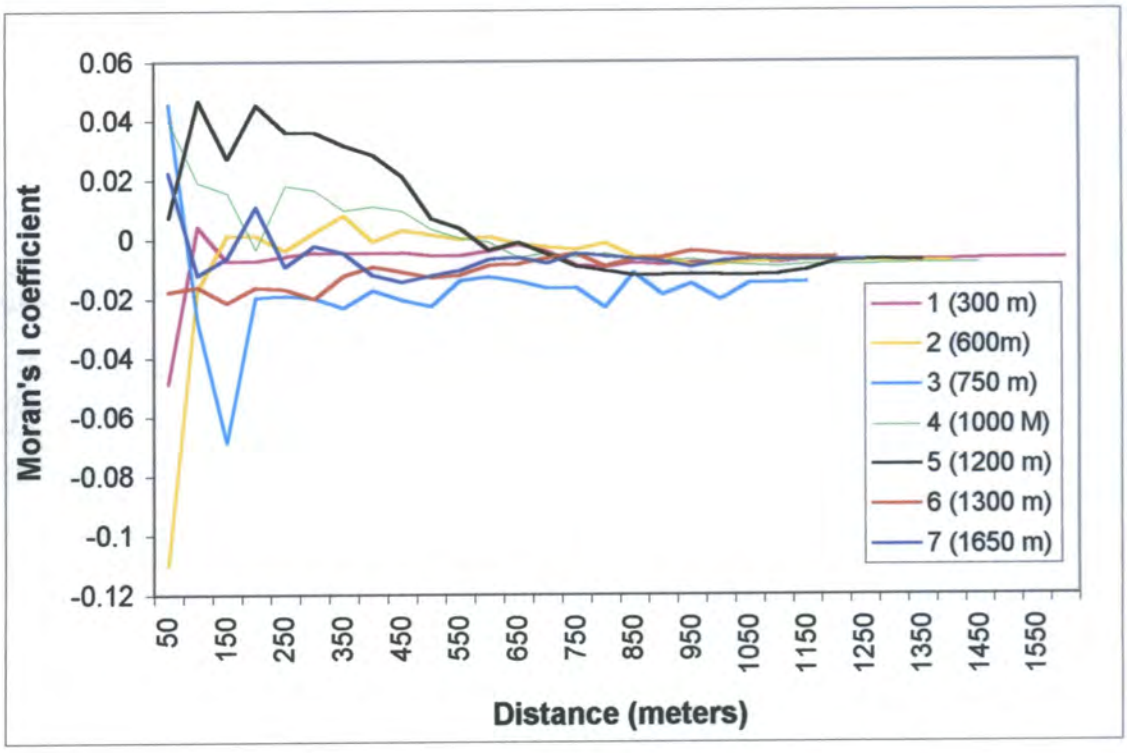


Figure 5.19. Correlogram of Moran's I for the residuals in Model 1

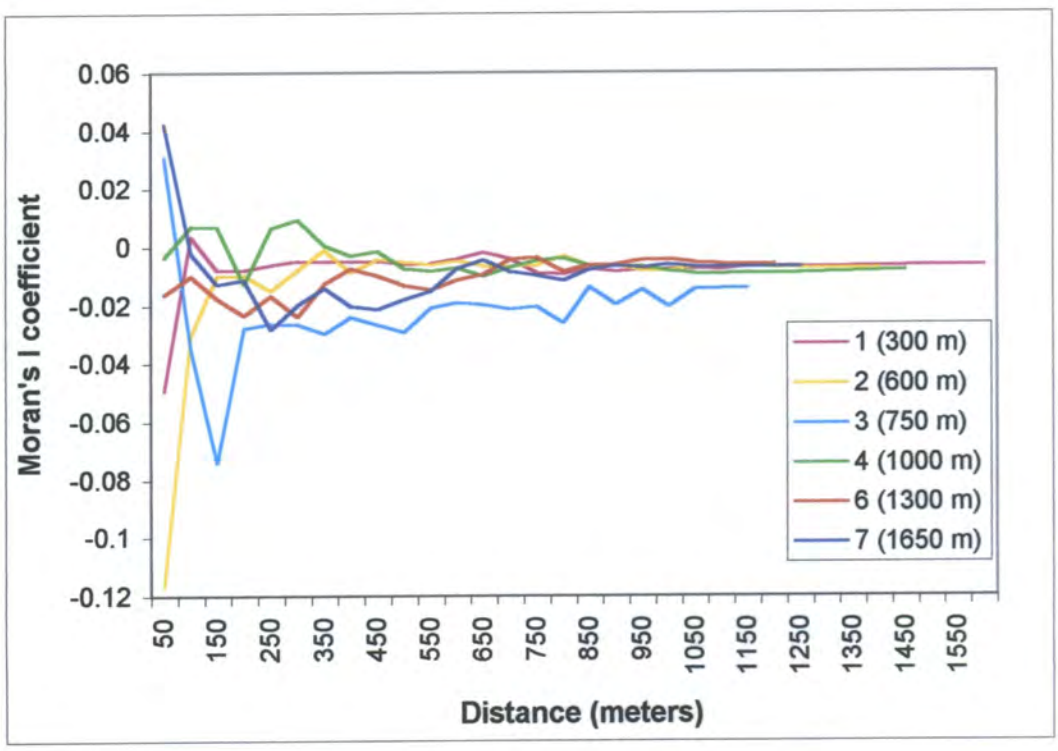


Figure 5.20. Correlogram of Moran's I for the residuals in Model 2

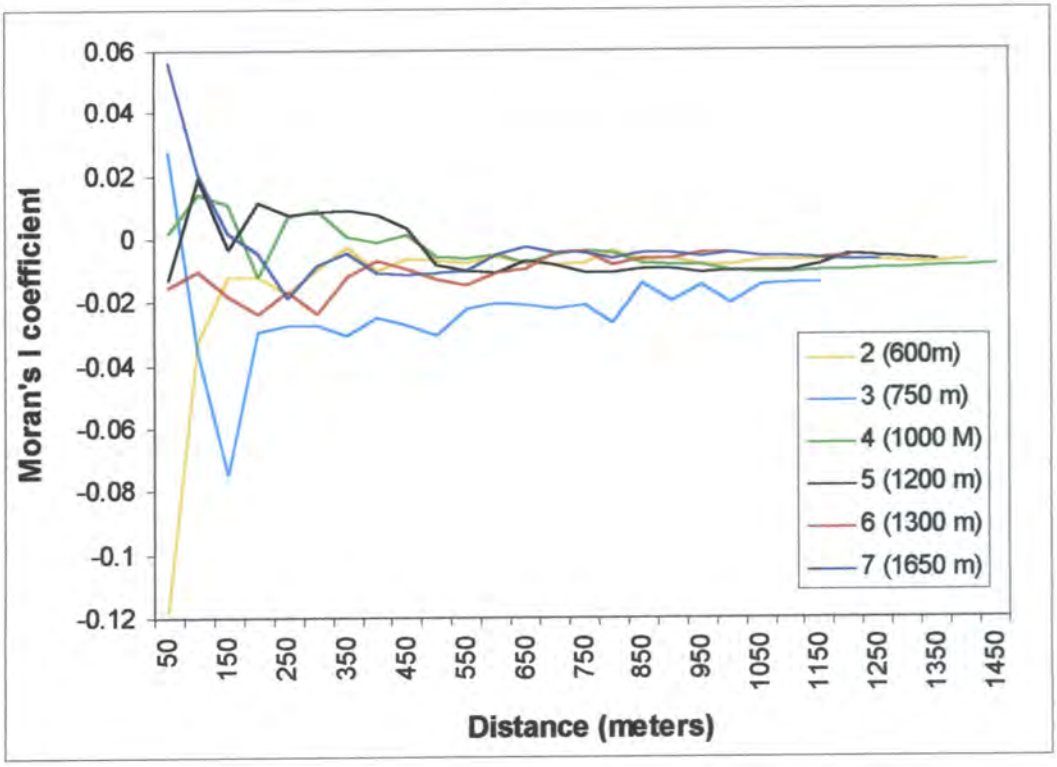


Figure 5.16 illustrates for each village how spatial autocorrelation declines with increasing distance at different altitudes. Positive spatial autocorrelation is significant across all distances for villages situated at 750 m, 1000 m, 1200 m and 1300 m. At 300 m and 600 m significant spatial autocorrelation is present up to 1000 m, whereas at 1650 m significant spatial autocorrelation is present up to 550 m. In Figure 5.17, for swampiness within 400 m of each household, again spatial autocorrelation decreases with increasing distance. Significant positive spatial autocorrelation is evident over all distances for villages at 300 m, 600 m, 1300 m and 1650 m. At 1000 m and 1200 m spatial autocorrelation is significant up to 650 m and 750 m distance respectively. In the correlogram investigating whether spatial autocorrelation is present for the prevalence of malaria in each household according to enlarged spleen (Figure 5.18), there is no general pattern. Positive spatial autocorrelation is not significant in any of the villages except between 300-400 m for the village located at 1000 m, and between 300-450 m at 1200 m elevation. Examining spatial autocorrelation for the residuals in Models 1 and 2 (Figures 5.19 and 5.20 respectively), again, there appears to be no general pattern, but spatial autocorrelation is not significant in any of the villages in either model.

5.5 Cluster analysis

The presence of spatial disease clustering in each of the survey villages was investigated. This analysis was carried out using either spleen rate or parasite rate to determine malaria cases or controls for three age categories. Table 5.9 summarises these results giving the significance level obtained by the most likely cluster. Villages without disease clusters have no significance level entered in the table.

Table 5.9 Significant clusters identified by the SaTScan software (Kulldorff *et al.*, 1998).

Village	Spleen rate			Parasite rate		
	All ages	< 10 years	< 5 years	All ages	< 10 years	< 5 years
1 - 300 m	$P < 0.05$			$P < 0.05$		
2 - 600 m	$P < 0.01$	$P < 0.05$		$P < 0.01$		
3 - 750 m					$P < 0.05$	
4 - 1000 m	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.01$
5 - 1200 m	$P < 0.001$	$P < 0.01$		$P < 0.01$	$P < 0.05$	
6 - 1300 m	$P < 0.01$	$P < 0.05$	$P < 0.05$			
7 - 1650 m						

Significant clusters are more evident for spleen rates than parasite rates. Village 4 at 1000 m has significant clusters in all age categories for both spleen and parasite rates, but there were no clusters identified in villages at 750 m and 1650 m. The most reliable criteria for identifying clusters in this investigation are spleen rates for all ages. However, the most likely clusters identified using these criteria for the villages at 300 m and 1300 m included only 2 and 3 houses respectively. Figures 5.21, 5.22 and 5.23 display the clusters identified by SaTScan (Kulldorff *et al.*, 1998) for the villages at 600 m, 1000 m and 1200 m respectively, together with the *Swamp* grid that illustrates the location of flat areas where water accumulates. Table 5.10 summarises the characteristics of these clusters.

The cluster detected in the village at 600 m includes only 13 houses within a radius of 111.32 m, whereas in the villages at 1000 m and 1200m there are more houses and a bigger cluster radius. The significance level for these clusters was also greater at $P < 0.001$, compared to $P < 0.01$ at 600 m.

Figure 5.21 Malaria cluster in the village at 600 m identified by SaTScan

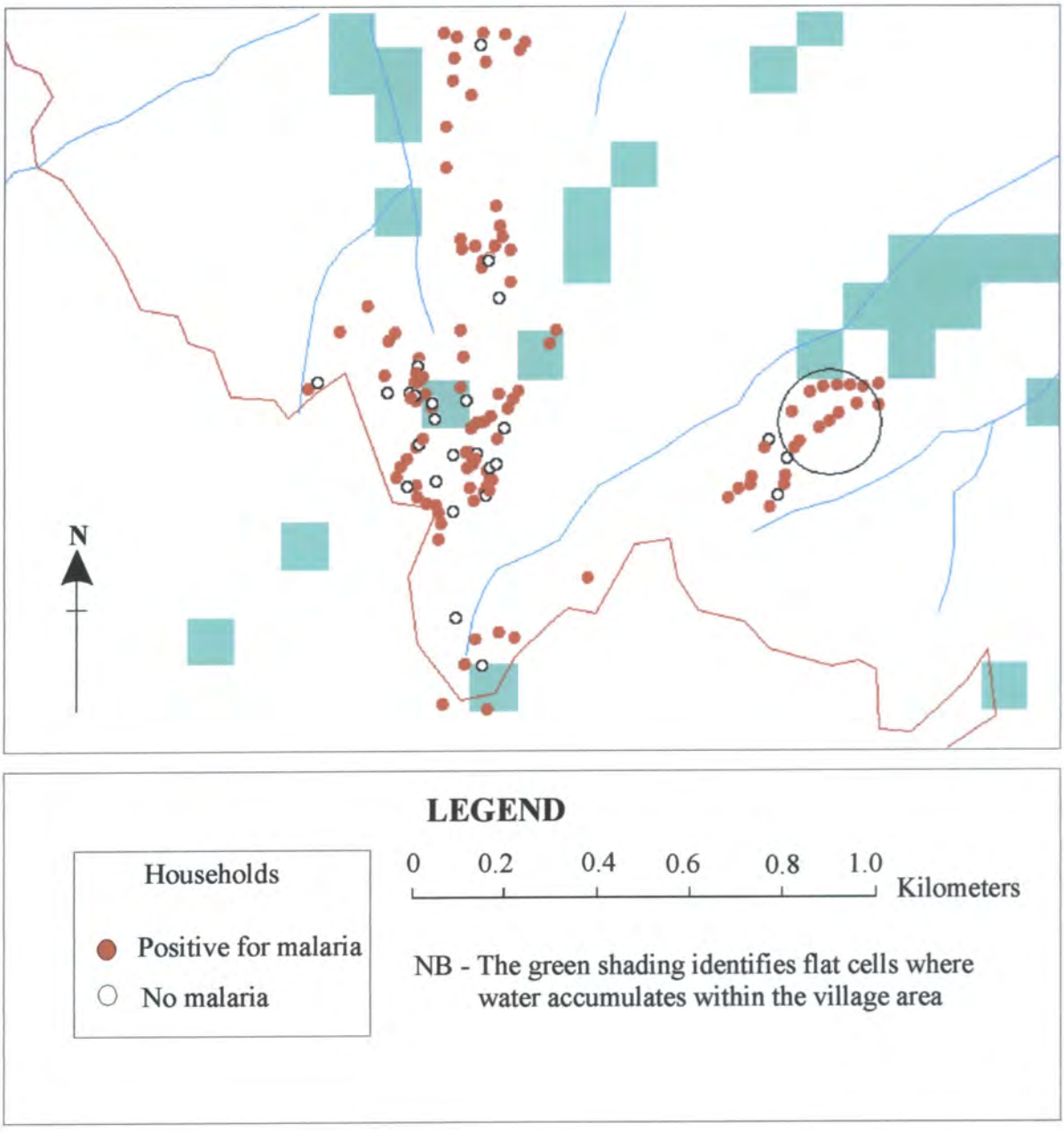


Figure 5.22 Malaria cluster in the village at 1000 m identified by SaTScan

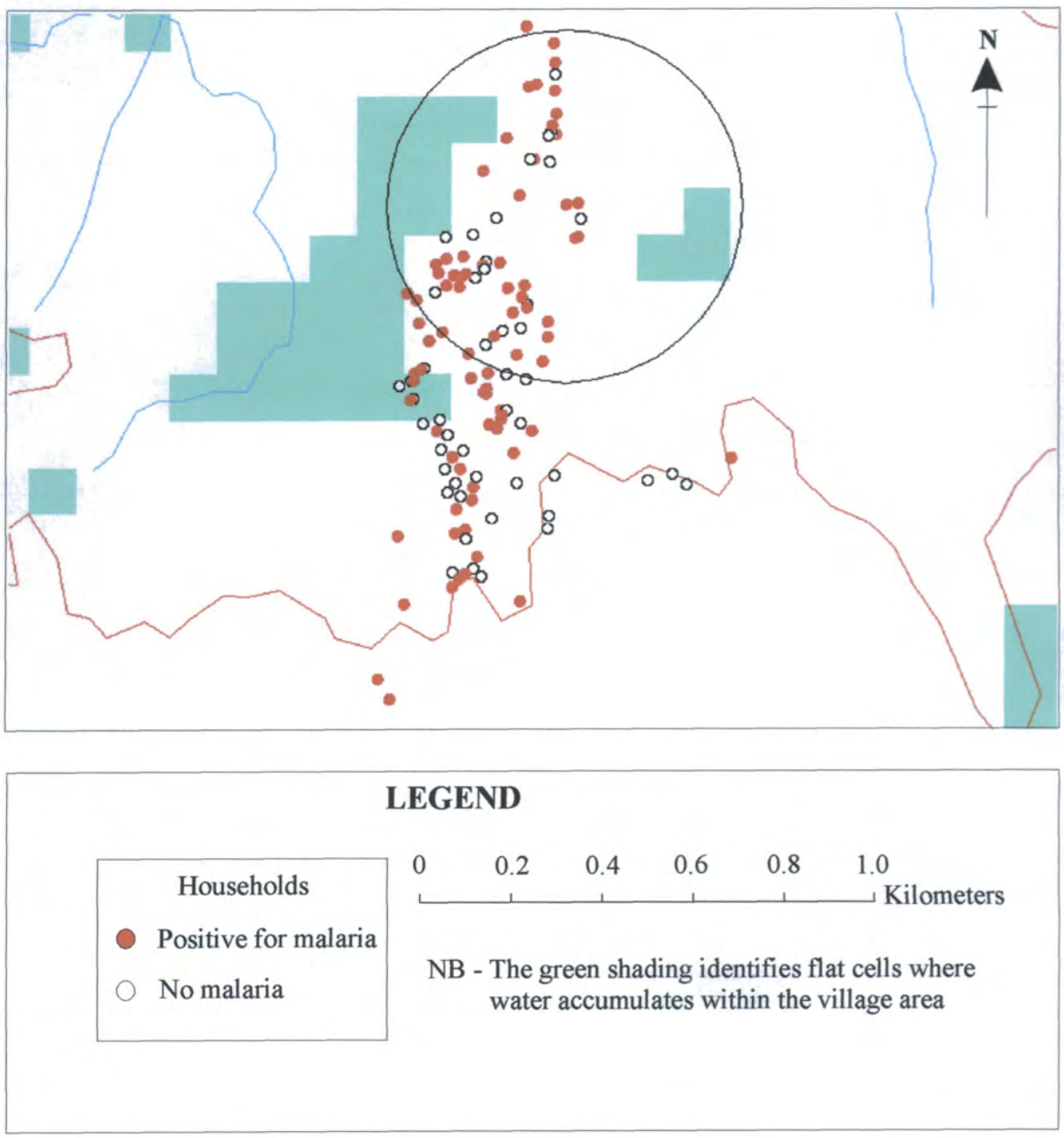


Figure 5.23 Malaria cluster in the village at 1200 m identified by SaTScan

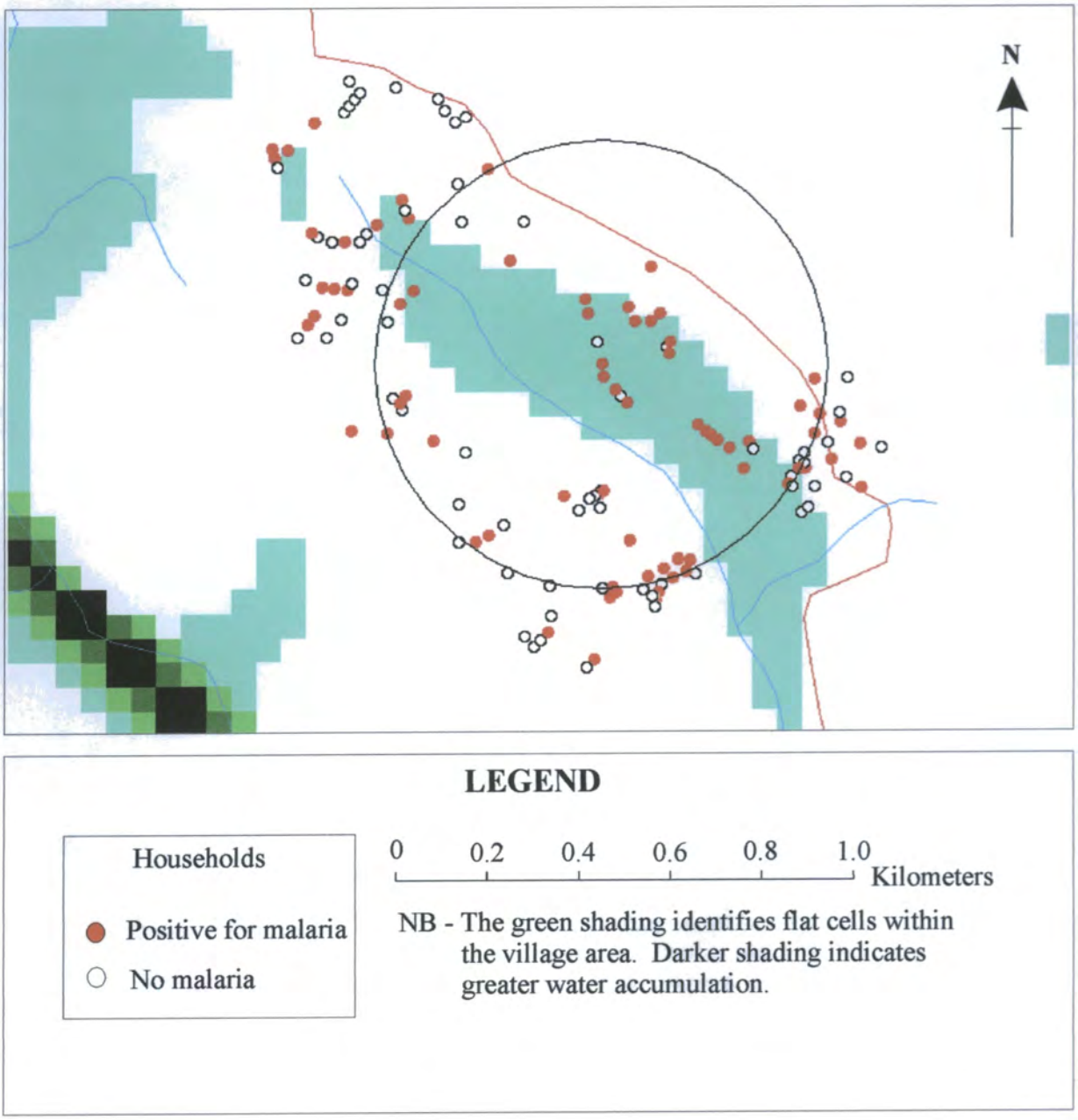


Table 5.10. Features of significant clusters for villages at 600 m, 1000 m and 1200 m identified by SaTScan (Kulldorff *et al.*, 1998)

Village	600 m	1000 m	1200 m
Houses in census	140	128	141
Village population	702	612	788
Total cases in village	354	168	124
Cluster radius	111.32 m	382.26 m	456.95 m
Houses in cluster	13	59	67
Number of cases:			
Observed	49	111	90
Expected	32.78	75.22	60.58
P-value	0.01	0.001	0.001

Cases and controls were classified using spleen rate for all ages.

Univariate analysis was carried out to compare the houses within the clusters with those outside the clusters, using the Mann-Whitney *U* test to see if there were any significant differences in the environmental variables. These results are shown in Table 5.11 below.

There was a significant ($P < 0.001$) difference in altitude comparing the cluster and non-cluster houses in all three villages; cluster houses were located at a lower median altitude than non-cluster houses. There was less distinction in slope values, clusters houses in villages at 600 m and 1200 m were situated on slightly flatter terrain than non-cluster houses, but at 1000 m the opposite was true. Flow accumulation values were significantly higher at 250 m and 500 m radii for the cluster houses. In the village at 1000 m at closer radii (0 – 250 m) there were no flat or swampy cells present, at 300 m – 400 m radii there were more swamp and flat cells near non-cluster houses, whereas at 500 m the opposite was true. In villages at 600 m and 1200 m there were more flat and swamp cells at all radii for the cluster houses compared to non-cluster houses, these differences were particularly significant ($P < 0.001$) at 1200 m. Figure 5.24 displays this phenomenon well; a flat area where water accumulates is situated in the middle of this village. However, at 600 m and 1000 m (Figures 5.22 and 5.23 respectively), there is little evidence of flat areas where water accumulates, near or within the clusters

Table 5.11. Comparing the geographical variables for the cluster and non-cluster houses.

	600 m			1000 m			1200 m		
	median			median			median		
	cluster	non cluster	Sig (2 tailed)	cluster	non cluster	Sig (2 tailed)	cluster	non cluster	Sig (2 tailed)
Altitude(m)	480	600	< 0.001	1013	1046	< 0.001	1186	1208	< 0.001
Gradient	7	11	< 0.001	12	10	0.001	7	9	0.029
Flowacc	0	1	0.002	2	16	< 0.001	1	1	0.782
Flow50	1	5	0.050	12	80	< 0.001	16	13	0.050
Flow100	68	17	0.029	35	198	0.014	55	68	0.007
Flow250	1071	313	< 0.001	2304	2604	0.892	1658	546	< 0.001
Flow500	3145	2056	< 0.001	8516	8005	0.009	5653	3542	0.005
Flat50	0	0	0.947	0	0	0.005	1	0	< 0.001
Flat100	1	0	0.466	0	0	0.189	4	1	0.001
Flat250	11	5	0.001	7	5	0.002	30.5	21	< 0.001
Flat500	34	21	< 0.001	73	41	< 0.001	96	82	< 0.001
Swamp			1.0	0	0	0.021	0	0	< 0.001
Swamp100	0	0	0.148	0	0	0.020	51	0	< 0.001
Swamp200	16	2	0.004	0	0	0.020	518.5	33	< 0.001
Swamp300	117	2	< 0.001	123	406	0.018	1526	305	< 0.001
Swamp400	204	2	< 0.001	583	881	0.025	2317	746	< 0.001
Swamp500	320	2	< 0.001	1411	1397	0.240	3106.5	1667	< 0.001

6 Discussion

6.1 Prevalence and altitude

6.1.1 Malaria biology

Altitude is by far the most important topographic variable influencing malaria. Highland areas surrounded by holoendemic lowlands have been and still are regarded as a refuge from the heat, humidity and threat of diseases at lower altitudes, and have been inhabited widely as a result. The Usambara Mountains for example are a densely populated region in north-eastern Tanzania where the current population averages over 300 people/km^2 (Adams *et al.*, 1996). The reason why altitude plays such an important role in determining malaria infection is due to its effect on temperature. In free atmosphere the average temperature decrease with height is 0.6°C per 100 m (Adams *et al.*, 1996). In the Usambara's there was a progressive decrease in mean temperature with increasing altitude of 0.7°C per 100 m in the warm season and 0.6°C per 100 m in the cold season (Bødker, 2000). For the seven villages investigated in this study the annual mean temperature varied between 24.7°C in the lowest village at 300 m and 15.7°C at 1650 m. However, there was a large seasonal variation in all villages. In January, the peak of the warm season, mean temperatures varied between 5.3°C and 7.1°C higher than during the cold season in July (Bødker, 2000). Indoor temperatures were also recorded in each village and were 2.6°C higher than atmospheric temperatures (Bødker, 2000).

Temperature plays a vital role in the biology of malaria, affecting the development rates and survival of the poikliothermic mosquito and parasite. The optimum temperature for the development of *P. vivax* and *P. falciparum* in the *Anopheles,* and hence the transmission of infection is about 28°C. Below 15°C the development of *Plasmodia* in the *Anopheles* does not occur (Gilles & Warrell, 1993). Likewise, the reproductive potential of the mosquito is enhanced greatly with increasing temperature. The female mosquito takes a blood meal every 2-5 days, and a small rise in temperature can make a big difference, as demonstrated by *An. gambiae* and *An. funestus* that feed every 2 days at 25°C, whereas at lower temperatures they will feed every 3 days (Gilles, 1953). Mosquito larval development time is also dependent on temperature. The cycle from

egg to adult may take 20 days at 20°C, whereas it only takes about 7 days at 31°C, although this cycle does vary between species (Gilles & Warrell, 1993). The most crucial factor governing the intensity of transmission however, is the survival of the adult *Anopheles*. This also varies between species, but is chiefly influenced by external factors, particularly temperature, humidity and predation. Under favourable environmental conditions the female *Anopheles* may survive up to a month or longer, but most species live between 10-14 days (Knell, 1991). The female mosquito has no chance of transmitting malaria unless she lives through the time taken for the sporongonic development of the malaria parasite, which may take between 8 and 25 days depending on temperature and parasite species (Gilles & Warrell, 1993).

In 1996 Bødker (2000) found that *An. gambiae* accounted for 93% of the malaria vectors caught in light traps in the Usambara's, with the remaining 7% being *An. funestus*. The vector populations declined with increasing altitude with the exception being at 1000 m, where there was a proportionally high population of *An. funestus*. This was probably attributed to the presence of a small stream running through the middle of the village, providing suitable breeding habitat in the dense vegetation along its edges. In the lowest village the accumulated annual man biting rate (MBR) was 4230, compared with only 2 at 1650 m, and linear regression of the $\log_{10}$ annual MBR for each village suggests the rate dropped by 50% for every 126 m increase in altitude (Bødker, 2000). Larval development times were estimated to be about 10 days at 300 m and 25 days at 1650 m in the warm season. In the cold season larval development times displayed a marked difference, taking approximately 18 days at 300 m compared with over 100 days at 1650 m (Bødker, 2000). Due to the low numbers of mosquitoes caught in the spray catches it was not possible to calculate the duration of the gonotrophic cycle. However the theoretical duration of the gonotrophic cycle within the temperature limits found in the villages, assuming the mosquitoes are resting at indoor temperatures 2.6°C higher than atmospheric temperatures differed greatly, from at least 2 days in the lowlands compared to 3 days in the highlands in the warm season. In the cold season this varied from 3 days at 300 m to 7 days at 1650 m, illustrating the dramatic effect small temperature changes have on the gonotrophic cycle. The estimated length of sporogony varied greatly along the altitudinal transect and ranged from less than 8 days in the lowlands in the warm season to not at all in the highlands during the cold season (Bødker, 2000).

6.1.2 Malaria prevalence in the Usambara Mountains

This study was concerned with predicting the risk of malaria infection in each of the seven villages in relation to the topography, namely elevation, flat areas and water inundation. The unit of investigation in this case was every household involved in the malaria survey conducted by Bødker (2000), rather than each individual. Therefore, for this research malaria prevalence was measured in each village by household rather than by individual, which is the conventional way of calculating disease prevalence. Hence, if one individual in a household is positive for malaria (due to enlarged spleen and/or with parasites in the blood) the household was considered positive for malaria, and so the spleen rates and parasite rates for each village refer to the proportion of positive houses in each village.

Prevalence rates generally decreased with rising elevation. This trend was more pronounced with regard to spleen rates, with an initial rise followed by a progressive decrease from 600 m to 1650 m. Spleen rates are a cruder measure of endemic malaria compared with parasite rates, but Bødker (2000) has shown that they are a reliable measure of malaria prevalence in the West Usambara's. Splenic enlargement is a characteristic feature of malaria infections in endemic areas, and a good indicator of repeated attacks and prolonged exposure to malaria parasites, whereas the parasite rate conveys the intensity of transmission at the time of the survey (Gilles & Warrell, 1993). The parasite rates were high in all age groups from 300 m to 1300 m, dropping considerably at 1650 m. An epidemic occurred in the village located at 1200 m just before the survey, explaining the highest parasite rate along the altitudinal transect (87% for the population as a whole and 67% for children under 5 years). The spleen rates for this village were considerably lower (52% for the population as a whole and 32% for children under 5 years), indicating that this village is on the fringe of transmission. At this altitude the population are not exposed to malaria infection enough to acquire protective immunity from the disease. When favourable environmental conditions arise such as raised temperatures, malaria epidemics may occur, as could have been the case for this village. Both parasite rates and spleen rates were higher for the population as a whole compared to children under 10 and 5 years, indicating that the population is immune naïve, adults being as susceptible to disease as children. The temperature range in this village is only just conducive to mosquito and parasite development and survival rates. The average temperature was about 18°C, with the minimum and maximum mean temperatures being around 15°C and 22°C respectively.

Parasite rates were also high at 1000 m, especially for all ages. In this village there were a higher proportion of *An. funestus* present compared to *An. gambiae*. *An funestus* prefer permanent water for breeding in, especially with vegetation along the edges, creating shaded habitats. Due to the nature of their breeding places its season is often different from *An. gambiae* which thrives in temporary unshaded pools. Thus, *An. gambiae* tend to be more numerous during and directly after the rainy season. The presence of *An. funestus* at 1000 m exposes the population to a more regular and constant risk of infection, resulting in the high parasite rate found during the survey, but the lower spleen rate implies that transmission is seasonal, reliant on temperatures that are more favourable. The mean annual temperature at 1000 m was 19.4°C (mean minimum 16°C and maximum 23.7°C), which are within the limits of malaria transmission, but the low mean minimum temperature will limit mosquito survival and sporogony so that malaria transmission is likely to take place when temperatures become warmer.

At 300 m, parasite rates were lower in the whole survey population compared to villages situated between 600 m and 1300 m, despite an entomological inoculation rate (EIR) of 91 infective bites per year per person compared with EIRs of 9.7 and 0.08 at 600 m and 1300 m respectively (Bødker, 2000). The lower parasite rate was probably a result of high immunity at the lowland village. However, in this village the mosquito vector will be a very noticeable pest and so the lower parasite rate (and spleen rate compared to the village at 600 m) could be also be a consequence of the population using various protective measures such as bed-nets, insecticides and chemoprophylaxis. In such holoendemic areas the population will have acquired immunity to malaria supported by the fact that children under 5 years have a higher parasite and spleen rate compared to children under 10 years and for the population as a whole.

Spleen rates were highest in the village located at 600 m, where annual temperatures range between 19.6°C and 27.4°C, favourable for mosquito and parasite development all year round. However, the annual MBR and EIR were considerably lower at 716 and 9.7 (Bødker, 2000) respectively compared to the lowland village. There was less disparity between the spleen and parasite rates for this altitude; both rates were the same for the population as a whole, at 81%. The high parasite and spleen rates in this village could be a result of infection contracted in the lowlands and the accidental

transportation of mosquitoes from lower elevations. The decline in mosquito abundance suggests that mosquitoes pose less of a nuisance and therefore the population may not employ protective and control methods as extensively as in the lowland village, plus their immunity status will be lower. Similar reasons could explain the high parasite and spleen rates at 750 m, especially considering the lower annual MBR of 276 and EIR of 1.7, however, temperatures at this altitude are still favourable for malaria transmission throughout the year, 18.9°C being the minimum mean temperature and 26.6°C the maximum (Bødker, 2000). At 1300 m the parasite rates were still surprisingly high for the population as a whole compared to spleen rates (77% and 19 % respectively), but for children under 10 and under 5 years both rates were lower, considerably so for spleen rates signifying that this area is hypoendemic. The high parasite rate in the whole survey population infers that adult individuals are acquiring malaria away from the village, possibly as a consequence of visiting and working in the holoendemic lowlands. Temperatures in this village range from 15.2°C to 21.8°C and although it is possible for the transmission of malaria very few mosquitoes were found with the annual MBR being 9 and EIR just 0.08 (Bødker, 2000), malaria transmission is therefore likely to be highly seasonal. In the highest village at 1650 m spleen rates were very low, only 8% for the population altogether and 6% for children under 10. Parasite rates were lowest in this village, but still high in the population as a whole at 44% considering that the annual MBR was only 2, with 0.03 infective bites expected for each individual annually (Bødker, 2000). In children under 5 years the parasite rate was only 19%, suggesting that adults may be moving to areas of high endemicity and contracting malaria.

According to the classification of endemicity adopted by the WHO (1951) based on spleen rates, villages at 300m and 600 m are holoendemic and are thus areas of highest endemicity. The village at 750 m is hyperendemic, malaria becoming less endemic and more seasonal. Mesoendemic villages are located at 1000 m and 1200 m, where transmission is generally for a short time following on from the rainy season. In mesoendemic areas the development of immunity is delayed so adults as well as children display clinical symptoms. Such areas are on the boundaries of malaria transmission, and seasonal epidemics may occur, as with the village at 1200 m. The two highest villages located at 1300 m and 1650 m are classified as hypoendemic. At these altitudes malaria transmission is normally absent but may occur following extremely favourable environmental conditions. The immunity of the population in

such areas is practically non-existent, and if transmission takes place it may rapidly achieve epidemic status with many deaths occurring among all age groups within the population. The Usambara's do experience cooler temperatures compared to other highland regions of East Africa at similar altitudes and so these classifications of endemicity cannot be extrapolated to other African highland areas (Kenworthy, 1966).

Although temperature and hence altitude is the main abiotic factor governing malaria transmission in the highlands, the presence of mosquito breeding sites is also a limiting factor. All villages in the survey received between 1000 mm and 1300 mm of rainfall in 1996, except for the village located at 1000 m, which received 1920 mm (Bødker, 2000). The proportionally higher abundance of *An. funestus* could be explained by higher precipitation levels at this altitude, resulting in the permanent stream running through the village. Most rainfall is received between April and May with peak transmission occurring after these long rains between May and June in the lowland villages. Although peak transmission occurs during the cold season, vector numbers are at their highest and temperatures, although not optimal, still allow sporogony. In the two highland villages transmission peaks in January when temperatures are at their highest despite low vector densities at this time (Bødker, 2000). The cross sectional malaria survey was carried out between November and December, at the end of the transmission season in the lowlands, and so prevalence rates measured in the two highland villages may represent old infections.

Although altitude is a strong indicator of malaria prevalence rates there are many other factors that are important for transmission that are generally not as easy to quantify. The location of a village in a highland area will have some influence. A village situated in a valley will be sheltered, but with cold air accumulating at night, the nearby upper slopes will be warmer. However, at higher elevations conditions become harsher for the survival of the mosquito vector. Insolation levels are influenced by aspect; with shaded slopes being cooler and wetter than sunny slopes, and so villages at the same altitude, but situated at different aspects may have their own distinctive microclimate. In addition, villages situated on steeper slopes will have fewer flat areas that may provide suitable breeding habitats if inundated with water compared to villages situated in valley bottoms, often with accompanying stream networks. Rainfall distribution is another crucial parameter that is affected by altitude and aspect. Rainfall affects the suitability of breeding sites and hence is responsible for the marked seasonal cycles in

mosquitoes and malaria transmission. Land cover and land use also have an effect on vector densities. In some cases humans altering their environment generally for agricultural purposes can significantly increase the number of available mosquito breeding sites and hence the risk of malaria. The construction of drainage channels and local dams for irrigation purposes has been shown to increase malaria prevalence (Mutuwatte *et al.*, 1997; Ghebreyesus *et al.*, 1999), likewise has the reclamation of natural swamp for agriculture (Lindblade *et al.*, 2000). Rice fields are also major breeding habitats for mosquitoes (Wood *et al.*, 1992; Gbakima, 1994; Lindsay *et al.*, 1995; Mutuwatte *et al.*, 1997). Deforestation can affect malaria transmission in a variety of ways. On a large scale deforestation has been blamed for increasing local temperatures (Hamilton & Macfadyen, 1989) and holes left in the ground from uprooting trees provide potential unshaded breeding sites for *Anopheles* following rainfall. Various social, economic and behavioural factors are also important, such as house construction, the presence of livestock, and the use of bednets, insecticides and antimalarial drugs. Parasite resistance to the valuable antimalarial chloroquine has been blamed on the misuse of the drug and consequently the resurgence and emergence of malaria in many vulnerable areas (Bødker *et al.*, 2000). Another aspect relates to the movement of people as either carriers of the parasite or by transporting the infected vector from the highly endemic lowlands to the highlands. The highlands are also under pressure from burgeoning human populations. As people migrate to the highlands they may bring their parasites with them, and in establishing new settlements will create conditions suitable for malaria transmission.

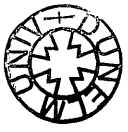
6.2 Validation of models

6.2.1 Logisitic regression analysis between villages

The two logistic regression models developed calculating malaria risk throughout the study villages as displayed in Figures 5.10 and 5.11 are based on spleen rates rather than parasite rates. Model 1 incorporates only altitude in the logistic regression equation whereas Model 2 uses altitude and the amount of 'swampiness' within 400 m of each household to predict malaria prevalence. Both models show a linear decrease in risk with rising elevation (Figures 5.12 and 5.13), but the predicted risk values in Model 2 are higher for comparable altitudes in Model 1 and cover a greater range from 1000 m upwards. The 'swampiness' variable increases malaria risk for houses situated on or

near swampy ground, resulting in higher predicted prevalence rates for those houses compared to others further away on steeper and drier ground. Hence, the greater range in risk values compared to Model 1. This is especially evident for those villages situated at 1200 m, 1300 m and 1650 m, highlighting the importance of such areas to malaria transmission at higher altitudes if favourable conditions arise.

The predicted risk values in Model 2 closely reflect the range of spleen rates for all ages during the 1996 malaria survey in each village, except for the village at 300 m, where the predicted values were far greater than those found (67% of households had at least one member with an enlarged spleen, whereas Model 2 predicted that between 90-100% would be positive). Model 1 also predicts risk in this village to be higher (between 80-90%) than prevalence rates found in any age category. If all the physical, social, economic and behavioural attributes of each village were identical, the only difference between villages being altitude then the lowest village would be expected to display the highest prevalence rates. As it is the lowest village has the highest abundance of mosquitoes and consequently the inhabitants will probably have acquired immunity and may be using bed nets, insecticides and antimalarials to combat the nuisance and threat, thus, resulting in lower prevalence rates than those encountered in villages up to 750 m. The higher predicted risk values in Model 2 compared to Model 1 take into account that the houses in this village are located on predominantly flat terrain which is likely to experience water inundation after rainfall. In the villages situated at 600 m and 750 m, the predicted risk values for Model 2 are greater than for Model 1; 0.801-0.9 and 0.701-0.8 respectively for Model 2, and 0.701-0.8 and 0.601-0.7 for Model 1, with the recorded prevalence rates being 0.81 and 0.78 respectively for the population altogether. The proximity of these villages to the flat lowlands has influenced prevalence risk rates explaining the higher values for Model 2 even though the villages themselves are situated on hilly ground, especially the village at 750m where there are very few flat areas at all. At 1000 m 63% of households had at least one member with an enlarged spleen, which is higher than the predicted values in Model 1 (40-50%), whereas Model 2 predicts a greater range of prevalence values (40-70%). In the centre of this village there is a potential swampy area surrounding the permanent stream. This area is irrigated and used for the cultivation of food crops and could explain the higher proportion of *An. funestus*. Prevalence risk is higher in this area (between 60-70%) compared to the village fringes where the terrain is steeper. The epidemic village situated at 1200 m is located in a flat valley where the level terrain is utilised for the



cultivation of vegetables, taking advantage of water flowing through the area to irrigate the crops. This is a prime area for potential mosquito breeding sites. The houses are situated very close to this cultivated area and are at obvious risk from malaria when favourable conditions permit, as could have been the case with the recent epidemic. Model 1 predicts that malaria risk in this area is between 0.301 and 0.40, whereas spleen rates for the population as a whole were 0.52. Model 2, allowing for the 'swampiness' of the landscape predicts malaria risk to be between 0.301 and 0.7, the central swampy area being at greater risk (between 0.601 and 0.7) with risk decreasing towards the outskirts of the village where the houses are situated on higher, steeper ground. The two figures displaying these models illustrate their differences well; risk being consistent for the village area and across different altitudinal bands in Model 1 compared to the greater variability in risk for Model 2, where pockets of risk are highlighted at various altitudes. The same comparison is evident when viewing the villages located at 1300 m and 1650 m. The red 'hot spots' for potential risk are very conspicuous in Model 2. In these villages spleen rates were much lower at 0.19 at 1300 m and 0.08 at 1650 m. Model 1 predicted these villages to have higher prevalence rates than those found, whereas Model 2 again predicted a greater range in prevalence rates for these two villages, taking into consideration the proximity of flat areas where water may accumulate. However, the potential risk as calculated in Model 2 did not greatly increase for either village since they were out of range of major 'hot spots'.

Considering altitude as the only influential topographic variable underestimates prevalence rates, whereas the combination of altitude and the amount of 'swampiness' within 400 m of each household gives a better prediction of prevalence at the various altitude bands. The 'swampiness' factor allows for the variation in terrain and water accumulation across the landscapes, and adjusts prevalence rates in accordance with those factors. Thus, especially in the highland areas pockets of high risk are located within areas of low risk

The importance of altitude has already been discussed and for this reason and the fact that univariate analysis indicated that there were significant differences between malaria positive and negative households it was the only independent variable used in the logistic regression analysis for Models 1 and 3. The differences between these 2 models were the age categories and number of villages used for analysis. Model 1 incorporated all the 811 households included in the survey, whereas Model 3 included 418

households with children under 5 years of age, except for the village at 1200 m. Model 3 predicted a higher percentage of cases and controls, with an overall predictive ability of 78% compared with 73% for Model 1. The village at 1200 m was not included in Model 3 because it substantially reduced the predictive ability of the model, probably because of the higher prevalence rates than would be expected at this altitude. However, Model 1 included the whole survey population and nearly twice as many households as Model 3, and thus was considered the most reliable and robust model, portraying the relationship between altitude and malaria prevalence well.

A variety of other topographic variables were also investigated together with altitude to see how they affected malaria risk. Altitude had to be included in all the models because of its overriding importance. Initially, the amount of flatness and the amount of water accumulating in the landscape were examined together and separately. The interest in these variables relates to their potential for producing suitable mosquito breeding sites; pooling may occur in flat areas, and areas of high water accumulation may also develop pools. The water accumulation factor calculates the amount of flow across a landscape. Water will flow downwards along the steepest gradient forming drainage channels along the way, and so the areas of highest flow will be drainage channels or stream networks; not the type of habitat suitable for *An. gambiae* breeding sites. However, the amount of flow accumulating within 100 m, 250 m and 500 m of each household was investigated using logistic regression analysis since the univariate analysis indicated that there were significant differences between positive and negative households. However, the models produced were unreliable. Univariate analysis found that there were only significant differences between case and control houses within 500 m of flat areas and so Model 4 incorporates altitude and the amount of flatness within 500 m of households with children under 5 years of age. This model predicted 81% of the cases and 82% of the controls, however only 346 households were included in the model; households in villages located at 300 m and 1200 m were not included.

Just because an area is flat does not mean that water will flow into it and accumulate. The 'swampiness' parameter combines the flat and flow accumulation attributes to identify flat areas where water accumulates as likely mosquito breeding habitats. The swamp variable was examined using univariate analysis at various radii from each household. Significant differences were found between case and control houses within 200 m, 300 m and 400 m radii from swampy areas. However, the amount of

swampiness within 400 m produced the best result in the logistic regression analysis and Model 2 was the product, incorporating altitude and the amount of 'swampiness' within 400 m of each household included in the survey, except for the lowland village at 300 m. This model correctly predicts 78% of the cases and 75% of the control houses. The village at 300 m was not included due to the lower prevalence rates found there compared to the next two villages along the transect.

6.2.2 Spleen rates and parasite rates

All four models included spleen rates as the malaria indicator. Parasite rates were investigated in logistic regression analysis, but the resulting risk models were not reliable for any of the topographic criteria. Earlier univariate analysis comparing the topographic variables for cases and controls using parasite rates for individual villages and for villages combined showed fewer significant differences compared with spleen rates. However, altitude was highly significant across all age groups for both spleen and parasite rates for all villages combined, with positive households being situated at lower altitudes than households without malaria, confirming the important influence altitude has on malaria prevalence. Spleen rates however, proved to be a better delineator of altitude, declining steadily with increasing elevation, whereas parasites rates remained high in the survey population up to 1300 m. This could be the result of the cross-sectional malaria survey being carried out at the end of the transmission season when parasite rates would be at their highest. Although they are a cruder measure of malaria prevalence, spleen rates do provide a more historic picture of malaria transmission, since an enlarged spleen signifies the accumulation of repeated and prolonged attacks over time. Parasite rates on the other hand are a measure of the intensity of malaria transmission at the time of the survey and are a snap-shot of malaria prevalence. In areas where transmission is seasonal, parasite rates will fluctuate, whereas spleen enlargement is an immune response to a malaria attack, growing over time. In highly endemic areas spleens tend to become enlarged in children, reducing in size after acquired immunity (Gilles & Warrell, 1993). For individuals living in holoendemic areas both parasite and spleen rates will be high, with increasing altitude both rates should decline steadily. However, individuals travelling to highly endemic areas from areas of low transmission may become infected and malaria parasites will be present in their blood, but their spleen may not be enlarged. This could partly explain why parasite rates were particularly high, compared to spleen rates in the two highest

villages. Spleen rates therefore provide a fuller picture and a more comparable measure of malaria prevalence in highland areas where transmission varies with altitude. Spleen palpation is also a quicker and more efficient than taking blood, and in the Usambara's Bødker (2000) has shown spleen rates to be a reliable method of assessing malaria prevalence in a population.

6.2.3 Investigations within villages

Although the focus of this research was to assess the effect topography has on malaria between villages in a highland region as a whole, the effect within villages was also investigated. Firstly, through univariate analysis comparing case and control households to see if there were any significant differences in their topographic variables. Only those villages located at 750 m, 1000 m and 1200 m, the mid-altitude and villages with unstable malaria were investigated further. Results in Figure 5.15 show the difference in predicted risk of an enlarged spleen between case and control households in Model 2 is more pronounced in these three villages, indicating that topography is important. The two lowland holoendemic villages suffer the highest malaria burden, conditions being favourable for year round transmission, that the topographic variables identified in this research (flatness, water accumulation and swampiness) do not play such an important part. In the two highland hypoendemic villages however, malaria is very rare and because there were so few cases the univariate analysis did not find any significant differences between case and control houses and their topographic attributes. It seems that villages in the mid-altitude range where malaria transmission is more seasonal topographic characteristics become more relevant. These three villages were then investigated further via logistic regression analysis to see if models could be developed predicting risk in these villages. Models 5, 6 and 7 were the result of that investigation. Again, spleen rates were used as the malaria indicator due to its reliability and for comparable purposes. Model 6 was the best predictive model for the village at 1000 m and incorporated altitude and the amount of flatness within 250 m of each household with one or more member under 5 years old. Only altitude was included in Model 5 for the village at 750 m and proved to be a reasonable predictor of cases more than controls for households with at least one child under 5 years. The least reliable was Model 7 for the village at 1200 m, where altitude and the amount of flatness within 500 m were included for each household with at least one member under the age of 10. Whether or not these models could be used as a basis to predict malaria within villages at such altitudes would require further investigation.

No one village is the same and so such models probably only reflect the village it is modelled on. In these three models the whole survey population were not included. Households with children under 5 or 10 years old provided the better models, possibly because they are more likely to sleep in the household they are associated with in the survey, whereas the adults may admit to sleeping in one household, but in fact sleep somewhere else. Also, the children are less likely to travel from the village compared to the adults and so their spleen rates may be a better representation of malaria prevalence for their village.

Malaria clusters were also examined within villages, but only three significant clusters emerged for villages located at 600 m, 1000 m and 1200 m. The houses within and outside the cluster were compared to see if there were differences in their topographic variables. The epidemic village at 1200 m and the village at 600 m displayed significant differences between cluster houses and non-cluster households for all the topographic variables. Households within the clusters were at lower altitude, situated on flatter terrain, located nearer areas of higher water accumulation, and were surrounded by more swampy areas. Figure 5.22 displays the 13 houses included in the cluster at 600 m overlaid with the swamp grid and stream networks. The cluster houses were located on a ridge, which is not depicted well in the figure, with 2 temporary streams running either side of the ridge and a swampy area below to the north-east. There were no clusters found in the main body of the village, and from Figure 5.22 it is clear that there are very few potential swamp areas within or near the village. However, in Figure 5.24, for the village at 1200 m the cluster radius was large (457 m) and encompasses 67 households. The cluster is positioned in the middle of a substantial swamp area with a stream running through the middle. This is the valley bottom where crops are cultivated, making use of the wetter soil conditions, and clearly is a potential site for mosquito breeding habitat. From this figure and these results it does appear that the topography in this village could influence malaria transmission. At 1000 m differences between cluster households and non-cluster households were mainly found for altitude, and flatness. Flow and swampiness only became significant within a radius of 500 m. There are 59 houses within the cluster as shown in Figure 5.23 that contain 111 individuals with enlarged spleen, the expected number of cases according to the scan statistic is 75. This cluster is highly significant, but within the cluster there are very few swampy areas, most of them are outside the cluster helping to account for swampiness becoming significant at a radius of 500 m.

6.3 Model limitations

6.3.1 Morbidity and mortality

More people in the world die and suffer from malaria than any other parasitic disease. This research has been concerned with examining the risk of infection in a highland area based on prevalence, using data collected on parasite and spleen rates. Prevalence rates indicate the intensity of transmission within a population at the time of the malaria survey, and provide a snap-shot of malaria infection. However, the relationship between prevalence and mortality and morbidity within a population may not be linear. Not everyone infected with the malaria parasite becomes ill or dies. In areas of stable endemicity repeated exposure to the parasite leads to the acquisition of specific immunity and so serious problems are restricted to young children, pregnant women and other non-immune individuals. Malaria in older individuals generally causes a mild febrile illness (Gilles & Warrell, 1993). With rising elevation malaria becomes more seasonal until it is rare. Communities where malaria transmission is unstable and exposure is low have little if any immunity to the disease. In these areas adults are as susceptible as children to the severe symptoms, and epidemics can occur when favourable climatic conditions arise. Morbidity and mortality rates are difficult to assess. The morbidity rate is generally based on the number of recorded admissions or attendances at hospitals and dispensaries. Obviously, this method will not deliver the true picture of morbidity in a population since many of the symptoms of malaria can be mistaken for other medical conditions (influenza, anaemia, diarrhoea etc.) and many of the indigenous people may live too far away or are too poor to seek medical help. Mortality is probably even more difficult to determine for similar reasons. The risk models produced in this research only indicate the risk of infection along the altitudinal transect and do not consider how the populations concerned are affected by their varying malaria burden. The village at 1200 m had experienced a recent epidemic; records for the number of morbidities and mortalities were not available.

6.3.2 Seasonality

As previously mentioned peak transmission may occur at different times in the highlands compared to the lowlands. The lowland peak is generally between May and June closely following the long rains in April and May. This is the coldest time of the year, but temperatures are still favourable for malaria transmission and the vector

population soars due to the increased availability of breeding sites. In the highlands however, peak transmission is thought to be in January when temperatures for the region are highest, thus allowing for the increase in survival of the mosquito and the likelihood that sporogony will take place and hence malaria transmission. Since the malaria survey was carried out between November and December 1996, at the end of the transmission season in the lowlands, but before the peak transmission in the highlands, prevalence rates will not be entirely comparable.

6.3.3 Measuring prevalence using household

The household was used to measure malaria prevalence rather than the individual because individuals in each household are highly spatially correlated and should not be used in regression analysis as they reduce the models reliability. However, by classifying a household as being positive for malaria if one or more member has an enlarged spleen and/or has parasites in their blood fails to take into account how infective each household is. For example, a household of 5 people is considered positive for malaria whether there is just one individual with enlarged spleen or parasites or the entire household presents such signs. The proportion of malaria in each household is not considered. Weighting the infectivity of each household and incorporating this factor into the logistic regression analysis may create a better or more reliable model.

6.3.4 Spatial autocorrelation

Spatial autocorrelation is a feature of geographical data where near things are related or are similar to each other. Spatial autocorrelation was measured for the variables *altitude* and *swamp400* included in Models 1 and 2 produced by logistic regression analysis. For a continuous variable like altitude it comes as no surprise that positive spatial autocorrelation exists. Elevation generally decreases or increases gradually or remains the same within short distances, except for rapid changes in height with the location cliff faces. Spatial autocorrelation was investigated for the position of each household within each of the seven villages. Significant positive spatial autocorrelation was found in each of the villages, which decreased with increasing distance. Likewise the amount of swampiness was also positively spatially autocorrelated, with flat areas where water is likely to accumulate being situated together rather than scattered randomly throughout the landscape and so the same spatial association was found with *swamp400*.

Spatial correlations complicate logistic regression modelling of the relationship between disease prevalence and environmental parameters. Failure to allow for spatial correlation results in overstating the significance of regression effects, leading to overestimation of the true degrees of freedom. There are tests available to account for the spatial correlation of data sets and recent epidemiological research has included such tests (Kitron *et al.*, 1996; Thomson *et al.*, 1999). However, a clear and consistent approach to significance testing for multivariate spatial processes has yet to emerge (Lennon *et al.*, 2000). Hence, there have been very few ecological studies that have included this aspect in their analysis. Although spatial autocorrelation was measured and found for the topographical parameters included in Models 1 and 2 it was not taken into consideration and so the significance levels of these variables will be reduced. However, spatial analysis of malaria prevalence for enlarged spleen for each household did not find spatial autocorrelation within any of the villages and significant spatial autocorrelation was not found in the residuals for Models 1 and 2 confirming that both models are reliable and robust (Cliff & Ord, 1981).

6.3.5 Limitations within the GIS

To map the location of the households in each village a GPS was used. The accuracy of each house position captured in this way is limited up to 100 m because of Selective Availability (SA) imposed on the technology by the United States Air Force. There were two tiers of accuracy implemented: authorised (military) and unauthorised (civilian), the military had receivers that are more accurate (Letham, 1995). To compensate for this loss of accuracy 300 fixes were taken for the position of each household mapped in this way. These fixes were then averaged giving a final position reading that should locate each household up to 25 m (August *et al.*, 1994). Therefore, although the households have been mapped with a reasonable degree of accuracy their actual position will be + or – 25 m. However, in 2000 SA was discontinued and so it is now possible to map geographical features with much greater accuracy.

The hydrological models produced in this research allowed the analysis of flow within each village catchment. However, ARC/Info only allows the modelling of surface run-off and therefore through-flow and base-flow are not accounted for. The soil properties of each village area are also missing from the models. Soil type plays an important role in determining whether water falling as rainfall is likely to become waterlogged creating pools or drain away. However, although the hydrological models produced may be

crude they do reveal where rainfall is likely to flow overland, delineating stream networks and where it will accumulate in flatter areas at the base of slopes.

6.4 Why is this research relevant?

Malaria continues to be one of the most serious risks to human health in the developing world, with 40% of the world's population at risk from this disease (WHO, 2001). Controlling malaria in sub-Saharan Africa where around 90% of the clinical cases and between 1-2 million deaths occur each year (WHO, 1997) is a daunting task. Malaria biology is complex. The *Plasmodium* and *Anopheles* thrive in a variety of habitats and under different environmental conditions depending on the species. Knowledge of the physical limits for the survival and development of the different parasite and vector species, the environments they prefer and their behavioural characteristics are essential to tackling the disease. A wealth of knowledge has been and continues to be gathered about malaria throughout the world through scientific research. Recent developments in computer technology have allowed complicated statistical analysis of data collected concerning malaria, enabling further understanding of the disease. Advances in GIS have provided a valuable tool to comprehend the spatial dimensions of malaria from local to global scales. Identifying environmental parameters allowing for the transmission of malaria and subsequently which communities, regions or nations are at risk will lead to more efficient and effective targeting of intervention methods. In sub-Saharan Africa where resources are scarce it is vital that the appropriate control measures (bed-nets, insecticides and chemoprophylaxis) and health services are aimed at the areas where they will do the most good.

There have been many attempts to map the global distribution of malaria throughout the world, showing its geographical limits and giving estimates of the number of people at risk. Although, wealthy and often more cooler nations in the developed world are not currently at risk from malaria or have eradicated it, there is growing concern that global warming will shift the current limits of malaria transmission. Most attempts at mapping the global distribution have therefore been concerned with assessing the changing pattern of malaria and the potential populations facing this risk with the advent of climate change (Martens *et al.*, 1995; Martens *et al.*, 1999; Rogers & Randolph, 2000). In sub-Saharan Africa the need for reliable, up-to-date malaria distribution maps at a variety of spatial scales is essential now. The MARA/ARMA collaboration has been

developed with this aim in mind (MARA/ARMA, 1998). Maps have already been developed on a continental level that show the risk of malaria transmission and its seasonality (Snow *et al.*, 1999b). Regional models have so far been produced for Kenya (Omumbo *et al.*, 1998), showing predicted levels of transmission intensity and prevalence rates. The distribution of the *An. gambiae* complex has also been mapped (Coatzee *et al.*, 2000). The African highlands and their populations are recognised as being particularly vulnerable to increased malaria transmission due to environmental change and over the past 50 years there has been a rise in the number of malaria epidemics occurring (Lindsay & Martens, 1998). The reasons for these epidemics have been blamed on climatic change, the movement of people, population increases, the decline of health services, growing resistance to antimalarial drugs and environmental degradation and it is probable that one or more, or all these factors may be responsible. Lindsay and Martens (1998) have reviewed the African highland malaria situation and forecast that future global warming is likely to increase the risk of epidemics occurring. A warming climate could push malaria transmission higher up leading to increased transmission in populations that have no acquired immunity.

The current research examining the role of topography in the Usambara Mountains with regard to malaria prevalence adds another dimension to our understanding of this disease. Bødker (2000) has shown that altitude is a good predictor of malaria prevalence in the West Usambara's as a whole and this is supported by the present study that models and predicts the risk of malaria in the seven villages using altitude in Model 1. The inclusion of the swampiness parameter with altitude in Model 2 improves the predictive potential, showing that above 1000 m there are pockets of malaria risk where the terrain is flat and where water is likely to accumulate. In the villages from 1000 m upwards the risk of malaria increases for households that are nearer these flat areas. However, malaria at such elevations is unstable, and when temperatures become favourable the risk of transmission increases and often epidemics result. Investigating the effect of topography on the villages individually found that there could be a mid-altitude range where topographic variables such as flat areas and the amount of swampiness in or near each village become more important. According to the villages studied this altitude range is around 750 m to 1200. However, to be sure of this further investigation of comparable villages need to be undertaken to confirm the altitude band where the shape of the land becomes more important. These findings have several important implications for malaria control. Firstly, using the logistic regression

equation in Model 1 that identifies the potential and varying risk of malaria with altitude, the potential risk of malaria could be projected to the whole of the Usambara's. Secondly, the equation produced for Model 2, could be used to identify areas that may be risk hot-spots that are near potential swamps for the Usambara's. Thirdly, if the relationship between malaria and topography holds true to other highland areas in sub-Saharan Africa then potential risk could be modelled on a continental scale. However, it should be remembered that the Usambara's experience a cooler climate than many other highland areas in East Africa (Kenworthy, 1966) and so the predictions of risk based on Models 1 and 2 cannot be extrapolated to these other highland areas. Lastly, and most importantly, understanding how malaria prevalence varies with altitude and terrain will enable better targeting of control initiatives: making sure that those households near potential swamp areas are protected with bednets and insecticides, monitoring the potential swamp areas for the presence of mosquito larvae and thus use larvicides if warranted, enabling those villages most at risk to have better access to health care, and ultimately ensure that villages are not located in or near risk hot spots.

6.5 Future work

As with any research many questions may be answered, but the answers lead to more questions and the need for more research. The main model produced, Model 2 needs to be validated by carrying out further malaria surveys in the seven villages to assess the seasonality of transmission and varying prevalence over time and throughout the village. Likewise, vector densities and their infectivity should be monitored over a year. The potential swampy areas identified by the GIS should be investigated over the course of a year, especially after rainfall to see if they are indeed susceptible to water inundation and are therefore mosquito breeding sites.

The model could also be refined in several ways. Rainfall distribution could be included within the GIS. The amount of rainfall falling within a catchment area would enable the quantification of the amount of water flowing over the terrain in the hydrological model. A soil component could be added to the GIS. Soil type will affect how water flows through the terrain and whether it is likely to remain on the surface or drain away. The potential swampy areas already identified by the GIS could be visited and soil surveys carried out. This would not be a difficult and time-consuming task if this was done for the seven study villages, but would not be feasible if the model was

extrapolated to the whole of the Usambara's. The soil information could then be incorporated in the GIS and soil type weighted depending on the water retaining or draining properties. Other topographical components that could easily be investigated since they are the result of simple manipulations of the DEMs are aspect and exposure. Aspect affects insolation levels and rainfall patterns and may influence the survival and development rates of the *Plasmodia* and *Anopheles* and the availability of vector breeding sites. The exposure of a village will be related to its altitude, position within the terrain, and in relation to wind direction. An exposed site would have a harsher environment, i.e. colder, often drier and would be harder for *Anopheles* to survive in compared to a more sheltered village.

Satellite remote sensing images of the study area of NDVI could easily be incorporated in the GIS and would identify vegetation types and hence soil moisture properties of the landscape. These data could be used to augment Model 2 or as an alternative method of assessing potential swampiness, if it provided a better identifier of malaria hot spots in the Usambara Mountains. The advent of a high resolution (90 m) satellite remote sensing data of 80% earths topography will enable the development of DEMs of the Usambara's without going through the time-consuming digitising process. This data is due to be released from November 2001 to October 2002 as part of the Shuttle Radar Topography Mission; an international project spearheaded by the National Imagery and Mapping Agency (NIMA) and the National Aeronautics and Space Administration (NASA, 2000). It would therefore be feasible to use the equation from Model 2 to identify malaria risk in the Usambara Mountains and possibly the nearby Pare Mountains that share a similar climate. The definitive aim for this project would be to refine the model so that risk maps could then be developed for highland areas of sub-Saharan Africa to aid in targeting malaria control activities in these vulnerable communities.

6.6 Conclusion

The general goal of this research was to investigate the role of topography to the risk of malaria in the western Usambara's, Tanzania. Using GIS technology surface models were produced for the seven villages situated along an altitudinal transect which are ecologically similar apart from temperature. Thus, the survey villages cover the range of malaria intensities experienced in sub-Saharan Africa, from the holoendemic

lowlands to the hypoendemic uplands. It was found that altitude is the most important topographic variable because of its close correlation with temperature. Temperature is the most significant environmental variable controlling the development and survival of the malaria parasite and vector, and hence their geographical distribution. However, flat areas where water is likely to accumulate also affects malaria prevalence, especially in those villages on the fringe of transmission between altitudes of 750 m and 1200 m. The inclusion of the swampiness parameter with altitude improves the predictive ability of the model and helps to identify villages most at risk from epidemics when environmental conditions such as raised temperature and rainfall occur. The village at 1200 m that experienced a malaria epidemic was found to have a large malaria cluster where houses were located near a flat raised river bed used for the cultivation of crops. The houses within the malaria cluster were identified as being at greater risk of malaria than those outside the cluster due to their proximity to this flat area where water is likely to accumulate. In the lowland villages terrain and hydrological characteristics are not as important determinants of malaria prevalence since transmission intensity is so high due to extensive mosquito breeding sites and optimal temperatures and humidity's. Conversely, in the highland villages temperature becomes an important limiting factor, slowing down the development rates of the mosquito and parasite and reducing their survival. Although, malaria was prevalent in these villages, the numbers of mosquitoes found were so few making transmission unlikely. If temperatures did increase in these villages coinciding with higher vector densities due to the availability of breeding sites then terrain may become important. However, peak transmission in these areas tends to be during the warm season, when temperatures are more conducive to the development rates and survival of the *Plasmodium* and *Anopheles*, but the warm season occurs some time after the long rains and so the availability of breeding sites may limit mosquito abundance. During the rainy season mosquitoes are likely to be more abundant, but temperatures too low for their development and survival.

It is important to recognise that a range of environmental and human factors influence the transmission of malaria. Producing a model that encompasses all the known parameters would be overcomplicated and impossible. One of the hardest parameters to model is the behavioural element of people, but physical variables such as temperature and rainfall can be quantified and the distribution of malaria mapped in relation to them. Altitude and terrain are physical variables that can also be quantified, but as far as we know this is the first time this has been attempted with regard to predicting malaria

prevalence in a highland area. DEMs are a valuable tool for quantifying many landscape features of mountain regions that are likely to affect biological systems such as malaria. Models of infection risk based on these parameters could provide a rapid and accurate technique for identifying malaria hot spots in upland areas. The African highlands have been identified as a distinct unit where communities living in areas of unstable transmission are susceptible to devastating epidemics. The highlands are often characterised by their remoteness and lack of health services. It is therefore hoped that this research will contribute to the initiative launched by HIMAL and help define epidemic-prone highland regions so that interventions can be targeted to those communities most at risk.

References

Adams, W. M., Goudie, A. S. & Orme, A. R., Eds. (1996). *The Physical Geography of Africa*, Oxford University Press.

Alles, H. K., Mendis, K. N. & Carter, R. (1998). "Malaria mortality rates in South Asia and in Africa: implications for malaria control." *Parasitology Today* **14**: 369-375.

Anderson, M. G. & Kneale, P. E. (1982). "The influence of low-angled topography on hillslope soil water convergence and stream discharge." *Journal of Hydrology* **57**: 65-80.

Asenso-Okyere, W. K. (1994). *Socioeconomic factors in malaria control* World Health Forum, Volume 15

August, P., Muchaud, J., Labash, C. & Smith, C. (1994). "GPS for environmental applications: accuracy and precision of location data." *Photogrammetric Engineering & Remote Sensing* **60**: 41-45.

Bamford, E. J., Dunne, L., Taylor, D. S., Symon, B. G., Hugo, G. J. & Wilkinson, D. (1999). "Accessibility to general practitioners in rural South Australia: a case study using geographical information system technology." *Medical Journal of Australia* **171**: 614-616.

Beck, L. A., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Rejmankova, E., Ulloa, A., Meza, R. A., Roberts, D. R., Paris, J. F., Spanner, M. A., Washino, R. K., Hacker, C. & Legters, J. (1994). "Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission." *American Journal of Tropical Medicine* **51**: 271-280.

Berry, L. (1971). *Tanzania in maps*. London, University of London Press.

Bhatnagar, V. N. (1954). "Altitude and malaria." *Bulletin of the National Society of Indian Malaria, Mosquitoes and Disease* **2**: 132-138.

Bødker, R. (2000). *Variation in malaria risk in the Usambara Mountains, Tanzania* Zooligical Institude, University of Copenhagen: PhD Thesis.

Bødker, R., Kisinza, W., Malima, R., Msangeni, H. & Lindsay, S. (2000). "Resurgence of malaria in the Usambara mountains, Tanzania, an epidemic of drug resistance parasites." *Global Change & Human Health* **1**: 134-153.

Brown, M., Harper, J. L. & Townsend, C. R. (1991). *Geographic Information Systems: a guide to the technology*. New York, Van Nostrand Reinhold.

Burt, T. P. & Butcher, D. P. (1985). "Topographic controls of soil moisture distribution." *Journal of Soil Science* **36**: 469-486.

Butler, D. (1997). "Briefing malaria: Time to put malaria control on the global agenda." *Nature* **386**: 535-541.

Chrisman, N. (1991). *Exploring geographic information systems*, John Wiley & Sons.

Connor, S. J., Thomson, M. C., Flasse, S. P. & Perryman, A. H. (1998). "Environmental information systems and malaria risk mapping and epidemic forecasting." *Overseas Development Institute: Disasters* **22**: 39-56.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes: models and applications*. London, Pion Limited.

Coatzee, M., Craig, M. & le Sueur, D. (2000). "Distribution of African malaria mosquitoes belonging to the *Anopheles gambiae* complex." *Parasitology Today* **16**: 74-77.

Craig, M. H., Snow, R. W. & le Sueur, D. (1999). "A climate-based distribution model of malaria transmission in sub-Saharan Africa." *Parasitology Today* **15**: 105 - 111.

Dean, A. G. (1995). *Epi Info* 6: A word-processing database and statistics program for public health. The Division of Surveillance and Epidemiology, Epidemiology Program Office, Centre for Disease Control and Prevention, Atlanta

DeMers, M. N. (1997). *Fundamentals of geographic information systems*. London, John Wiley & Sons.

Dobson, M. J., Snow, R. W. & Malowany, M. J. (1999). *Roll back malaria: the history of malaria and its control in twentieth-century East Africa* The Welcome Trust, The Welcome Trust Review: A selection of research projects and major initiatives, Vol. 8

Dunn, C. & Kingham, S. (1996). Modelling air quality and the effects on health in a GIS framework. *Innovations in GIS 3*. D. Parker, Taylor & Francis: Chapter 17.

Dunn, C. E., Woodhouse, J., Bhopal, R. S. & Acquilla, S. D. (1995). "Asthma and factory emissions in northern England: addressing public concerns by combining geographical and epidemiological methods." *Journal of Epidemiology and Community Health* **49**: 395-400.

ESRI (1991). *ARC/Info Users Guide* 6.0. Environmental Systems Research Institute, Inc., Redlands, California

Fontaine, R. E., Najjar, A. E. & Prince, J. S. (1961). "The 1958 malaria epidemic in Ethiopia." *American Journal of Tropical Medicine and Hygiene* **10**: 795-803.

Fowler, J., Cohen, L. & Jarvis, P. (1998). *Practical statistics for field biology*, Wiley & Sons Ltd.

Gabinaud, A. (1987). "Ecological mapping to support mosquito control on the French Mediterranean coast." *Parasitology Today* **3**: 317-320.

Gamage-Mendis, A. C., Carter, R., Mendis, C., de Zoysa, A. A. P. K., Herath, P. R. J. & Mendis, K. N. (1991). "Clustering of malaria infections within an endemic population: Risk of malaria associated with the type of housing construction." *American Journal of Tropical Medicine* **45**: 77-85.

Garnham, P. C. C. (1948). "The incidence of malaria at high altitudes." *Journal of the National Malaria Society* **7**: 275-284.

Gbakima, A. A. (1994). "Inland valley swamp rice development: malaria, schistosomiasis, onchocerciasis in South Central Sierra Leone." *Public Health* **108**: 149-157.

Ghebreyesus, T. A., Haile, M., Witten, K. H., Asefaw, G., Yohannes, A. M., Yohannes, M., Teklehaimanot, H. D., Lindsay, S. W. & Byass, P. (1999). "Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey." *British Medical Journal* **319**: 663-666.

Gilles, H. M. & Warrell, D. A., Eds. (1993). *Bruce-Chwatt's essential mararíology.* New York, Oxford University Press.

Gilles, M. T. (1953). "The duration of the gonotrophic cycle in *Anopheles gambiae* and *Anopheles funestus*, with a note on the efficiency of hand catching." *East African Medical Journal* **30**: 129-135.

Gilles, M. T. (1961). "Studies on the dispersal and survival of *Anopheles Gambiae Giles* in East Africa, by means of marking and release experiments." *Bulletin of Entomological Research* **52**: 92-127.

Gilles, M. T. & deMeillon, B. (1968). *The Anophelinae of Africa south of the Sahara.* Johannesburg, Publications of the South African Institute for Medical Research No. 54.

Githeko, A. K., Lindsay, S. W., Ulisses, E. C. & Patz, J. A. (2000). "Climate change and vector-borne diseases: a regional analysis." *Bulletin of the World Health Organization* **78**: 1136 - 1147.

Goodchild, M., F. (1986). *Spatial autocorrelation.* Ontario, University of Western Ontario.

Haines-Young, R., Green, D. R. & Cousins, S. H. (1993). *Landscape ecology and GIS.* London, Taylor & Francis Ltd.

Hamilton, A. C. & Bensted-Smith, R., Eds. (1989). *Forest conservation in the East Usambara Mountains, Tanzania*, IUCN.

Hamilton, A. C. & Macfadyen, A. (1989). Climatic change in the East Usambaras. *Forest Conservation in the East Usambaras.* A. C. Hamilton & R. Bensted-Smith. Gland: Switzerland, International Union for the Conservation of Nature: 103-107.

Hay, S. I., Snow, R. W. & Rogers, D. J. (1998). "Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**: 12-20.

Heisch, M. D. & Harper, J. O. (1949). "An epidemic of malaria in the Kenya highlands transmitted by *Anopheles funestus*." *Journal of Tropical Medicine and Hygiene* **50**: 187-190.

Hicken, B. W., Maguire, D. J. & Strachan, A. J. (1991). *Introduction to GIS: The Arc/Info method*, Midlands Regional Research Laboratory.

Hightower, A. W., Ombok, M., Otieno, R., Odhiambo, R., Oloo, A. J., Lal, A. A., Nahlen, B. L. & Hawley, W. A. (1998). "A geographic information system applied to a malaria field study in Western Kenya." *American Journal of Tropical Medicine and Hygiene* **58**: 266-272.

Hilton, K. (1991). Spaceship Earth. *Geographical Magazine*. Supplement: 1-31.

Hirsch, A. (1883). *Handbook of geographical and historical pathology, Volume 1, Acute Infective Diseases*. London, The New Sydenham Society.

Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A. & Maskell, K. (1996). *Climate change 1995. The science of climate change. Contribution of WGI to the second assessment report of the Intergovernmental Panel on Climate Change.*, Cambridge University Press.

Hugh-Jones, M. (1989). "Applications of remote sensing to the identification of the habitats of parasites and disease vectors." *Parasitology Today* **5**: 244-251.

Indaratna, K., Hutubessy, R., Chupraphawan, S., Sukapurana, C., Tao, J., Chunsutthiwat, S., Thimasarn, K. & Crissman, L. (1998). "Application of geographical information systems to co-analysis of disease and economic resources: dengue and malaria in Thailand." *The Southeast Asian Journal of Tropical Medicine and Public Health* **29**: 669-684.

Jones, A., Duck, R., Weyers, J. & Reed, R. (2000). *Practical skills in environmental science*. Harlow, Essex, Prentice Hall.

Kenworthy, J. M. (1966). "Temperature conditions in the tropical highland climates of East Africa." *East African Geographical Review* **4**: 1-11.

Kirkwood, B. R. (1988). *Essentials of medical statistics*. Oxford, Blackwell Scientific Publications.

Kitron, U., Otieno, L. H., Hungerford, L. L., Odulaja, A., Brigham, W. U., Okello, O. O., Joselyn, M., Mohamed-Ahmed, M. M. & Cook, E. (1996). "Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS." *Journal of Animal Ecology* **65**: 371-380.

Kitron, U., Pener, H., Costin, C., Orshan, L., Greenberg, Z. & Shalom, U. (1994). "Geographic information system in malaria surveillance: mosquito breeding and imported cases in Israel." *American Journal of Tropical Medicine and Hygiene* **50**: 550-556.

Knell, A. J., Ed. (1991). *Malaria*. Oxford, Oxford University Press.

Knox, G. (1964). "Epidemiology of childhood leukaemia in Northumberland and Durham." *British Journal of Preventative and Social Medicine* **18**: 17-24.

Kovats, R. S., Haines, A., Stanwell-Smith, R., Martens, P., Menne, B. & Bertollini, R. (1999). "Climate change and human health in Europe." *British Medical Journal* **318**: 1682-1685.

Kovats, R. S. (2000). "El Niño and human health." *Bulletin of the World Health Organization* **78**: 1128 - 1135.

Kulldorff, M., Rand, K., Gherman, G., Williams, G. & DeFrancesco, D. (1998). *SaTScan* 2.1: Software for the spatial and space-time scan statistics. National Cancer Institute, Bethesda, MD

Knell, A. J., Ed. (1991). *Malaria.* Oxford, Oxford University Press.

Lennon, J. J., Greenwood, J. J. D. & Turner, J. R. G. (2000). "Bird diversity and environmental gradients in Britain: a test of the species-energy hypothesis." *Journal of Animal Ecology* **69**: 581-598.

Lepers, J. P., Deloron, P., Andriamagatiana-Rason, M. D., Ramanamirija, J. A. & Coulanges, P. (1990). "Newly transmitted *Plasmodium falciparum* malaria in the central highland plateaux of Madagascar: assessment of clinical impact in a rural community." *Bullietin of the World Health Organisation* **68**: 217-222.

Lepers, J. P., Fontenille, D., Rason, M. D., Chougnet, C., Astagneau, P., Coulanges, P. & Deloron, P. (1991). "Transmission and epidemiology of newly transmitted faciparum malaria in the central highland plateaux of Madagascar." *Annals of Tropical Medicine and Parasitology* **85**: 297-304.

Letham, L. (1995). *GPS made easy.* Calgary, Rocky Mountain Books.

Lindblade, K. A., Walker, E., D., Onapa, A. W., Katungu, J. & Wilson, M. L. (2000). "Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda." *Tropical Medicine and International Health* **5**: 263-274.

Lindsay, S. W., Armstrong Schellenberg, J. R. M., Zeiler, H. A., Daly, R. J., Salum, F. M. & Wilkins, H. A. (1995). "Exposure of Gambian children to *Anopheles gambiae* malaria vectors in an irrigated rice production area." *Medical and Veterinary Entomology* **9**: 50-58.

Lindsay, S. W. & Birley, M. H. (1996). "Climate change and malaria transmission." *Annals of Tropical Medicine and Parasitology* **90**: 573-588.

Lindsay, S. W., Bødker, R., Malima, R., Msangeni, H. A. & Kisinza, W. (2000). "Impact of 1997-8 El Niño on highland malaria in Tanzania." *The Lancet* **355**: 989-990.

Lindsay, S. W. & Martens, P. (1998). "Malaria in the African Highlands: past, present and future." *Bulletin of the World Health Organisation* **76**: 33-45.

Lindsay, S. W., Parson, L. & Thomas, C. J. (1998). "Mapping the ranges and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *An. arabiensis*, using climate data." *Proceedings of the Royal Society of London. B.* **265**: 847-854.

Lindsay, S. W. & Thomas, C. J. (2000). "Mapping and estimating the population at risk from lymphatic filariasis in Africa." *Transactions of the Royal Society of Tropical Medicine and Hygeine* **94**: 37-45.

Linthicum, K. J., Bailey, C. L., Tucker, C. J., Mitchell, K. D., Logan, T. M., Davies, F. G., Kamau, C. W., Thande, P. C. & Wagateh, J. N. (1990). "Application of polar-orbiting, meteorological satellite data to detect flooding of Rift Valley Fever virus vector mosquito habitats in Kenya." *Medical and Veterinary Entomology* **4**: 433-438.

Loevinsohn, M. E. (1994). "Climatic warming and increased malaria incidence in Rwanda." *The Lancet* **343**: 714-18.

Manga, L., Fondjo, E., Carnevale, P. & Robert, V. (1993). "Importance of low-dispersion of *Anopheles gambiae* on malaria transmission in hilly towns in South Cameroon." *Journal of Medical Entomology* **30**: 936-938.

MARA/ARMA (1998). *Toward an atlas of malaria risk in Africa* Mapping Malaria Risk in Africa (MARA) Collaboration, Durban December 1998First Technical Report, 1 www.mara.org.za/

Marchant, P., Eling, W., van Gemert, G.-J., Leake, C. J. & Curtis, C. F. (1998). "Could British mosquitoes transmit falciparum malaria?" *Parasitology Today* **14**: 344-345.

Martens, P. (1999). "How will climate change affect human health?" *American Scientist* **87**: 534-541.

Martens, P. & Hall, L. (2000). "Malaria on the move: human population movement and malaria transmission." *Emerging Infectious Diseases* **6**: 103-109.

Martens, P., Kovats, R. S., Nijhof, S., de Vries, P., Livermore, M. T. J., Bradley, D. J., Cox, J. & McMichael, A. J. (1999). "Climate change and future populations at risk of malaria." *Global Environmental Change* **9**: 89-107.

Martens, W. J. M., Niessen, L. W., Rotmans, J., Jetten, T. H. & McMichael, A. J. (1995). "Potential impact of global climate change on malaria risk." *Journal of the International Institute of Environmental Health Services* **103**: 458-464.

Matola, Y. G., White, G. B. & Magayuka, S. A. (1987). "The changed pattern of malaria endemicity and transmission at Armani in the eastern Usambara mountains, north-eastern Tanzania." *Journal of Tropical Medicine and Hygiene* **90**: 127-134.

McKenzie, F. E. (2000). "Why model malaria?" *Parasitology Today* **16**: 511-516.

Mitchell, C. (1991). *Terrain evaluation.* London, Longman Scientific & Technical.

Moore, I. D., O'Loughlin, E. M. & Burch, G. J. (1988). "A contour-based topographic model for hydrological and eological applications." *Earth Surface Processes and Landforms* **13**: 305-320.

Morrow, B. H. (1999). "Identifying and mapping community vulnerability." *Disasters* **23**: 1-18.

Mutuwatte, L. P., Konradsen, F., Renault, D., Sharma, S. K., Gulati, O. T. & Kumara, W. A. U. (1997). *Water-related environmental factors and malaria transmission in Mahi Kadana, Gujarat, India.* International Irrigation Management Institute, Colombo, Sri Lanka Working Paper, 41

Najera, J. A., Bernhard, H. L. & Hammer, J. (1992). *Malaria: New patterns and perspectives.* Washington D. C., World Bank.

NASA (2000) *www.jpl.nasa.gov/srtm/ - Shuttle Radar Topography Mission* 2000 19 October

Norusis, M., J. (1994). *SPSS Advanced Statistics 6.1.* Chicago, SPSS Inc.

Omumbo, J., Ouma, J., Rapuoda, B., Craig, M. H., le Sueur, D. & Snow, R. W. (1998). "Mapping malaria transmission intensity using geographical information systems (GIS): an example from Kenya." *Annals of Tropical Medicine & Parasitology* **92**: 7-21.

Patz, J. A., Epstein, P. R., Burke, T. A. & Balbus, J. M. (1996). "Global climate change and emerging infectious diseases." *Journal of the American Medical Association* **275**: 217-223.

Patz, J. A. & Lindsay, S. W. (1999). "New challenges, new tools: the impact of climate change on infectious diseases." *Current Opinion in Microbiology* **2**: 445-451.

Patz, J. A., Strzepek, K., Lele, S., Hedden, M., Green, S., Noden, B., Hay, S. I., Kalkestein, L. & Beler, J. C. (1998). "Predicting key malaria transmission factors, biting and entomological inoculation rates, using modelled soil moisture in Kenya." *Tropical Medicine and International Health* **3**: 818-827.

Petrie, G. & Kennie, T. J. M., Eds. (1990). *Terrain modelling in surveying and civil engineering.* London, Whittles.

Price, M. F. & Heywood, D. I., Eds. (1994). *Mountain environments and geographical information systems.* London, Taylor & Francis Ltd.

Pullar, D. & Springer, D. (2000). "Towards integrating GIS and catchment models." *Environmental Modelling and Software* **15**: 451-459.

Rejmankova, E., Roberts, D. R., Pawley, A., Manguin, S. & Polanco, J. (1995). "Predictions of adult *Anopheles albimanus* densities in villages based on distances to remotely sensed larval habitats." *American Journal of Tropical Medicine and Hygiene* **53**: 482-488.

Rogers, D. J. & Randolph, S. E. (2000). "The global spread of malaria in a future, warmer world." *Science* **289**: 1763 - 1766.

Sawada, M. (1999). "ROOKCASE: An Excel 97/2000 Visual Basic (VB) Add-in for Exploring Global and Local Spatial Autocorrelation." *Bulletin of the Ecological Society of America* **80**: 231-234.

Sharma, V. P. & Aruna, S. (1997). "Role of geographic information system in malaria control." *Indian Journal of Medical Research* **106**: 198-204.

Smith, P. A. (1994). "Autocorrelation in logistic regression modelling of species' distributions." *Global Ecology and Biodiversity Letters* **4**: 47-61.

Smith, T., Charlwood, J. D., Takken, W., Tanner, M. & Spiegelhalter, D. J. (1995). "Mapping the densities of malaria vectors within a single village." *Acta Tropica* **59**: 1-18.

Snow, R. W., Armstrong Schellenberg, J. R. M., Peshu, N., Forster, D., Newton, C. R. J. C., Winstanley, P. A., Mwangi, I., Waruiru, C., Warn, P. A., Newbold, C. & Marsh, K. (1993). "Periodicity and space-time clustering of severe childhood malaria on the coast of Kenya." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **87**: 386-390.

Snow, R. W., Craig, M., Deichmann, U. & Marsh, K. (1999a). "Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population." *Bulletin of the World Health Organization* **77**: 624-640.

Snow, R. W., Craig, M. H., Deichmann, U. & le Sueur, D. (1999b). "A preliminary continental risk map for malaria mortality among African children." *Parasitology Today* **15**: 99-104.

Snow, R. W., Marsh, K. & le Sueur, D. (1996). "The need for maps of transmission intensity to guide malaria control in Africa." *Parasitology Today* **12**: 455-457.

Sokal, R. R. & Oden, N. L. (1978). "Spatial autocorrelation in biology 1. Methodology." *Biological Journal of the Linnean Society* **10**: 199-228.

Tanser, F. & Wilkinson, D. (1999). "Spatial implications of the tuberculosis DOTS strategy in rural South Africa: a novel application of geographical information system and global positioning system technologies." *Tropical Medicine and International Health* **4**: 634-638.

Thomas, C. J. & Lindsay, S. W. (2000). "Local-scale variation in malaria infection amongst rural Gambian children estimated by satellite remote sensing." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **94**: 159-163.

Thomson, M., Connor, S., Bennett, S., D'Alessandro, U., Milligan, P., Aikins, M., Landerock, P., Jawara, M. & Greenwood, B. (1996). "Geographical perspectives on bednet use and malaria transmission in the Gambia, West Africa." *Social Science and Medicine* **43**: 101-112.

Thomson, M. C., Connor, S. J., D'alessandro, U., Rowlingson, B., Diggle, P., Cresswell, M. & Greenwood, B. (1999). "Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results." *The American Society of Tropical Medicine and Hygiene* **61**: 2 - 8.

Tim, U. S. (1995). "The application of GIS in environmental health sciences: opportunities and limitations." *Environmental Research* **71**: 75-88.

Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region." *Economic Geography* **46**: 234-240.

Washino, R. & Wood, B. L. (1994). "Application of remote sensing to arthropod vector surveillance and control." *American Journal of Tropical Medicine and Hygiene* **50**: 134-144.

Weibel, R. H., M. (1991). Digital terrain modelling. *GIS 1: Principles*. D. J. Maguire & M. F. Goodchild. London, Longman Scientific & Technical. **I: Principals:** Chapter 19.

Welch, J., B., Olson, J. K., Yates, M. M., Benton, A. R. & Baker, R. D. (1989). "Conceptional model for the use of aerial colour infrared photography by mosquito control districts as a survey technique for *Psorophora columbiae* oviposition habitats in Texas ricelands." *Journal of the American Mosquito Control Association* **5**: 369-373.

Wellcome Trust (1999). *Wellcome News: Research and funding from the Welcome Trust* The Welcome Trust, Issue 21 Q4

Wise, S. (2000). "Assessing the quality for hydrological applications of digital elevation models derived from contours." *Hydrological Processes* **14**: 1909-1929.

Wolock, D. M. & McCabe, G. (2000). "Differences in topographic characteristics computed from 100- and 1000-m resolution digital elevation model data." *Hydrological Processes* **14**: 987-1002.

Wood, B. L., Beck, L. R., Washino, R. K., Hubbard, K. A. & Salute, J. S. (1992). "Estimating high mosquito-producing rice fields using spectral and spatial data." *International Journal of Remote Sensing* **13**: 2813-2826.

World Health Organisation, W. H. (1951). *Report of the malaria conference in equatorial Africa* WHO, Geneva, Kampala 1950 Technical Report Series 38

World Health Organisation, (1997). "World malaria situation in 1994." *Weekly Epidemiologial Records* **72**: 269-76.

World Health Organisation. (2001) *www.who.int/aboutwho/en/preventing/malaria.htm* 2001 13 February

APPENDIX A

A selection of coverages, grids and surface models displaying a variety of topographical features for each village area

Figure A1. Vector coverages for roads, streams, contours and house positions of villages located at 300 m, 600 m, 750 m and 1000 m.

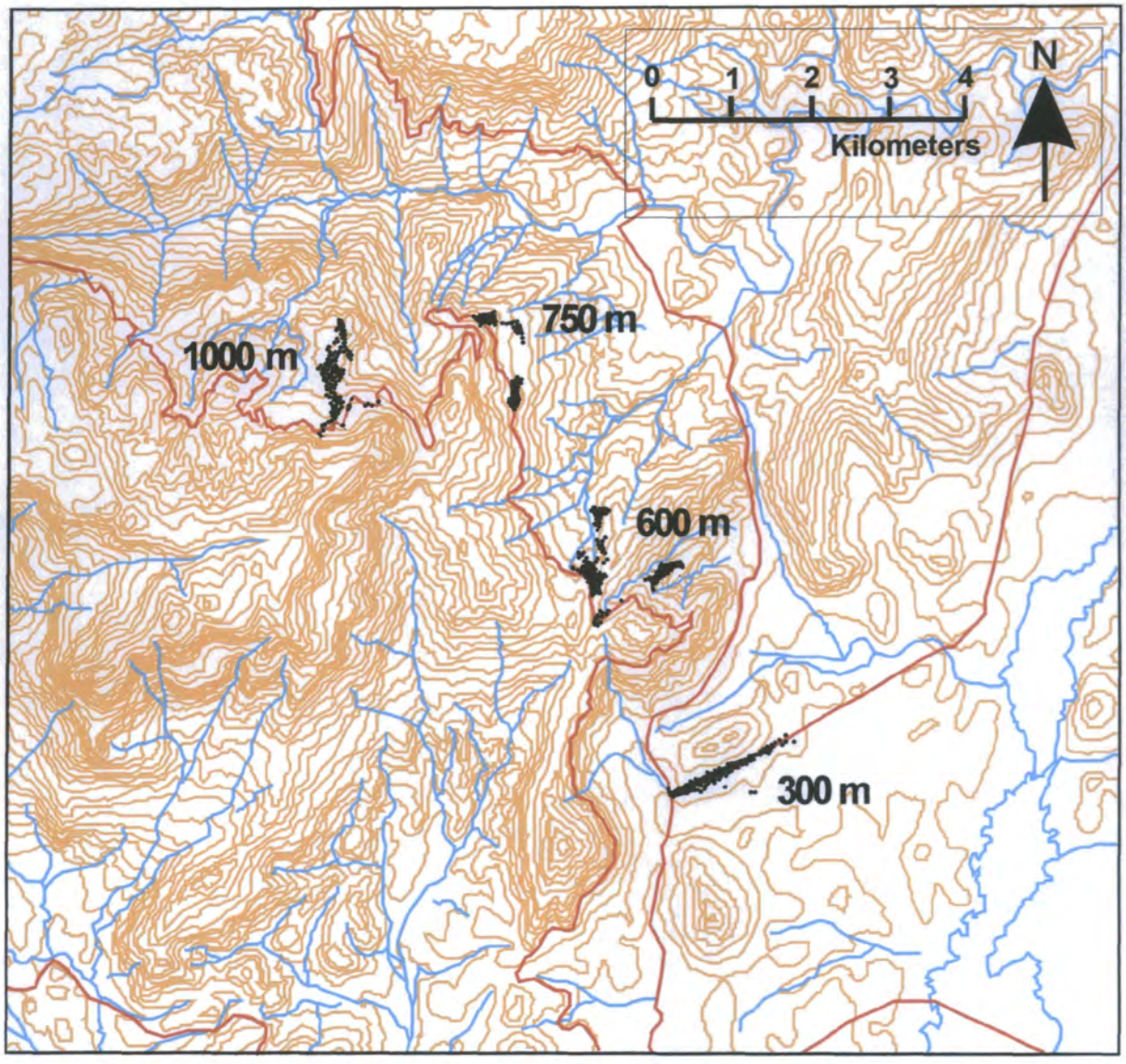


Figure A2. **Vector coverages for roads, streams, contours and house positions for the village located at 1200 m**

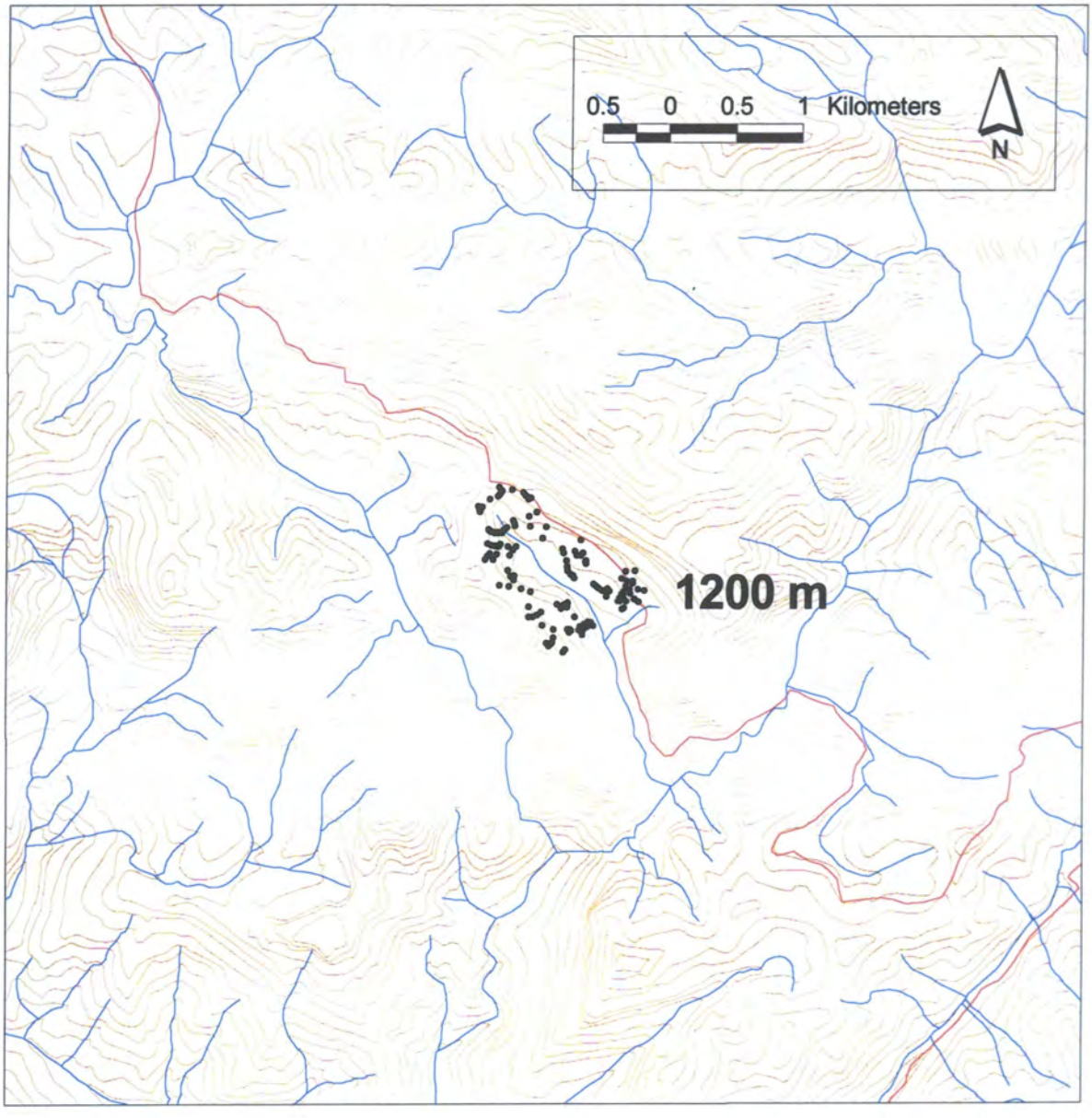


Figure A3. **Vector coverages for roads, streams, contours and house positions for the village located at 1300 m**

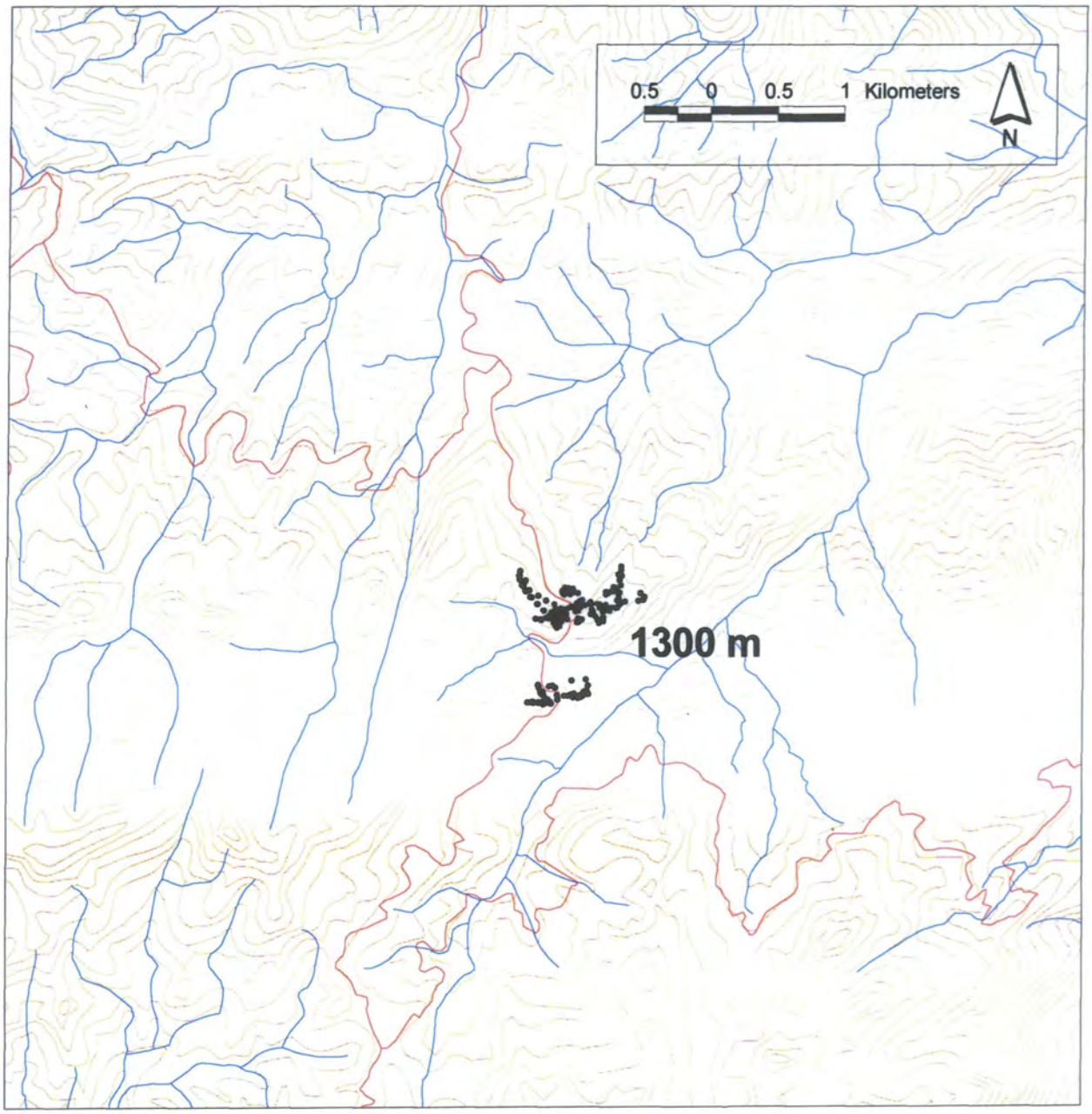
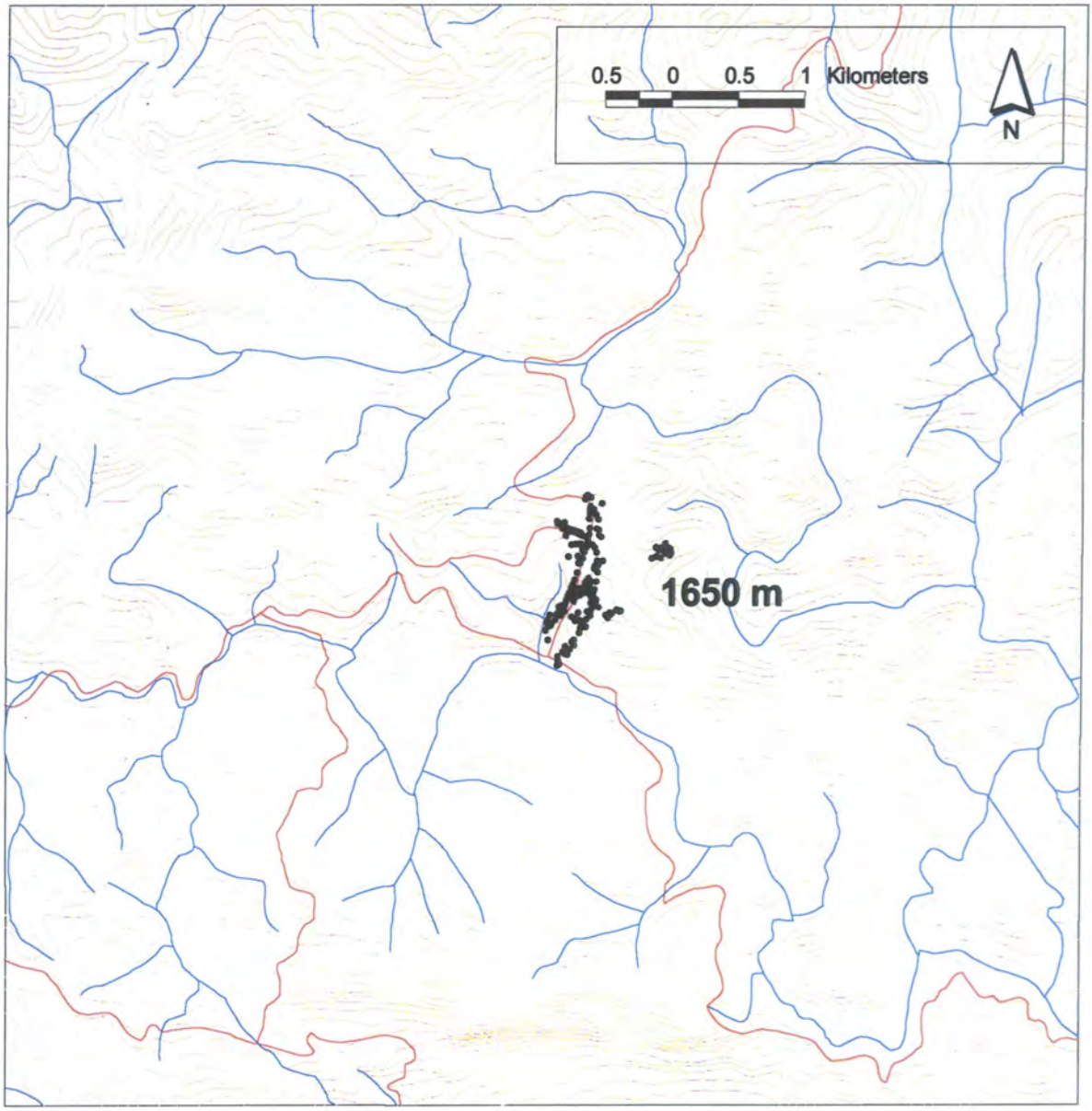


Figure A4. **Vector coverages for roads, streams, contours and house positions for the village located at 1650 m**



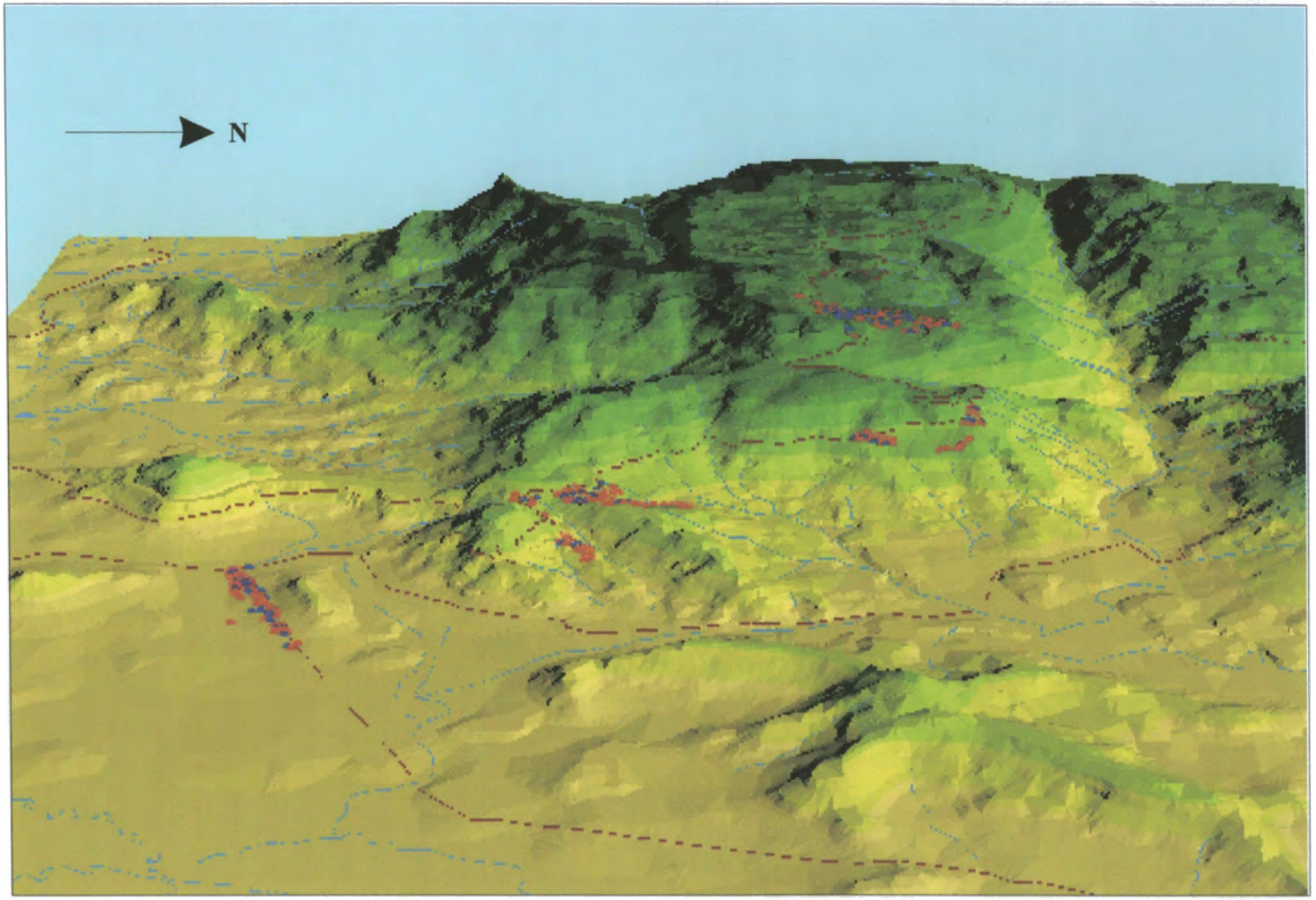


Figure A5. Surface representation showing the surrounding terrain for the villages situated at 300 m, 600 m , 750 m and 1000 m overlaid with stream and road network coverages.

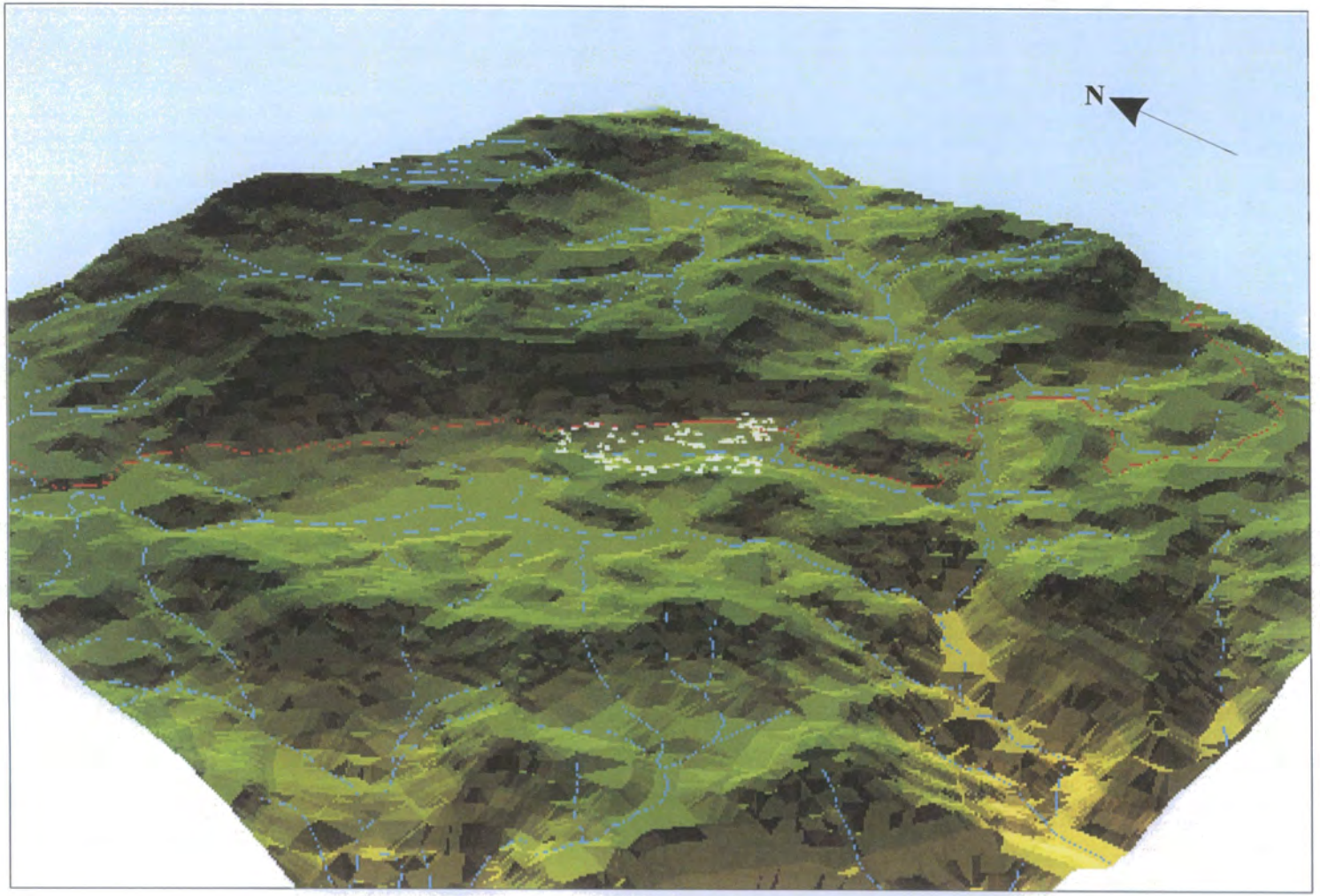


Figure A6. Surface representation showing the surrounding terrain for the village located at 1200 m, overlaid with stream and road network coverages.

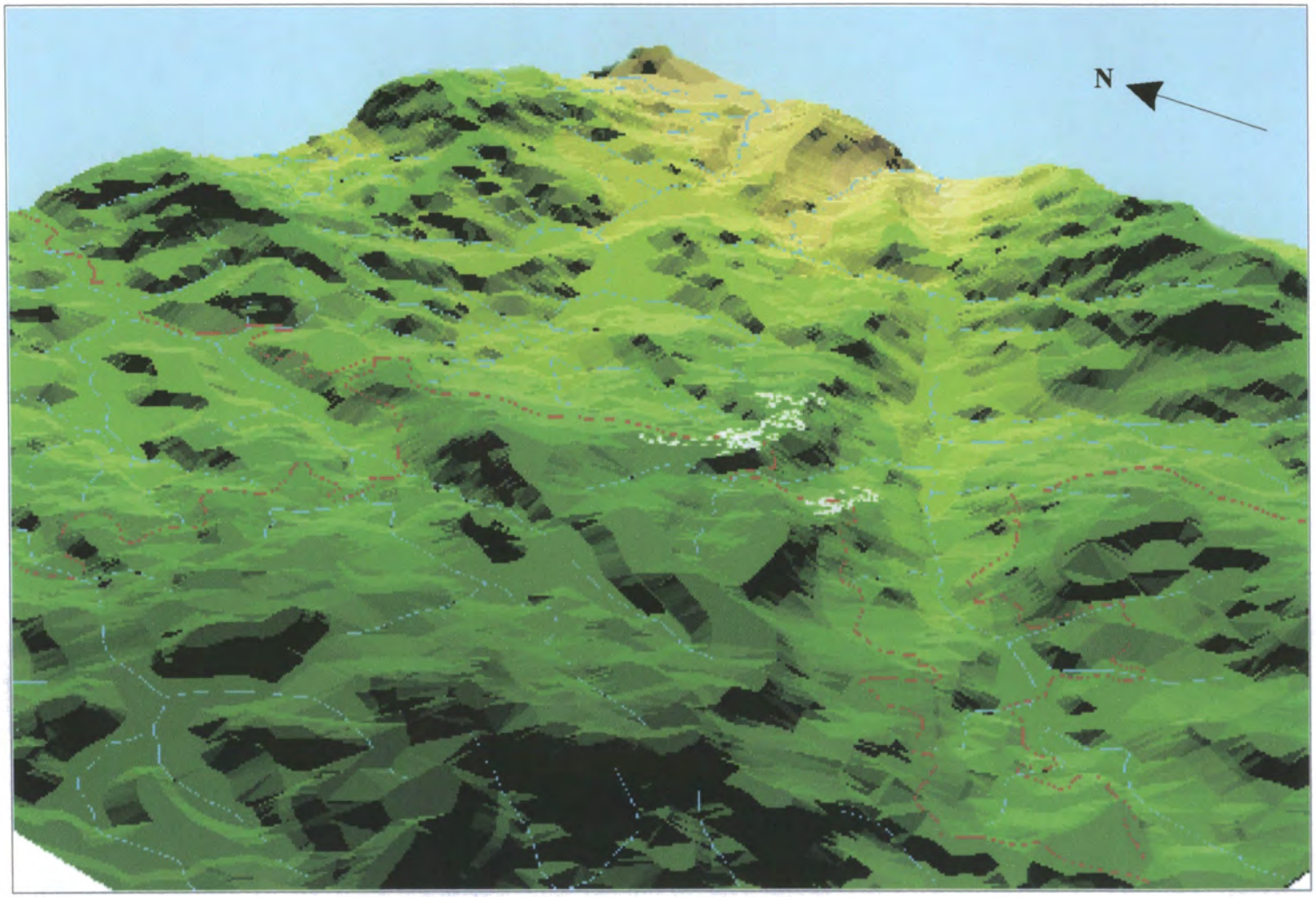


Figure A7. Surface representation showing the surrounding terrain for the village located at 1300 m, overlaid with stream and road network coverages.

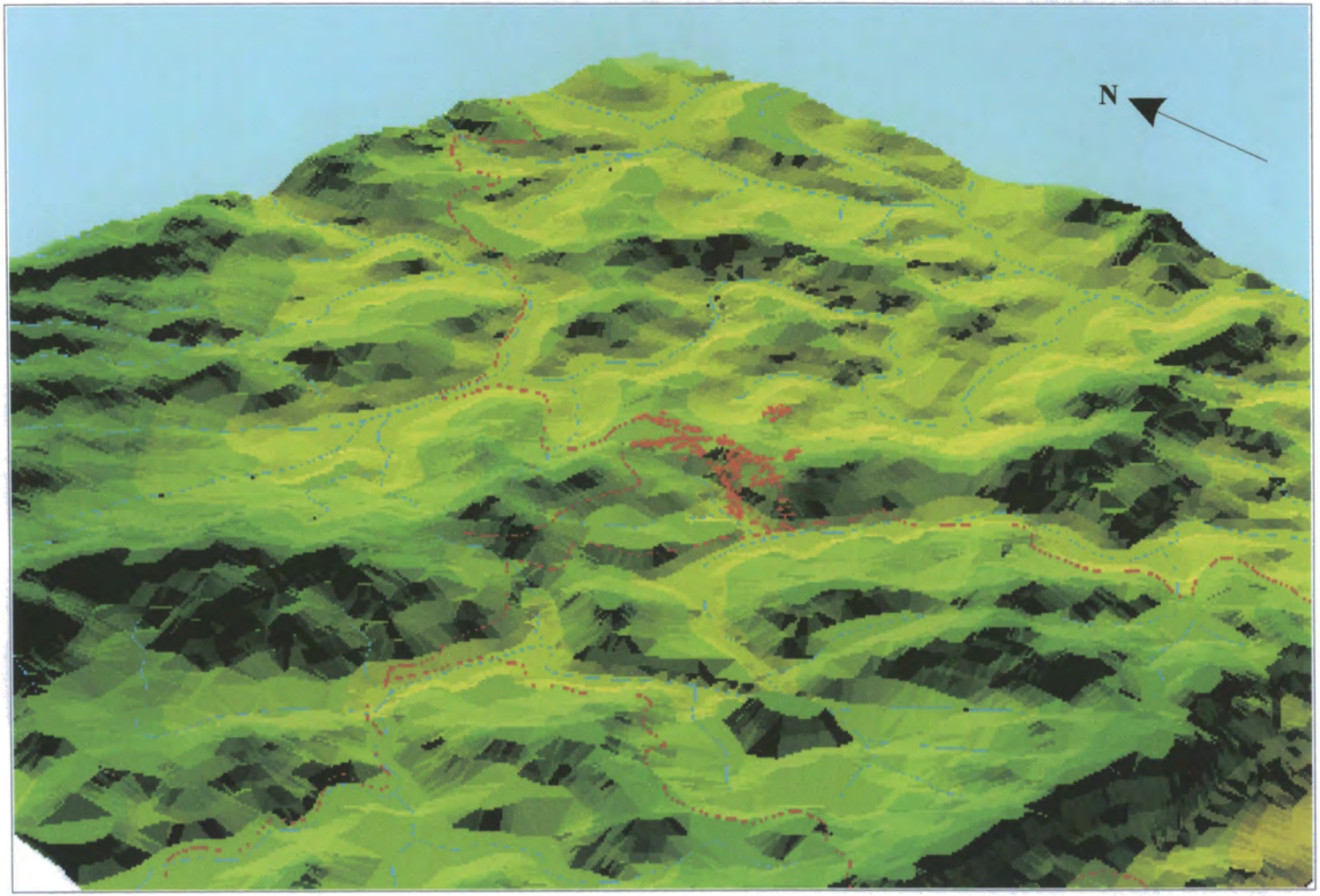


Figure A8. Surface representation showing the surrounding terrain for the village located at 1650 m, overlaid with stream and road network coverages.

Figure A9. Surface representation displaying flow direction for villages situated at 300 m, 600 m, 750 m and 1000 m.

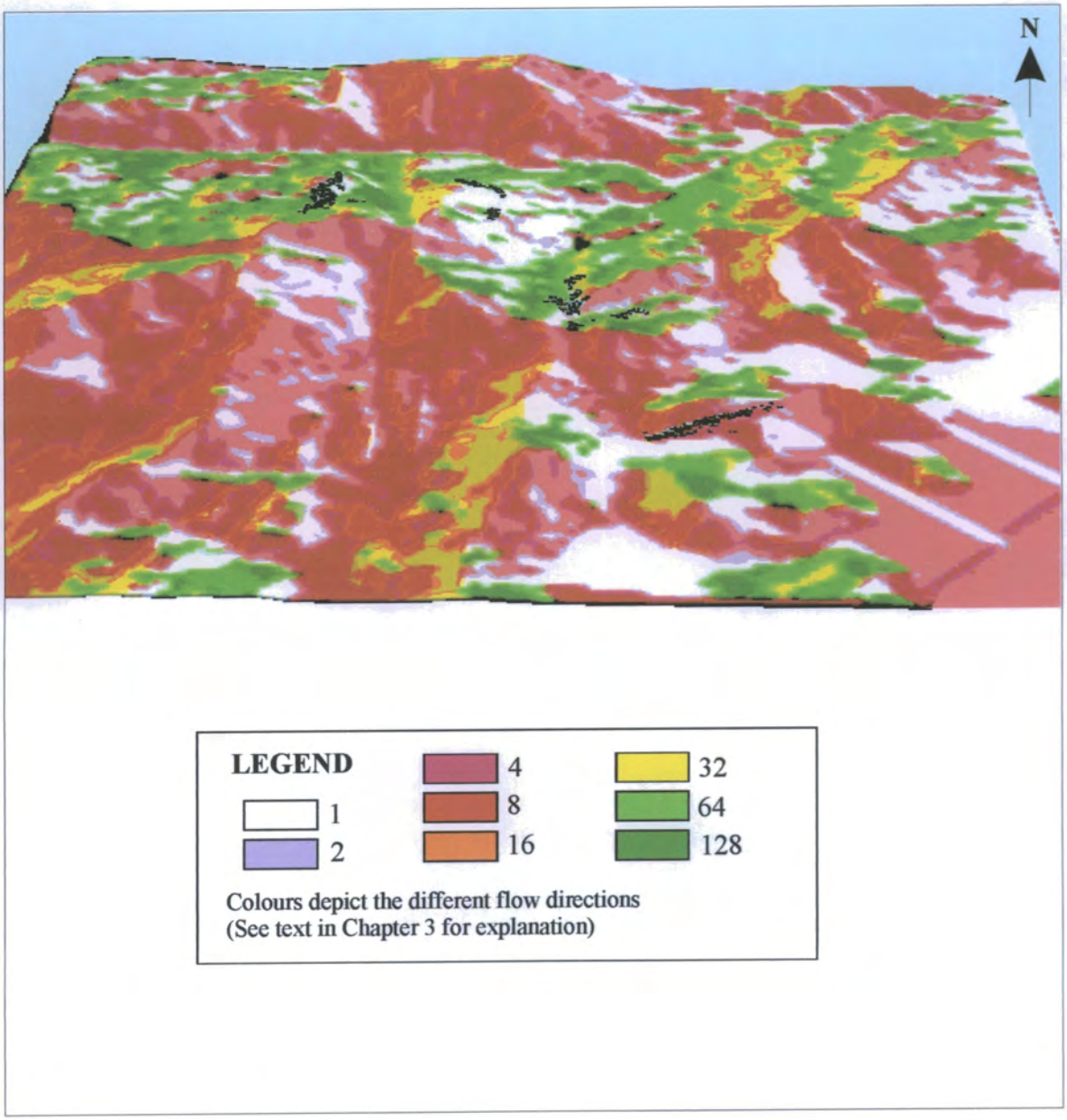


Figure A10. Surface representation displaying flow direction for the village area at 1200 m, overlaid with stream and road networks.

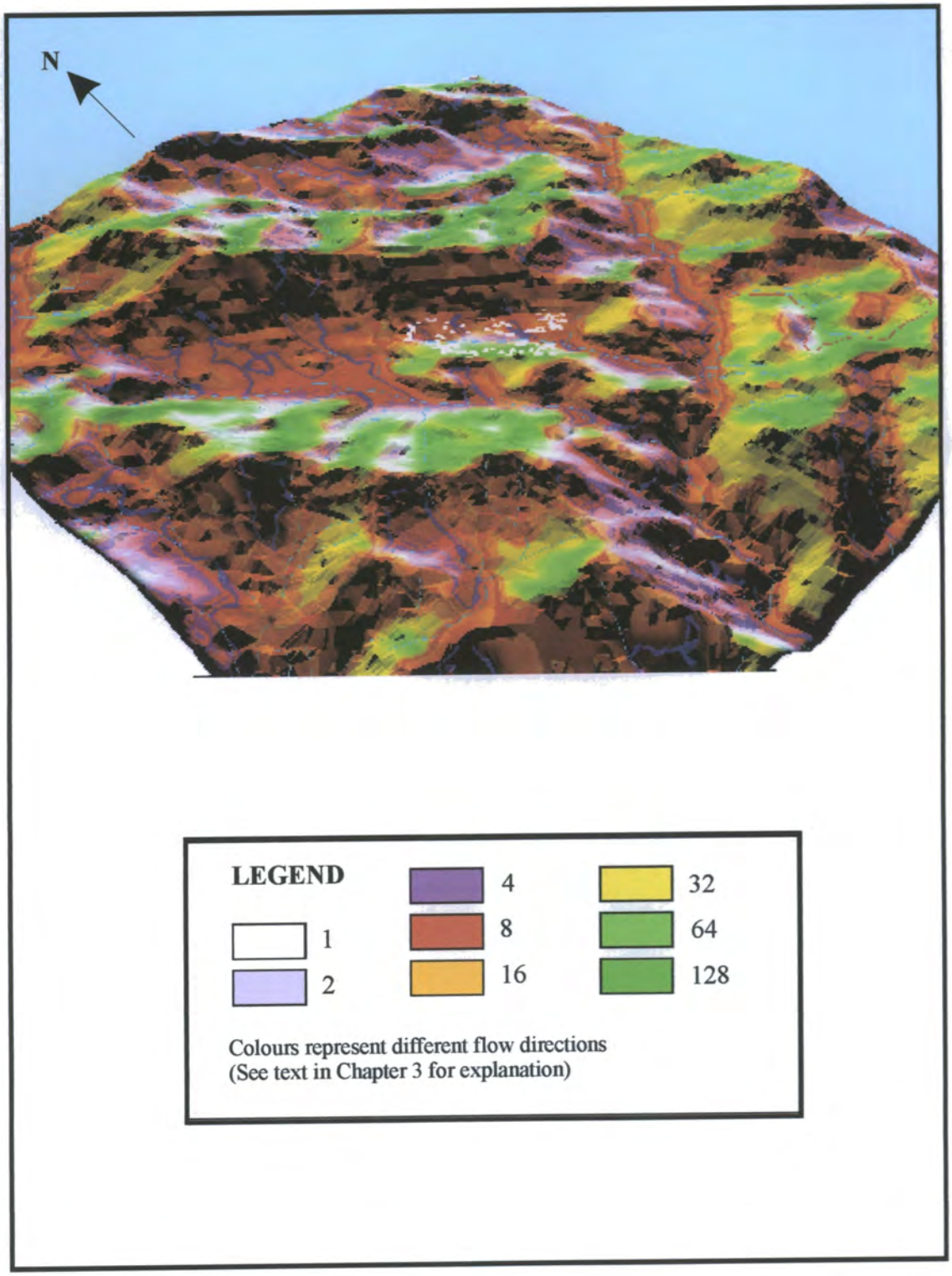


Figure A11. Surface representation displaying flow direction for the village area at 1300 m, overlaid with stream and road networks.

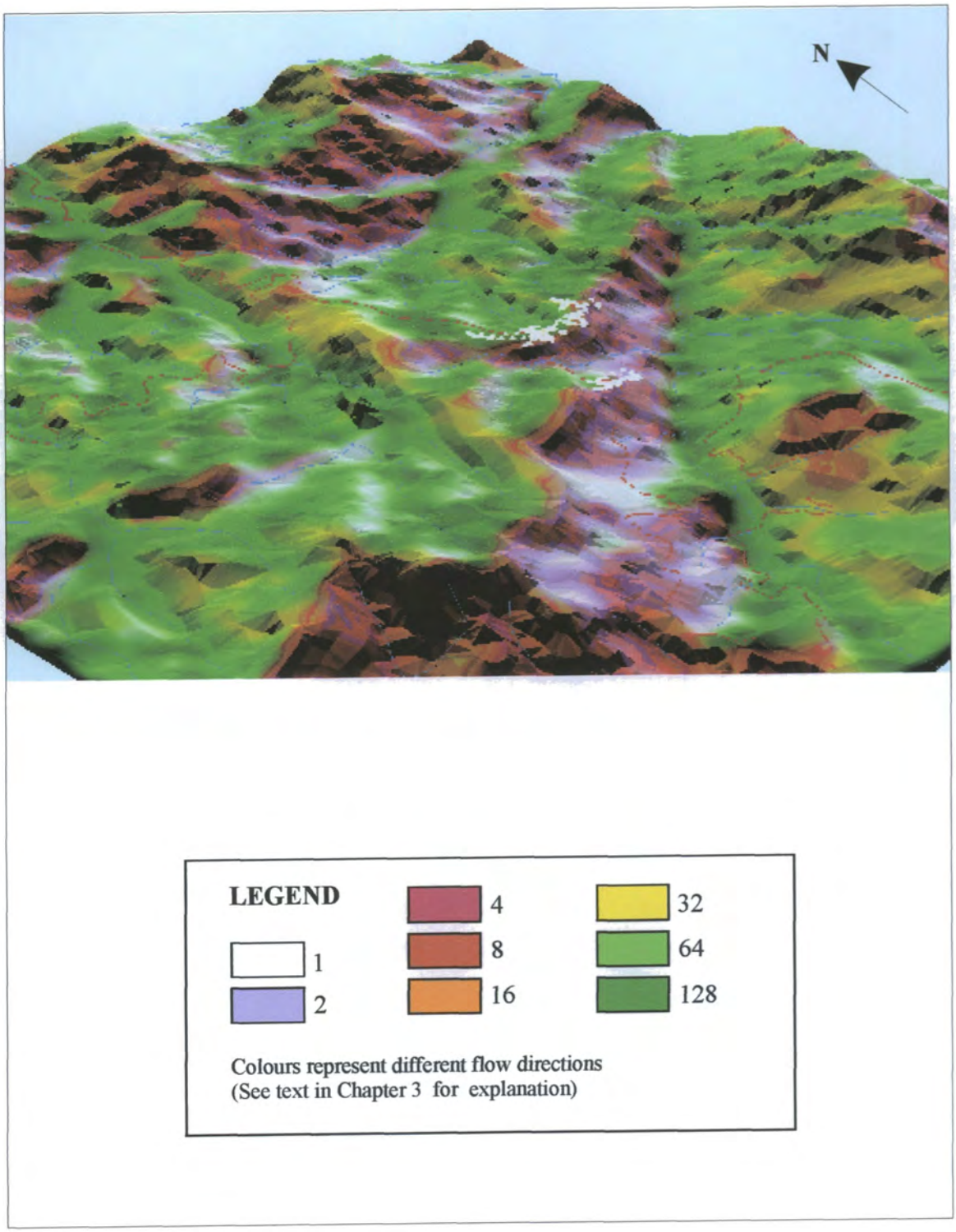
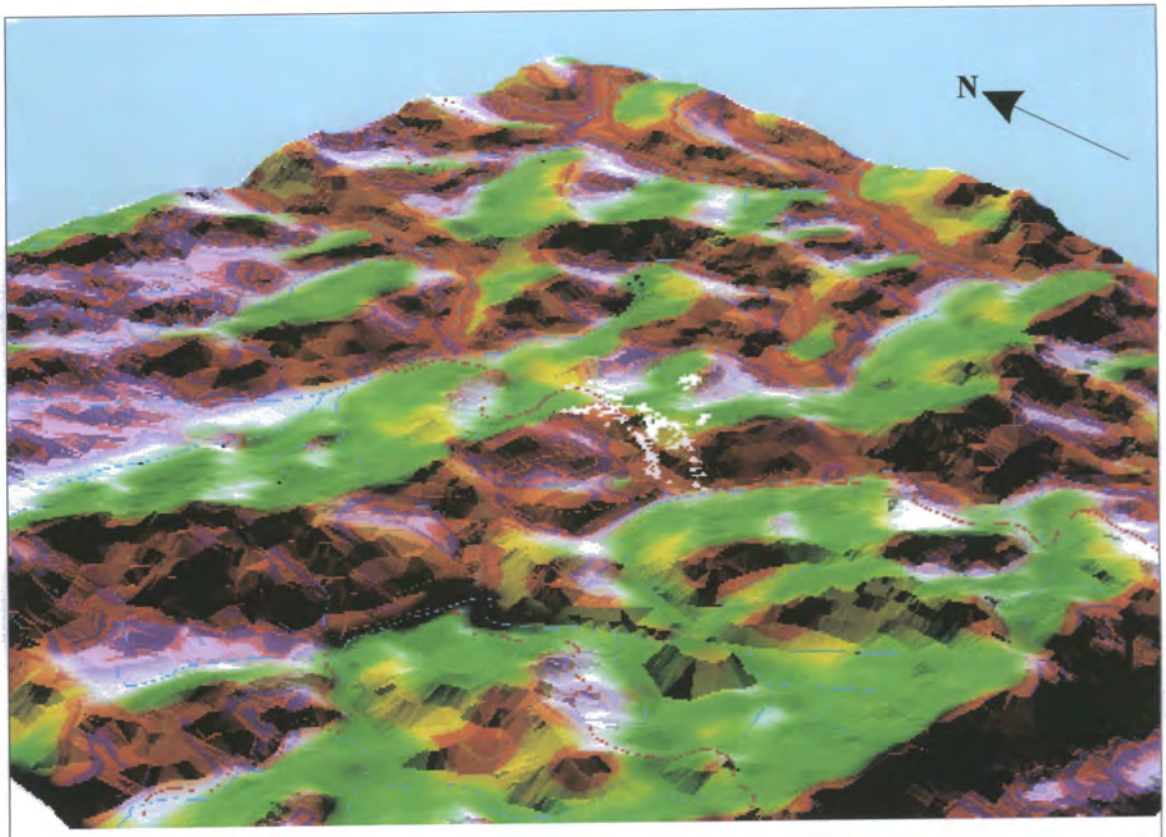


Figure A12. Surface representation displaying flow direction for the village area at 1650 m, overlaid with stream and road networks.



LEGEND

1	4	32
2	8	64
	16	128

Colours represent different flow directions
(See text in Chapter 3 for explanation)

Figure A13. **Grid showing accumulated flow overlaid with stream, road and house position coverages for villages located at 300 m, 600 m, 750 m and 1000 m**

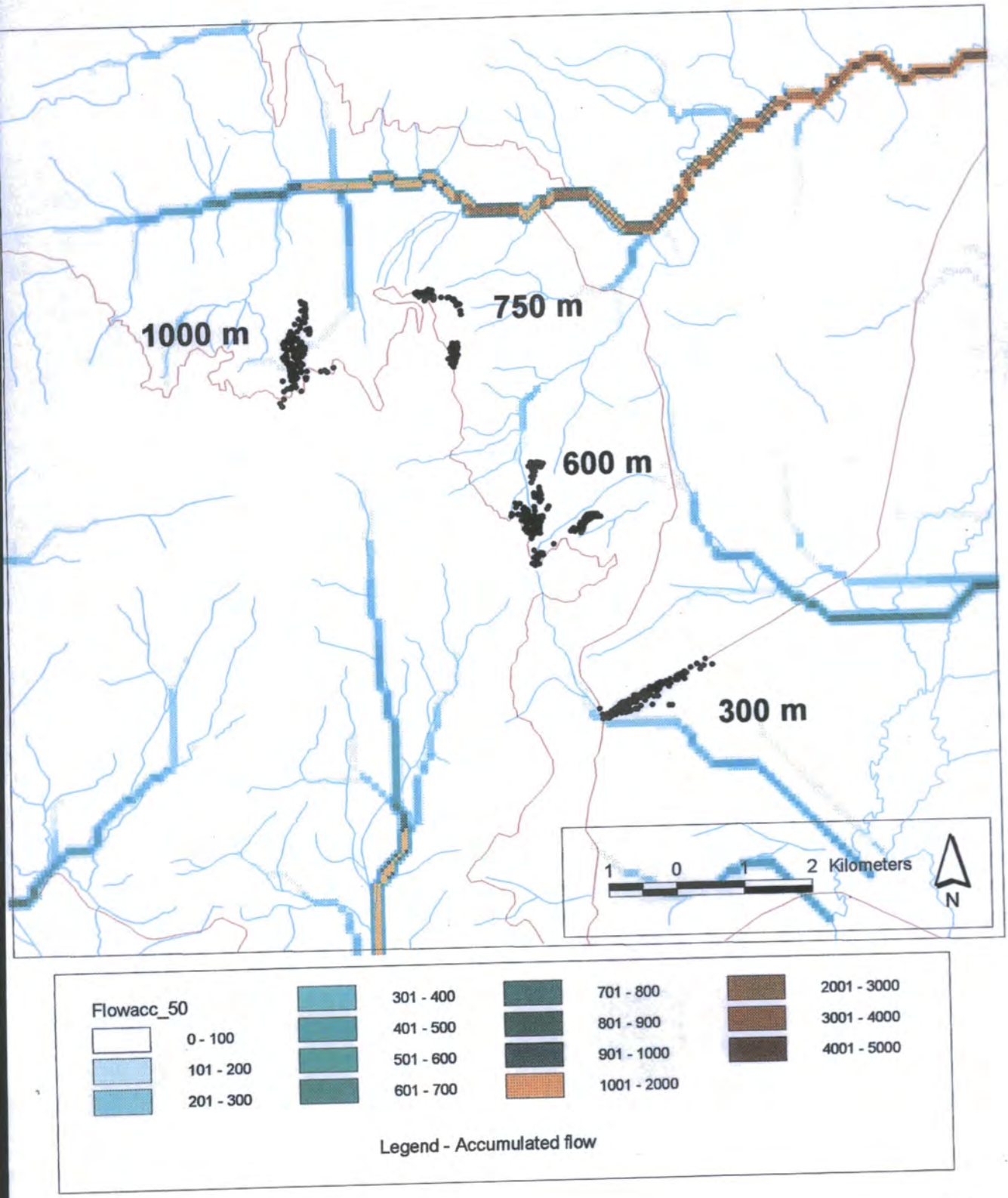


Figure A14. Grid showing accumulated flow overlaid with stream, road and house position coverages for the village located at 1200 m

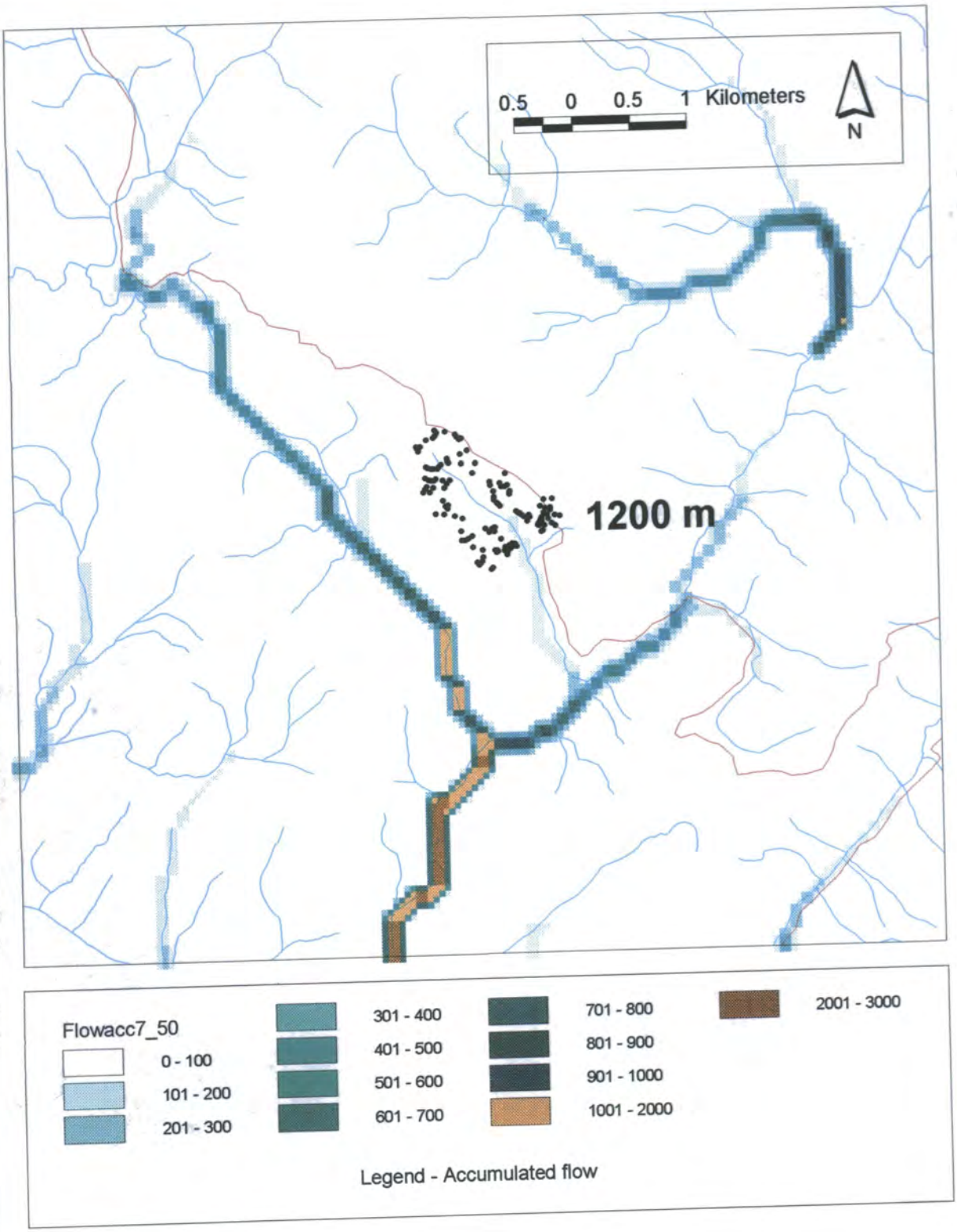


Figure A15. Grid showing accumulated flow overlaid with stream, road and house position coverages for the village located at 1300 m

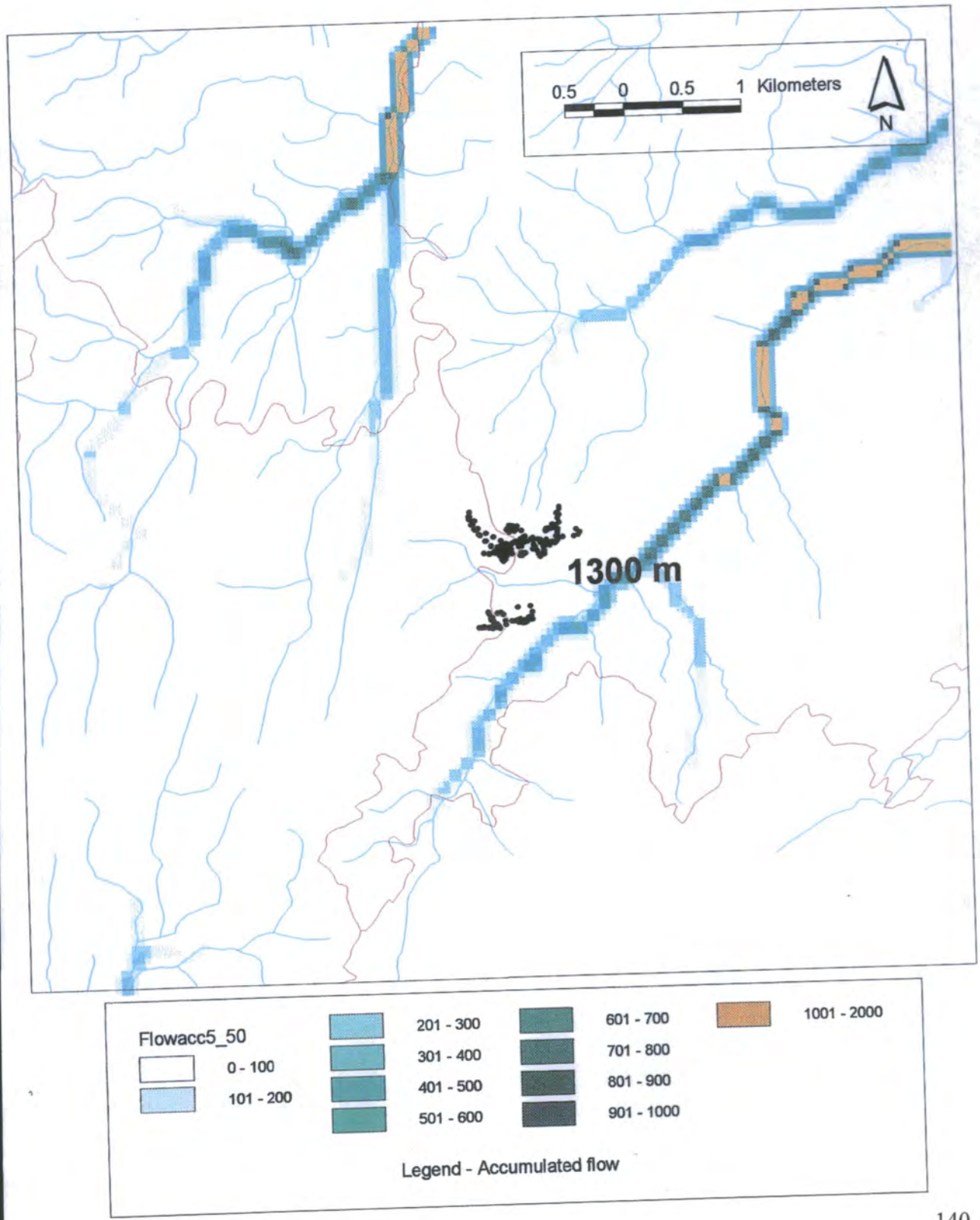


Figure A16. Grid showing accumulated flow overlaid with stream, road and house position coverages for the village located at 1650 m

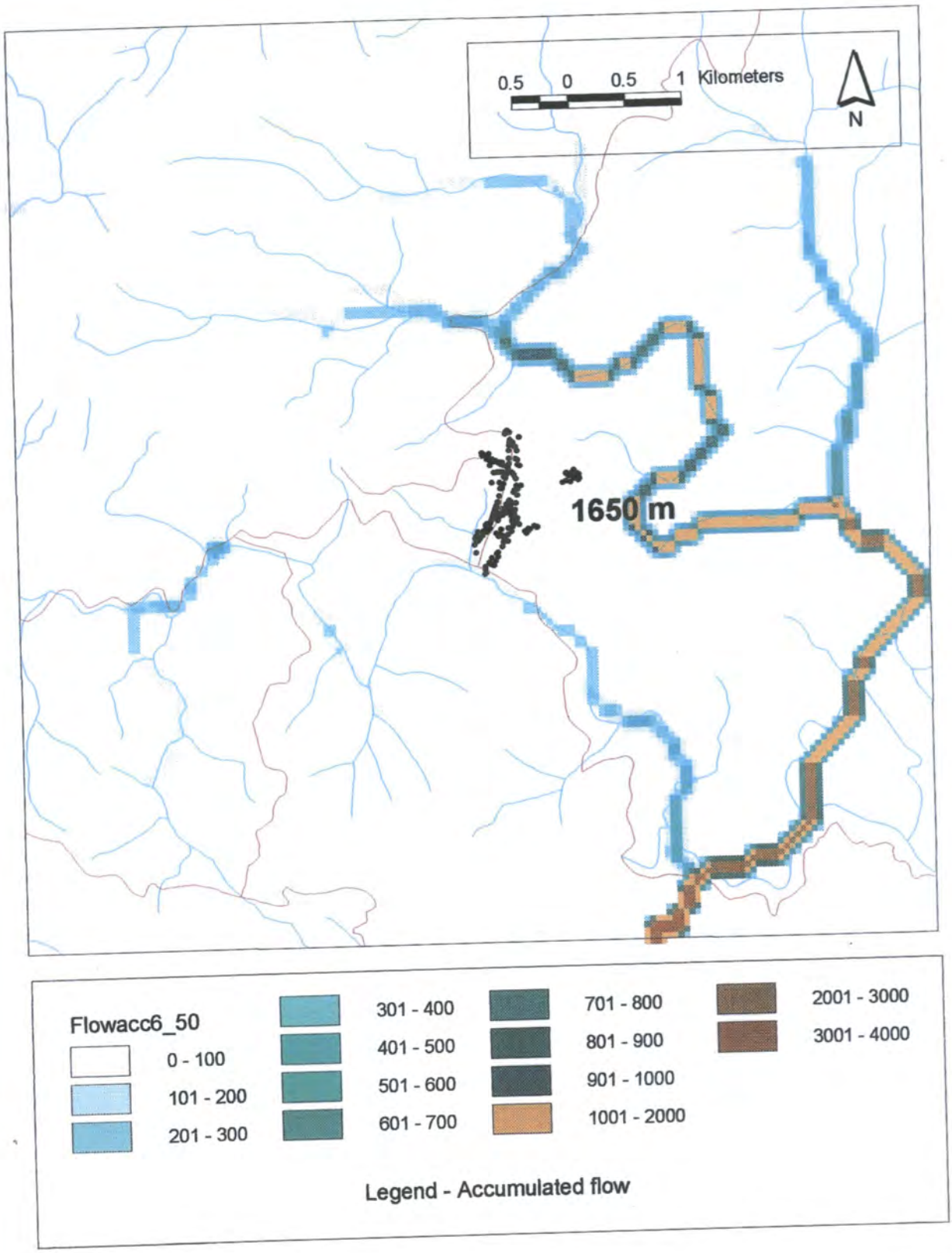


Figure A17. **Grid displaying slope angles overlaid with stream, road and house position coverages for villages located at 300 m, 600 m, 750 m, and 1000 m**

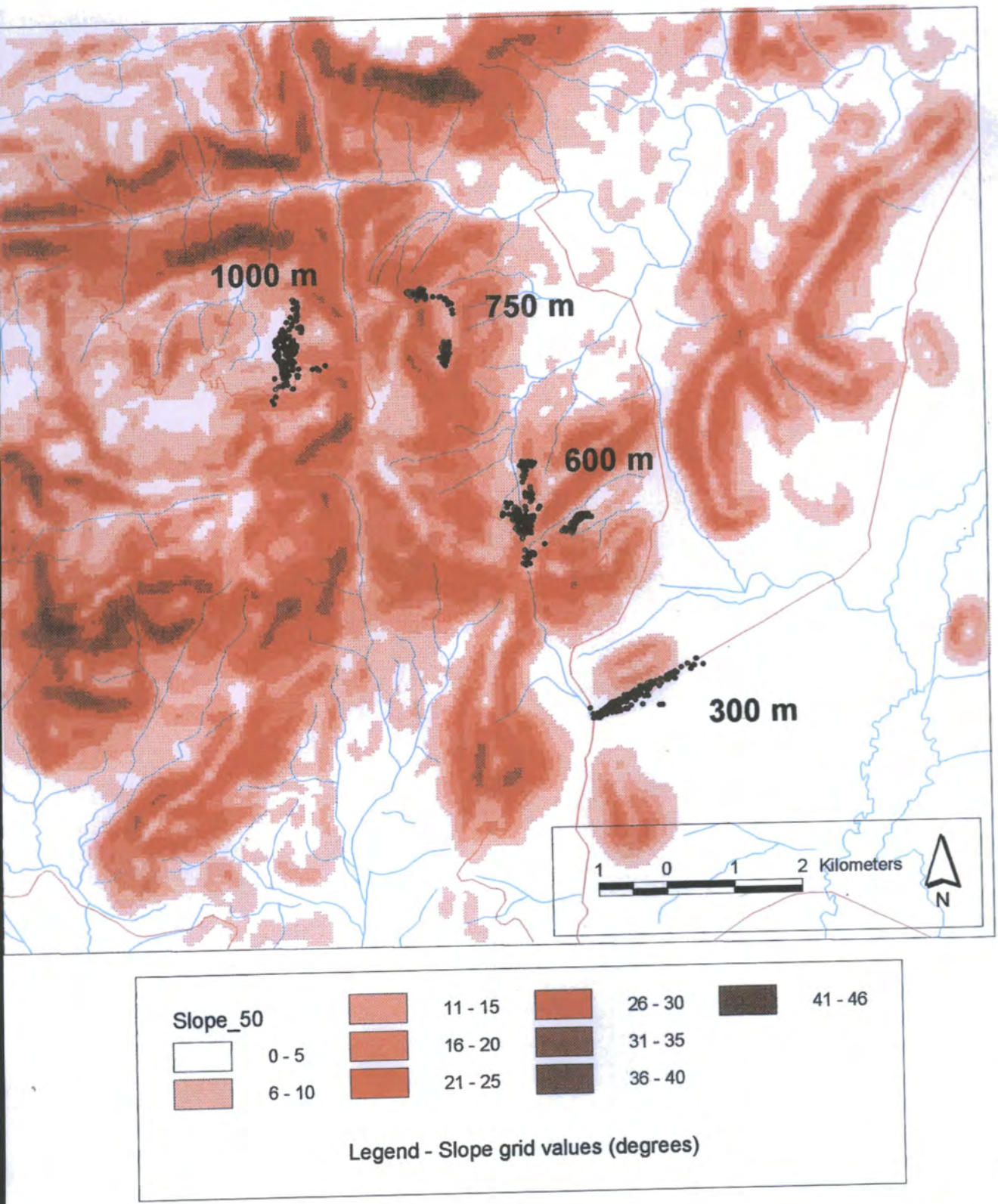


Figure A18. Grid displaying slope angles overlaid with stream, road and house position coverages for the village located at 1200 m

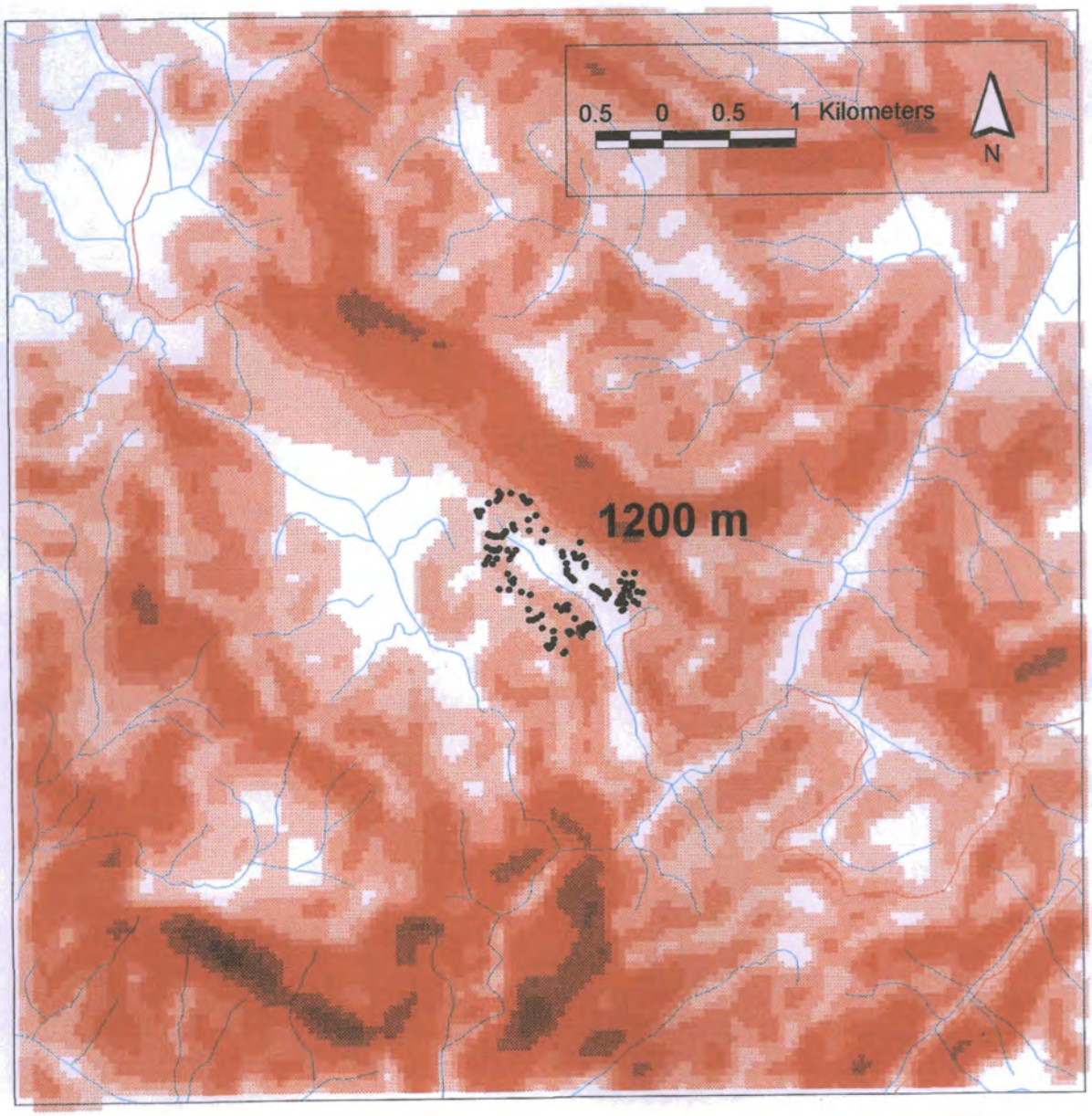


Figure A19. Grid displaying slope angles overlaid with stream, road and house position coverages for the village located at 1300 m

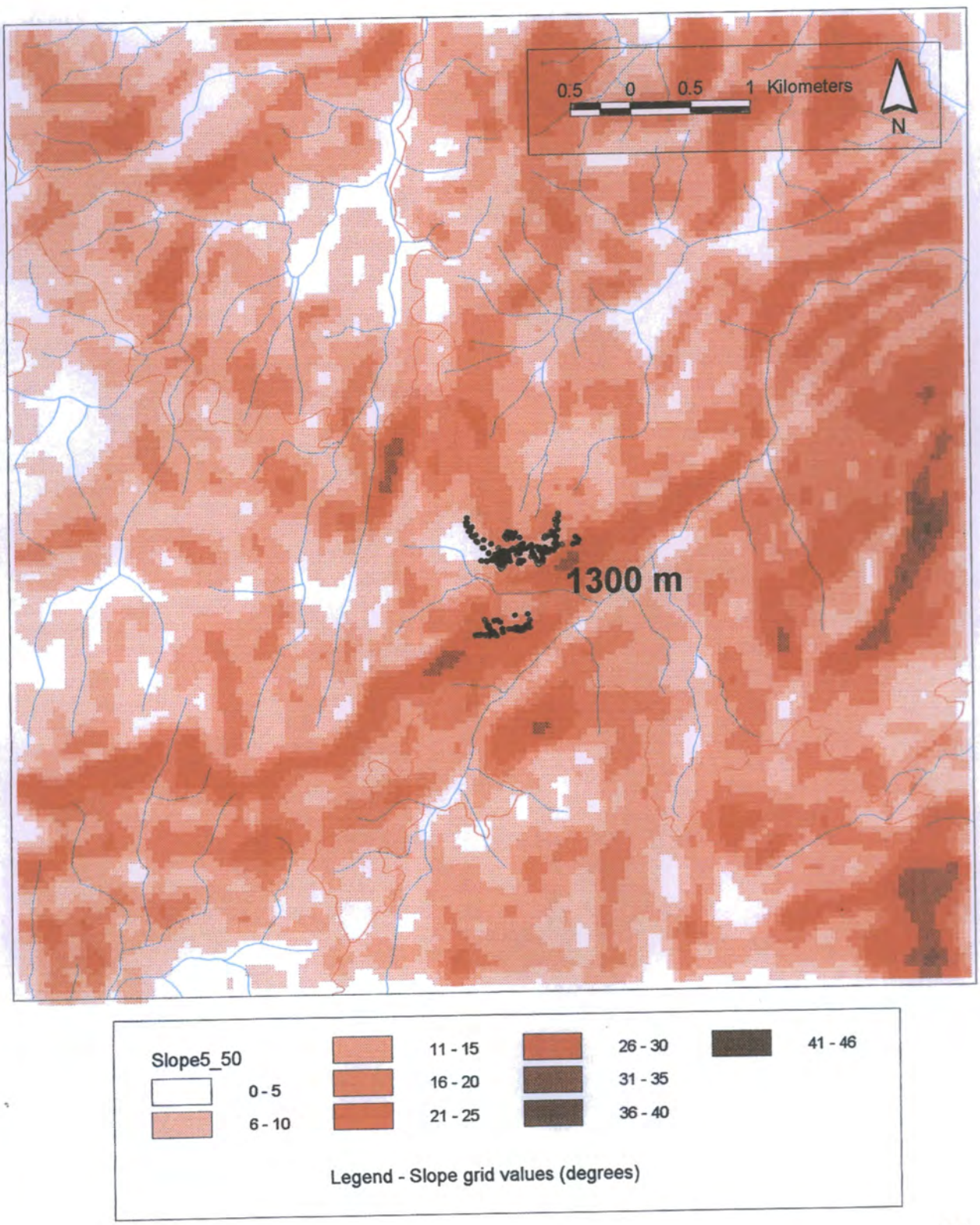


Figure A20. Grid displaying slope angles overlaid with stream, road and house position coverages for the village located at 1650 m

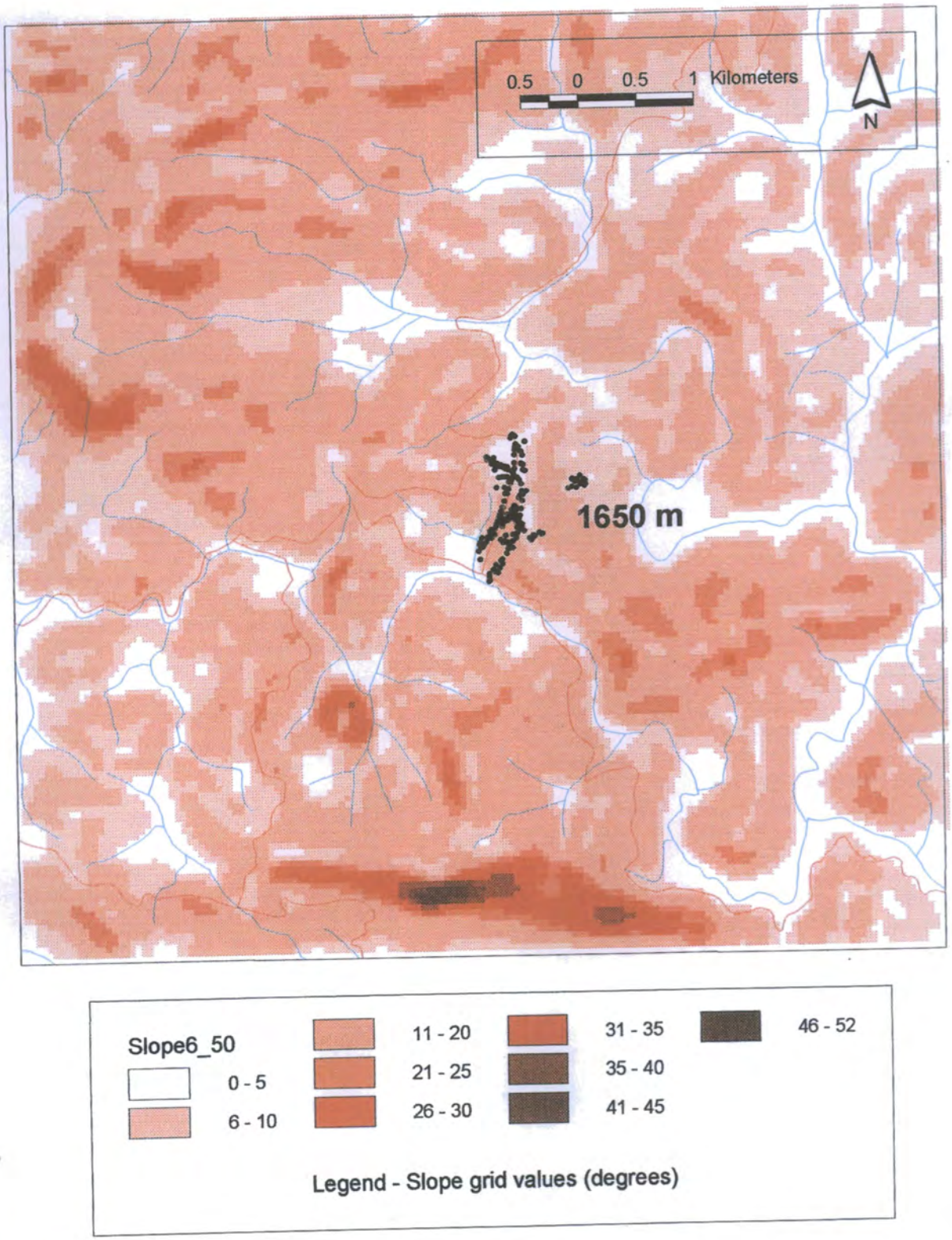
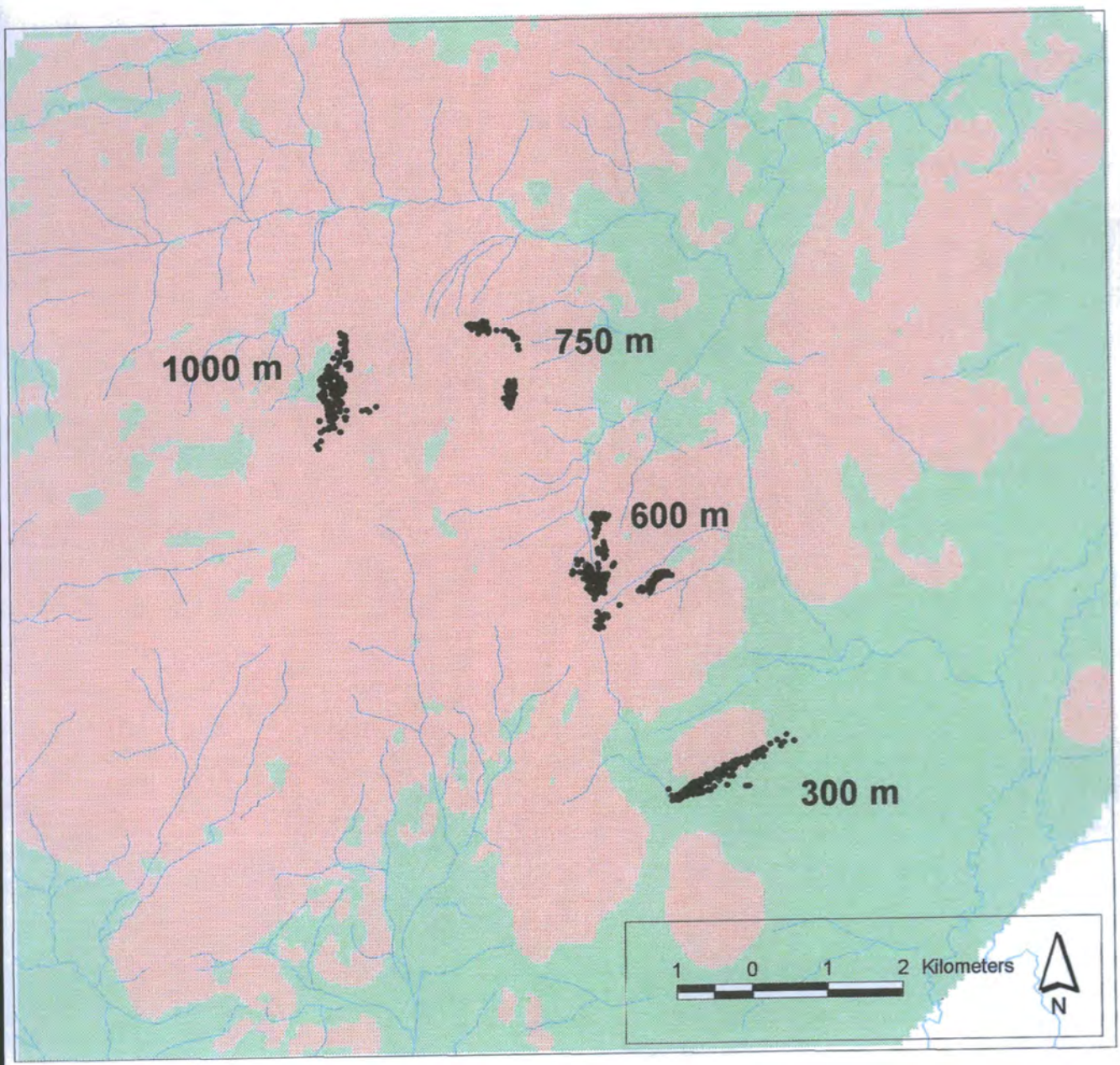


Figure A21. Grid displaying flat cells overlaid with stream and house position coverages for villages located at 300 m, 600 m, 750 m and 1000 m



Legend - Grid displaying flat cells

Flat

Non flat cells (> 5 degrees)

Flat cells (<=5 degrees)

Figure A22. Grid displaying flat cells overlaid with stream and house position coverages for the village located at 1200 m

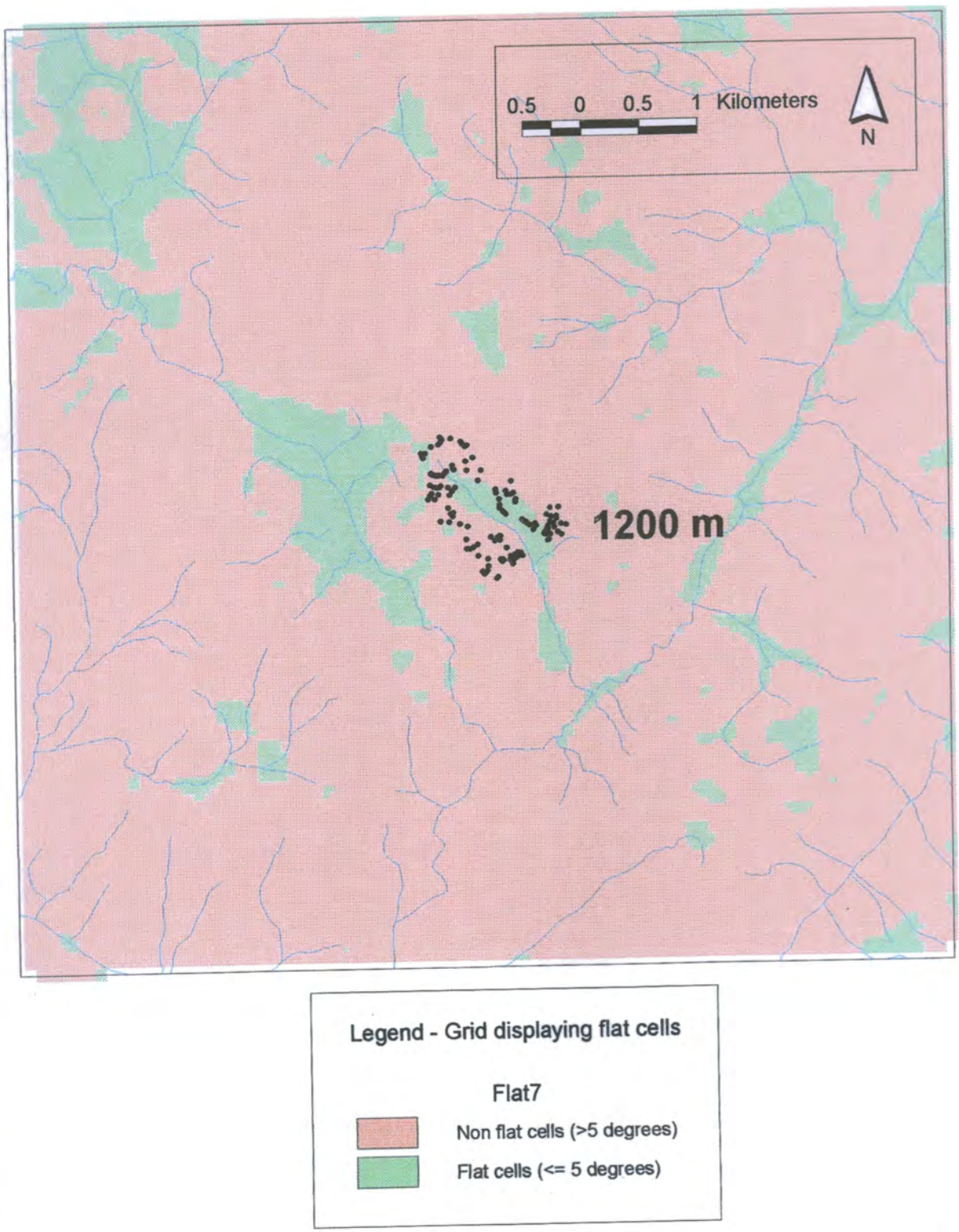


Figure A23. Grid displaying flat cells overlaid with stream and house position coverages for the village located at 1300 m

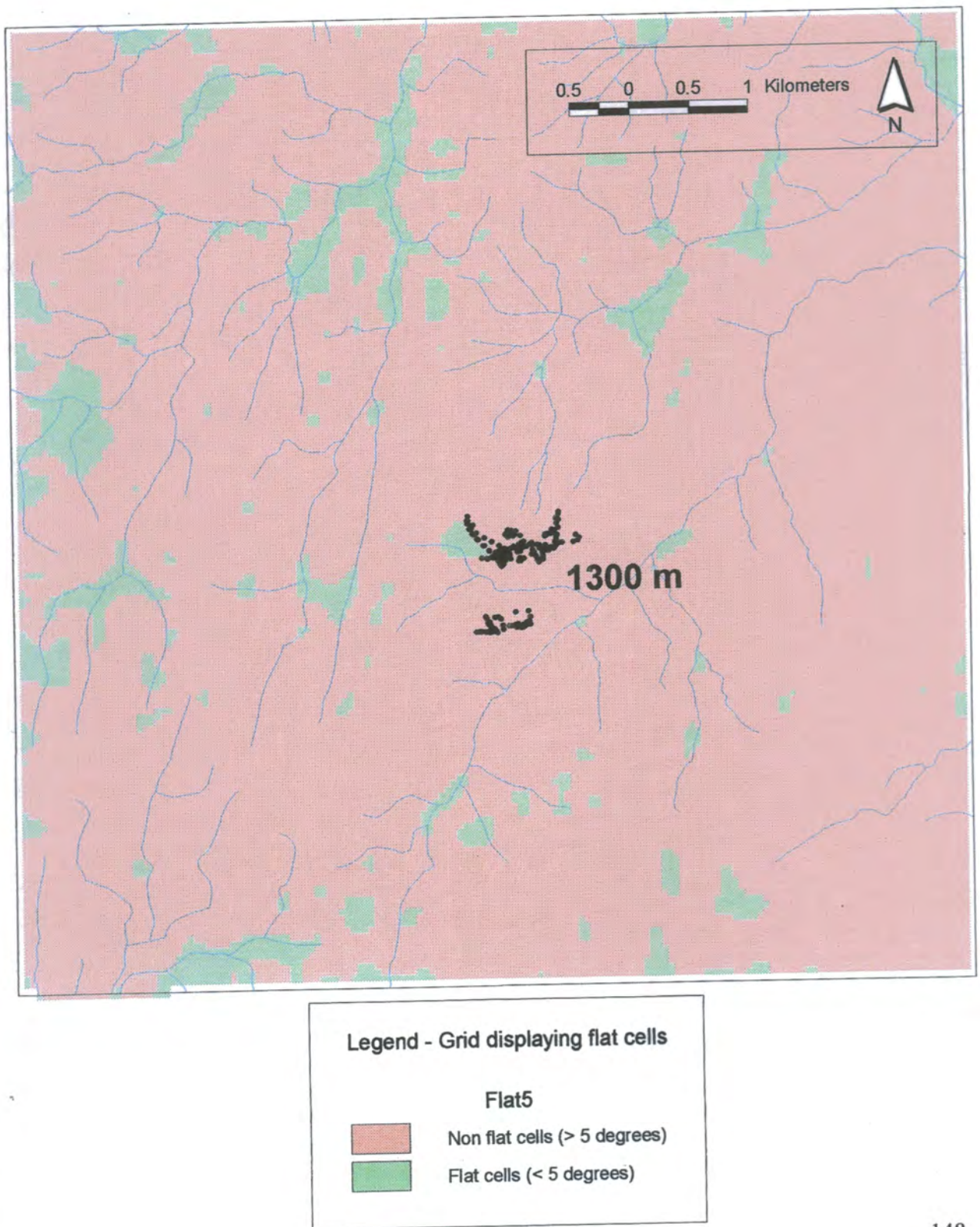


Figure A24. Grid displaying flat cells overlaid with stream and house position coverages for the village located at 1650 m

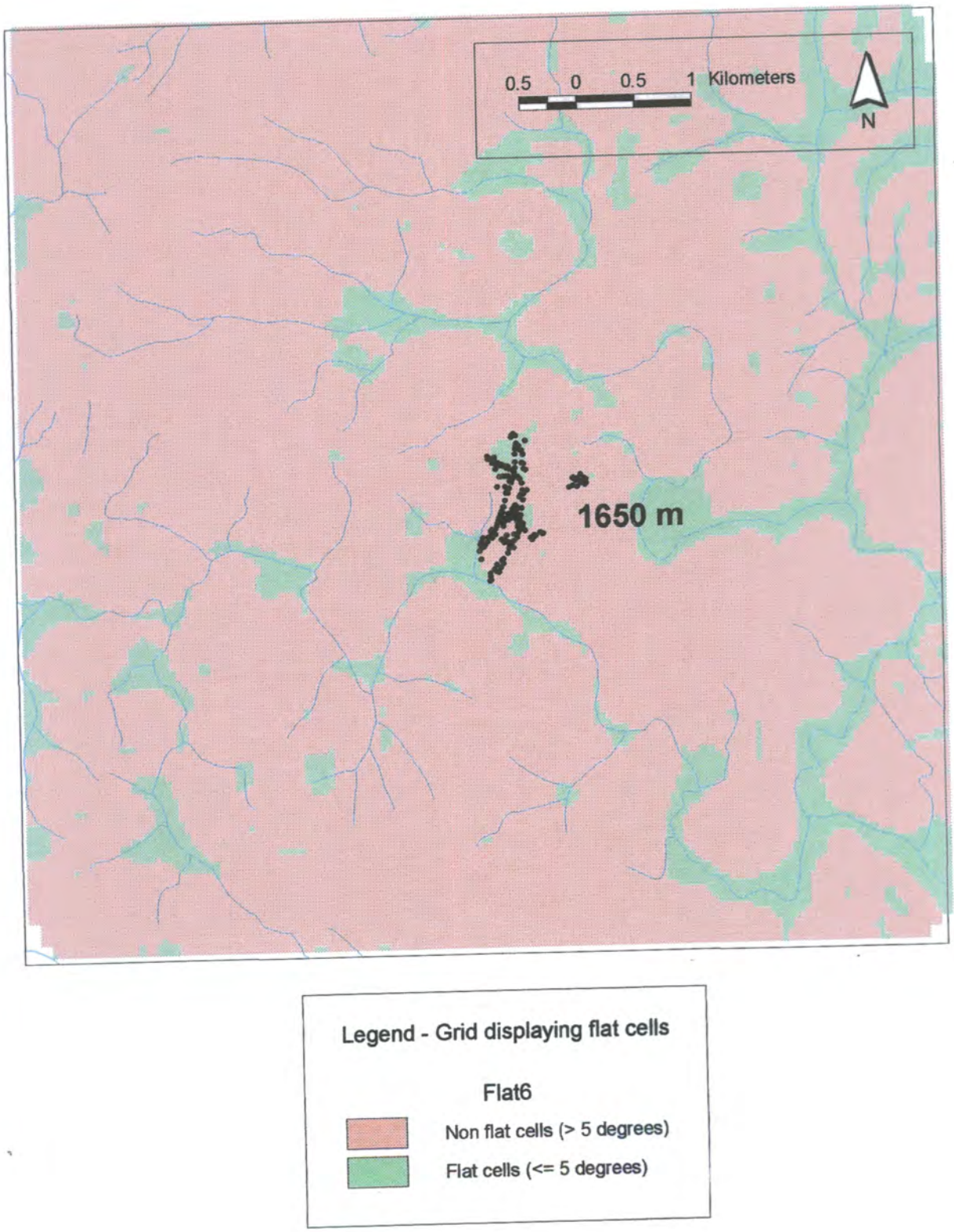


Figure A25. Surface representation displaying flat areas and their flow accumulation value for villages situated at 300 m, 600 m, 750 m and 1000 m, overlaid with the stream network coverage.

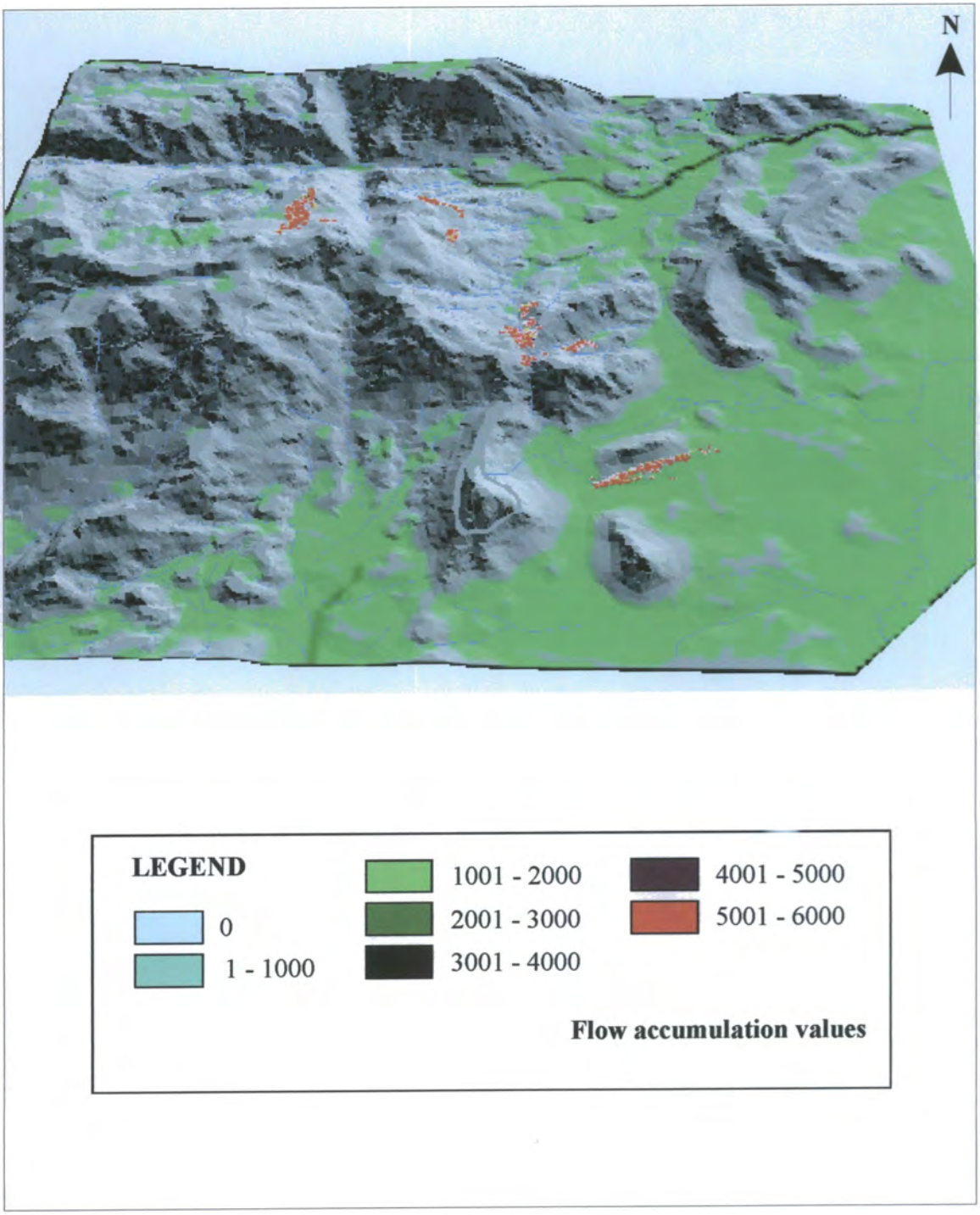
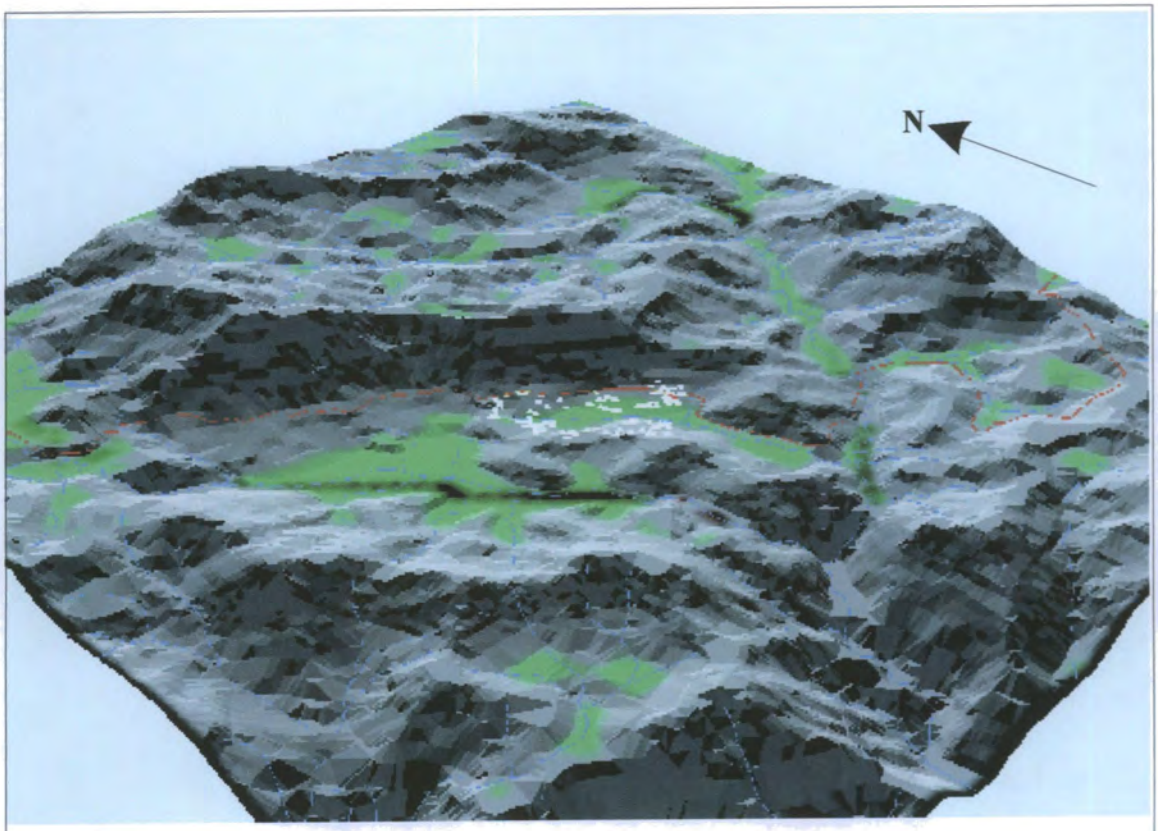


Figure A26. Surface representation displaying flat areas and their flow accumulation value for the village area located at 1200 m, overlaid with the stream and road network coverages.



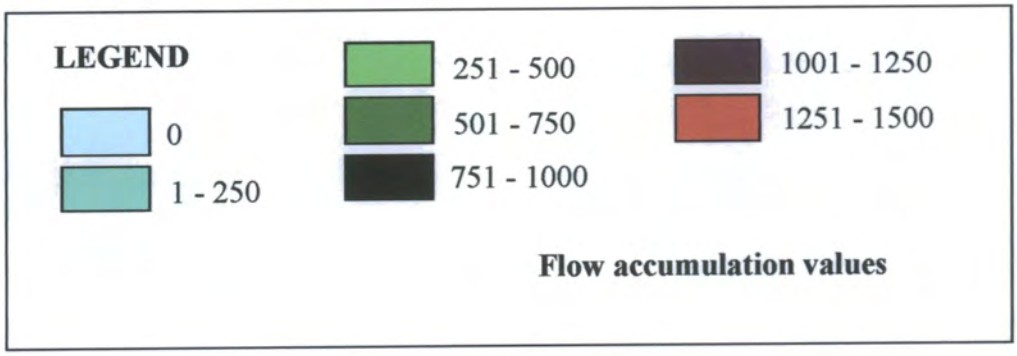


Figure A27. Surface representation displaying flat areas and their flow accumulation value for the village area at 1300 m, overlaid with the stream and road network coverages.

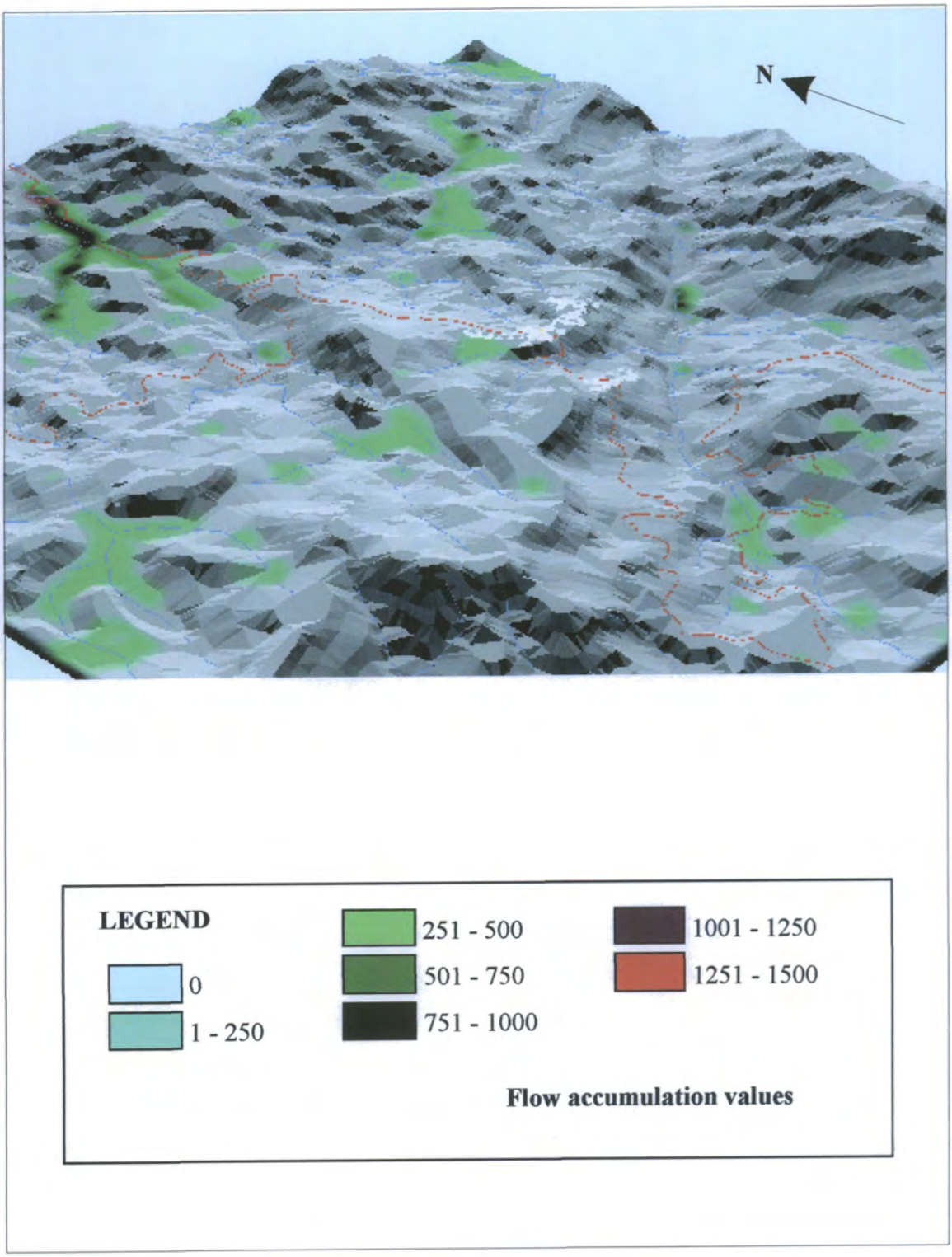
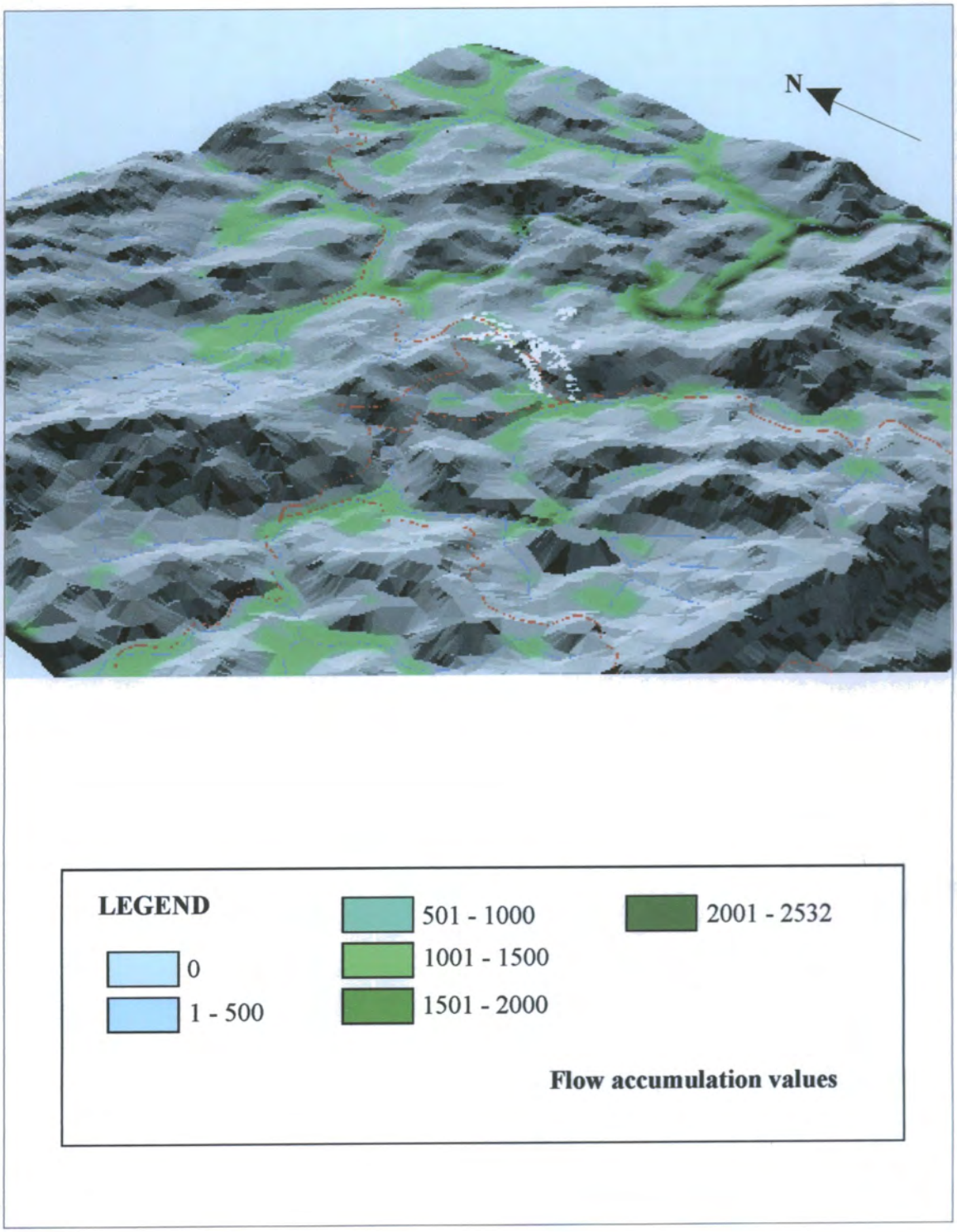


Figure A28. Surface representation displaying flat areas and their flow accumulation value for the village area at 1650 m, overlaid with stream and road network coverages.



APPENDIX B

Tables displaying results from non-parametric analysis carried out to compare all the topographic independent variables for the cases and controls

Table B1 - Comparisons between cases and controls using the Mann-Whitney *U* test, determined by enlarged spleen, for households of all ages

I V*	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (51)		CASES (104)			CONTROLS (26)		CASES (114)		
	median	IQ range	median	IQ range	Sig.**	median	IQ range	median	IQ range	Sig.**
Altitude (m)	335.0	11.00	335.0	10.00	0.229	600.0	14.00	596.0	60.75	0.055
Slope (°)	6.0	2.00	5.0	2.00	**0.028**	11.0	4.25	11.0	5.50	0.476
Flowacc	2.0	6.00	2.0	6.00	0.913	0.5	2.25	0.0	2.00	0.828
Flow50	13.0	22.00	12.5	23.75	0.779	4.5	11.25	5.0	11.00	0.662
Flow100	39.0	41.00	48.0	73.00	0.962	17.0	23.75	18.0	36.50	0.686
Flow250	376.0	3280.00	763.0	5637.75	0.080	276.0	357.50	319.0	611.50	0.148
Flow500	12678.0	14004.00	14938.0	7181.00	0.078	1953.5	760.00	2163.0	1299.00	0.088
Flat50	2.0	3.00	4.0	4.00	0.123	0.0	0.25	0.0	0.00	0.590
Flat100	8.0	6.00	8.0	7.00	0.189	0.0	1.50	0.0	1.00	0.908
Flat250	47.0	31.00	59.5	28.00	0.067	4.5	2.00	6.0	6.00	0.128
Flat500	195.0	61.00	215.5	67.00	0.090	18.5	7.25	21.0	13.25	0.066
Swamp	0.0	5.00	0.5	6.00	0.425	0.0	0.00	0.0	0.00	1.000
Swamp100	30.0	70.00	43.5	161.00	0.226	0.0	0.00	0.0	0.00	0.698
Swamp200	147.0	792.00	292.0	2364.75	0.121	2.0	2.00	2.0	2.00	0.974
Swamp300	1294.0	6075.00	3409.0	7015.50	0.103	2.0	2.00	2.0	2.00	0.194
Swamp400	6506.0	10272.00	9707.0	9504.50	0.096	2.0	0.00	2.0	90.25	0.153
Swamp500	12559.0	13412.00	14790.0	7646.50	0.086	2.0	0.00	2.0	318.00	0.022
	VILLAGE 3 (750M)					**VILLAGE 4 (1000m)**				
	CONTROLS (17)		CASES (60)			CONTROLS (47)		CASES (79)		
Altitude (m)	781.0	67.50	757.0	61.75	0.395	1040.0	52.00	1029.0	32.00	0.091
Slope (°)	18.0	12.00	17.0	10.00	0.350	11.0	4.00	10.0	4.00	0.859
Flowacc	1.0	3.50	1.0	4.00	0.741	5.0	17.00	5.0	18.00	0.757
Flow50	3.0	18.50	11.5	20.75	0.402	35.0	106.00	45.0	198.00	0.864
Flow100	24.0	53.50	24.5	57.25	0.445	163.0	543.00	138.0	759.00	0.854
Flow250	119.0	394.50	287.5	397.00	0.956	2555.0	2907.00	2574.0	2969.00	0.726
Flow500	681.0	1388.50	1323.0	1319.75	0.499	8005.0	1677.00	8453.0	2499.00	0.387
Flat50					1.000	0.0	0.00	0.0	0.00	0.473
Flat100					1.000	0.0	0.00	0.0	1.00	0.637
Flat250					1.000	5.0	13.00	6.0	20.00	0.289
Flat500	4.0	11.00	0.0	6.75	0.230	52.0	46.00	56.0	40.00	0.126
Swamp					1.000	0.0	0.00	0.0	0.00	0.521
Swamp100					1.000	0.0	0.00	0.0	0.00	0.671
Swamp200					1.000	0.0	123.00	1.0	132.00	0.547
Swamp300					1.000	231.0	549.00	230.0	549.00	0.828
Swamp400					1.000	583.0	738.00	761.0	712.00	0.565
Swamp500					1.000	1397.0	961.00	1397.0	870.00	0.888
	VILLAGE 5 (1200m)					**VILLAGE 6 (1300m)**				
	CONTROLS (67)		CASES (74)			CONTROLS (137)		CASES (33)		
Altitude (m)	1208.0	45.00	1190.0	31.00	**0.002**	1360.0	35.50	1360.0	28.00	0.549
Slope (°)	9.0	6.00	8.0	7.25	0.052	15.0	10.00	11.0	10.00	0.245
Flowacc	1.0	5.00	1.0	7.00	0.484	0.0	0.00	0.0	0.00	0.117
Flow50	15.0	43.00	16.0	111.25	0.141	0.0	2.00	0.0	2.00	0.649
Flow100	54.0	100.00	72.0	342.75	**0.048**	3.0	10.50	2.0	6.50	0.346
Flow250	671.0	1355.00	1485.5	1755.50	**0.012**	126.0	400.00	91.0	446.00	0.727
Flow500	4641.0	3786.00	5044.0	3737.00	0.243	2587.0	6422.00	2511.0	7905.00	0.334
Flat50	0.0	1.00	1.0	4.00	**0.010**	0.0	1.00	0.0	1.00	0.130
Flat100	1.0	4.00	4.0	7.25	**0.009**	0.0	3.00	0.0	3.00	0.444
Flat250	20.0	18.00	25.0	22.25	**0.002**	8.0	8.00	7.0	8.00	0.946
Flat500	83.0	25.00	91.5	14.50	**0.001**	13.0	30.00	11.0	17.00	0.356
Swamp	0.0	0.00	0.0	8.00	**0.002**	0.0	0.00	0.0	0.00	0.392
Swamp100	0.0	27.00	1.0	245.25	**0.014**	0.0	0.00	0.0	0.00	0.784
Swamp200	73.0	302.00	237.0	929.00	**0.002**	0.0	0.00	0.0	0.00	0.694
Swamp300	386.0	1415.00	1195.0	1563.00	**0.001**	0.0	7.00	0.0	0.00	0.254
Swamp400	915.0	2072.00	2200.0	1915.25	**0.005**	0.0	96.50	0.0	0.50	0.107
Swamp500	2197.0	2404.00	2926.0	2109.00	**0.031**	50.0	265.00	26.0	229.50	0.514

Table B1 continued

I V*	VILLAGE 7 (1750m)					ALL VILLAGES				
	CONTROLS (131)		CASES (12)			CONTROLS (476)		CASES (476)		
	median	IQ range	median	IQ range	Sig.**	median	IQ range	median	IQ range	Sig.**
Altitude (m)	1640.0	65.00	1640.0	26.50	0.443	1344.0	552.00	751.0	670.50	**< 0.0001**
Slope (°)	7.0	10.00	6.0	10.50	0.867	10.0	9.00	9.5	7.00	0.258
Flowacc	0.0	1.00	0.0	1.75	0.745	0.0	2.00	1.0	5.00	**< 0.0001**
Flow50	1.0	5.00	2.0	5.25	0.885	3.0	16.00	9.0	30.00	**< 0.0001**
Flow100	6.0	30.00	8.0	14.50	0.770	12.0	54.00	30.0	101.75	**< 0.0001**
Flow250	145.0	421.00	142.5	468.25	0.754	291.0	697.25	523.5	2045.50	**< 0.0001**
Flow500	1774.0	3791.00	1503.0	4923.50	0.853	2635.0	5548.25	3409.5	6714.50	**0.024**
Flat50	1.0	4.00	1.0	4.00	0.776	0.0	1.00	0.0	1.00	0.965
Flat100	3.0	8.00	3.5	9.00	0.924	1.0	4.00	1.0	5.00	0.493
Flat250	18.0	25.00	20.0	25.00	0.833	11.0	22.00	10.0	29.00	0.933
Flat500	68.0	21.00	69.0	25.75	0.683	51.0	59.00	48.5	84.00	0.241
Swamp	0.0	0.00	0.0	0.00	0.243	0.0	0.00	0.0	0.00	**0.001**
Swamp100	0.0	2.00	0.0	2.75	0.575	0.0	2.00	0.0	11.00	**0.016**
Swamp200	2.0	6.00	2.0	6.00	0.997	0.0	58.25	2.0	220.25	**< 0.0001**
Swamp300	9.0	114.00	7.0	336.00	0.745	9.0	324.50	121.5	965.25	**< 0.0001**
Swamp400	54.0	527.00	31.5	1638.75	0.918	123.0	822.75	381.0	2082.25	**0.004**
Swamp500	480.0	2798.00	290.5	3043.75	0.974	448.5	1929.00	733.0	3091.25	**0.041**

***IV** independent variable
****Sig.** Asymptotic significance (2-tailed)

Values highlighted in grey are significant

Table B2 - Comparisions between cases and controls using the Mann-Whitney *U* test, determined by pfcount, for households of all ages

I V*	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (41)		CASES (114)			CONTROLS (27)		CASES (113)		
	median	IQ range	median	IQ range	Sig.**	median	IQ range	median	IQ range	Sig.**
Altitude (m)	335.0	9.00	335.0	10.00	0.368	600.0	36.00	596.0	41.50	0.113
Slope (°)	5.0	2.00	5.0	2.00	0.524	11.0	5.00	11.0	5.00	0.367
Flowacc	2.0	6.00	2.0	6.00	0.387	1.0	2.00	0.0	2.00	0.461
Flow50	17.0	21.50	12.0	24.00	0.271	5.0	10.00	4.0	11.00	0.384
Flow100	48.0	48.00	37.5	72.25	0.439	18.0	25.00	18.0	32.50	0.939
Flow250	404.0	5402.00	702.0	5633.50	0.362	247.0	449.00	319.0	534.50	0.182
Flow500	12678.0	8384.50	14938.0	8229.50	0.380	1997.0	907.00	2163.0	1168.00	0.274
Flat50	4.0	4.00	4.0	4.00	0.687	0.0	1.00	0.0	0.00	0.309
Flat100	8.0	8.00	8.0	7.00	0.889	0.0	3.00	0.0	1.00	0.614
Flat250	47.0	32.50	57.0	28.50	0.251	5.0	4.00	6.0	5.50	0.265
Flat500	195.0	68.50	215.0	56.25	0.227	18.0	8.00	21.0	13.00	0.147
Swamp	1.0	5.50	0.0	5.25	0.643	0.0	0.00	0.0	0.00	1.000
Swamp100	42.0	96.00	42.0	163.00	0.814	0.0	0.00	0.0	0.00	0.310
Swamp200	147.0	2174.50	258.0	2369.00	0.478	2.0	2.00	2.0	2.00	0.700
Swamp300	1358.0	6826.00	3305.5	6902.00	0.387	2.0	2.00	2.0	2.00	0.605
Swamp400	6624.0	9923.00	9395.0	9810.50	0.392	2.0	19.00	2.0	78.00	0.822
Swamp500	12559.0	8590.00	14790.0	8622.00	0.436	2.0	131.00	2.0	303.00	0.210

Table B2 continued

I V*	VILLAGE 3 (750M)					VILLAGE 4 (1000m)				
	CONTROLS (11)		CASES (66)			CONTROLS (107)		CASES (21)		
	median	IQ range	median	IQ range	Sig.**	median	IQ range	median	IQ range	Sig.**
Altitude (m)	781.0	53.00	757.0	69.00	0.105	1040.0	59.00	1029.0	33.00	0.077
Slope (°)	18.0	15.00	17.0	10.00	0.322	11.0	5.50	10.0	4.00	0.739
Flowacc	0.0	4.00	1.0	4.00	0.306	3.0	16.50	5.0	18.00	0.155
Flow50	3.0	21.00	11.5	21.75	0.310	26.0	71.00	49.0	198.00	0.207
Flow100	11.0	64.00	25.0	53.75	0.199	99.0	203.00	157.0	759.00	0.455
Flow250	119.0	381.00	278.0	399.00	0.959	2127.0	2447.00	2574.0	2928.00	0.176
Flow500	711.0	1352.00	1251.0	1315.25	0.304	7570.0	3064.50	8409.0	1625.00	0.111
Flat50					1.000	0.0	0.00	0.0	0.00	0.547
Flat100					1.000	0.0	0.00	0.0	1.00	0.356
Flat250					1.000	5.0	13.00	6.0	15.00	0.355
Flat500	4.0	11.00	0.0	6.25	0.169	48.0	60.50	56.0	36.00	0.302
Swamp					1.000	0.0	0.00	0.0	0.00	0.979
Swamp100					1.000	0.0	0.00	0.0	0.00	0.412
Swamp200					1.000	0.0	8.00	1.0	170.00	0.240
Swamp300					1.000	123.0	499.00	243.0	549.00	0.193
Swamp400					1.000	583.0	898.50	752.0	634.00	0.221
Swamp500					1.000	1122.0	1164.00	1397.0	870.00	0.543

	VILLAGE 5 (1200m)					VILLAGE 6 (1300m)				
	CONTROLS (19)		CASES (122)			CONTROLS (39)		CASES (133)		
Altitude (m)	1210.0	40.00	1194.5	37.00	0.285	1360.0	36.00	1360.0	34.00	0.605
Slope (°)	10.0	7.00	8.0	6.00	0.200	15.0	9.00	14.0	10.00	0.438
Flowacc	1.0	10.00	1.0	6.00	0.749	0.0	0.00	0.0	0.00	0.330
Flow50	10.0	38.00	15.5	47.25	0.522	1.0	2.00	0.0	2.00	0.931
Flow100	42.0	67.00	68.0	218.00	0.252	3.0	11.00	3.0	7.00	0.588
Flow250	469.0	1490.00	1079.0	1651.00	0.212	122.0	459.00	130.0	409.00	0.910
Flow500	3954.0	4689.00	4743.0	3891.00	0.823	2539.0	6273.00	2573.0	6422.00	0.785
Flat50	0.0	2.00	0.5	3.00	0.289	0.0	0.00	0.0	1.00	0.406
Flat100	1.0	5.00	2.5	6.25	0.250	0.0	1.00	0.0	3.00	0.449
Flat250	19.0	17.00	24.0	17.25	0.286	6.0	8.00	8.0	8.00	0.491
Flat500	87.0	25.00	90.0	20.00	0.393	12.0	28.00	12.0	30.00	0.545
Swamp	0.0	0.00	0.0	0.00	0.527	0.0	0.00	0.0	0.00	0.345
Swamp100	0.0	51.00	0.0	89.50	0.225	0.0	0.00	0.0	0.00	0.043
Swamp200	27.0	238.00	166.0	826.00	0.233	0.0	0.00	0.0	0.00	0.127
Swamp300	386.0	1441.00	928.0	1542.50	0.128	0.0	1.00	0.0	8.50	0.443
Swamp400	1313.0	2200.00	1972.5	2039.00	0.255	0.0	63.00	0.0	102.00	0.667
Swamp500	2313.0	2497.00	2773.0	2249.75	0.305	123.0	265.00	26.0	265.00	0.256

	VILLAGE 7 (1750m)					ALL VILLAGES				
	CONTROLS (81)		CASES (63)			CONTROLS (239)		CASES (718)		
Altitude (m)	1640.0	36.00	1640.0	66.00	0.652	1254.0	1023.00	1040.0	720.00	**< 0.0001**
Slope (°)	7.0	11.50	7.0	10.00	0.995	9.0	10.00	10.0	8.00	0.113
Flowacc	0.0	1.00	0.0	1.00	0.529	0.0	3.00	1.0	4.00	0.204
Flow50	2.0	5.50	1.0	5.00	0.446	4.0	16.00	6.0	27.00	**0.034**
Flow100	6.0	38.00	7.0	15.00	0.651	15.0	54.00	24.5	87.25	**0.009**
Flow250	119.0	513.50	145.0	267.00	0.687	278.0	760.00	474.0	1543.00	**< 0.0001**
Flow500	1600.0	3712.00	1837.0	4088.00	0.268	2539.0	5884.00	3035.5	6431.00	**0.025**
Flat50	1.0	4.00	1.0	4.00	0.922	0.0	3.00	0.0	1.00	0.092
Flat100	3.0	9.00	3.0	6.00	0.812	1.0	6.00	1.0	4.00	0.060
Flat250	18.0	25.50	13.0	23.00	0.270	12.0	28.00	10.0	25.00	0.025
Flat500	68.0	15.00	68.0	23.00	0.491	67.0	63.00	45.0	73.00	0.094
Swamp	0.0	0.00	0.0	0.00	0.573	0.0	0.00	0.0	0.00	0.780
Swamp100	0.0	1.00	0.0	2.00	0.420	0.0	2.00	0.0	4.00	0.730
Swamp200	2.0	6.00	2.0	6.00	0.968	2.0	82.00	2.0	153.00	0.313
Swamp300	9.0	158.50	15.0	92.00	0.328	9.0	423.00	30.0	594.75	0.383
Swamp400	40.0	477.00	54.0	1127.00	0.368	115.0	1238.00	251.0	1297.75	0.970
Swamp500	352.0	2538.00	539.0	3053.00	0.482	480.0	2828.00	623.0	2241.50	0.568

Table B3 - Comparisons between cases and controls using the Mann_Whitney *U* test, determined by enlarged spleen, and under 10 households

IV *	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (32)		CASES (67)			CONTROLS (14)		CASES (81)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	335.0	9.50	335.0	15.00	0.542	603.0	43.50	600.0	34.50	0.301
Slope (°)	5.0	2.00	5.0	2.00	0.926	11.5	4.25	11.0	7.00	0.996
Flowacc	1.0	4.75	2.0	6.00	0.453	1.0	2.25	0.0	2.00	0.804
Flow50	11.0	26.00	13.0	24.00	0.875	4.5	10.25	5.0	11.00	0.546
Flow100	33.5	68.50	39.0	57.00	0.672	17.0	43.50	18.0	37.50	0.825
Flow250	2934.0	5799.00	645.0	5493.00	0.198	276.0	657.75	319.0	556.50	0.475
Flow500	15351.0	5209.00	14319.0	10968.00	0.244	1877.5	1138.50	2107.0	1273.50	0.462
Flat50	4.0	4.00	4.0	4.00	0.417	0.0	0.00	0.0	0.00	0.603
Flat100	9.0	6.75	8.0	8.00	0.364	0.0	1.00	0.0	1.00	0.491
Flat250	60.0	27.75	51.0	30.00	0.126	4.5	2.25	6.0	6.00	0.051
Flat500	227.0	70.50	202.0	59.00	0.099	18.0	6.50	21.0	14.00	0.098
Swamp	0.0	6.75	0.0	5.00	0.974	0.0	0.00	0.0	0.00	1.000
Swamp100	36.0	144.00	36.0	92.00	0.378	0.0	0.00	0.0	0.00	0.741
Swamp200	784.0	2593.50	213.0	2246.00	0.138	2.0	2.00	2.0	2.00	0.686
Swamp300	5672.5	7551.00	3142.0	6846.00	0.160	2.0	0.50	2.0	1.00	0.803
Swamp400	10481.0	8016.25	8686.0	10241.00	0.173	2.0	34.00	2.0	98.50	0.771
Swamp500	15265.0	5314.00	14231.0	10915.00	0.279	2.0	157.75	2.0	327.00	0.287
	VILLAGE 3 (750M)					VILLAGE 4 (1000m)				
	CONTROLS (21)		CASES (32)			CONTROLS (47)		CASES (45)		
Altitude (m)	796.0	59.00	751.0	71.25	**0.019**	1045.0	52.00	1027.0	40.00	**< 0.0001**
Slope (°)	22.0	10.50	17.0	12.00	0.127	10.0	4.00	10.0	3.50	0.484
Flowacc	1.0	3.00	1.0	10.50	0.799	6.0	17.00	6.0	26.50	0.674
Flow50	4.0	11.00	9.5	41.50	0.507	35.0	67.00	70.0	375.50	0.267
Flow100	19.0	33.00	25.0	62.25	0.476	129.0	203.00	213.0	950.50	0.394
Flow250	116.0	392.50	411.5	403.75	0.262	2102.0	2403.00	3113.0	2716.50	0.075
Flow500	679.0	1315.50	1819.0	1399.75	0.066	7889.0	3706.00	8516.0	2137.00	**0.004**
Flat50					1.000	0.0	0.00	0.0	0.00	0.543
Flat100					1.000	0.0	0.00	0.0	1.00	0.326
Flat250					1.000	5.0	12.00	6.0	13.00	0.095
Flat500	5.0	10.50	0.0	8.25	0.065	41.0	48.00	64.0	37.00	**0.002**
Swamp					1.000	0.0	0.00	0.0	0.00	0.183
Swamp100					1.000	0.0	0.00	0.0	0.00	0.718
Swamp200					1.000	0.0	170.00	0.0	123.00	0.867
Swamp300					1.000	231.0	550.00	230.0	453.00	0.674
Swamp400					1.000	560.0	810.00	761.0	634.00	0.346
Swamp500					1.000	1122.0	1043.00	1397.0	834.00	0.470
	VILLAGE 5 (1200m)					VILLAGE 6 (1300m)				
	CONTROLS (61)		CASES (53)			CONTROLS (110)		CASES (17)		
Altitude (m)	1203.0	38.00	1186.0	30.00	**0.012**	1360.0	37.00	1360.0	18.00	0.246
Slope (°)	8.0	5.00	8.0	8.00	0.354	16.0	9.75	11.0	11.00	0.145
Flowacc	1.0	6.00	1.0	7.00	0.685	0.0	0.00	0.0	0.00	0.065
Flow50	16.0	45.00	16.0	114.50	0.433	0.5	2.25	0.0	1.50	0.289
Flow100	57.0	153.00	69.0	362.50	0.125	3.0	11.00	2.0	6.50	0.069
Flow250	584.0	1313.00	1648.0	1700.00	**0.004**	211.0	426.50	90.0	258.50	0.030
Flow500	3914.0	3789.00	5091.0	4151.50	0.124	2599.0	10211.00	1929.0	1424.50	0.019
Flat50	0.0	1.50	1.0	4.00	0.180	0.0	0.00	0.0	1.00	0.143
Flat100	2.0	5.00	4.0	7.50	0.169	0.0	1.00	0.0	3.50	0.583
Flat250	21.0	18.00	25.0	22.50	**0.047**	6.0	8.00	7.0	4.50	0.953
Flat500	87.0	24.50	91.0	14.00	**0.024**	12.0	30.25	9.0	11.00	0.254
Swamp	0.0	0.00	0.0	8.00	0.051	0.0	0.00	0.0	0.00	0.577
Swamp100	0.0	65.50	1.0	247.50	0.209	0.0	0.00	0.0	0.00	0.175
Swamp200	73.0	456.00	264.0	911.00	**0.009**	0.0	0.00	0.0	0.00	0.058
Swamp300	381.0	1429.50	1273.0	1449.00	**0.001**	0.0	8.25	0.0	0.00	0.039
Swamp400	902.0	2016.00	2220.0	1907.00	**0.002**	0.0	123.00	0.0	0.00	0.013
Swamp500	2185.0	2131.50	3015.0	2341.00	**0.016**	108.0	265.00	0.0	50.00	0.039

Table B3 Continued

IV *	VILLAGE 7 (1750m)					ALL VILLAGES				
	CONTROLS (102)		CASES (7)			CONTROLS (387)		CASES (302)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	1640.0	66.00	1640.0	14.00	0.736	1320.0	516.00	705.0	673.00	**< 0.0001**
Slope (°)	7.5	12.00	6.0	8.00	0.449	11.0	10.00	9.0	7.00	**0.032**
Flowacc	0.0	1.00	0.0	1.00	0.487	0.0	3.00	1.0	5.00	**0.005**
Flow50	1.5	6.25	1.0	3.00	0.326	4.0	21.00	8.0	29.25	**< 0.0001**
Flow100	7.0	35.25	8.0	15.00	0.715	13.0	69.00	30.0	103.75	**< 0.0001**
Flow250	145.0	484.25	142.0	237.00	0.462	364.0	1009.00	514.5	1858.25	**< 0.0001**
Flow500	1837.0	4111.50	1173.0	956.00	0.121	2690.0	5612.00	3211.0	6683.25	0.400
Flat50	1.0	4.00	1.0	4.00	0.531	0.0	1.00	0.0	1.00	0.715
Flat100	3.0	9.00	4.0	9.00	0.744	1.0	5.00	1.0	5.25	0.721
Flat250	16.0	26.00	27.0	22.00	0.743	10.0	22.00	11.0	28.00	0.159
Flat500	68.0	21.00	68.0	26.00	0.446	47.0	65.00	55.5	83.50	**0.028**
Swamp	0.0	0.00	0.0	1.00	**0.050**	0.0	0.00	0.0	0.00	**0.004**
Swamp100	0.0	2.00	0.0	2.00	0.826	0.0	2.00	0.0	11.00	0.127
Swamp200	1.0	6.00	2.0	6.00	0.917	0.0	100.00	2.0	195.00	**< 0.0001**
Swamp300	9.0	174.75	7.0	9.00	0.249	9.0	386.00	121.5	965.00	**0.001**
Swamp400	96.0	1150.50	9.0	45.00	0.093	133.0	962.00	448.0	2194.50	**0.007**
Swamp500	539.0	2872.75	81.0	343.00	0.094	480.0	1929.00	734.0	3095.00	0.068

Table B4 - Comparisons between cases and controls using the Mann-Whitney *U* test, determined by pfcount for households of children under 10

IV *	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (28)		CASES (71)			CONTROLS (15)		CASES (80)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	335.0	11.00	335.0	14.00	0.317	600.0	39.00	600.0	32.50	0.261
Slope (°)	5.0	2.00	5.0	2.00	0.378	11.0	4.00	11.0	7.00	0.534
Flowacc	1.0	5.75	2.0	6.00	0.519	1.0	2.00	0.0	2.00	0.930
Flow50	13.0	26.00	12.0	24.00	0.315	4.0	11.00	5.0	11.00	0.680
Flow100	39.0	261.75	32.0	57.00	0.196	13.0	75.00	18.0	35.75	0.959
Flow250	1889.5	6160.25	702.0	5487.00	0.296	291.0	642.00	319.0	581.25	0.358
Flow500	14350.5	11965.25	14319.0	6941.00	0.759	1910.0	1053.00	2135.0	1286.25	0.472
Flat50	4.0	4.00	4.0	4.00	0.134	0.0	0.00	0.0	0.00	0.884
Flat100	9.0	8.25	8.0	8.00	0.175	0.0	1.00	0.0	1.00	0.887
Flat250	60.0	33.00	51.0	30.00	0.245	5.0	2.00	6.0	6.00	0.104
Flat500	222.0	82.25	210.0	59.00	0.237	18.0	8.00	21.0	14.00	0.134
Swamp	0.5	6.75	0.0	5.00	0.356	0.0	0.00	0.0	0.00	1.000
Swamp100	42.5	428.25	28.0	90.00	0.175	0.0	0.00	0.0	0.00	0.680
Swamp200	552.5	3612.50	258.0	2237.00	0.212	2.0	2.00	2.0	2.00	0.490
Swamp300	4363.0	7694.25	3142.0	6839.00	0.382	2.0	0.00	2.0	1.50	0.822
Swamp400	9386.5	10221.50	8686.0	9298.00	0.423	2.0	19.00	2.0	106.75	0.684
Swamp500	14273.0	11582.50	14231.0	7598.00	0.807	2.0	131.00	2.0	331.50	0.369

Table B4 continued

IV *	VILLAGE 3 (750M)					VILLAGE 4 (1000m)				
	CONTROLS (9)		CASES (44)			CONTROLS (28)		CASES		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	median
Altitude (m)	804.0	67.50	754.0	67.75	0.059	1061.0	61.50	1029.0	34.00	**0.001**
Slope (°)	26.0	10.50	17.0	12.50	0.051	10.0	5.00	10.0	3.00	0.343
Flowacc	1.0	2.00	1.0	6.00	0.215	7.0	16.00	5.0	18.00	0.603
Flow50	3.0	11.00	9.5	25.00	0.119	32.0	60.25	70.0	264.00	0.245
Flow100	11.0	32.00	25.0	59.75	0.125	112.0	207.75	213.0	892.00	0.121
Flow250	113.0	399.00	242.5	397.00	0.454	774.0	2519.50	2763.0	2495.00	**0.009**
Flow500	612.0	1338.50	1074.0	1328.75	0.094	7100.5	3906.00	8516.0	1575.00	**< 0.0001**
Flat50					1.000	0.0	0.00	0.0	0.00	0.621
Flat100					1.000	0.0	0.00	0.0	2.00	0.190
Flat250					1.000	3.5	10.25	7.0	13.00	**0.004**
Flat500	10.0	11.00	0.0	8.25	0.106	33.5	46.25	63.0	38.00	**< 0.0001**
Swamp					1.000	0.0	0.00	0.0	0.00	0.740
Swamp100					1.000	0.0	0.00	0.0	0.00	0.513
Swamp200					1.000	0.0	111.75	1.0	170.00	0.201
Swamp300					1.000	1.0	479.50	281.0	549.00	**0.037**
Swamp400					1.000	409.0	881.25	761.0	479.00	**0.030**
Swamp500					1.000	880.0	1113.75	1454.0	719.00	**0.015**

	VILLAGE 5 (1200m)					VILLAGE 6 (1300m)				
	CONTROLS (23)		CASES (91)			CONTROLS (49)		CASES (79)		
Altitude (m)	1203.0	36.00	1194.0	37.00	0.713	1360.0	37.00	1360.0	35.00	0.410
Slope (°)	9.0	7.00	8.0	6.00	0.884	16.0	7.00	15.0	10.00	**0.049**
Flowacc	2.0	12.00	1.0	6.00	0.116	0.0	0.00	0.0	0.00	0.868
Flow50	42.0	123.00	15.0	47.00	0.129	1.0	2.50	0.0	2.00	0.242
Flow100	77.0	163.00	58.0	220.00	0.417	4.0	11.00	2.0	7.00	0.120
Flow250	671.0	1982.00	1084.0	1647.00	0.577	194.0	427.00	154.0	430.00	0.746
Flow500	2893.0	4615.00	4730.0	3877.00	0.539	2609.0	7306.50	2573.0	10768.00	0.282
Flat50	1.0	3.00	0.0	4.00	0.305	0.0	0.00	0.0	1.00	**0.023**
Flat100	4.0	7.00	2.0	8.00	0.466	0.0	1.00	0.0	3.00	0.196
Flat250	25.0	21.00	24.0	17.00	0.745	6.0	8.00	7.0	8.00	0.353
Flat500	90.0	13.00	90.0	20.00	0.932	12.0	35.00	11.0	30.00	0.692
Swamp	0.0	0.00	0.0	0.00	0.947	0.0	0.00	0.0	0.00	0.741
Swamp100	2.0	87.00	0.0	97.00	0.257	0.0	0.00	0.0	0.00	0.151
Swamp200	222.0	808.00	153.0	765.00	0.935	0.0	0.00	0.0	0.00	0.647
Swamp300	432.0	1327.00	966.0	1542.00	0.325	0.0	6.00	0.0	1.00	0.824
Swamp400	806.0	2315.00	2090.0	2004.00	0.231	0.0	96.50	0.0	84.00	0.720
Swamp500	1411.0	2429.00	2773.0	2188.00	0.162	26.0	254.50	50.0	265.00	0.942

IV *	VILLAGE 7 (1650 M)					ALL VILLAGES				
	CONTROLS (83)		CASES (27)		Sig.**	CONTROLS (237)		CASES (455)		Sig.**
Altitude (m)	1640.0	32.00	1640.0	67.00	0.697	1349.0	605.50	1034.0	636.00	**< 0.0001**
Slope (°)	7.0	15.00	12.0	8.00	0.409	10.0	11.00	10.0	8.00	0.728
Flowacc	0.0	1.00	0.0	1.00	0.582	0.0	3.00	1.0	4.00	0.062
Flow50	1.0	6.00	3.0	6.00	0.191	3.0	22.00	6.0	27.00	**0.023**
Flow100	5.0	36.00	11.0	28.00	0.301	17.0	65.50	25.0	98.00	**0.014**
Flow250	92.0	546.00	169.0	216.00	0.210	313.0	712.00	494.0	1746.00	**< 0.0001**
Flow500	1696.0	4148.00	2532.0	4287.00	0.155	2659.0	5633.00	3354.0	6533.00	**0.045**
Flat50	1.0	4.00	1.0	1.00	0.294	0.0	3.00	0.0	1.00	0.092
Flat100	3.0	9.00	3.0	3.00	0.833	1.0	6.00	1.0	4.00	0.170
Flat250	18.0	25.00	13.0	21.00	0.136	11.0	27.50	10.0	25.00	0.279
Flat500	69.0	18.00	67.0	21.00	0.266	57.0	62.00	45.0	73.00	0.755
Swamp	0.0	0.00	0.0	0.00	0.554	0.0	0.00	0.0	0.00	0.692
Swamp100	0.0	2.00	0.0	3.00	0.342	0.0	2.00	0.0	4.00	0.823
Swamp200	3.0	6.00	0.0	6.00	0.264	1.0	89.50	2.0	159.00	0.235
Swamp300	9.0	200.00	15.0	93.00	0.944	9.0	423.00	81.0	733.00	0.208
Swamp400	54.0	1221.00	120.0	1121.00	0.144	123.0	982.50	265.0	1497.00	0.270
Swamp500	352.0	2869.00	917.0	3091.00	0.123	449.0	2074.00	733.0	2574.00	0.430

Table B5 - Comparisons between cases and controls using the Mann-Whitney *U* test, determined by enlarged spleen, and under 5 households

IV *	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (19)		CASES (53)			CONTROLS (15)		CASES (55)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	335.0	11.00	335.0	12.50	0.832	600.0	29.00	600.0	36.00	0.914
Slope (°)	5.0	2.00	5.0	2.00	0.910	12.0	5.00	11.0	5.00	0.531
Flowacc	0.0	2.00	2.0	6.00	0.084	1.0	1.00	0.0	2.00	0.882
Flow50	7.0	26.00	13.0	24.00	0.725	3.0	11.00	5.0	11.00	0.994
Flow100	33.0	103.00	39.0	64.50	0.774	13.0	29.00	17.0	33.00	0.897
Flow250	2934.0	5713.00	702.0	5612.00	0.255	319.0	651.00	313.0	421.00	0.280
Flow500	15386.0	3599.00	14319.0	6941.00	0.309	2173.0	1506.00	2107.0	1245.00	0.710
Flat50	4.0	4.00	4.0	4.00	0.496	0.0	0.00	0.0	1.00	0.359
Flat100	9.0	6.00	8.0	8.00	0.410	0.0	0.00	0.0	3.00	0.185
Flat250	60.0	28.00	51.0	31.00	0.282	6.0	4.00	6.0	7.00	0.516
Flat500	227.0	71.00	202.0	59.00	0.239	21.0	12.00	21.0	16.00	0.667
Swamp	0.0	6.00	0.0	6.00	0.873	0.0	0.00	0.0	0.00	1.000
Swamp100	36.0	194.00	43.0	117.00	0.636	0.0	0.00	0.0	0.00	0.508
Swamp200	784.0	2560.00	258.0	2304.00	0.250	2.0	2.00	2.0	2.00	0.943
Swamp300	5821.0	6736.00	3142.0	6890.00	0.215	2.0	66.00	2.0	2.00	0.093
Swamp400	10696.0	5272.00	8686.0	9379.50	0.239	2.0	131.00	2.0	77.00	0.377
Swamp500	15265.0	3623.00	14231.0	7598.00	0.325	2.0	494.00	2.0	318.00	0.937
	VILLAGE 3 (750M)					**VILLAGE 4 (1000m)**				
	CONTROLS (13)		CASES (24)			CONTROLS (30)		CASES (31)		
Altitude (m)	804.0	51.00	751.0	83.25	**0.020**	1055.5	52.50	1027.0	42.00	**< 0.0001**
Slope (°)	26.0	8.50	17.0	12.00	**0.023**	10.0	4.25	10.0	4.00	0.873
Flowacc	1.0	2.00	1.0	6.00	0.298	9.5	15.00	5.0	27.00	0.760
Flow50	3.0	10.50	10.5	43.75	0.208	43.0	239.00	70.0	433.00	0.740
Flow100	11.0	31.50	25.0	60.00	0.174	136.0	748.75	184.5	939.00	0.665
Flow250	113.0	185.50	331.5	403.75	0.141	2203.0	2650.50	3113.0	3724.00	0.387
Flow500	619.0	694.50	1549.5	1384.00	**0.016**	7947.0	4174.75	8516.0	1737.00	**0.015**
Flat50					1.000	0.0	0.00	0.0	0.00	0.922
Flat100					1.000	0.0	0.00	0.0	2.00	0.160
Flat250					1.000	3.5	12.00	6.0	13.00	**0.009**
Flat500	9.0	9.00	0.0	6.00	0.024	37.0	48.25	63.0	35.00	**0.002**
Swamp					1.000	0.0	0.00	0.0	0.00	0.525
Swamp100					1.000	0.0	0.00	0.0	0.00	0.879
Swamp200					1.000	0.0	231.00	1.0	123.00	0.617
Swamp300					1.000	147.0	482.50	230.0	447.00	0.289
Swamp400					1.000	583.0	767.00	728.0	679.00	0.479
Swamp500					1.000	1126.5	736.50	1397.0	1149.00	0.603
	VILLAGE 5 (1200m)					**VILLAGE 6 (1300m)**				
	CONTROLS (59)		CASES (28)			CONTROLS (90)		CASES (9)		
Altitude (m)	1203.0	40.00	1186.0	34.75	**0.021**	1358.0	32.00	1360.0	26.50	0.869
Slope (°)	8.0	7.00	8.0	7.50	0.234	16.0	10.50	10.0	7.00	0.086
Flowacc	1.0	6.00	1.0	7.00	0.941	0.0	0.00	0.5	2.00	0.171
Flow50	15.0	47.00	15.5	118.75	0.489	0.5	2.00	0.0	1.50	0.459
Flow100	56.0	191.00	126.0	418.75	0.092	3.0	11.00	2.0	5.00	0.114
Flow250	721.0	1337.00	1658.0	1842.75	**0.017**	174.0	414.50	90.0	240.50	0.287
Flow500	3593.0	3818.00	5672.0	3005.50	0.096	2608.0	10696.00	2026.0	1304.50	0.051
Flat50	0.0	3.00	0.5	4.00	0.387	0.0	0.00	0.0	1.00	0.179
Flat100	1.0	4.00	3.0	8.75	0.196	0.0	1.00	0.0	3.50	0.480
Flat250	21.0	18.00	26.5	27.75	**0.045**	6.0	8.00	7.0	2.00	0.371
Flat500	87.0	24.00	92.5	9.50	**0.026**	11.0	30.00	13.0	23.50	0.595
Swamp	0.0	0.00	0.0	17.75	**0.030**	0.0	0.00	0.0	0.00	0.653
Swamp100	0.0	87.00	0.0	341.75	0.285	0.0	0.00	0.0	0.00	0.354
Swamp200	84.0	506.00	325.0	1035.75	**0.011**	0.0	0.00	0.0	0.00	0.206
Swamp300	568.0	1123.00	1400.0	1374.75	**0.005**	0.0	6.00	0.0	0.50	0.454
Swamp400	915.0	2018.00	2317.0	1720.50	**0.004**	0.0	88.50	0.0	15.50	0.245
Swamp500	2185.0	2322.00	3159.0	1659.50	**0.020**	26.0	265.00	0.0	118.50	0.331

Table B5 continued

IV *	VILLAGE 7 (1750m)					ALL VILLAGES				
	CONTROLS (74)		CASES (5)			CONTROLS (300)		CASES (205)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	1640.0	58.50	1640.0	53.00	0.510	1315.5	307.50	633.0	689.00	**< 0.0001**
Slope (°)	7.0	12.00	6.0	12.00	0.936	11.0	10.75	9.0	7.00	**0.010**
Flowacc	0.0	1.00	0.0	1.50	0.706	0.0	3.00	1.0	6.00	**0.017**
Flow50	2.0	9.00	1.0	4.50	0.395	3.5	21.75	8.0	27.00	**0.001**
Flow100	7.0	36.00	8.0	16.50	0.687	13.0	73.75	30.0	100.00	**< 0.0001**
Flow250	223.5	483.50	142.0	389.00	0.573	365.0	1001.25	507.0	1875.00	**0.001**
Flow500	1837.0	3803.50	1310.0	2993.00	0.286	2690.0	5212.75	3145.0	6957.50	0.266
Flat50	1.0	4.00	1.0	2.50	0.830	0.0	1.00	0.0	1.00	0.434
Flat100	3.0	8.00	3.0	6.50	0.554	1.0	4.00	1.0	5.00	0.351
Flat250	15.5	24.25	13.0	19.50	0.621	10.0	20.75	11.0	31.00	0.073
Flat500	68.0	21.00	67.0	41.00	0.426	47.0	64.75	53.0	112.00	**0.013**
Swamp	0.0	0.00	0.0	0.50	0.391	0.0	0.00	0.0	0.00	**0.007**
Swamp100	0.0	2.00	0.0	1.00	0.522	0.0	1.75	0.0	15.00	**0.049**
Swamp200	1.0	6.25	2.0	9.00	0.806	0.0	108.00	2.0	184.50	**< 0.0001**
Swamp300	12.0	174.75	7.0	216.00	0.473	9.0	426.75	117.0	1239.50	**0.004**
Swamp400	96.0	978.00	9.0	852.50	0.219	128.5	913.00	360.0	2327.00	**0.019**
Swamp500	539.0	2841.50	81.0	1715.50	0.135	539.0	1927.00	733.0	3633.00	0.061

Table B6 - Comparisons between cases and controls using the Mann-Whitney *U* test, determined by pfcount for households of children under 5

IV *	VILLAGE 1 (300m)					VILLAGE 2 (600m)				
	CONTROLS (17)		CASES (56)			CONTROLS (10)		CASES (61)		
	median	IQ range	median	IQ range	Sig. **	median	IQ range	median	IQ range	Sig. **
Altitude (m)	335.0	11.00	335.0	13.75	0.864	603.0	61.75	600.0	32.50	0.298
Slope (°)	5.0	2.00	5.0	2.00	0.734	11.5	6.50	11.0	6.00	0.665
Flowacc	0.0	5.50	2.0	6.00	0.240	0.0	1.00	0.0	2.00	0.379
Flow50	13.0	31.50	12.5	24.50	0.647	3.0	12.75	5.0	11.00	0.670
Flow100	39.0	208.00	37.5	57.00	0.535	11.5	104.00	18.0	27.50	0.823
Flow250	3016.0	5985.00	702.0	5544.25	0.457	308.5	725.75	313.0	457.00	0.895
Flow500	15316.0	4503.50	14981.0	6927.00	0.629	2063.0	1303.50	2107.0	1246.50	0.914
Flat50	4.0	4.00	4.0	4.00	0.399	0.0	0.25	0.0	0.50	0.816
Flat100	9.0	7.50	8.0	7.00	0.414	0.0	1.50	0.0	2.50	0.701
Flat250	60.0	33.50	55.5	27.75	0.700	5.5	4.25	6.0	6.00	0.473
Flat500	222.0	85.50	214.0	50.00	0.657	21.0	10.75	21.0	15.00	0.613
Swamp	1.0	7.50	0.0	5.00	0.431	0.0	0.00	0.0	0.00	1.000
Swamp100	43.0	379.50	39.0	98.00	0.517	0.0	0.00	0.0	0.00	0.880
Swamp200	813.0	3368.50	258.0	2270.00	0.426	1.0	2.00	2.0	2.00	0.410
Swamp300	5524.0	7318.50	3287.0	6831.00	0.477	2.0	16.50	2.0	1.00	0.481
Swamp400	10266.0	6565.50	9395.0	9269.75	0.453	2.0	2.00	2.0	78.00	0.788
Swamp500	15265.0	4514.00	14836.0	7569.50	0.633	2.0	302.00	2.0	318.00	0.747

Table B6 continued

IV*	VILLAGE 3 (750M)					VILLAGE 4 (1000m)				
	CONTROLS (8)		CASES (29)			CONTROLS (21)		CASES (42)		
Altitude (m)	804.0	44.00	767.0	77.00	0.072	1046.0	49.50	1028.0	38.50	**0.001**
Slope (°)	26.5	8.75	18.0	12.00	**0.044**	10.0	5.50	10.0	3.25	0.676
Flowacc	0.5	2.50	1.0	6.00	0.310	12.0	14.50	5.0	26.25	0.924
Flow50	3.0	11.00	6.0	25.00	0.178	45.0	239.00	70.0	283.00	0.965
Flow100	9.5	39.50	24.0	59.50	0.176	133.0	790.00	204.0	945.00	0.878
Flow250	113.0	267.25	207.0	393.00	0.374	2102.0	2761.50	2831.5	2667.50	0.246
Flow500	615.5	959.25	897.0	1355.50	0.052	7889.0	4076.00	8516.0	1699.00	**0.021**
Flat50					1.000	0.0	0.00	0.0	0.00	0.267
Flat100					1.000	0.0	1.00	0.0	1.25	0.986
Flat250					1.000	4.0	13.00	6.0	12.25	0.065
Flat500	9.5	9.75	0.0	7.50	0.086	33.0	49.00	59.0	39.25	**0.009**
Swamp					1.000	0.0	0.00	0.0	0.00	0.206
Swamp100					1.000	0.0	0.00	0.0	0.00	0.362
Swamp200					1.000	0.0	336.50	0.0	123.00	0.936
Swamp300					1.000	123.0	736.00	255.5	457.50	0.412
Swamp400					1.000	538.0	1001.50	740.0	645.25	0.418
Swamp500					1.000	1081.0	1107.00	1397.0	846.00	0.502
	VILLAGE 5 (1200m)					VILLAGE 6 (1300m)				
	CONTROLS (29)		CASES (59)			CONTROLS (55)		CASES (46)		
Altitude (m)	1201.0	40.50	1191.0	37.00	0.653	1356.0	35.00	1360.0	40.50	0.970
Slope (°)	8.0	8.00	8.0	7.00	0.993	16.0	7.00	13.5	11.75	0.081
Flowacc	2.0	10.00	1.0	6.00	0.348	0.0	0.00	0.0	0.00	0.580
Flow50	15.0	115.50	16.0	46.00	0.426	1.0	2.00	0.0	2.00	0.213
Flow100	44.0	250.00	68.0	213.00	0.926	4.0	10.00	2.5	6.00	0.061
Flow250	628.0	1655.50	1108.0	1521.00	0.180	130.0	360.00	182.5	430.00	0.576
Flow500	4641.0	3249.50	4656.0	4185.00	0.739	2583.0	10211.00	2582.0	13435.00	0.981
Flat50	1.0	3.50	0.0	3.00	0.611	0.0	0.00	0.0	1.00	0.175
Flat100	3.0	7.50	2.0	7.00	0.758	0.0	1.00	0.0	3.00	0.442
Flat250	20.0	24.00	24.0	15.00	0.335	6.0	8.00	6.5	8.00	0.529
Flat500	87.0	23.50	90.0	17.00	0.132	10.0	33.00	12.0	30.00	0.176
Swamp	0.0	0.00	0.0	0.00	0.885	0.0	0.00	0.0	0.00	0.888
Swamp100	0.0	208.50	0.0	97.00	0.813	0.0	0.00	0.0	0.00	0.782
Swamp200	55.0	798.00	180.0	821.00	0.241	0.0	0.00	0.0	0.00	0.376
Swamp300	532.0	1438.50	1223.0	1540.00	0.079	0.0	1.00	0.0	8.25	0.729
Swamp400	1401.0	1918.00	2170.0	1880.00	0.135	0.0	66.00	0.0	107.25	0.754
Swamp500	2471.0	2242.50	2773.0	2301.00	0.250	0.0	216.00	79.0	265.00	0.366
IV*	VILLAGE 7 (1750m)					ALL VILLAGES				
	CONTROLS (64)		CASES (15)			CONTROLS (204)		CASES (308)		
Altitude (m)	1640.0	55.25	1636.0	67.00	0.753	1346.0	504.75	994.0	625.00	**< 0.0001**
Slope (°)	7.0	11.75	12.0	9.00	0.174	11.0	12.00	10.0	8.00	0.606
Flowacc	0.0	1.00	1.0	1.00	0.773	0.0	3.00	1.0	4.00	0.054
Flow50	1.0	6.75	3.0	10.00	0.438	3.0	21.00	6.0	27.00	**0.019**
Flow100	6.5	34.25	11.0	35.00	0.595	12.0	68.50	26.0	100.75	**0.008**
Flow250	145.0	490.75	293.0	346.00	0.351	314.0	715.25	505.0	1731.00	**< 0.0001**
Flow500	1774.0	3717.25	2109.0	4462.00	0.464	2685.5	5479.75	3225.0	6672.75	0.078
Flat50	1.0	4.00	0.0	1.00	0.108	0.0	3.00	0.0	1.00	0.284
Flat100	3.0	8.75	2.0	3.00	0.282	1.0	5.00	1.0	4.00	0.552
Flat250	15.5	26.50	13.0	16.00	0.495	10.0	22.75	11.0	25.75	0.806
Flat500	68.0	20.00	67.0	35.00	0.603	53.5	63.00	46.0	76.75	0.221
Swamp	0.0	0.00	0.0	0.00	0.534	0.0	0.00	0.0	0.00	0.433
Swamp100	0.0	1.75	0.0	4.00	0.346	0.0	1.75	0.0	7.00	0.271
Swamp200	2.5	6.75	0.0	6.00	0.426	0.0	50.75	2.0	169.25	**0.012**
Swamp300	9.0	192.25	15.0	165.00	0.674	9.0	420.75	99.0	957.75	**0.014**
Swamp400	54.0	882.50	117.0	1602.00	0.579	117.0	961.00	321.5	1900.75	0.079
Swamp500	480.0	2735.75	648.0	3110.00	0.590	475.5	2089.75	733.0	2909.00	0.174