Making Tracks:

Simulating Prehistoric Human Travel Networks.

Northland, New Zealand.

M.Sc. Dissertation

Andrew Standley

Submitted in part fulfilment of the requirements for the degree of
Master of Science in Geographical Information Science
Faculty of Economics and Business Administration
Vrije Universiteit Amsterdam
The Netherlands
March 2015

Making Tracks: Simulating Prehistoric Human Travel Networks. Northland, New Zealand.

ABSTRACT

The pattern of prehistoric human travel paths and the spatial relationship with archaeological sites in New Zealand is not well understood. Archaeological surveys have focused on individual sites mainly in coastal areas with less attention paid to large inland areas and corridors connecting settlements or culturally significant sites. The use of Geographic Information Systems (GIS) to model potential pedestrian travel routes has been limited to coarse national scale isotropic least cost path analysis between known obsidian resource locations. This thesis uses a GIS to model natural travel corridors and paths across a 5,230 square kilometre area of the Northland region to form a prehistoric human travel path network. The modelled network is validated by testing the significance of the distribution of archaeological sites associated with human travel routes prior to European settlement. This research asks**,** are inland Pa sites positioned in close proximity to least cost overland travel paths? Supporting hypotheses state that Pa sites are unevenly distributed and strongly clustered in close proximity to least cost paths.

Existing GIS techniques for modelling regional least cost primary travel corridors and secondary branching paths were replicated and enhanced to create a complex network of paths originating from the edge of the study area, navigating through a simulated prehistoric travel friction surface. Using Pa site density grouped by travel cost intervals leading away from the path network, a Chi Squared test of significance strongly rejected the null hypothesis confirming that Pa sites are unevenly distributed. Although the Pa site distribution test produced a weak Gain statistic, travel time analysis shows that the majority of Pa sites are highly accessible from the least cost travel path network. This adapted method has not previously been used in New Zealand and the results provide an original contribution to the archaeological and geospatial science body of knowledge.

Future GIS research may refine this approach by developing a more realistic prehistoric land cover dataset and by addressing current limitations associated with GIS based representations of human travel speed and time. The performance of the model presented in this study has the potential to improve as new Pa sites are discovered. This research provides a method for narrowing down future archaeological site exploration and is an initial step towards gaining a greater understanding of prehistoric human travel networks within New Zealand.

Keywords: Geographic Information Systems (GIS), path distance, isotropic, anisotropic, least cost path, Pa site, Chi Squared, Multinomial Goodness of Fit, Gain Statistic.

DISCLAIMER

The results presented are based on my own research at the Faculty of Economics and Business Administration of the Vrije Universiteit Amsterdam. All assistance received from other individuals and organisations has been acknowledged and full reference is made to all published and unpublished sources. This dissertation has not been submitted previously for a degree at any institution.

Place, Date:

(Signature):

Name student: Andrew Standley

ACKNOWLEDGEMENTS

I wish to thank my Wife Nicola and our children Ayla and Joel for their understanding and support throughout this Masters programme and my previous post graduate research. I would also like to thank my employer Opus International Consultants Ltd for funding this Masters course.

The initial idea for this research topic came from a predictive archaeological risk modelling project based in the Bay of Islands area in 2012 for the New Zealand Transport Agency. I would like to acknowledge the co-author of that project report, Nick Cable, a Senior Archaeologist who recommended further research into the inclusion of social and cultural variables such as travel routes for predicting archaeological sites in New Zealand.

Finally, special thanks and appreciation for the excellent guidance and advice provided by my Tutors, Dr. Philip Verhagen and Dr. Niels van Manen.

CONTENTS

FIGURES

TABLES

CHAPTER 1: INTRODUCTION

The Northland region of New Zealand has a rich history of human occupation. Prehistoric human travel across the region would have been necessary to access coastal and inland natural resources and for social, economic and political interaction between communities. Archaeological surveys have focused mainly on clusters of settlement and other land use evidence within the coastal environment leaving vast inland areas largely unsurveyed. Research concerning prehistoric human movement patterns across inland areas of New Zealand has received far less attention than site specific analysis even though the body of archaeological knowledge recognises the social and cultural significance of transport networks. As the indigenous Maori population did not document trails until European arrival in New Zealand, evidence of trails is limited to isolated historic written accounts and maps. The regional prehistoric human travel network across Northland remains unknown. Geographic Information Systems (GIS) has been used in archaeological research in New Zealand to model potential least cost routes between known obsidian sources, however the one example of this was at a national scale using a cost weighted isotropic method at a coarse resolution.

This thesis builds on this previous research by introducing a more detailed method of defining low cost travel paths at a regional scale. An existing method of modelling natural travel corridors (Whitley and Hicks 2003) was applied to a simulated prehistoric New Zealand environment (representing the period between 1200AD and 1769AD) prior to European settlement, also referred to as the *Prehistoric* period. Once the natural travel corridor network was mapped, the focus then moved to measuring the spatial relationship between the modelled travel network and one particular archaeological site type, inland Pa sites which Archaeologists believe to have been positioned for accessibility to overland trails. The GIS method selected was based on the following knowledge and assumptions;

1. The archaeological body of knowledge suggests that prehistoric people would have sought out the most efficient travel routes.
2. Travel time is the most popular travel cost measure applied in archaeological research mainly due to time based effort being easier to understand than calorie consumption.
3. ArcGIS Path Distance and Cost Path methods guarantee the globally optimal least cost route between known origins and destinations, assuming a travel friction surface is used to represent a realistic historic terrain.
4. Archaeologists have stated that they believe inland Pa sites were staged along travel routes as well as serving other defensive purposes.

This research has used publicly available spatial information which does not cause ethical issues in itself. However, this research derives new information from what is known, and this could be considered sensitive to indigenous people if validated by future research for example. The recent production of historic accounts and associated GIS data required for The Waitangi Treaty negotiations counters this risk as all the data is published and in the public domain post settlement.

1.1 PROBLEM STATEMENT

Knowledge of prehistoric human movement using overland travel paths across the Northland region is currently limited to historic written accounts and maps created by early European Surveyors, Missionaries and the Military, sometimes with the assistance of Maori guides. These historic accounts focus on specific expeditions rather than a broad network of interconnecting routes. Archaeologists know that Maori used the river networks extensively for inland travel, however, it is also acknowledged that overland on foot travel was still essential for longer distances. The lack of knowledge of the regional pattern of overland travel paths warrants further investigation. This research postulates that least cost travel paths represent a plausible prehistoric regional human travel network and that inland Pa sites are positioned in close proximity to these paths.

1.2 RESEARCH PURPOSE

The purpose of this quantitative research was to use a GIS to replicate and enhance an existing method (Whitley and Hicks 2003) of modelling a regional network of prehistoric natural human travel paths in a New Zealand setting. The statistical significance and strength of the relationship between the dependent variable (Pa sites) and the independent variable (least cost paths) was explored using Pearson's Chi Squared test and a Gain statistic (Kvamme 1988). The significance of this spatial relationship is that a simulated prehistoric human travel network may provide an explanation for the siting of inland Pa sites.

1.3 RESEARCH QUESTION

This research attempts to understand the relationship between Pa sites and natural least cost travel paths by asking the following question;

Are inland Pa sites positioned in close proximity to least cost overland travel paths?

In order to answer the research question, the least cost path network was modelled and the statistical tests were applied in order to test the following hypotheses;

1) Use of a Chi Squared test to determine if inland Pa sites are unevenly distributed across the landscape. This will test the following hypothesis;

There is a difference between observed and expected Pa site distribution.

2) Test the least cost network model's performance using a Gain statistic to measure the accuracy and precision of Pa site distribution relative to modelled low cost travel paths. This will test the second hypothesis;

Inland Pa site density measured relative to regional least cost paths will result in a Gain statistic > 0.5.

1.4 OBJECTIVES

The following objectives are designed to answer the research question;

1. To determine suitable travel cost variables.
2. To replicate a method of defining a network of least cost travel paths using a perimeter based origin-destination technique (Whitley and Hicks 2003).
3. Use a statistical significance test on two forms of path distance classification (equal travel time interval and equal area interval) to determine whether Pa sites are normally distributed relative to natural travel routes.
4. Apply a Gain statistic test to determine a relative measure of the accuracy and precision of observed Pa site distribution.
5. Review objectives 1-4 to answer the research questions and suggest future areas of research to expand on the findings from this research.

1.5 RESEARCH DESIGN

This study follows a quantitative research route by partly replicating and enhancing existing research methods on the use of a GIS to model natural least cost travel corridors representing potential prehistoric human movement patterns at a regional scale (Whitley and Hicks 2003). An attempt has been made in this research to generalise these methods by applying the concept of unrestricted natural travel corridors to a New Zealand setting. A replication approach is justified because the original method uses universal landform based rules verified using statistical analysis of archaeological site proximity to modelled overland paths. As such, the method should be valid anywhere and should provide some value for archaeological research in New Zealand. This research follows the positivist paradigm associated with the use of a GIS to represent spatially enabled vector and raster information required to represent theoretical travel paths. An instrumental rationale assumes a connection between spatial analyses derived from the use of a GIS with logical spatial reasoning (Malczewski 2004).

This research study has five key stages;

1. **A travel cost surface*:*** a raster based representation of accumulated anisotropic travel cost which allows for the direction of travel and surface distance. Graph theory (Dijkstra 1959) is then applied using least cost path techniques to develop the following path networks.
2. **Primary least cost paths** which take the form of node and link based weighted networks using multiple origins and destinations. The result is a Primary travel network dataset of the most efficient overland arterial paths crossing the region.

   Once the primary travel network has been defined, a hydraulic flow modelling technique was used to create additional paths described below.

3. **Secondary branching paths** using a flow accumulation method over an anisotropic surface originating at the primary travel network. The result is a secondary path draining towards primary paths. A threshold value limits the extent of the secondary paths. Secondary and primary paths then combine to form an extended natural travel network.

   Both the primary and combined travel networks are then tested by exploring inland Pa site distribution relative to both networks.

4. **Pearson's Chi Squared test of significance** was used to measure if observed and expected Pa site populations are evenly distributed across the study area. Once Pa site distribution is understood, the following test was applied.
5. A **Gain statistic** is used to test the strength of the correlation between observed and expected Pa site distributions relative to travel paths.

1.6 THESIS STRUCTURE

Chapter 2 provides a literature review which is structured to briefly introduce the overarching subject of GIS based predictive archaeological modelling. A brief background is useful to explain how this research area has evolved from the sole use of environmental variables for predicting areas suitable for a particular land use to the gradual introduction of more complexity offered by social and cultural influences. A summary of how cultural and social variables have been incorporated into GIS is provided before focusing on one of these variables and the core subject for this research: Least Cost Path modelling and path distance catchment analysis representing potential human movement patterns at a regional scale. The key elements of cost surface composition and path finding algorithms are presented along with various methods used to determine regional travel corridor networks including the method that this study attempted to replicate. Statistical methods are then reviewed for testing hypotheses designed to measure the distribution of inland Pa sites relative to these theoretical least cost travel routes and the strength of the relationship. Chapter 3 describes supporting archaeological information covering the Prehistoric New Zealand landscape, documented native trails and Pa site characteristics.

Chapter 4 covers the methods used starting with an introduction to the study area's natural character followed by the replicated methods used to form the primary travel arteries and secondary paths. Finally, statistical methods used to test hypotheses related to the significance and strength of Pa site distribution relative to travel networks is presented.

Chapter 5 examines the resulting travel cost surface, the primary and secondary natural travel paths and includes a brief test of the effect that river costs can have on least cost path results. The original method's use of equal travel time interval classification of path distance leading way from paths is compared with an alternative equal area interval classification. The significance of Pa site distribution is confirmed using a Chi Squared test followed by a Gain statistic test which measures the strength of the model's accuracy and precision for each interval.

Chapter 6 discusses the research objectives and interprets the statistical results to answer the hypotheses and research question. Chapter 7 concludes this research and confirms the research findings and provides a number of ideas for future research. Figure 1 provides an overview diagram of the literature review structure described in Chapter 2.

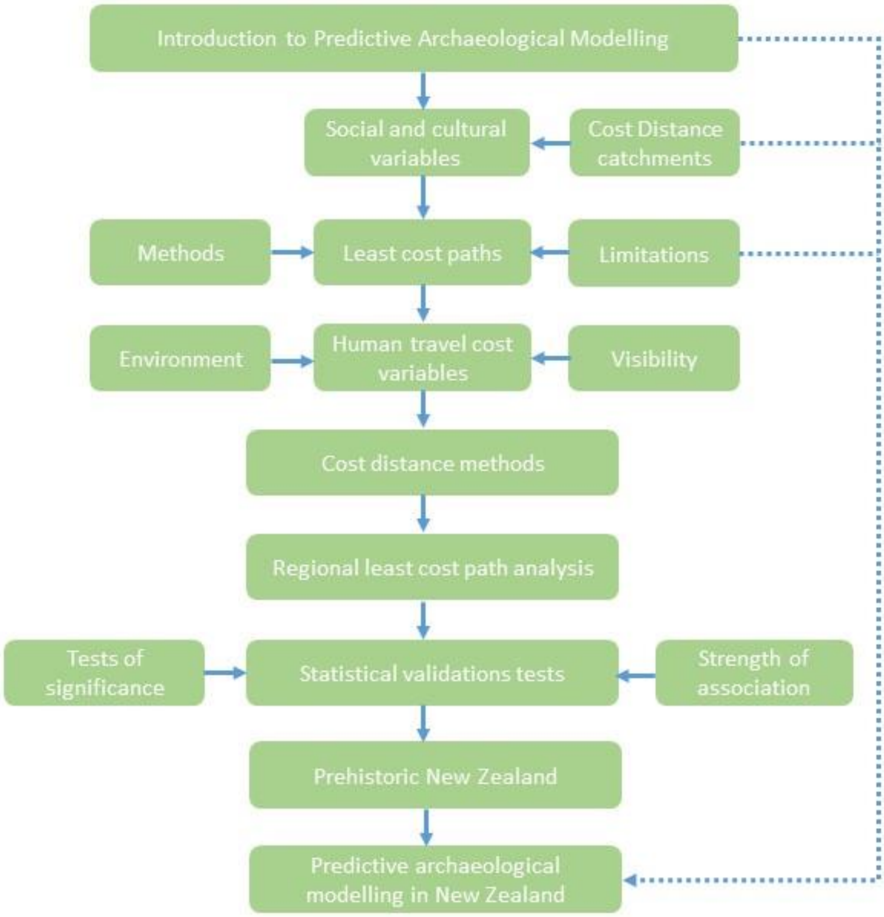


Figure 1: Literature review design

CHAPTER 2: LITERATURE REVIEW

The following literature review has been divided into sub sections in a logical sequence, starting with a general introduction to the overarching subject of GIS based predictive archaeological modelling and the use of environmental variables influencing the prediction of archaeological sites. Social and cultural variables are then discussed, leading into a specific example and the main subject of this research, an introduction to cost distance and least cost path analysis as a means of modelling a prehistoric network of optimal travel paths. Regional scale least cost path techniques are then discussed followed by an overview of statistical significance and strength tests used to examine the relationship between least cost path networks and inland Pa sites. The final part of the literature review introduces the Prehistoric New Zealand environment and briefly describes Pa site characteristics which are used as the dependent variable at the model validation stage.

2.1 AN INTRODUCTION TO PREDICTIVE ARCHAEOLOGICAL MODELLING

A brief description of how Geographic Information Systems (GIS) has been used to predict archaeological site suitability is necessary to provide some context for this research. GIS based predictive models are used in several areas of research including geological mineral prospection, ecological habitat analysis and archaeological predictive modelling (Verhagen and Whitley 2011). Computer based predictive archaeological modelling was introduced in the United States in the late 1970’s to assist with managing archaeological risk on large areas of Government owned land (Kohler 1988). One of the first publications summarising predictive archaeological modelling progress up to the mid 1980’s introduced a definition that recognised two forms of predictive modelling, correlative and cognitive;

Predictive archaeological modelling attempts to predict the location of archaeological sites based either on a sample of that region or on fundamental notions concerning human behaviour. *(Kohler and Parker 1986).*

Correlative models typically use inductive methods such as logistic regression and weights of evidence to identify spatial relationships between known archaeological sites and environmental variables such as soils, surface rock, land cover or proximity to natural resources (Kohler and Parker 1986) (Kvamme 1988), (Wheatley & Gillings 2002). Predictive archaeological modelling is based on the assumption that archaeological sites are likely to be found in locations suitable for human habitation (Warren 1990). Models that are used to derive information regarding dependent and independent variable relationships are also referred to as empirical predictive moWestcottdels.

These data driven inductive methods rely on large dependent variable datasets such as archaeological site data and require acceptable levels of unbiased spatial and categorical site data accuracy. Inductive methods do not however explain why there may be a strong correlation between a particular soil type and archaeological sites, (Judge and Sebastian 1988), (Westcott and Brandon 2000) they can only confirm that a spatial pattern is present. Even though statistical observations are limited to the extent of the archaeological evidence, such methods can provide valuable information which can reinforce universal site proximity rules or suggest new rules.

Deductive methods are not dependent on the archaeological data that inductive methods require. Instead, a hypothesis is formed based on a combination of site suitability rules, (Dalla Bona 1994). Deductive methods offer the advantage of being able to be used in areas with little or no archaeological evidence, areas where inductive methods cannot operate due to lack of statistically significant numbers of archaeological sites. Recorded archaeological sites may be used for statistically testing the accuracy and precision of explanatory modelled data, but are not a required part of a deductive modelling process (Kohler and Parker 1986), (Verhagen & Whitley 2011). However, some deductive site suitability rules are fully or partially based on previous inductive knowledge, so generally predictive rules are a combination of inductive and deductive approaches, (van Leusen and Kamermans 2005). Inductive and deductive approaches to predictive modelling are generally constrained by ecological determinism, a reliance on available environmental spatial datasets from which modelling rules are derived (Judge and Sebastian 1988), (Gaffney & Van Leusen 1995).

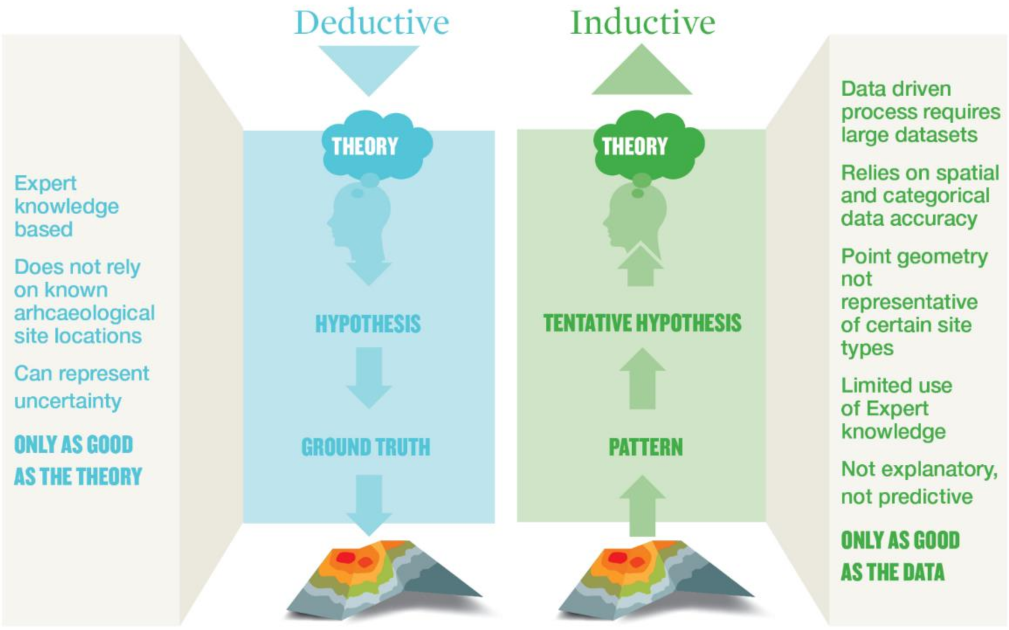


Figure 2: Opposing approaches to predictive archaeological model design

While inductive modelling methods are generally limited to such datasets, deductive methods have the ability to incorporate alternative influences on site suitability such as cultural and social attractions, (Wheatley 1995), (Verhagen and Whitley 2011).

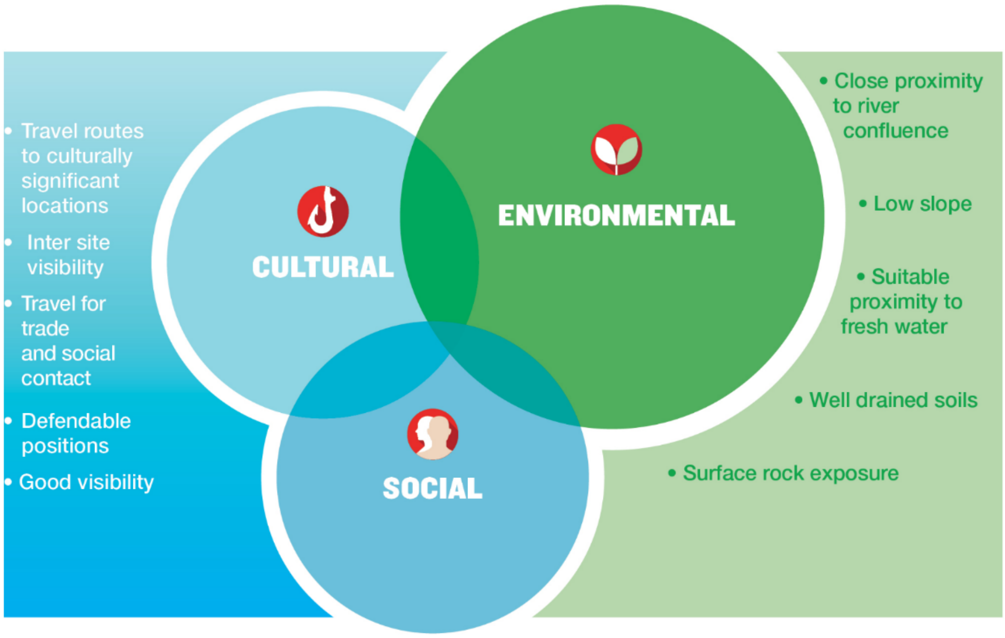


Figure 3: Types of predictive archaeological modelling variables

Much of the criticism by academic researchers on the use of predictive models (Wheatley 2003) revolves around the level of explanatory value (Gaffney and van Leusen 1995), theoretical limitations of GIS software (Harris and Lock 1990), or inappropriate use of statistical theory (Thomas 1978). The following section looks at the most popular ways of using a GIS to represent social and cultural influence by modelling human travel cost and overland movement patterns.

2.2 SOCIAL AND CULTURAL VARIABLES

Criticism of the over reliance on environmental variables in predictive archaeological modelling has been balanced with calls for the inclusion of social and cultural influences on settlement site selection to improve model performance, (Judge and Sebastian 1988); (Zubrow 1990); (Wheatley 2003), (van Leusen 2002), (Verhagen and Whitley 2011). Identifying and representing features within the landscape that have cultural value is more abstract than dealing with environmental datasets. GIS is able to replicate one of these social and cultural variables using travel cost distance analysis.

2.2.1 COST DISTANCE CATCHMENTS

Cost Distance is a computer representation of movement using cumulative travel cost units and a weighted travel friction surface and is considered one of the most significant social variables for predictive modelling (Howey 2007). Cost distance typically involves some way of representing resistance to movement by developing a cost surface in the form of a uniform grid based raster dataset. Cost represents anything that is thought to have impeded human travel such as slope, vegetation cover and wetlands or other factors that may have attracted or repelled movement such as cultural site attraction or sensitive areas which travellers may have avoided. Cost distance analysis originated from catchment analysis techniques (Limp 1990). In archaeological research, catchment zones are used to represent natural resource exploitation zones, foraging areas surrounding natural resources or known archaeological sites.

An early example (Gaffney and Stancic 1992) used catchment analysis to study the accessibility of defensive hill fort settlements on the island of Hvar, Croatia. This study involved a comparison between two dimensional concentric catchment zones based on typical foraging distances (5km and 10km) vs. a walking energy consumption cost catchment analysis. A slightly different use of cost distance (Hare 2004) is used in a study defining polity boundaries in the Yautepec valley, Mexico. This study used an anisotropic walking velocity method (Gorenflo and Gale, 1990) measuring pedestrian travel speed away from each centre, converted to a measure of energy consumption per kilometre (Gorenflo and Gale, 1996) to represent travel efficiency, forming cost catchment boundaries.

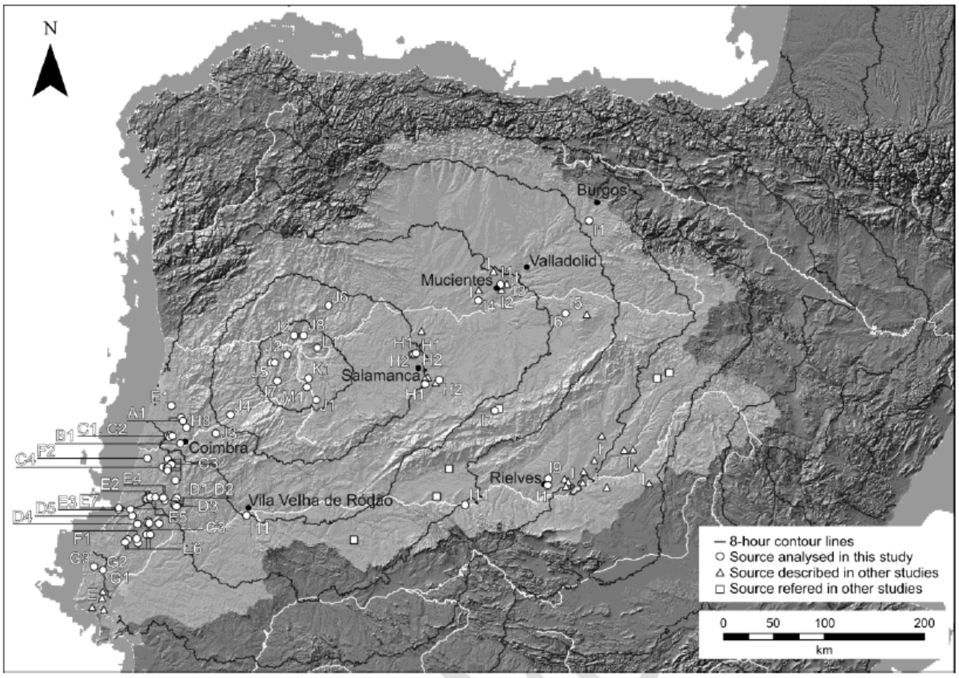


Figure 4: Anisotropic travel time catchment intervals, central Spain (Aubry 2012)

A more recent example of large scale cost distance catchments (Aubry et al 2012) applies path distance to a large portion of western Spain to form travel catchments at eight hour intervals from a single origin. Within New Zealand, archaeological research has explored evidence of obsidian use from known source locations in Northland, Coromandel, Taupo, the Bay of Plenty and offshore islands (Moore 1982, Brassey 1998).

Cost distance has also been used in the Northland area of New Zealand (McCoy et al 2010) to study obsidian supply zones across the region using a distance decay linear regression graph method to form likely primary and secondary supply zones. McCoy discusses the relative quality of obsidian sources, which may explain the need for long distance travel rather than accessing the nearest source. Recent research by the same author however suggests that Northland Maori mainly sourced obsidian from local mainland and offshore island sources. Evidence of high quality obsidian from Mayor Island in various locations around New Zealand shows that long distance exchange still took place and overland obsidian movement was more likely than by sea (McCoy and Carpenter 2014).

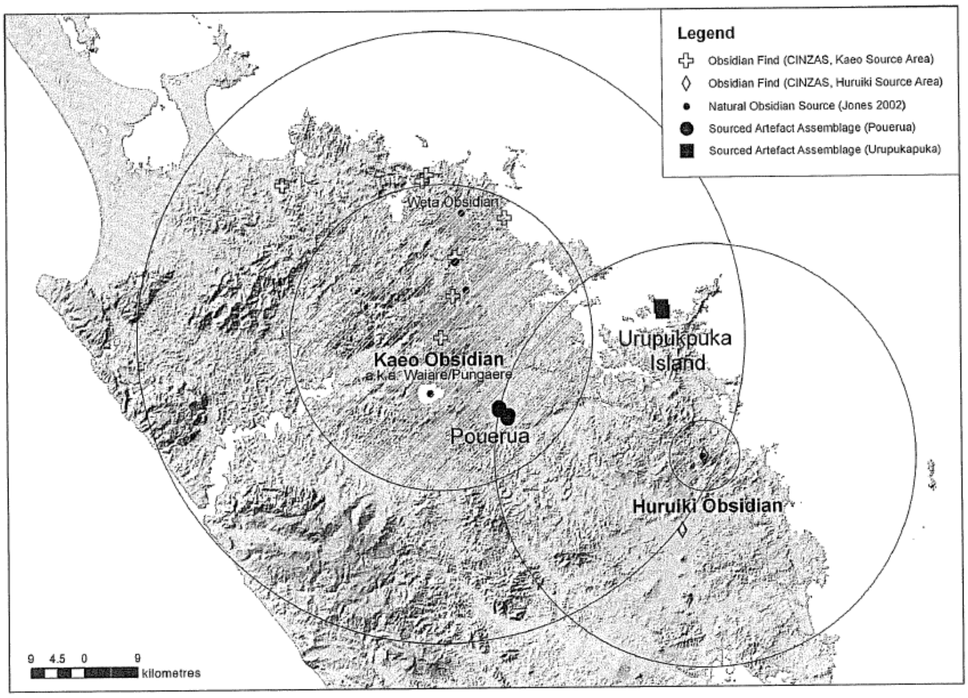


Figure 5: Obsidian supply zones, Northland, McCoy 2010

2.3 LEAST COST PATHS

The main predictive modelling method used in this research is ‘Least Cost Path’ (LCP). Least cost path analysis is a technique that seeks to select a route from an origin to a known destination based on minimum accumulated travel cost. Least Cost Path analysis originates from graph theory which is generally credited to Leonhard Euler. Euler’s publication ‘Seven Bridges of Konigsberg’ in 1735 (Alexanderson 2006), attempted to find an optimal route which crossed each of the seven bridges surrounding the town once and then return to the start location. Although the problem was not solvable, the concept of optimal paths and topology is the basis of least cost path capability in today’s GIS software. In a GIS, least cost path analysis determines the least costly route between an origin and destination using a directional node and link format while accounting for the weighted cost of movement along the links.

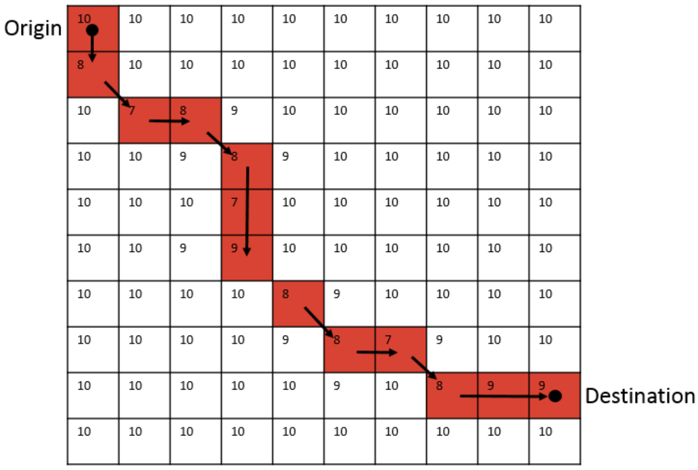


Figure 6: Least cost movement across a raster cost surface

A number of researchers applied this approach prior to GIS based examples (Ericson and Goldstein 1980), (Gorenflo and Gale, 1990), (Gaffney & Stančič, 1992).

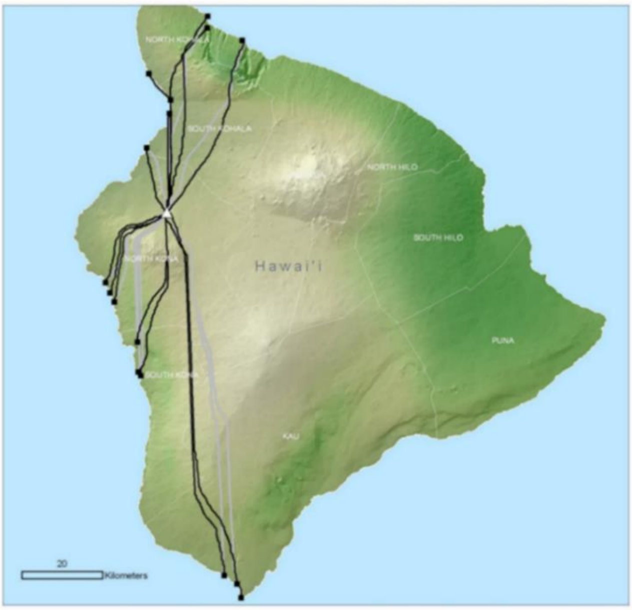


Figure 7: Least cost paths between obsidian sources, Hawaii (McCoy et al 2011).

Archaeological research on this topic has to assume that prehistoric people preferred to travel routes that optimised time or energy (Herzog 2012) as other reasons may be unknown or difficult to reproduce in a GIS. The guaranteed shortest path search algorithm (Dijkstra 1959) calculates this optimal shortest route by analysing route costs from one raster cell to all other cells in the cost surface dataset. ArcGIS applies this using the Cost Distance (isotropic) or Path Distance (anisotropic) tools to create a cost direction raster dataset and a backlink raster dataset which drains back to the origin from the destination point along the globally optimal route.

2.3.1 ISOTROPY VS. ANISOTROPY

Least cost path analysis generally follows one of two methods for modelling travel cost accumulation, isotropy or anisotropy. Isotropic analysis assumes that the cost of travelling across a terrain raster cell is equal regardless of the approach and exit direction with least cost paths controlled by other cost variables present. Travel direction however, is a critical element of the anisotropic method as the optimal path from A to B may not match the least cost path from B to A due to variations in land form (van Leusen 2002). Least cost path methods used for finding optimal routes for infrastructure such as pipelines use the isotropic method to represent the degree of building difficulty, for example the construction cost related to the type of vegetation cover, slope and other constraints (Rapaport and Snickars 1999). Anisotropy is mainly used to model human or animal movement which have biological limitations controlling the maximum slope or other surface friction they are able to travel over.

2.3.2 LEAST COST PATH LIMITATIONS

Dijkstra's shortest route algorithm can be time consuming when processing large numbers of paths. Dijkstra's method of examining all possible routes across the cost surface has been described as a model of global knowledge where prehistoric people would have developed knowledge of their landscape in order to discover optimal paths over time (van Leusen 2002). This is a reasonable assumption as this would have been a collective effort passed on through experience. However, GIS least cost path methods would not be suitable for attempting to model initial human movement into an area where the most efficient routes would not be known because the destination would not have been known. A lesser known alternative to Dijkstra's method is the A* method (Hart et al 1968). (White and Surface Evans 2012) describe the A* algorithm as replicating Dijkstra's method but with an additional function which models alternative routes which can be selected as detours at high cost locations along the globally optimal route.

The A* method also requires a smaller search area compared with Dijkstra's use of the whole study area resulting in faster processing times. However, A* will not always match least cost path results derived from Dijkstra's method. Surface Evans and White state that Dijkstra's method is used by all standard commercial GIS software. However, (Herzog and Posluschny 2008) counter this stating that although ArcGIS fully adopts Dijkstra's algorithm guaranteeing to find the optimal least cost path, GRASS GIS only partially implements Dijkstra's method in combination with a hydrology based drainage backlink method.

This difference between these two methods was tested by (Gietl et al 2008) who presented a summary of the differences in path distance algorithms used by ArcGIS 9.1, GRASS 6.1, and IDRISI 15. The experiment involved set origin and destination points and attempted to test anisotropic least cost path analysis using the same data and settings where possible. The comparison noted a limitation regarding the size of the digital elevation model (DEM) when using ArcGIS. However, this is easily resolved by significantly increasing the maximum number of unique values to render (in ArcGIS 10.1, this setting is under the *Customise/ArcMap Options/Raster Dataset tab*). Gietl also suggests that methods of compiling a cost surface differ and exchange of raster datasets is limited by ArcGIS having data formats that are not compatible with GRASS and IDRISI. Gietl's test shows large variations in paths chosen by each software.

A further example of the inconsistency between GIS software is the ArcGIS default slope calculation which is based on maximum slope change between raster cells whereas other GIS can determine average slope change. Average change in slope may be more suitable for least cost path modelling, avoiding the exaggeration of slope due to minority peak height values dominating the analysis. In addition, ArcGIS is restricted to an eight neighbouring cell search when defining slope compared with a 24 cell search applied by GRASS GIS.

Another limitation of least cost path analysis is the use of satellite quality elevation data. A digital elevation model resolution of 90m is too coarse for use in archaeological least cost path studies, (Monteleone 2013), missing critical detail in the terrain surface. However this may be the only data source available for some remote study areas. In addition, where available, the historic record should be taken into account regarding travel mode during the period being researched and comparison with any documented historic travel routes should be considered to validate GIS generated routes. Historic research may also indicate the presence of attracting and repelling cultural influences on travel routes which could be built into the cost surface in some way.

2.4 HUMAN TRAVEL COST VARIABLES

Archaeologists are interested in how cost distance and least cost path analysis can be used to identify hypothetical travel paths using modelled constraints to represent barriers to human movement such as steep slope, dense vegetation and water bodies, (White and Surface Evans 2012). The following section provides some examples of travel cost variables.

2.4.1 TERRAIN AS TRAVEL COST VARIABLE

Most studies use current digital terrain model data based on the assumption that the terrain surface is a reasonable representation of prehistoric land form. An exception to this could be where sea levels are known to have changed significantly. Historic maps have been used in an attempt to replicate a prehistoric terrain, for example (van Leusen 2002) uses an example of Fogliano in the Perugia region of Italy where 17th Century maps were used to recreate a historic terrain model which revealed some differences compared to the current landscape. However, examples of this type of comparison are rare.

2.4.2 VEGETATION AS A TRAVEL COST VARIABLE

Prehistoric vegetation cover data typically does not exist in GIS format so any attempt to include this type of movement cost will be based on assumptions related to today's environmental data. Unless the study area is considered to be unchanged since the era the researcher is interested in, vegetation cover is not often used. Spatial datasets representing vegetation in New Zealand are limited to current land cover (the New Zealand Land Cover dataset). This national dataset currently covers 33 classifications of modern vegetation types captured using SPOT 5 satellite imagery to produce a dataset suitable for use at 1:50,000 scale. Due to the amount of land use change associated with land clearance and drainage for farm use, today's vegetation data is of little use for predicting areas that may have presented a barrier to prehistoric human travel.

2.4.3 WATER AS A TRAVEL COST VARIABLE

River travel cost is a special case representing a partial barrier to overland human movement and an upstream and downstream medium for travel by water craft. Typical river network vector data is unlikely to hold river characteristics such as depth, width and seasonal flow, details needed to determine the realistic ease or difficulty of crossing a river or travelling up or down stream.

(Aubry 2012) states that river ford crossings are not possible to model, rivers have to be crossed at some point so do not present a realistic barrier without detailed river characteristics. Although actual river flow data may not be available, a GIS can be used to model accumulated drainage flow across a digital elevation model surface. ArcGIS provides tools for modelling flow direction and flow accumulation and other GIS such as GRASS and IDRISI have a similar process. The flow accumulation process involves filling depressions within the digital elevation model (DEM) surface, followed by a calculation of drainage flow direction (modelling which direction water will flow from each DEM cell). Flow accumulation (the sum of the number of raster cells draining to each cell) is applied and finally a threshold is introduced to eliminate low flows near the river source and simplify the river network. The edited flow accumulation raster can then be converted to a relative cost as the river flows downstream from its source, reflecting the relative effort required to cross a stream vs. the cost of attempting to traverse an estuary.

(Whitley and Hicks 2003) included this drainage flow cost approach in their regional study of historic pathways across a large study area in North Georgia, United States. In order to use the flow data as a cost surface, the flow cost values needed to be relative to overland human movement. Whitley and Hicks determined the relative time to cross a river at its maximum accumulated flow location, for example six times longer than walking the same distance on flat land, then divided all accumulated flow values by the maximum flow value and multiplied by a factor six. This then formed a simple river travel cost input to an anisotropic least cost path analysis of the study area. A weakness associated with GIS based flow accumulation is that it produces a single cell river network. Where single accumulated flow cells connect diagonally to a neighbouring flow cell, a least cost path can cross the rasterised river at that point without being allocated a cost (Van Leusen 2002), (Herzog 2010). To overcome this, the width of river cells can be increased to produce a three cell wide version of the river network. This river representation forces a least cost path algorithm to find genuine low cost river crossing locations.

2.4.4 VISIBILITY

Another cultural and social variable, site visibility has received more attention than cost distance in archaeological GIS research; (van Leusen 1993), (Lock and Harris 1996), (Llobera 1996), (Wheatley and Gillings 2002), (Verhagen and Jeneson 2012). Commercial GIS software has the ability to model visibility using a digital elevation model and observer points located on the terrain surface. Viewshed analysis determines which terrain raster cells are visible from each vertically offset observer cell. An alternative option is line of sight analysis which models direct sight lines which can be presented using 3D GIS software such as ArcScene.

ArcGIS allows the user to populate observer points with a number of fields containing specific field names representing visibility characteristics (observer and target vertical offset, view radius and angle). The quality of viewshed analysis however is only as good as the resolution and accuracy of the digital terrain dataset, (Westcott & Brandon 2000). A study on the island of Brac, (Stancic and Kvamme 1999) analysed intervisibility between hill forts using cumulative viewshed analysis from known fort locations. The resulting overlaps in these viewshed catchments suggested inter site visibility may have been a factor for site selection. A similar observation regarding hill fort distribution and prominence relates to the historic Ridgeway trail in Oxfordshire (Bell and Lock 2000). A review of this study (Lock and Pouncett 2010) suggests that viewshed analysis confirms that some hill forts where probably placed as navigation markers on the landscape. Visibility analysis was proposed as an influence while modelling potential routes for the Roman road 'Via Belgica' between Germany and the north east coast of France, while also considering slope limitations for wheeled travel on Roman roads (Verhagen and Jeneson 2012).

The inclusion of cultural influences on route selection has received far less attention than environmental variables. However, (Llobera 2000) provides a study of the effect that historic monuments may have had on human travel behaviour presented a way of representing the effect of cultural site attraction on travel networks. Llobera used an energy consumption method (Minetti 1995), and a cost surface representing viewshed areas around two monument locations. The viewshed area symbolised an area of influence surrounding each monument, which had a decaying attraction effect on least cost paths passing nearby. Llobera's descriptions of linear decay and alternatives such as exponential and step wise decay closely resembles the options for using fuzzy membership to classify raster data (Zadeh 1965).

2.5 COST DISTANCE METHODS

Human overland travel cost is generally measured in two ways;

- Walking velocity and derived travel time.
- Energy consumption and derived travel time.

Both methods are used to allocate a travel unit cost rate to traverse a raster cell allowing a cumulative travel cost to be calculated between origin and destination. A hiking calculation devised by (Tobler 1993) using empirical data from (Imhof 1950) who studied soldiers' walking speed and time is the most popular anisotropic method used in archaeological least cost path studies, (Herzog 2010), (Carballo and Pluckhahn 2007).

V = 6e (-3.5 x abs (s + 0.05))

(Tobler 1993).

Where;

V = walking velocity (Km/Hour)

e = base of natural logarithms

s = slope (degrees)

This walking velocity calculation can be used to form a vertical factor table representing the degree of difficulty when walking over variable terrain. The vertical factor table forms one of several inputs to the Path Distance tool in ArcGIS. A vertical factor table requires a two column list representing slope in degrees and the corresponding vertical factor. Once Tobler's walking speed has been calculated for a given slope value, the vertical factor can be derived by dividing one hour by the distance travelled in metres. The distribution curve this matrix produces estimates maximum walking velocity of approximately 6km per hour at 3 degrees downhill slope.

An older time based method is used as the basis of the *r.walk* function in GRASS GIS software; (White and Surface Evans 2012);

T = (a AL) + (b AU) + (c AD) + (d AS)

Where

T = time (seconds)	c = downhill speed
a = speed	AD = downhill altitude difference
AL = distance walked	d = speed down a steep slope
b = uphill speed	AS = altitude difference for steep slope
AU = uphill altitude difference (metres)	

(Langmuir 1984).

A number of studies have assessed different ways of measuring energy consumption across variable terrain. Pandolf's energy consumption method has been adapted by a number of researchers to overcome a limitation where downhill walking is concerned. The original equation produces negative energy consumption downhill, which cannot be the case as even walking down a gentle slope consumes energy (White 2012).

M watts = 1.5W + 2.0 (W+L) (L/W) 2+n (W+L) (1.5 V2 + 0.35 VG)

M watts = metabolic rate (watts)

W = travellers weight (Kg)

L = Load (Kg)

V = walking velocity (metres per second)

G = slope (%)

n = terrain coefficient *(Pandolf et al 1977)*

(Herzog 2010) points to another example that proposes optimal travel cost is at zero slope (Ericson and Goldstein 1980);

Cost = horizontal distance + (3.168 * upward vertical distance) + (1.2 * downward vertical distance)

Adapted from (De Silva and Pizziolo 2001).

De Silva and Pizziolo adapted Ericson and Goldstein's method to account for this weakness in order to conduct an anisotropic analysis of cost distance catchments and least cost routes thought to link settlements and other archaeological sites in the Campobasso province, southern Italy.

Tyler Bell and Gary Lock introduced a relative effort based method of calculating travel cost which uses the ratio of slope tangent, assuming the walkers' weight, gravity and raster cell size are constants.

Tan (slope - degrees) / Tan (1 degree)

(Bell and Lock 2000), (Bell et al 2002)

2.6 REGIONAL LEAST COST PATH ANALYSIS

The purpose of regional least cost path analysis is to identify a Prehistoric society's likely pattern of social and economic movement constrained by landform (White 2012). Identifying these movement patterns at a regional scale is an area of archaeological research that has grown in recent years due to its potential for explaining the siting of known archaeological sites and for identifying optimal paths which may be associated with undiscovered archaeological sites (Whitley and Hicks 2003), (Howey 2007), (Whitley and Burns 2007), (Verhagen 2013), (Murrieta-Flores, 2012), (White 2012). Regional least cost path analysis moves beyond individual travel routes towards the identification of the landscape's natural matrix of interconnecting low cost transit areas.

Travel networks connecting these low cost areas are formed by either modelling unrestricted paths from origins placed at entry points to a study area or by forcing least cost paths through known archaeological sites within the study area. To achieve this, analysis must move beyond point to point least cost path analysis to one to many, or many to many origin-destination patterns and broader corridor analysis where alternative routes to the optimal path are of interest. Archaeological research has focused on this regional approach to modelling cost based accessibility to form large scale hierarchical travel networks (Whitley and Hicks 2003), (Whitley and Burns 2007).

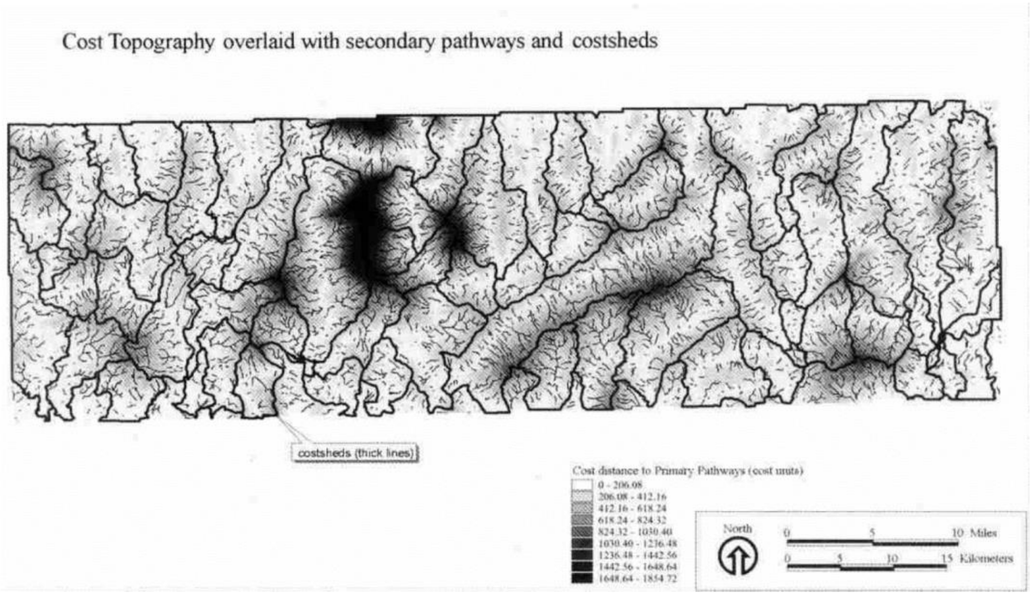


Figure 8: Primary and secondary least cost paths (Whitley and Hicks 2003)

These studies in (North Georgia covering 2,250km2) and (South Carolina covering an area of 6,500km2) respectively involved the placement of least cost path origin points around the study area perimeter at 1km intervals (and repeated at 2km, 5km, 10km and 20km intervals for comparison). These origin points also acted as least cost path destinations. Although Tobler's hiking equation is mentioned, the ArcGIS path distance method is not, so both examples are assumed to be isotropic, supported by a reference to the use of the cost distance tool in ArcGIS 3.2 in the earlier study.

A cumulative river cost method described in section 2.4.3 was used for both studies, however the South Carolina study also included soil potential as a variable travel cost. The one to many least cost paths were converted to vector lines and combined to form a regional primary travel network dataset. This optimal travel network was then used to derive secondary path catchments stemming outward from the primary routes into the surrounding landscape using flow accumulation techniques.

These secondary paths represent isotropic walking time catchments for accessing areas alongside the main primary travel routes. A recent study in south west Spain (Murrieta-Flores 2012) also used perimeter origin-destinations and applied raster and vector methods to identify natural travel corridors using IDRISI GIS software (figure 9).

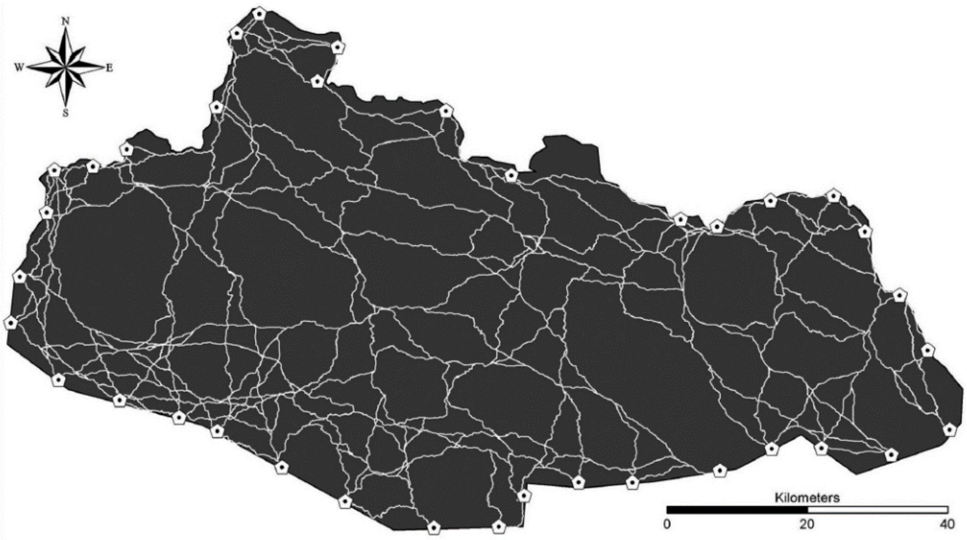


Figure 9: Regional primary least cost paths using perimeter origins (Murrieta-Flores 2012)

This study took the concept of regional travel corridor modelling further by introducing a form of anisotropic analysis, using a weighted river classification in the cost surface dataset and also used classified land morphology as a cost variable. An alternative approach to regional accessibility analysis is provided by White and Barber (2012) who used a uniform grid of origin-destination points at 300m intervals across the state of Oaxaca, in southern Mexico.

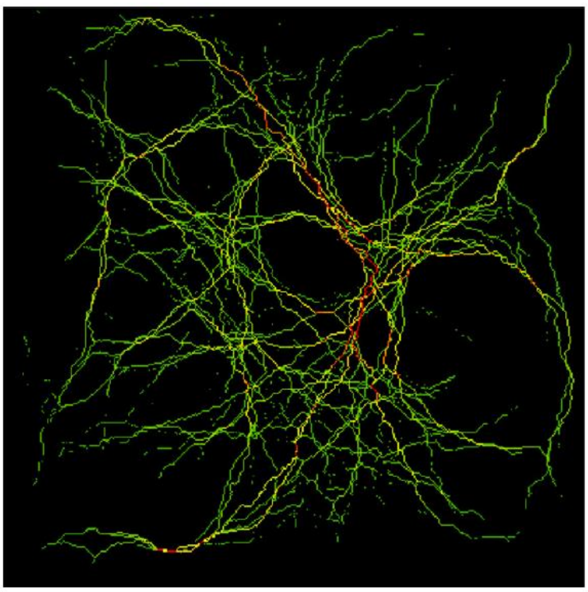


Figure 10: Anisotropic least cost path density (White and Barber 2012)

An anisotropic walking time calculation was performed using Tobler's hiking function followed by an energy consumption calculation (Pandolf 1977, adjusted by Santee 2001) producing an extremely dense least cost path matrix. Relative path density was measured using localised searches around each cell, with a radius of 100 pixels, and for comparison an area of 500 pixels. No noticeable difference was found by increasing the search radius surrounding each cell. The authors state that the results support the travel network literature in the region and highlight additional routes which have received little attention from archaeological researchers.

2.7 STATISTICAL VALIDATION TESTS

Once a least cost path network is created, it is useful to validate the resulting network by testing whether particular archaeological sites are in close proximity to the paths or not. In other words, are the sites evenly distributed across the landscape or clustered with easy access to the modelled travel paths. Statistical hypothesis testing is an established approach to this type of problem. The null hypothesis states that there is no difference between the observed and expected archaeological site distributions relative to paths. The alternative hypothesis claims that there is a difference between the observed and expected distributions, and the data varies enough from the expected distribution to allow the null hypothesis to be rejected. There are several statistical methods that can be used to measure the significance of observed data distribution compared with an expected distribution. These tests fall into two categories, parametric and non-parametric. Parametric methods are applied when the data distribution is known to be normally distributed. Non-parametric methods can operate without the need for this assumption and can also be applied to smaller samples but are considered less robust than parametric tests. The following are some examples of non-parametric and parametric tests.

2.7.1 NON PARAMETRIC TESTS OF SIGNIFICANCE

Pearson's Chi-squared test for statistical significance is a non-parametric test which has been used to test for significant close proximity of archaic sites of different time periods to natural travel corridors (Whitley and Hicks 2003). An isotropic cost distance surface was reclassified into equal intervals of 100 cost units, and the expected distribution was calculated using the area of each cost distance band as a percentage of the overall study area.

$$X^2 = \sum \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}$$

(Pearson's Chi Squared, Connolly and Lake 2006).

The Chi squared test produces a *p value* which represents the probability of a type 1 error (the false rejection of the null hypothesis). A general interpretation is that $p < 0.05$ is statistically significant (5% chance of error), and $p < 0.01$ is statistically highly significant (1% chance of error). A Confidence level is used with the p value (1-p), so small p values result in greater confidence. If the null hypothesis is rejected, the Chi Square test result will be greater than the critical value taken from a Chi Squared distribution table, where in this case Degrees of Freedom = the number of path distance intervals being tested minus 1. Chi Squared Distribution tables are not required these days as statistical software calculates Degrees of Freedom, Chi Square and the p value automatically from a given set of observed and expected data counts. An example of a Chi Squared distribution table is shown in table 10. An important rule when using Chi Squared is that expected site counts per interval must be > 1.5, but ideally >5 (Cochran 1952), (Yates, Moore and McCabe, 1999). If expected site counts fail this rule for more than 20% of the sample, an exact multinomial goodness of fit test can be used which simulates the required number of expected samples using Monte Carlo simulation to achieve a valid result (McDonald, 2009). The Chi Squared statistic will be low when the observed distribution does not differ significantly from expected distribution, and higher where the gap between observed and expected distribution increases.

An alternative non parametric test of significance is the Kolmogorov-Smirnov (K-S) test, which is a test of data normality and is used for hypothesis testing (Herzog 2013), (Murrieta-Flores 2012). The K-S test uses the maximum difference between the cumulative observed and expected data, known as the *Dmax* value. A critical value is obtained using widely available critical value tables for different confidence levels. Unlike the Chi Squared equation which uses real sample values, the Kolmogorov Smirnov test requires the sample range to be cumulative, normalised and standardised. If *Dmax* exceeds the critical value, the null hypothesis can be rejected. A further alternative test of normality is the Shapiro Wilks test, another nonparametric test well suited to small samples.

2.7.2 PARAMETRIC TESTS OF SIGNIFICANCE

Examples of parametric tests include the Student T test, and the Analysis of Variance (ANOVA) test. The T test compares the mean values of two sample datasets and determines whether the difference is statistically significant. The ANOVA test is related to the T test and is used when there are multiple sub groups present. A comparison of natural resources available within Euclidean and hiking cost distance based catchments surrounding shell mound sites was tested using a T test (Surface-Evans 2012).

2.7.3 STRENGTH OF ASSOCIATION TEST

The results from a test of significance such as Pearson's Chi Squared only confirms whether the sampled sites are evenly distributed. If the distribution is shown to be uneven, such tests cannot confirm the strength of the observed vs expected site distribution pattern. A Gain statistic (Kvamme 1988) can be used to provide a measure between 0 and 1 of how accurately and precisely a modelled set of zones captures known archaeological sites. The Gain statistic also provides a means of comparing one model's performance with another. A successful result will have both high precision and high accuracy. Accuracy is a measure of the proportion of archaeological sites that fall within each interval, and precision is a measure of each intervals proportional area. A Gain value > 0.5 represents moderate strength, and Gain values rarely exceed 0.9. A weakness of the Gain statistic is that the same Gain value can be achieved with different proportions of accuracy and precision.

CHAPTER 3: ARCHAEOLOGY

This chapter provides a brief introduction to Prehistoric New Zealand, covering land use change over time, historic trails, Pa site characteristics and examples of the use of GIS for predictive archaeological modelling.

3.1 PREHISTORIC NEW ZEALAND

New Zealand was the last area of the Pacific to be colonised (Anderson 1991). The earliest Polynesian settlement period is still debated by Archaeologists, with initial estimates between 800AD (Green 1975), (Sutton 1987), and 1200AD (Kinaston et al 2013), (Anderson 1991). These early settlers migrated from near Oceania (New Guinea, the Solomon Islands, Vanuatu, New Caledonia and Fiji) to remote Oceania (Samoa, Tonga, the Cook Islands and New Zealand) and brought with them crops such as Kumara (a type of sweet potato), Yam and Taro, adapting agricultural techniques to suit the cooler New Zealand climate. The majority of Maori settled in the North Island where the climate is more suitable than the South Island for crop cultivation. When Maori arrived in New Zealand, both islands were covered in dense native forest. Over a period of 500 years prior to European settlement, approximately 40% of the forested land was cleared mainly by seasonal burning (Landcare Research). Prior to 1500AD, settlements were thought to be mainly coastal, however following the eradication of the giant Moa, a main food source, and the loss of forest habitat, settlements began to appear inland (McFadgen 2007). A slightly earlier example of this is Maori settlement around the volcanic basalt and scoria fields at Pouerua around 1400AD (Sutton 1990).

Figure 12 provides an indication of changes in Maori society about 500 years ago as historic food sources diminished and Pa construction began. Early European survey records show vast areas of forest still remained in the late 18th Century. However, Europeans continued the deforestation throughout the 19th Century exporting native timber and introducing large scale agriculture.

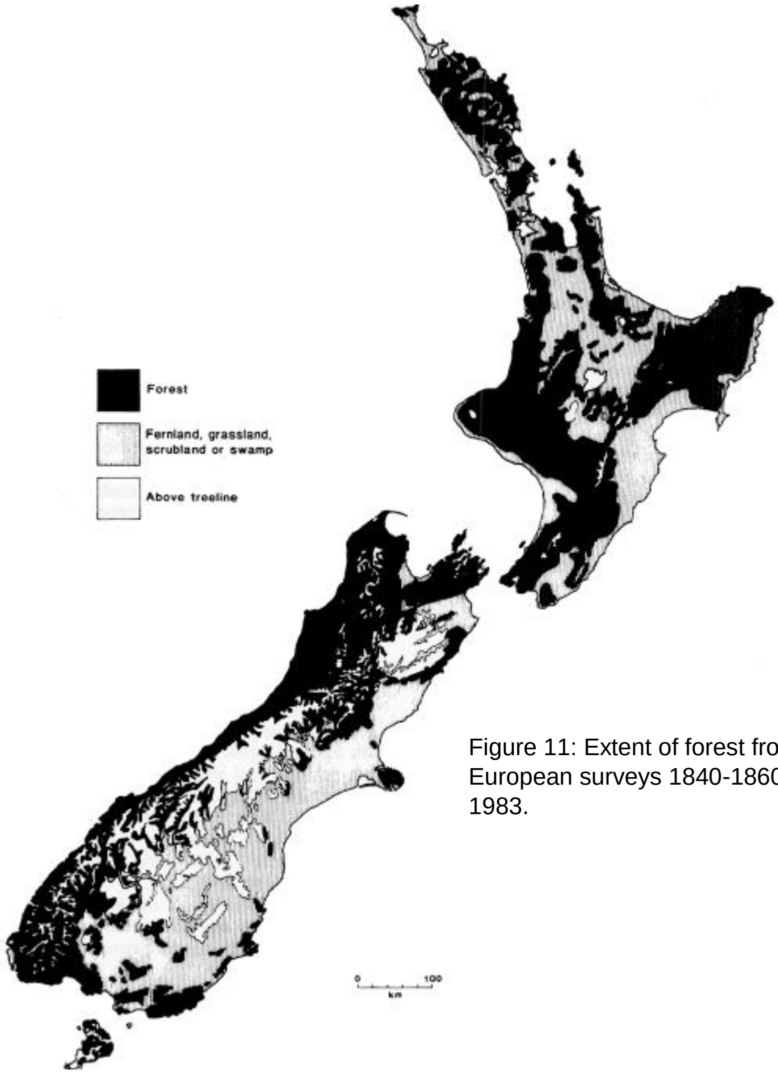


Figure 11: Extent of forest from first European surveys 1840-1860. McGlone 1983.

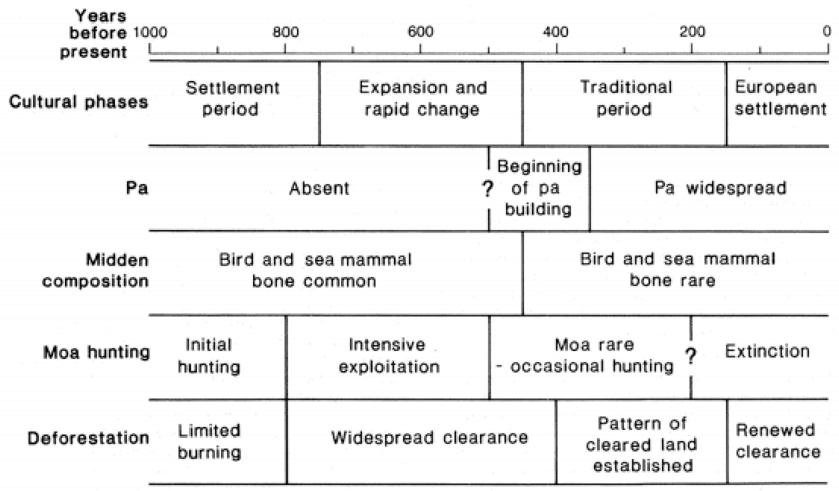


Figure 12: Cultural and environmental change since human settlement (McGlone 1989).

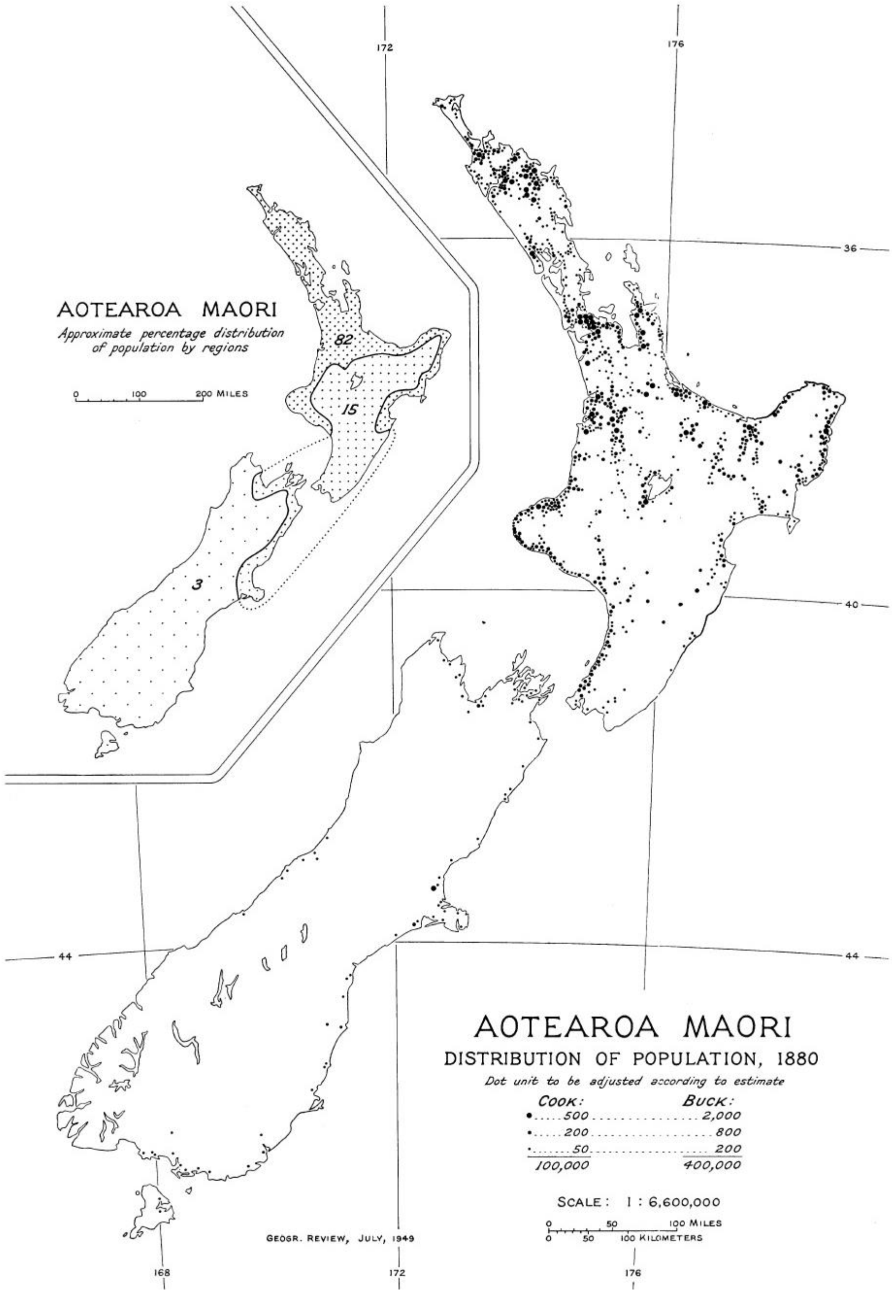
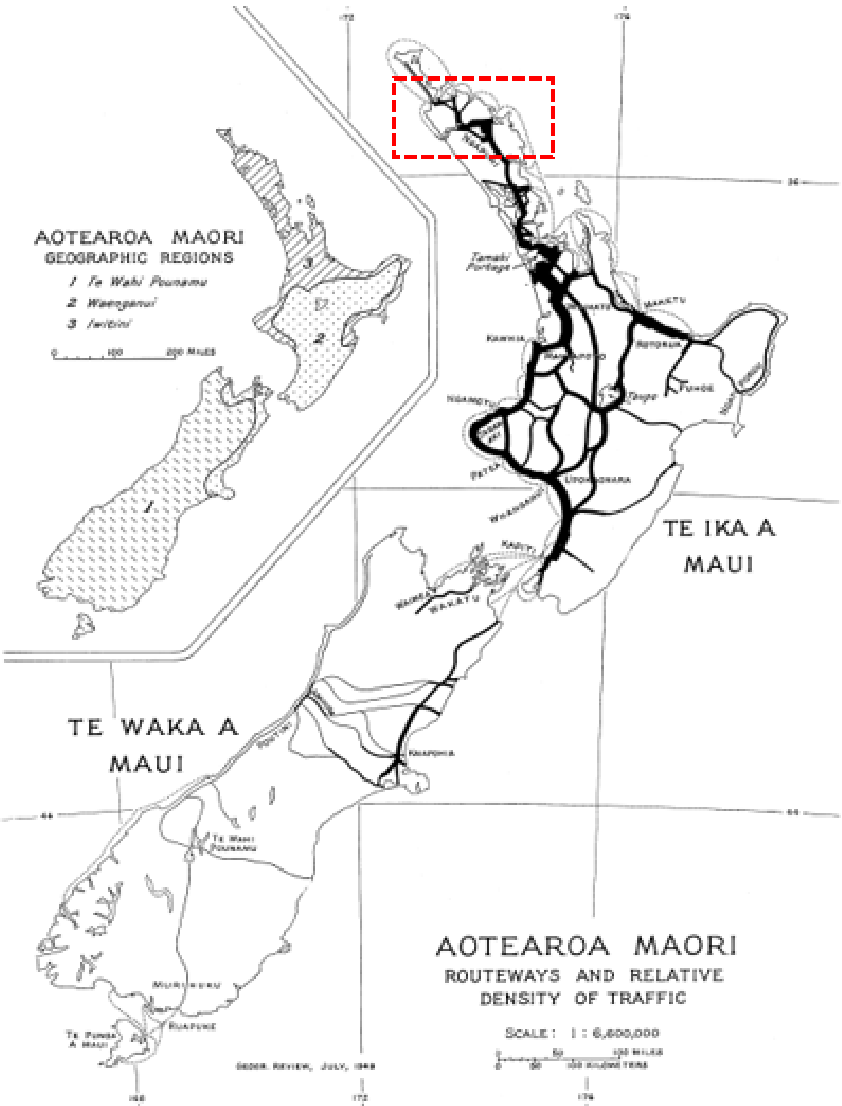


Figure 13: Estimated Maori population, 1880 (Cumberland 1949)

3.2 NATIVE TRAILS

Barton (1980) describes pre-European travel networks across the North and South islands linking natural resource locations and population centres, also reflected in Cumberland's trade routes map (figure 14). Coastal journeys using large waka canoes were common but inland travel was a safer and more direct option for Maori to maintain their social network. Inland travel involved navigating rivers by waka canoe and overland hiking following a well-established network of trails. The evidence for this comes from the experiences of European explorers attempting to move inland from the coast.

Maori Guides assisted British explorers by drawing maps based on topographic features including routes marking lakes, rivers and mountains. Location names were passed on orally as the Maori language was only documented from 1820 (Barton 1980). Some of the Maori maps exaggerated certain features or areas which reflected the importance of that area with inlets enlarged and distant lands reduced in size. Missionaries were among the first European settlers in New Zealand. Samuel Marsden established the Christian church in the Bay of Islands and travelled extensively across the north of the country as he visited Maori chiefs, introducing wheat grain and agricultural technology. Marsden's journeys across Northland in the early 1800's includes routes crossing the Kaipara Harbour on the west coast, extensive routes along the Northland east coast and several overland paths traversing the peninsula west of Whangarei and the Bay of Islands. Figure 14 indicates a likely travel route from the Bay of Islands to Hokianga Harbour within this research study area.



Indicative historic route between the Bay of Islands and Hokianga Harbour

Figure 14: Maori trade routes, (Cumberland 1949)

3.3 PA SITE CHARACTERISTICS

Pa sites are fortified defensive structures found on terraces, ridgelines, near swamps and staged along important travel routes (Leahy and Walsh 1976). Pa were used seasonally as Maori were highly mobile or permanently inhabited in areas where food and water were abundant. Pa sites were also used as observation points, temporary protected shelter in times of conflict or for supporting large settlement centres.

A typical settlement would include an undefended site called Kainga with a defended and elevated Pa situated in close proximity. However, Pa status within Maori society also influenced how these sites were used (Golson 1957). Good visibility to and from Pa sites was an important factor influencing Pa site position on the landscape (Reeler 1997). Golson (1957) classified upland Pa into three categories.

Hill Pa took advantage of peaks and natural vantage points to provide a protected site with good visibility (figures 15 and 16). A Promontory Pa (figure 17) is typically found on ridgelines with man-made defences such as ditches and palisade fences. A Ridge Pa has access limited to either side of the site along a ridgeline (figure 18).

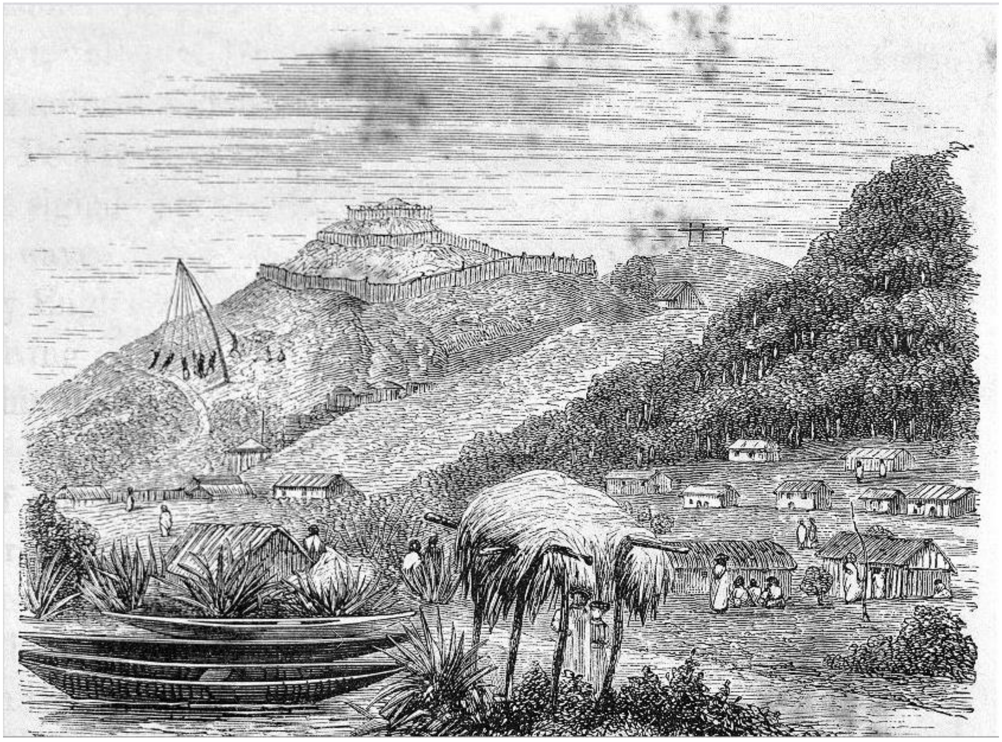


Figure 15: Native village with Pa. Bridge, Cyprian (Lieutenant-Colonel). 1807-1885.

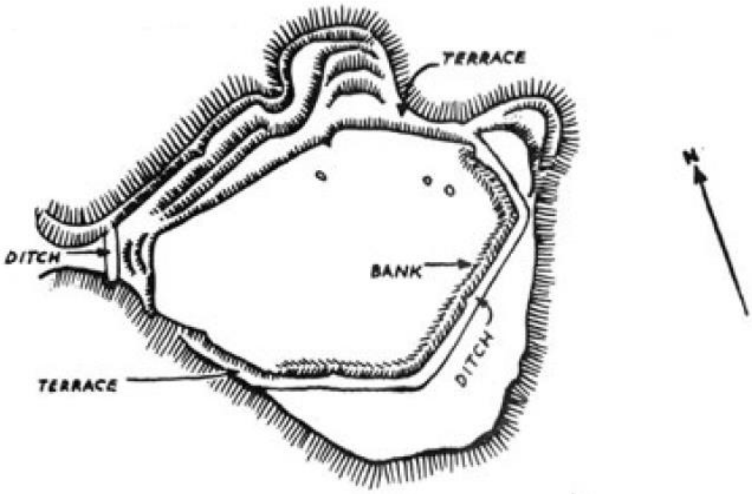


Figure 16: Okuratope Hill Pa, Bay of Islands (Golson 1957).

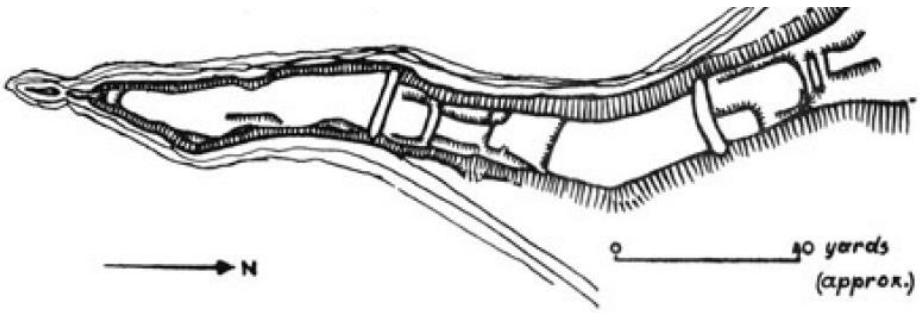


Figure 17: Tangitu Promontory Pa, Bay of Islands (Golson 1957)

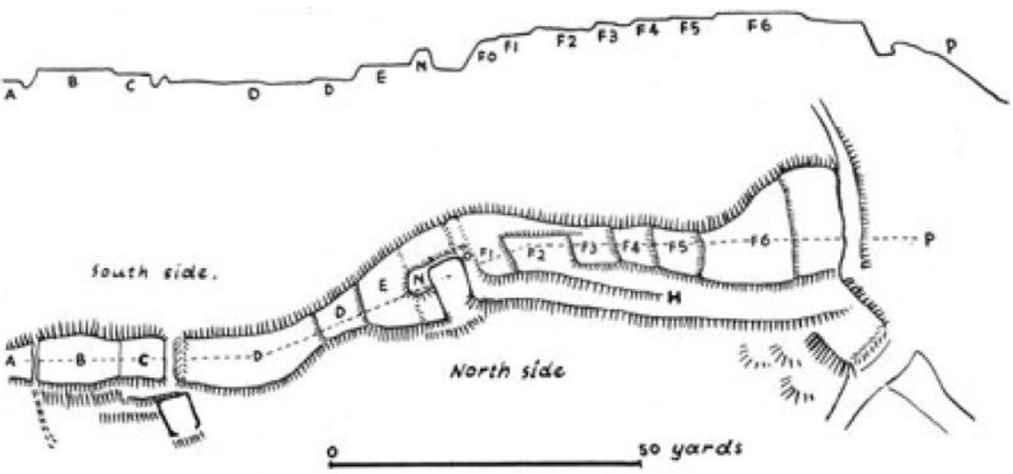


Figure 18: A Ridge Pa with profile, Kawakawa, Bay of Islands (Golson 1957)

3.4 PREDICTIVE ARCHAEOLOGICAL MODELLING IN NEW ZEALAND

Archaeological research in New Zealand has preferred inductive approaches to predictive archaeological modelling. A study of Pa site and pit distribution (Leathwick 2000) used logistic regression to determine correlations with environmental spatial datasets representing mean annual temperature, mean annual solar radiation, seasonal temperature range, rainfall, main soil types, surface rock types and distance to water sources. Leathwick focused on good accessibility to protein sources such as coastal fish supply and kumara growth areas. The results show correlations between Pa and pit locations and suitable soils (andesite and limestone), and close proximity to rivers, lakes and shoreline (within 1km), also areas that are protected from extreme winter temperatures while experiencing long warm periods throughout the year. Leathwick's results supported previous research that suggested pre-European Maori settlement was predominantly in the north and east of the country (Cumberland 1949, Gorbey 1970, Walton 2002, Moore 2012) with the majority of the early population being within the northern and central North Island regions (Northland, Waikato, Bay of Plenty, Taranaki, Gisborne and Hawkes Bay). The exception within the North Island being an absence of archaeological sites within the lower North Island interior (Walton 2002), although this is probably explained by the increased elevation and low seasonal temperature in the Tongariro National Park area just south of Taupo. The South Island prehistoric population was significantly lower than the North Island with just a few isolated areas favoured along the east coast near Christchurch and the Marlborough and Tasman regions at the top of the South Island. Leathwick acknowledges a number of elements of the study that could be improved;

- The spatial resolution of 1km for this national study could be improved significantly to match the available 20m topographic map resolution or more recently the 25m national digital elevation model.
- The spatial inaccuracy of the New Zealand Archaeological Association site data is also a constraint, typically mapped with a 100m radius buffer zone to reflect the accuracy of early survey methods.
- Only major rivers were included in the analysis, claiming that the use of lower river orders for suitable proximity analysis would have confused the results at a national scale.
- The development of an expert system at a regional or localised scale could be beneficial.

Following Leathwick's preliminary work, a follow up study (Arnold et al 2004) addressed some of Leathwick's recommendations by using a 100m x 100m map resolution.

Proximity to water source used rivers shown on the national 1:50,000 scale topographic maps (Land Information New Zealand), a refinement from Leathwick's use of the higher order river network. Water sources were classified into navigable river sections, distance to fresh water rivers, distance to coast, with the coastline suitability also considered. Both studies (Leathwick 2000) and (Arnold et al 2004) used Euclidean distance to water with no mention of the benefits of anisotropy, although this may not have been an option at a national scale due to the data quality and suitability for such analysis. Arnold does not refer to the uneven distribution of archaeological surveys across New Zealand, resulting in clusters of recorded sites along most of the coastline and significant gaps inland. These studies treated the archaeological record as a single environment and temporal period. Of more relevance to this research is the use of cost surface analysis of obsidian exchange across New Zealand (Scott 2007). Scott used ArcGIS to produce a least cost path network between 58 obsidian volcanic glass sources across the North and South Islands.

Figure 19: Least cost paths between obsidian sites with high sea cost (Scott 2007).

The method used ArcGIS 9 software and involved a resolution of 25km x 25km, and a cost weighted method combining some simple relative travel friction costs representing rivers and sea. The Cost Path tool then produced optimal paths through this cost surface producing high level isotropic least cost paths. Different scenarios were modelled using high and low marine travel cost for different obsidian sample time periods. Figure 19 shows one of the scenarios where high sea travel cost forced least cost paths overland and across narrow stretches of water.

CHAPTER 4: METHOD

This chapter provides an introduction to the study area's natural environment and formally recorded archaeology. Data sources and a description of the geo-processing methods followed are provided for the creation of the cost surface, and the execution of the path distance and least cost path processes. The extraction of the denser primary least cost paths and secondary path networks are also described. Finally, Chi Squared statistical significance testing is used to test a Pa site distribution hypothesis followed by a Gain statistic to measure the model's performance.

4.1 THE STUDY AREA

The study area (figure 20) covers 5,230 square kilometres spanning the Northland peninsula from east to west coast and falling mainly within the current Far North District, but also covering parts of Whangarei District (south-east) and Kaipara District (on the southern boundary).The Northland regional climate is sub-tropical due to the effect of the South Pacific flow bringing warm water down from the equator. The region experiences an average summer daytime temperature ranging from 22 to 30 degrees, and mean annual rainfall is approximately 1000 -1300mm in coastal areas. Elevation is relatively low compared with other regions with a maximum height of 767m located within the study area.

The study area has a rich history of Maori and European settlement and conflict, which is evident in the archaeological record. The study area was chosen because the Bay of Islands on the east coast was one of the most highly populated areas in New Zealand prior to European settlement. The significant amount of coastal archaeological sites on each side of the peninsula provides an ideal setting for this research. The following sections briefly describe the current and historic environment.

Figure 20: Study area

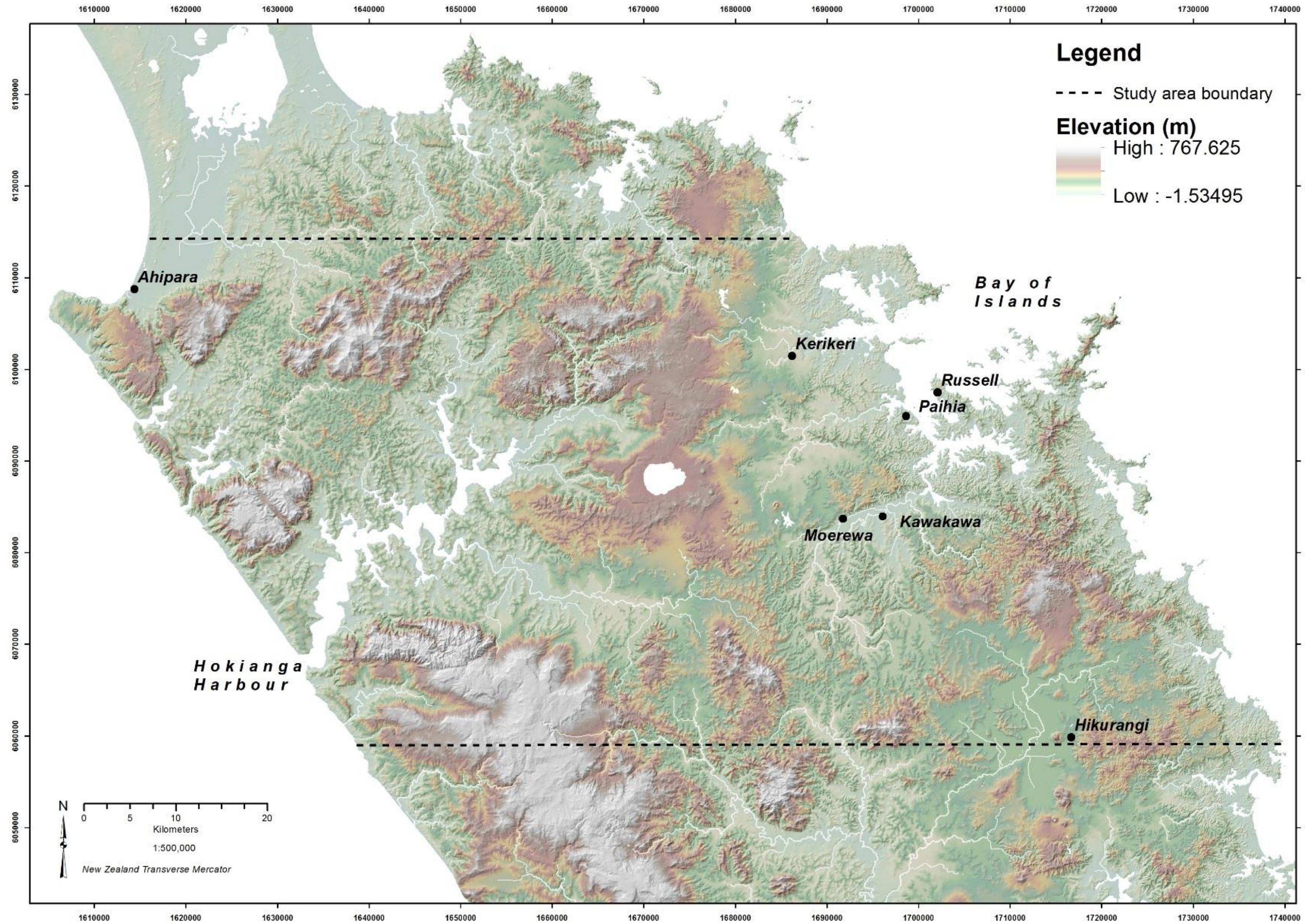


Figure 21: Elevation

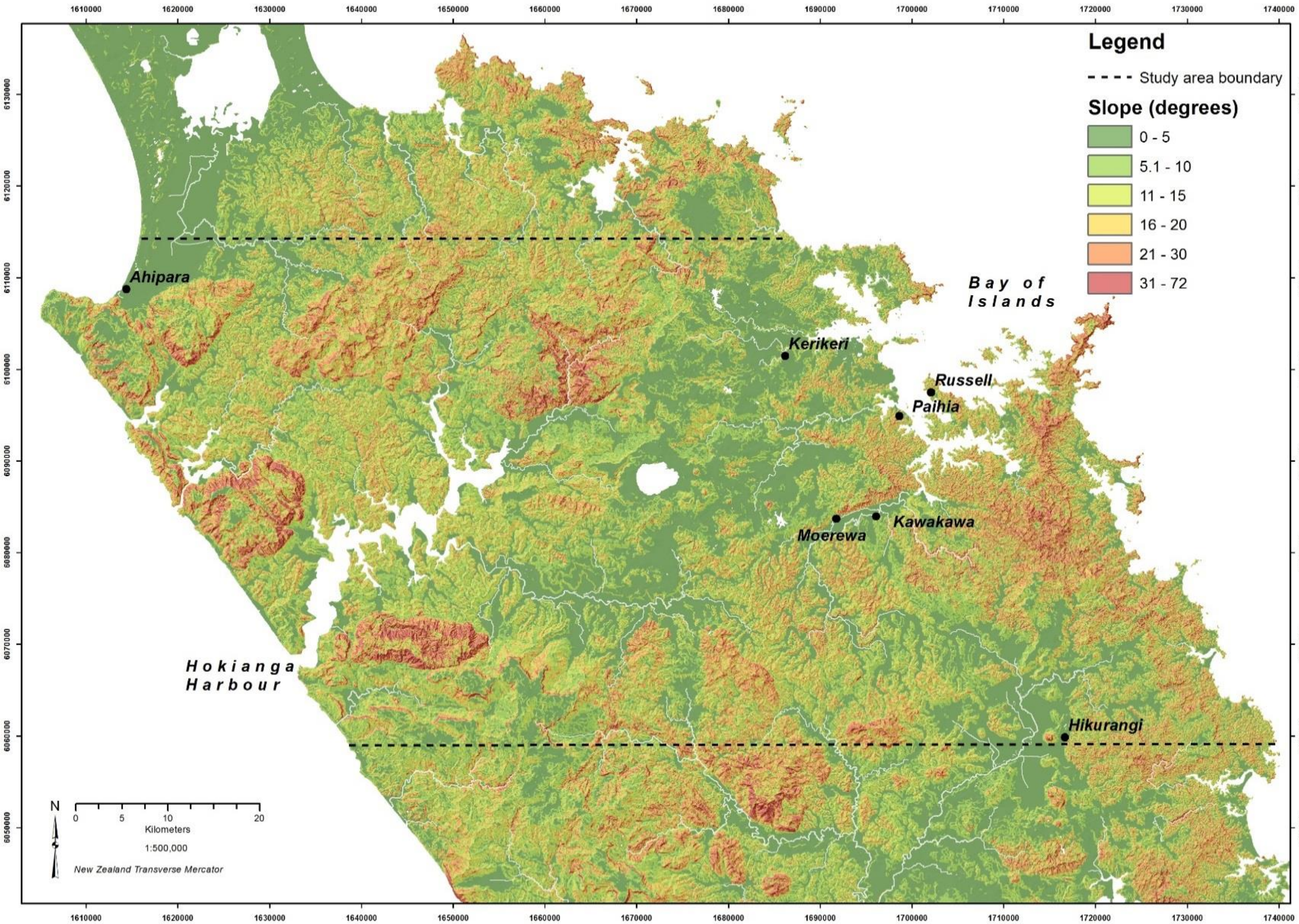


Figure 22: Slope

4.2 GEOLOGY

The Far North District area has 80 distinctive landscape features identified in the Inventory of Important Geological Sites and Landforms in the Northland Region, recorded by the Geological Society of New Zealand (Kenny and Hayward, 1993). Volcanic scoria rock cones are surrounded by basalt rock fields formed from lava flows 30,000 to 60,000 years ago (figure 23). In some cases, the cones had airborne ash deposited on them which would later become suitable for garden terracing (Sutton 2003).

Figure 23:Volcanic cone, Pouerua (K.L. Jones/Department of Conservation).

The study area geology contains a mixture of sedimentary and volcanic surface rock. Sandstone based greywacke dominates the coastal areas and large areas of basalt break up the inland plains, becoming more prominent to the north near Kerikeri. Northland's soils are typically thin and prone to erosion. See also figure 62: The New Zealand Fundamental Soils map.

4.3 HYDROLOGY

Northland has a network of relatively short rivers mainly draining to harbours. The Wairoa River is Northland's largest river with a catchment that covers approximately a third of the region. Hokianga Harbour is a drowned river system which was navigable up to 40 Km inland by early European settlers involved in the timber trade between 1822AD and 1900AD. Flooding has become more frequent due to land use change and erosion.

Figure 24: Inland view from Hokianga Harbour on the west coast.

With permission: Te Ara, the Encyclopedia of New Zealand.

Figure 25: Waitangi River, lowland area draining to the Bay of Islands (Land Air Water Aotearoa).

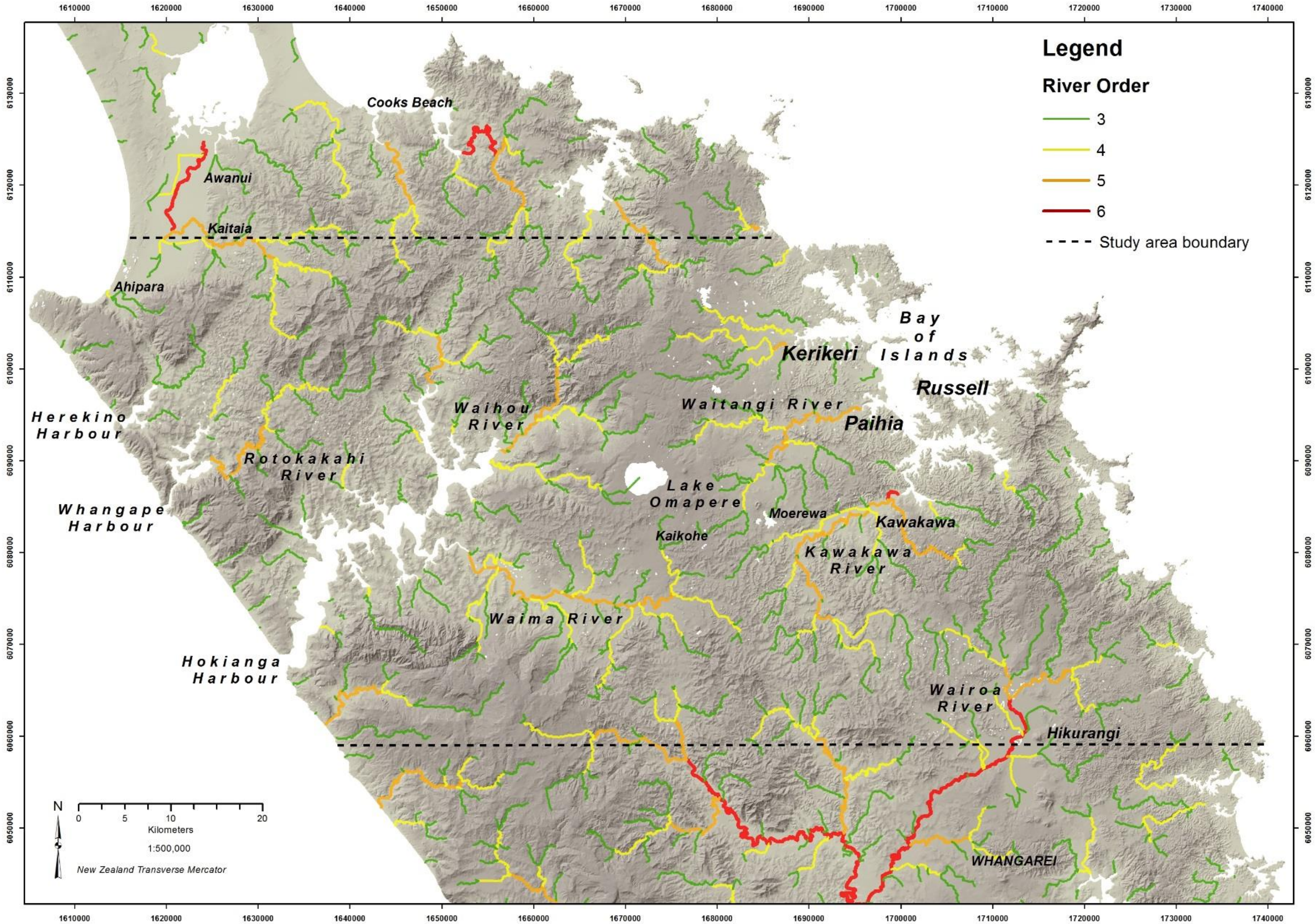


Figure 26: River network showing catchment order

4.4 THE NEW ZEALAND ARCHAEOLOGICAL ASSOCIATION SITE DATASET

The New Zealand Archaeological Association (NZAA) spatial database (ArchSite - archsite.org.nz/) provided a point dataset of 3,779 recorded archaeological sites within the study area (figure 27). These sites are classified into 35 site types (table 1). The spatial and categorical accuracy of the site records is variable. Some areas have had intensive archaeological field surveys conducted resulting in relatively accurate coordinates, while others are generally considered to be mapped within 100 metres of the real position. The site dataset does not contain a temporal classification. The point based site records typically represent visible surface evidence only, rarely the subsurface of a site that it may be associated with.

Site Type	Count	Site Type	Count	Site Type	Count
Agricultural/ pastoral	21	Health care	36	Pa – island / swamp	2
Artefact find	38	Historic - domestic	63	Pit/Terrace	1733
Botanical evidence	42	Historic - land parcel	8	Religious	6
Burial/ cemetery	56	Industrial	18	Shipwreck	1
Canoe building	1	Maori horticulture	235	Source site	7
Cave/ rock shelter	10	Midden/Oven	724	Timber milling	53
Commercial	8	Military (non-Maori)	11	Traditional site	16
Educational	4	Mining	7	Transport/ communication	42
Fishing	6	Mining - gold	2	Unclassified	36
Flax milling	3	Mission station	15	Whaling Station	2
Flour milling	2	Pa	563	Working area	2
Gum digging	4	Pa - gunfighter	2		
				Total	3779

Table 1: Summary of archaeological site types within the study area.

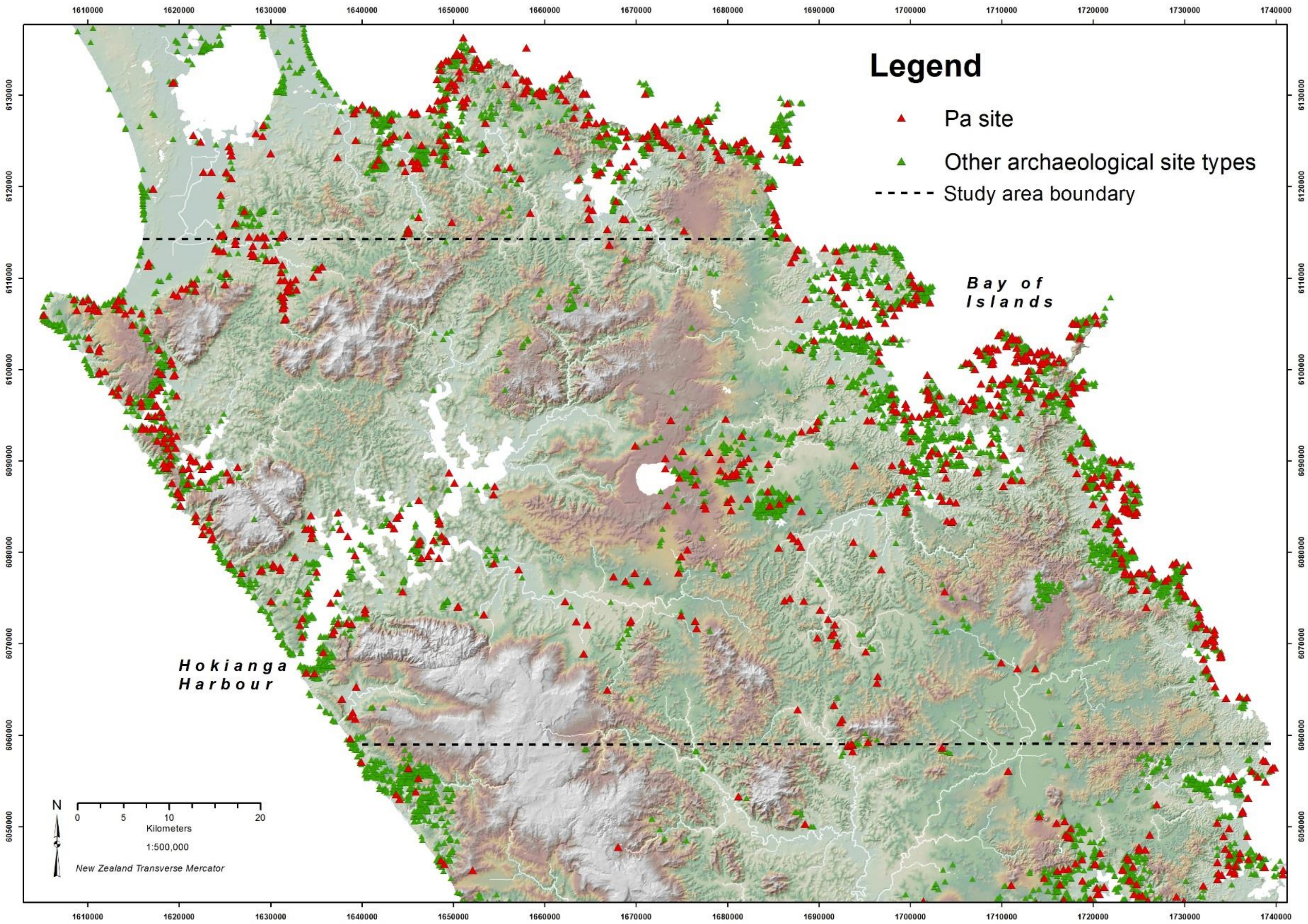


Figure 27: New Zealand Archaeological Association sites.

4.5 DATA SOURCES

Landcare Research NZ Ltd provides a 25m resolution Digital Elevation Model (DEM), the *New Zealand National Digital Elevation Model (North Island),* created in 2010 which was considered suitable for this research. The elevation data was sourced from 1:50,000 scale topographic data and was interpolated using 20m contours, water bodies and known height locations. The Digital Elevation Model was originally captured in NZ Map Grid 1949 projection and was converted to New Zealand Transverse Mercator 2000 projection using the Landcare Research GIS data portal for this research project. The DEM uses the New Zealand Geodetic Datum 2000, which includes height above mean sea level established by modern survey techniques. Landcare Research state planimetric accuracy (XY) as: 90% of well-defined points are within +/- 22m and vertical accuracy (Z): 90% of well-defined points are within +/- 10m.

Lakes were sourced from Land Information New Zealand's Topo 50 map series, also captured at 1:50,000. River data was sourced from the National Institute of Water and Atmospheric Research (NIWA) River Environment Classification dataset (REC v2.0). This data provides national river network coverage in shape file format for catchment order 1 to 7. This river network data was derived from a 30m DEM which is based on 20m vector contours and spot heights. Archaeological site data in point shape file format was provided by Opus International Consultants Ltd who have a licence to access the New Zealand Archaeological Association spatial dataset. All data was used in New Zealand Transverse Mercator 2000 projection.

Dataset	Format	Source
NZDEM North Island. Digital elevation model with 25 metre resolution	TIF	The Land Resource Information Systems Portal https://lris.scinfo.org.nz/layer/131-nzdem-north-island-25-metre/
NZ Mainland lake polygon. Digitised at 1:50,000 scale	Polygon Shape file	Land Information New Zealand https://data.linz.govt.nz/layer/293-nz-mainland-lake-polygons-topo-150k/
NIWA river order (REC v2.0).	Line Shape file	The National Institute of Water and Atmospheric Research (NIWA) http://www.niwa.co.nz/freshwater-and-estuaries/management-tools/river-environment-classification-0
New Zealand Archaeological Association site data	Point Shape file	http://archsite.org.nz/

Table 2: Spatial data sources

4.6 SOFTWARE

This research used the following software as stated;

- **ArcGIS 10.1 for Desktop** (sole GIS software used throughout this research)
- **ArcGIS Spatial Analyst and Model Builder 10.1** (required to operate the Path Distance and Least Cost Path models)
- **XLSTAT** http://www.xlstat.com/en/ (used to apply the Multinomial Goodness of Fit test and check Chi Squared results.
- **InDesign** (graphic design of figures used in the literature review)
- **Microsoft Excel** template for Chi Squared http://www.real-statistics.com/free-download/real-statistics-examples-workbook/

4.7 COST SURFACE COMPOSITION

In order to conduct anisotropic least cost path analysis, a cost surface is required to represent relative travel friction across the landscape. Two cost themes were considered;

- Lakes and rivers
- Historic wetlands.

The hydrology cost surface required lakes represented as a high travel cost and rivers as a gradually increasing cost as river flow increases. An assumption was made that current rivers and lakes are a reasonable representation of historic water features (approx. 1,000AD–1750AD). Current lake vector polygon data was converted to a raster dataset and reclassified with lake cells allocated a value of 99 and the remaining cells given a value of 1. The overland cost of travelling across a river required an initial three step geo-processing model to fill any digital elevation model depressions followed by the use of ArcGIS hydrology tools to calculate flow direction and flow accumulation across the DEM surface. The resulting dense flow network dataset required simplifying by eliminating drainage flows less than 200 cells. This was achieved using ArcGIS Raster Calculator with the following expression;

Con("%flow accumulation raster%" >= 200,"%flow accumulation raster%")

Introducing a relative flow cost across the river network involved each river cell value being divided by the highest flow value in the river network (Whitley and Hicks 2003). This provided a continuous flow scale from source to sea. River travel cost must be relative to over land travel time to realistically represent how much longer it would take to cross a river. To achieve this the flow data was converted to a scale between zero and one. The flow value at the most upstream raster cell was allocated a cost of zero, equivalent to a small stream which would not require additional time to cross.

The largest accumulated flow value is situated at the mouth of the Waima River which flows into Hokianga Harbour on the west coast. The last downstream cell was given a cost of one. The relative cost of traversing a river was considered to take approximately six times longer than travelling across flat dry land assuming swimming or wading across the river is practical in a tidal river (Balstrøm 2002).

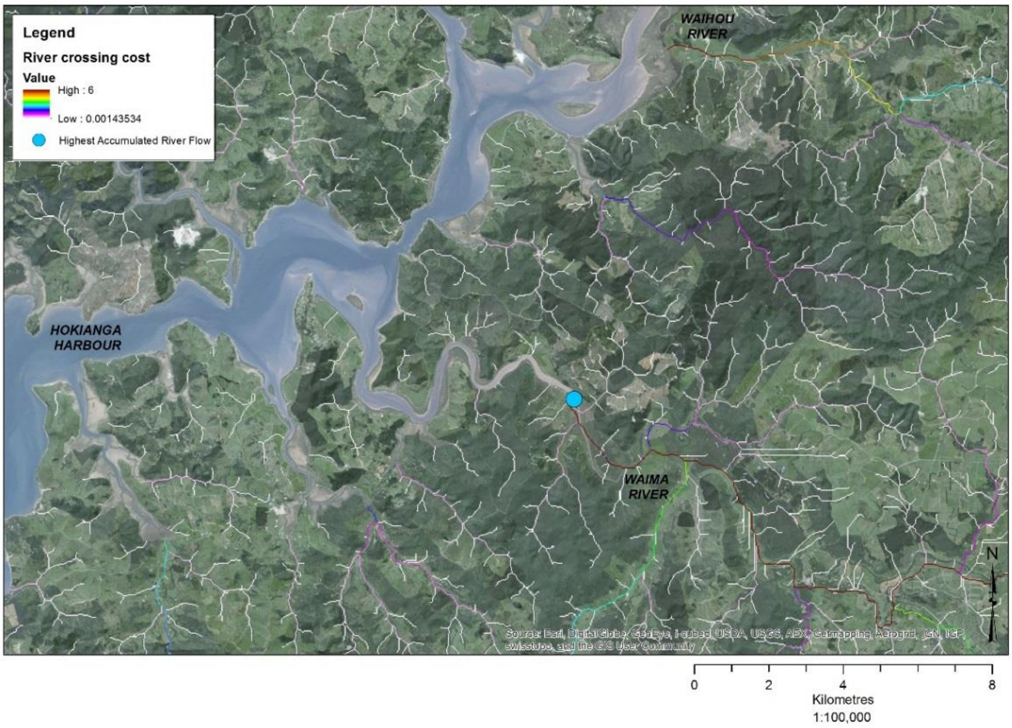


Figure 28: Maximum accumulated river crossing location

The zero to one flow data was then multiplied by this factor of six to produce a simple river crossing friction surface. As least cost paths are able to exploit single cell flow accumulation networks by crossing rivers at points where cells connect diagonally, the single cell network had its width increased. This was completed using Euclidean Allocation transferring the river cost to one cell either side of each river raster cell. Other travel cost variables were considered including data that may have indicated the presence of historic wetlands as a useful travel cost variable. The Authors of the only available spatial dataset representing historic wetlands in New Zealand - The Freshwater Ecosystems of New Zealand (Ausseil et al 2008), (figure 63), stated that the level of generalisation, assumptions and resolution of the data made it unsuitable for use as a travel cost variable. Alternative modern soil datasets were also considered as a potential travel cost variable for wetlands. There are several national spatial datasets available covering different soil characteristics. The first edition of the New Zealand Land Resource Inventory (NZLRI), was created at 1:63,360 scale between 1973 and 1979. Northland and several other regions had a second edition of the soils dataset developed at 1:50,000 scale post 1983.

The New Zealand Fundamental Soils dataset was an alternative option. This dataset is an enhancement of the NZLRI data with improved classifications but introduced very little spatial difference between the two datasets. Currently, the S-Map dataset is a further improvement on previous editions digitised at 1:25,000 scale and is being gradually developed and published. However this data does not yet cover the Northland region.

Following discussions with Landcare research and Northland Regional Council, these datasets were considered unsuitable as a travel cost variable due to the amount of land use change following mass land drainage for farm deforestation in the 1800's and the highly generalised nature of the soil type classification.

4.7.1 RIVER COST EFFECT ON LEAST COST PATHS

Once least cost paths had been created using the lake and river cost surface, a sample test was conducted to understand the effect that river cost has on the path of least resistance. The test involved path distance analysis with and without a cost surface input. Two locations were selected representing a simple river crossing scenario and a more complex version involving a river and tributaries.

4.8 LEAST COST PATH NETWORK AND STATISTICAL TESTING OVERVIEW

Figure 29 shows an overview of the key stages in the development of the primary and secondary paths and statistical analysis steps.

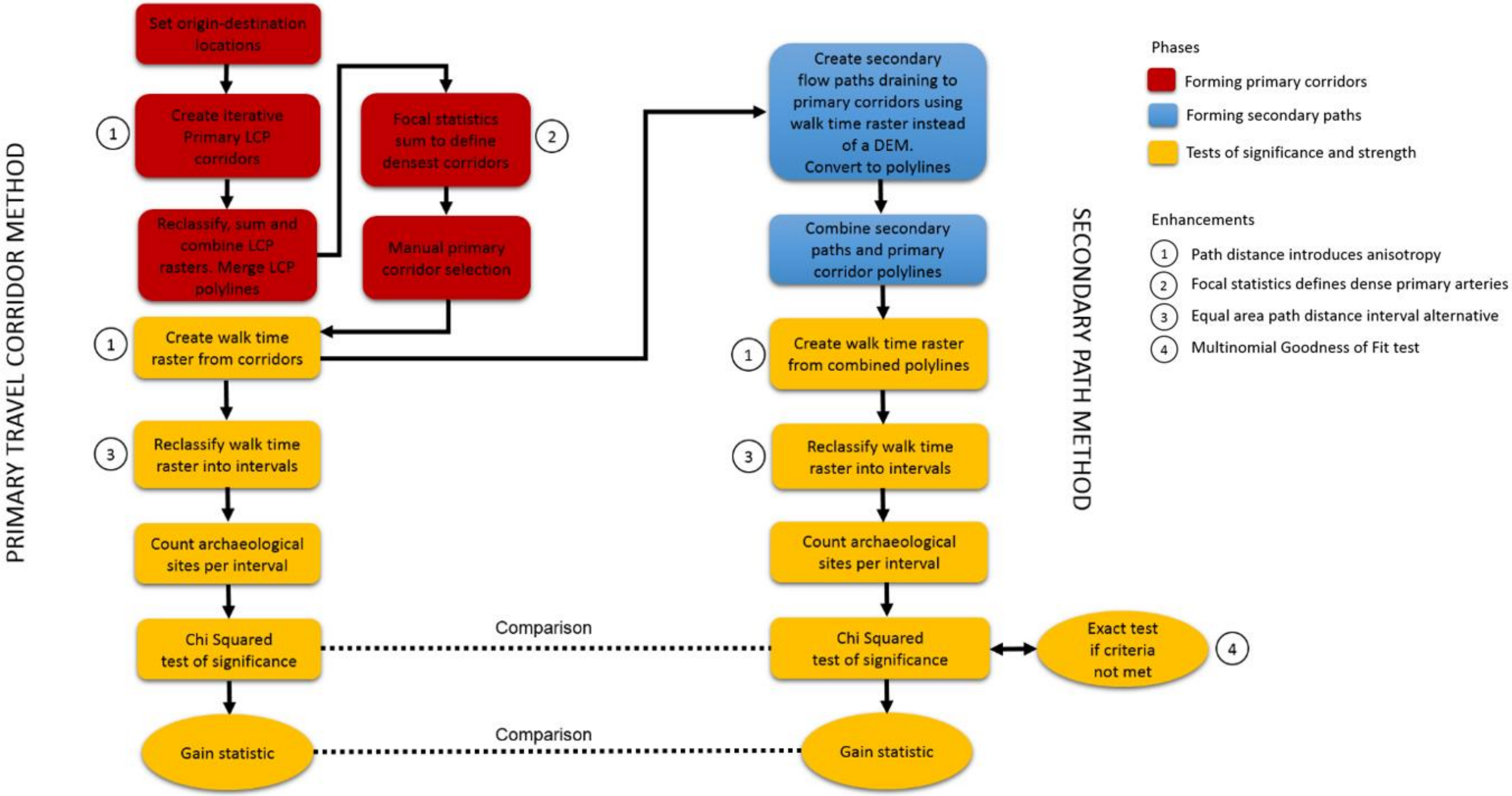


Figure 29: Overview process for defining travel paths and statistical analysis

4.8.1 PRIMARY TRAVEL CORRIDOR METHOD

Primary travel corridors were modelled using the ArcGIS Path Distance tool to produce an anisotropic based least cost path network. Figure 30 shows the main steps involved in defining the primary path network.

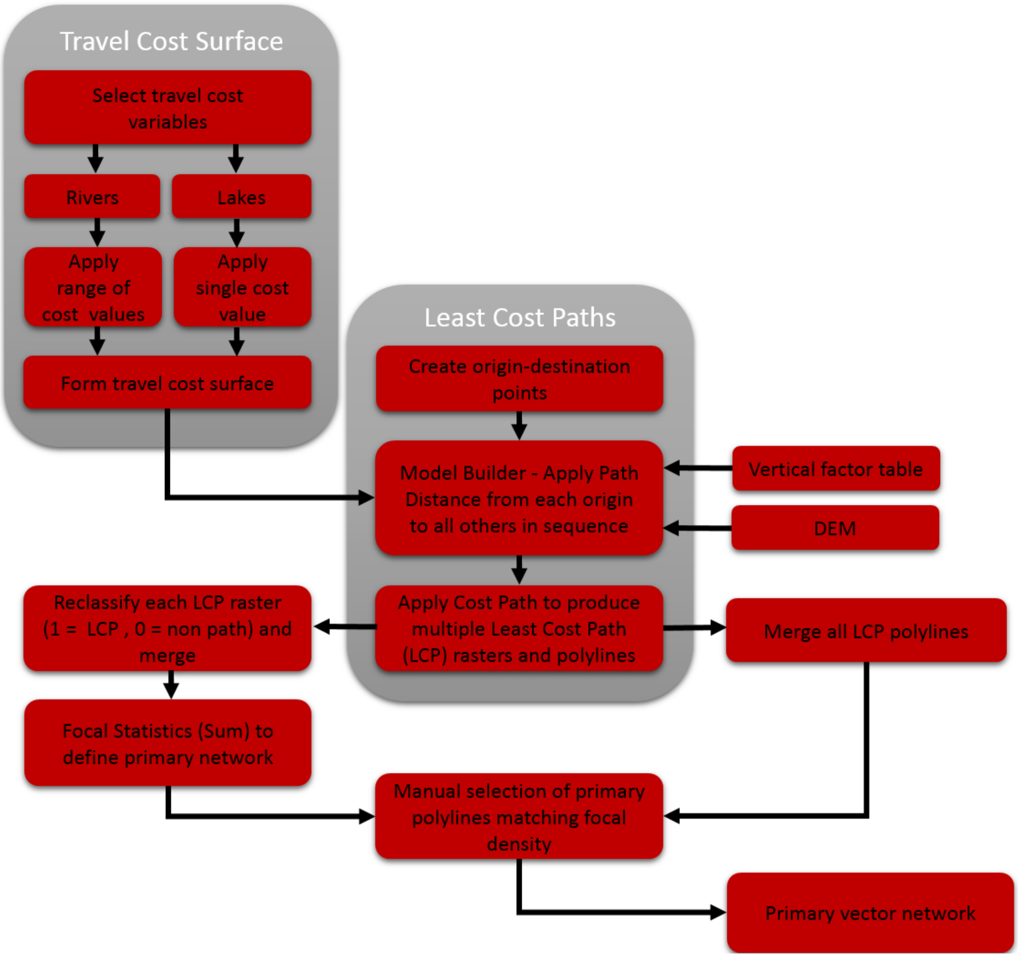


Figure 30: Objective 2 - Primary corridor workflow

A vertical factor table was required to perform path distance calculations. In this case the table factor represents the relative travel cost of terrain slope. Tobler's hiking algorithm (Tobler 1993) was used to create a text file with a column for slope in degrees and a second column which applies the following equation;

Hours to cross 1m = 0.000166666(exp (3.5*(ABS (Tan (Radians (slope)) + 0.05))))* (Tripcevich 2009).

Tripcevich provides an example (Table 3) showing a shortened vertical factor table using the above equation. ArcGIS will accept this short version by interpolating between values so not every degree value is required.

Slope (degrees)	Tobler's Vertical Factor	Speed (Km/hr)
-90	-1.000000	-0.001
-80	-1.000000	-0.001
-70	2.099410	0.000
-50	0.009065	0.110
-30	0.001055	0.947
-10	0.000259	3.856
-5	0.000190	5.262
-4	0.000179	5.596
-3	0.000168	5.950
-2	0.000176	5.692
-1	0.000187	5.354
0	0.000199	5.037
5	0.000270	3.708
10	0.000368	2.717
30	0.001498	0.668
50	0.012863	0.078
70	2.979204	0.000
80	-1.000000	-0.001
90	-1.000000	-0.001

Table 3: Example of a vertical factor table and equivalent walking speed

Figure 31 shows that the maximum walking speed using Tobler's hiking equation to form the vertical factor table is 5.95 km/hr on a downhill slope of 3 degrees.

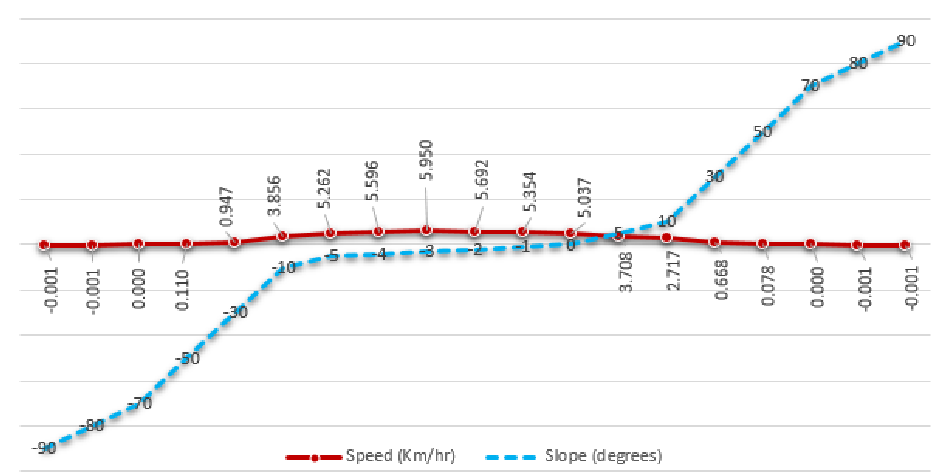


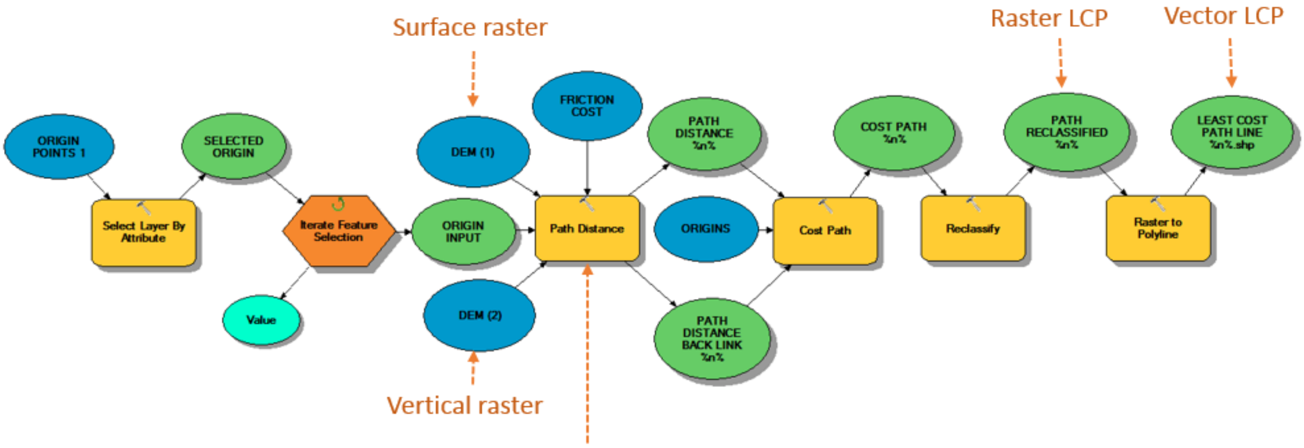
Figure 31: Profile showing Tobler's calculation of walking speed vs slope

The ArcGIS Path Distance tool requires the following data inputs

Path Distance tool setting	Data
Input Source Feature	Origin point shape file with a sequential index value
Input Cost Raster	Lake and river cost raster
Input Surface Raster	Digital Elevation model
Input Horizontal Raster	N/A
Horizontal Factor	N/A
Input Vertical Raster	Digital Elevation model
Vertical Factor	Text file with two columns representing slope in degrees and the corresponding walking velocity derived using Tobler's hiking function.
Maximum distance	N/A

Table 4: ArcGIS Path Distance settings

ArcGIS Model Builder was used to develop a geo-processing model to produce one to many least cost paths from each origin point to all other points (destinations). An iteration process was used to loop through the origin point data in an ordered sequence following a unique identifier index. Each least cost path raster was then reclassified to a (1, 0) format where 1 represented the path and 0 represented non path cells. Finally the least cost path was also converted to a polyline format as an alternative to the classified raster format for further analysis.



Slope (degrees)	Tobler's Vertical Factor	Speed (Km/hr)
-90	-1.000000	-0.001
-80	-1.000000	-0.001
-70	2.099410	0.000
-50	0.009065	0.110
-30	0.001055	0.947
-10	0.000259	3.856
-5	0.000190	5.262
-4	0.000179	5.596
-3	0.000168	5.950
-2	0.000176	5.692
-1	0.000187	5.354
0	0.000199	5.037
5	0.000270	3.708
10	0.000368	2.717
30	0.001498	0.668
50	0.012863	0.078
70	2.979204	0.000
80	-1.000000	-0.001
90	-1.000000	-0.001

Figure 32: Path distance and least cost path generation

4.8.2 LEAST COST PATH PATTERN TESTS

Three origin-destination patterns were tested for modelling the primary least cost corridors to confirm whether start and end positions change the pattern and density of primary natural travel corridors. Least cost paths were processed using ArcGIS Model Builder from each origin to all destination points in each case.

- **Method 1** involved origin-destination points set around the study area perimeter at 1Km intervals (figure 33), replicating Whitley and Hicks 2003 method. The shape of the study area required more origin-destination points on the east and west coast compared with the artificial straight line border on the northern and southern boundaries. Harbours were classed as complete barriers to overland travel and were excluded from the study area.

- **Method 2** involved a uniform grid of 225 origin-destination points at 5Km intervals, (figure 34).

- **Method 3** involved 100 origin-destination points selected randomly (figure 35) using the ArcGIS create random points tool.

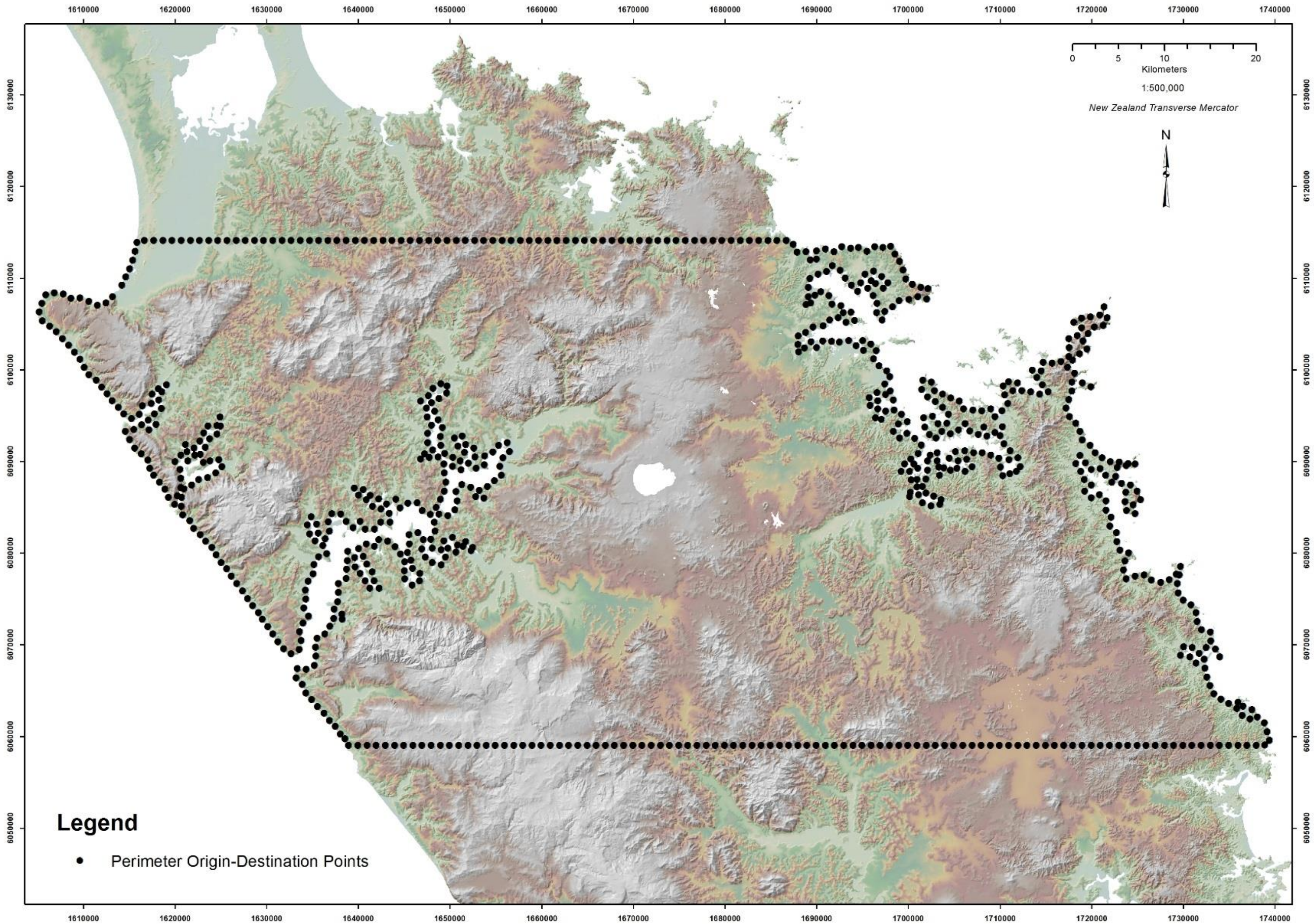


Figure 33: Method 1: Perimeter origin-destinations at 1km intervals

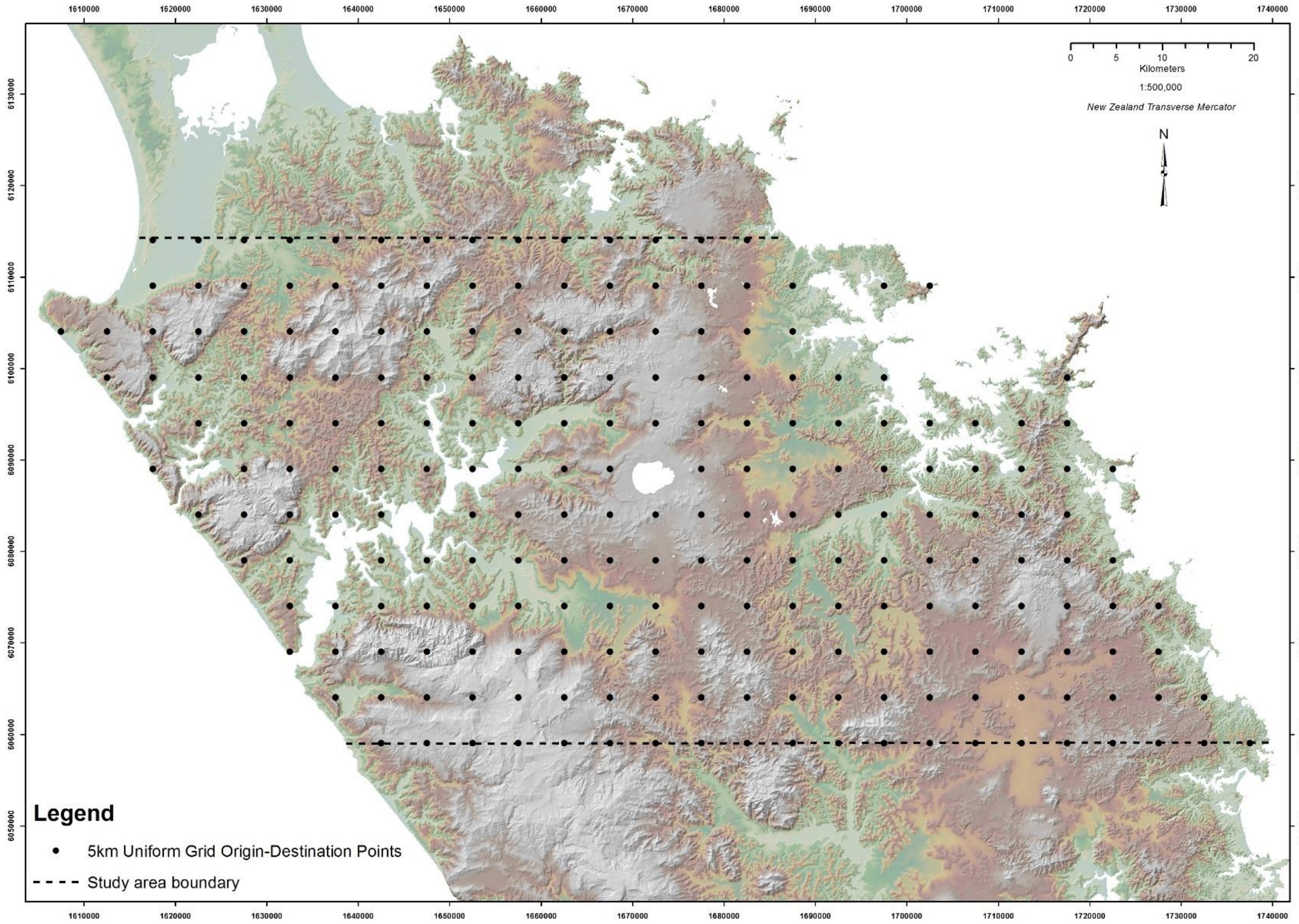


Figure 34: Method 2: A uniform grid of origin-destinations at 5km intervals

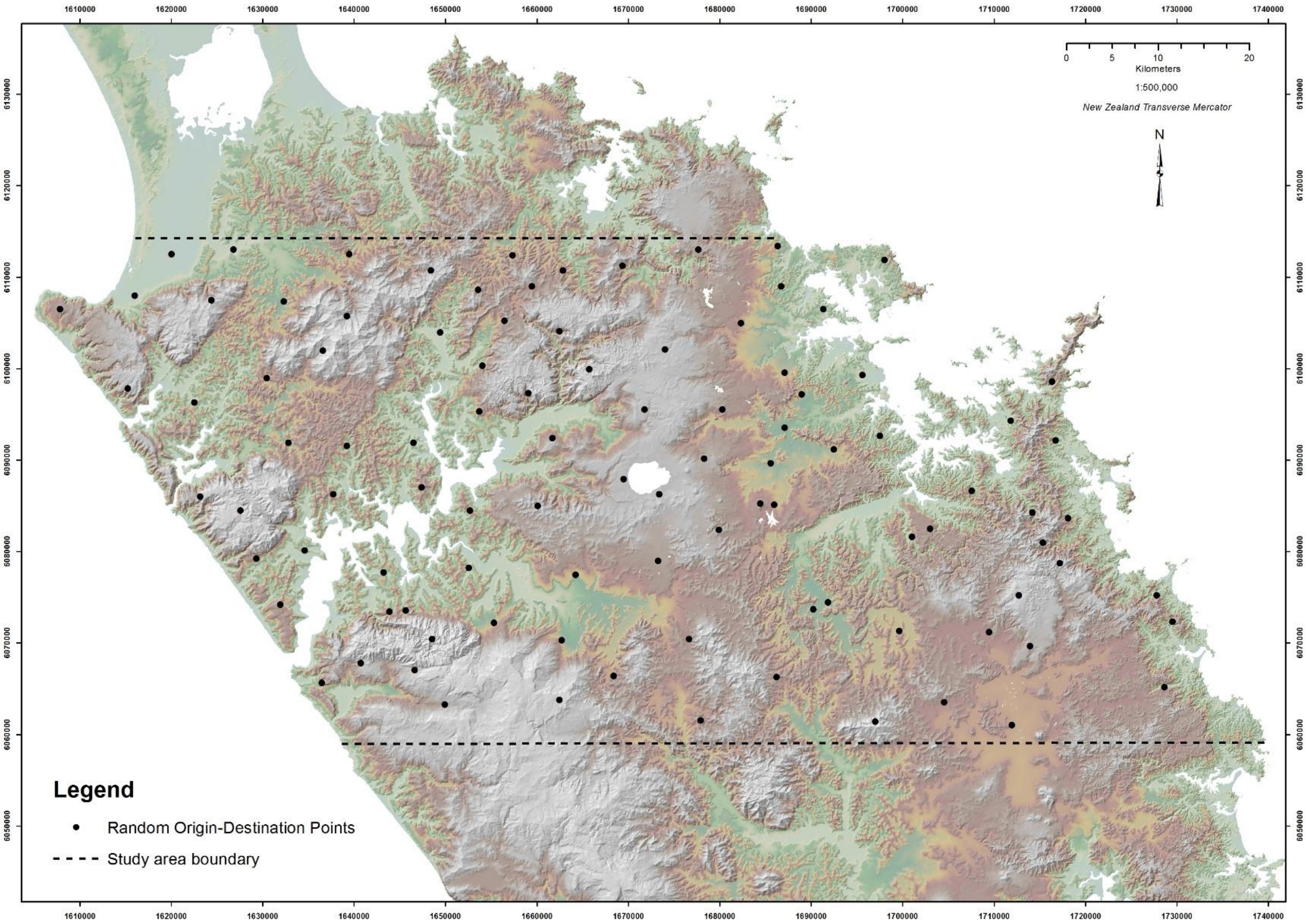


Figure 35: Method 3: Random origin-destinations

4.8.3 PRIMARY CORRIDOR DEFINITION

All reclassified least cost path raster datasets were summed using the ArcGIS Cell Statistics tool. This produced a single raster dataset with cells representing the total number of least cost paths passing through them from all origin-destination routes. The resulting raster dataset required a further step to help define the densest routes. ArcGIS Focal Statistics was used to count the number of least cost path values present within a circular four cell radius (100m) search around each raster cell. This method picks up adjacent paths to form a broad corridor. This process was repeated for search radius 250m and 500m for comparison.

The primary least cost path corridors needed to be identified within the focal statistics raster dataset by testing different raster classification methods available within ArcGIS 10. The least cost path model also produced a vector line for each path, merged into one master network dataset. This vector line data was used to select paths that matched the densest corridors recorded in the 250m radius focal statistics raster dataset. Least cost paths following the edge of the study area were not considered for the remaining analysis due to potentially unrealistic edge effects. Corridors near the edge of the study area were considered to start from where routes converged.

4.8.4 SECONDARY PATH METHOD

Figure 36 shows the ArcGIS workflow that was followed to produce secondary paths and combine them with the primary corridor network. As secondary paths have no known destination they must be formed using an alternative method to least cost path. This stage follows Whitley and Hicks's use of Cost Distance (adapted to Path Distance) and overland flow accumulation. Normally the Flow Direction tool used to create a stream network would use the terrain (DEM) as its surface. In this case the DEM is replaced with the output from the Path Distance tool providing an anisotropic cost distance raster for the secondary paths to navigate across. A number of tests were conducted to determine a suitable upstream drainage threshold value.

Thresholds of 100, 200, 300 and 400 were modelled. Once a suitable threshold was selected (200 upstream cells), the secondary paths were converted to a Strahler order (values 1 to 6) and reclassified to provide a background value of zero. The primary raster network was also reclassified and allocated a value of 7 (the equivalent of the largest Strahler order ensuring that the primary network was classified) and also given a background value of zero. The primary and secondary path raster datasets were then combined using the Fuzzy OR operator which merges two raster datasets retaining the maximum coincident input values.

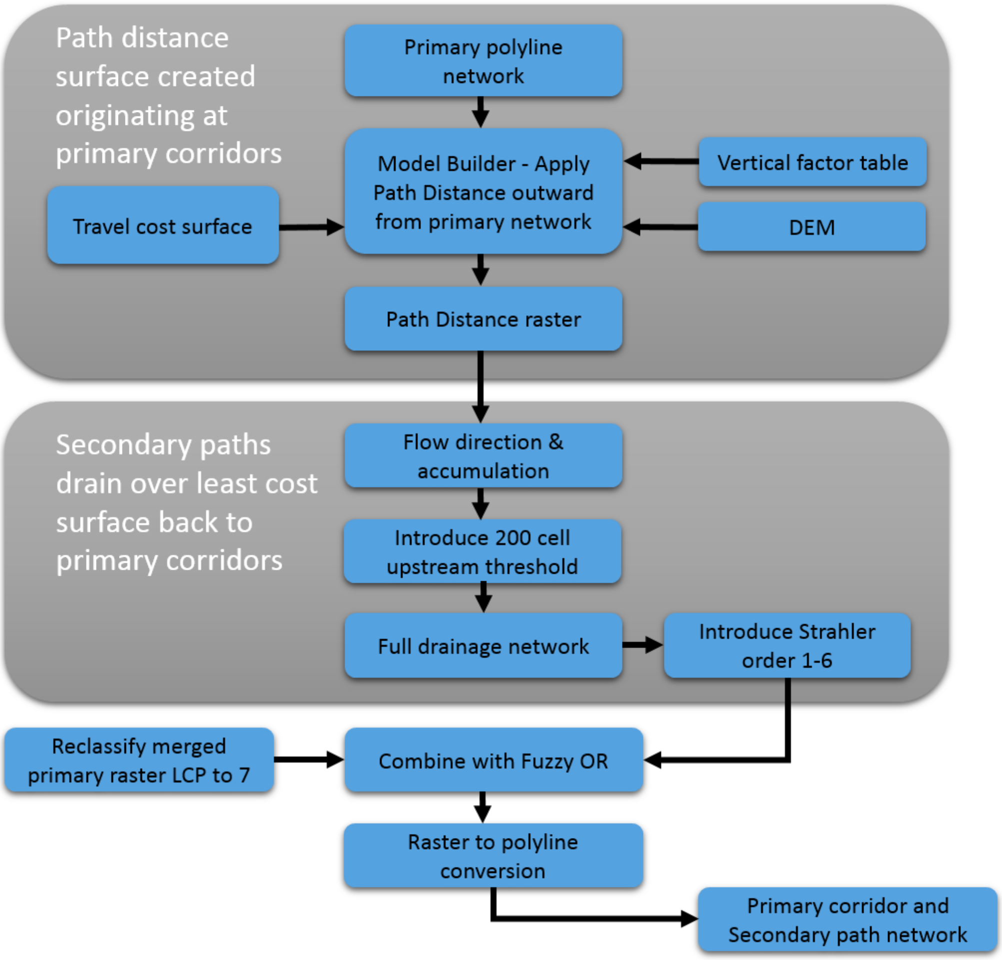


Figure 36: Objective 2 – Secondary path workflow

The Strahler order was only used as a visual check that the merge had been completed successfully. The combined raster network dataset was converted to a vector polyline for statistical analysis.

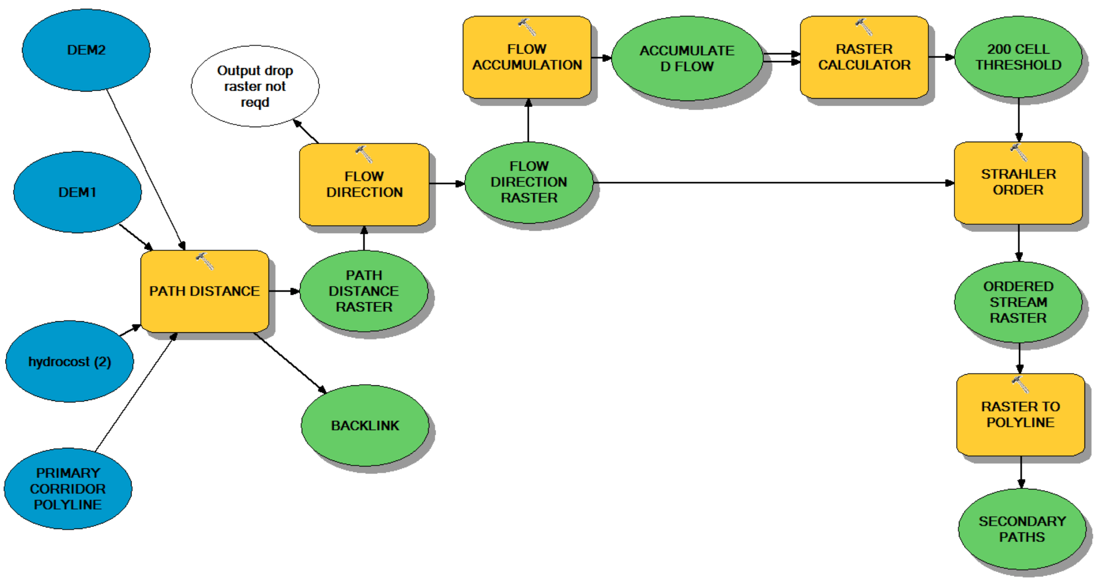


Figure 37: Secondary path geo-processing steps using ArcGIS Model Builder

The primary and secondary path raster datasets were then combined using the Fuzzy OR operator which merges two raster datasets retaining the maximum coincident input values. Finally, the combined raster network dataset was converted to a vector polyline dataset for statistical analysis.

4.9 PA DISTRIBUTION RELATIVE TO PATHS

4.9.1 CLASSIFICATION OF INLAND PA

As Pa sites are thought to have a connection with inland trails, this site type seemed suitable to analyse the spatial relationship between Pa distribution and least cost paths. Of the 3,779 recorded archaeological sites that fall within the study area, 567 are classified as some form of Pa site. Pa sites located offshore were removed for this analysis, leaving a dataset of 527 Pa sites. The clustering of mainland Pa sites within the study area suggested that they are located within two distinct environments, coastal and inland. As the majority of archaeological sites (including Pa) are located in the coastal environment, it was necessary to define a boundary between these two zones to allow the study to focus on inland sites only, as coastal Pa sites would have been positioned for other reasons such as easy access to coastal resources rather than being positioned near to cross country trails.

An arbitrary 30 minute path distance walking travel time catchment from the coastline was formed (figure 38). This classification resulted in 346 Pa being classified as coastal and 181 as inland sites. The inland Pa sites are not evenly distributed. In the southern and north western areas inland Pa appear to be generally located in low lying areas, whereas the central area has Pa sites clustered around basalt fields at Pouerua and near to Lake Omapere (central large lake) and Kawakawa, an area that has had a significant amount of archaeological surveys.

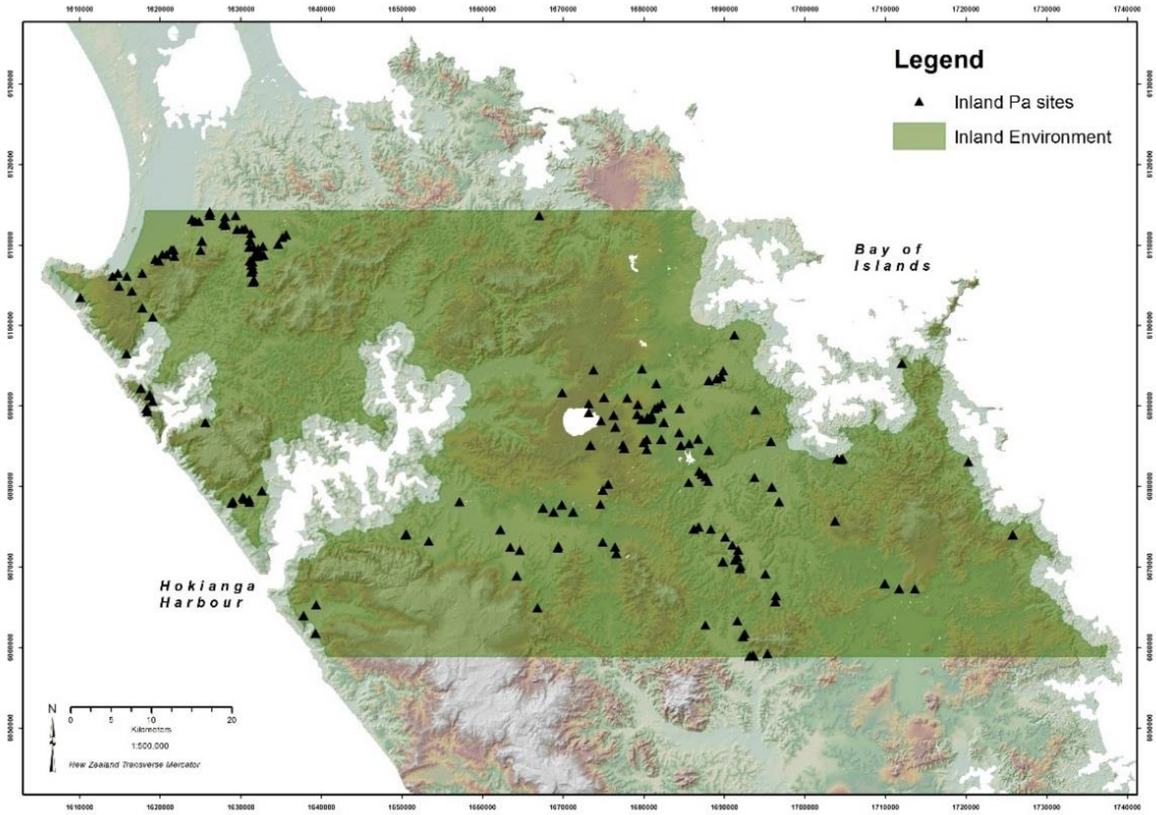


Figure 38: The inland environment

4.9.2 PATH DISTANCE CLASSIFICATION METHOD

In order to answer the research question;

Are inland Pa sites positioned in close proximity to least cost overland travel paths?,

a Chi Squared test of significance was used to test the following hypothesis;

- *Null hypothesis: There is no difference between observed Pa site distribution relative to regional travel paths and the expected background site distribution'.*
- *Alternative Hypothesis: There is a difference between observed Pa site distribution relative to regional travel paths and expected background site distribution'.*

Figure 39 shows the steps followed to test Pa distribution significance for both networks using equal time and equal area path distance intervals. The multinomial goodness of fit test is only an option if the expected site counts required for the Chi Squared test are below the minimum number required.

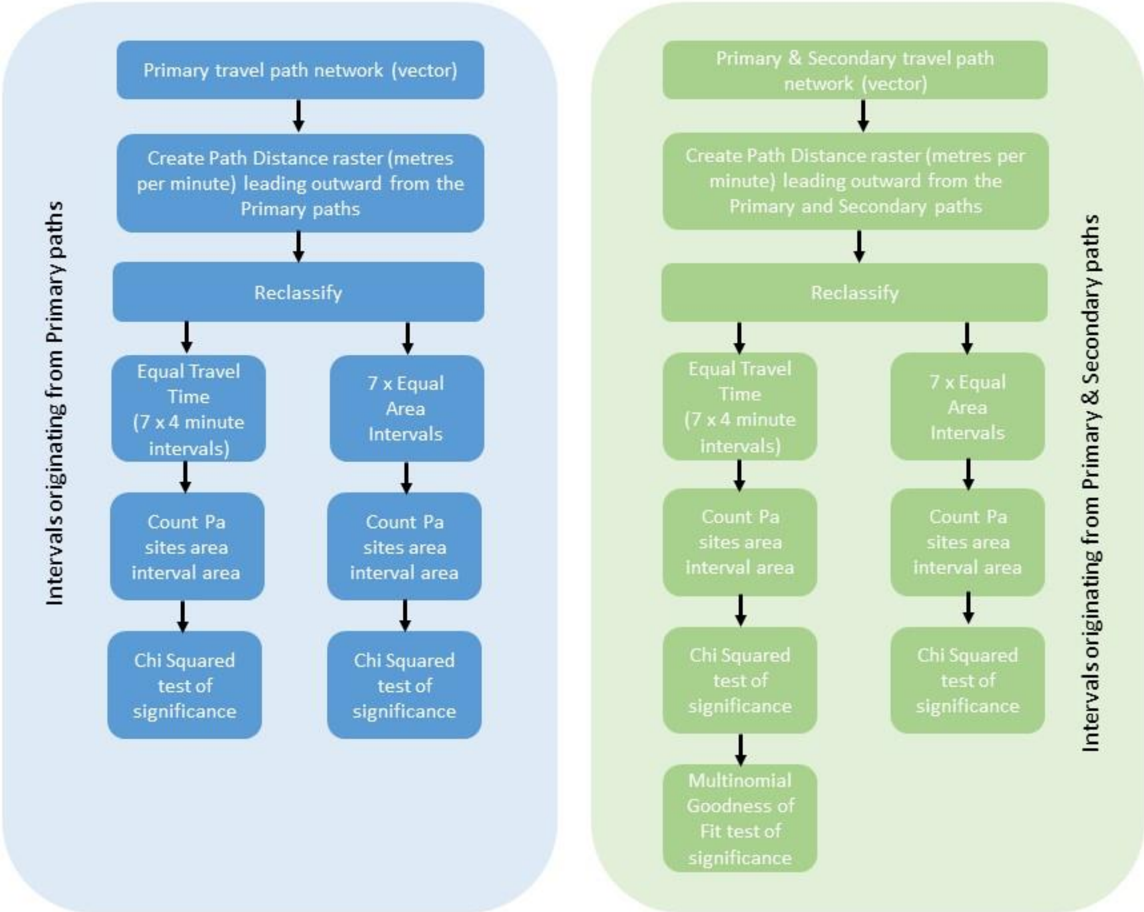


Figure 39: Method followed to create travel cost distance intervals

4.9.3 PATH DISTANCE CLASSIFICATION BY EQUAL TRAVEL COST INTERVAL

The primary and combined travel networks were used as origins to form anisotropic walking time cost surfaces leading away from the path network. The original method (Whitley and Hicks 2003) classified an isotropic walking time cost distance raster dataset into equal intervals of 100 travel cost units. This research used path distance and replicated the original equal travel cost classification method by reclassifying the raster into 4 minute travel time intervals. The inland Pa sites were then counted within each travel time interval.

4.9.4 PATH DISTANCE CLASSIFICATION BY EQUAL AREA INTERVAL

A second travel time classification was then used in order to improve the original method by producing equal area intervals which will not create areas so low that expected site counts are below the minimum required for a Chi Squared test. To create equal area classification, two ArcGIS tools were used.

Firstly, the Slice tool was attempted which claims to produce equal area classes. Unfortunately this is not the case, with some intervals being allocated twice the area of others. The second approach was to use the inbuilt quantile classification in ArcGIS which divides the input raster into classes of equal cell counts. As cell size is known, this is a suitable method for creating equal area intervals.

4.9.5 PEARSON'S CHI SQUARED TEST OF SIGNIFICANCE

Pearson's Chi Squared test of significance was calculated for the Pa site counts classified by equal travel time using Microsoft Excel following a template; (Zaiontz 2014) to apply the following equation;

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where;

O = the number of observed sites per interval

E = the expected proportion of expected sites per interval

The expected background site distribution is proportional to the travel time interval area containing the sites. For example, if the 4 minute interval zone area covers 10% of the whole study area, then assuming an even distribution, the expected site count for this interval would 18.1 (10% of the 181 sites). As anisotropic travel time intervals produce variable surface distances, no assumption could be made about the site data having a normal distribution, as a result, only a non-parametric test could be used.

The equal travel time method of producing intervals to measure Pa site distribution produced small interval areas with very low expected site counts, requiring a Multinomial Goodness of Fit test. XLSTAT statistics software was used to run the test and required two columns of input data;

1. Observed site counts per interval (integer)
2. Expected site counts (as a proportional value between 0 and 1).

The multinomial goodness of fit test uses a Monte Carlo simulation method to overcome the problem of the low expected site counts by repeating the test using random samples to obtain higher expected counts to produce a valid result. In this case, the test used 10,000 simulations. The test produces an adjusted p value from the Monte Carlo simulation process which will be slightly different to the p-value produced by a standard Chi Squared test on the original sample data.

4.9.6 GAIN STATISTIC

Kvamme's gain statistic is used to measure model performance. In this case observed and expected site counts from the Chi Squared test of significance were used as variables to calculate the Gain statistic which will always fall between 0 and 1.

1 – (Pa / Ps)

(Kvamme 1988)

Where

Pa = % path distance interval area

Ps = % of observed sites per interval

CHAPTER 5: RESULTS

This chapter summarises the results produced by the methods described in the previous chapter. The application of multiple least cost path analysis with various origin-destination configurations has revealed a natural least cost path network across the Northland peninsula. This chapter presents the results of each stage of the path modelling process followed by statistical tests of the significance and strength of the relationship between Pa and travel paths.

5.1 COST SURFACE

The river cost surface was developed using a series of geo-processing steps in ArcGIS Model Builder. Figure 40 shows the result of a flow accumulation raster with all cells that have 200 cells or less draining to them removed. This produced a single cell representation of the river network. The accumulated flow raster dataset was then divided by the maximum accumulated flow value (836040) to produce a new raster dataset with the same relative data range between zero and one.

This dataset was then multiplied by 6 to represent the relative effort of crossing rivers compared to traversing dry flat land. This value was chosen as an approximate cost factor and is relatively low as rivers in this region are generally not considered a barrier to overland travel. To manage the risk of least cost paths crossing single river cells diagonally, avoiding the full crossing cost, the river cost raster was converted to a three cell wide river network. This was achieved using the ArcGIS Euclidean Allocation tool, which copied the accumulated and weighted cost value to one neighbouring cell either side of the original flow. The result of widening the river cells is shown in figure 41. Lakes were allocated a single cost value of 99 for all lake cells and combined with the accumulated river cost raster to form the hydro cost dataset.

5.2 WATER COST EFFECT ON LEAST COST PATHS

In order to check if the hydro cost surface had an effect on least cost paths, a test was performed in a valley where the least cost path had to cross a main river and its tributaries to reach the destination point (figure 42). The costed and uncosted least cost paths follow an almost identical route until both reach the main river. At this point the costed path crosses the river once and remains on the southern bank until it reaches the destination avoiding accumulating additional river crossing costs. In contrast, the path calculated without the hydro cost surface runs down the centre of the valley crossing the river several times. The result suggests that the hydro cost surface does have a minor effect on least cost routes in this localised test.

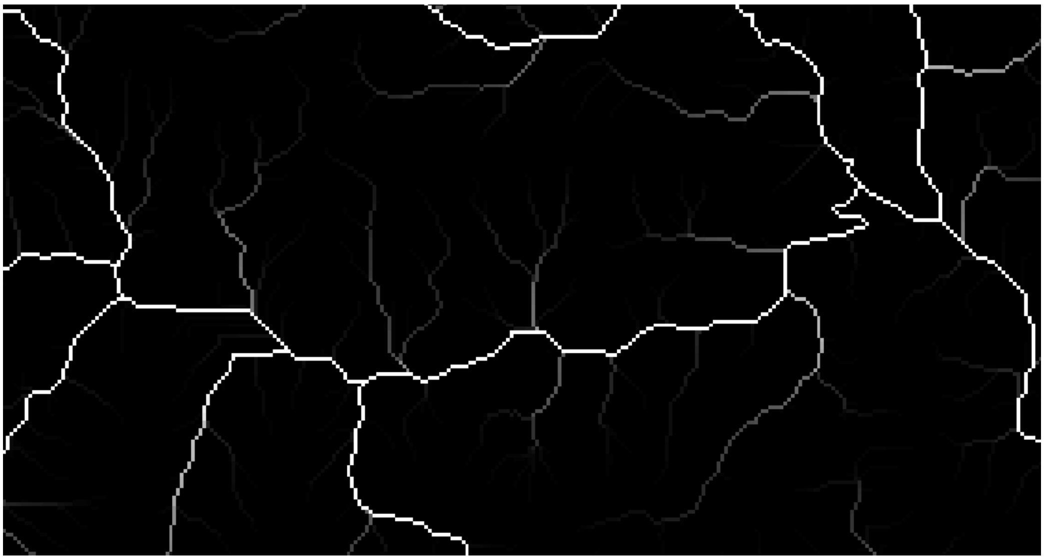


Figure 40: Flow accumulation with upstream threshold applied

Figure 41: Triple cell costed river network

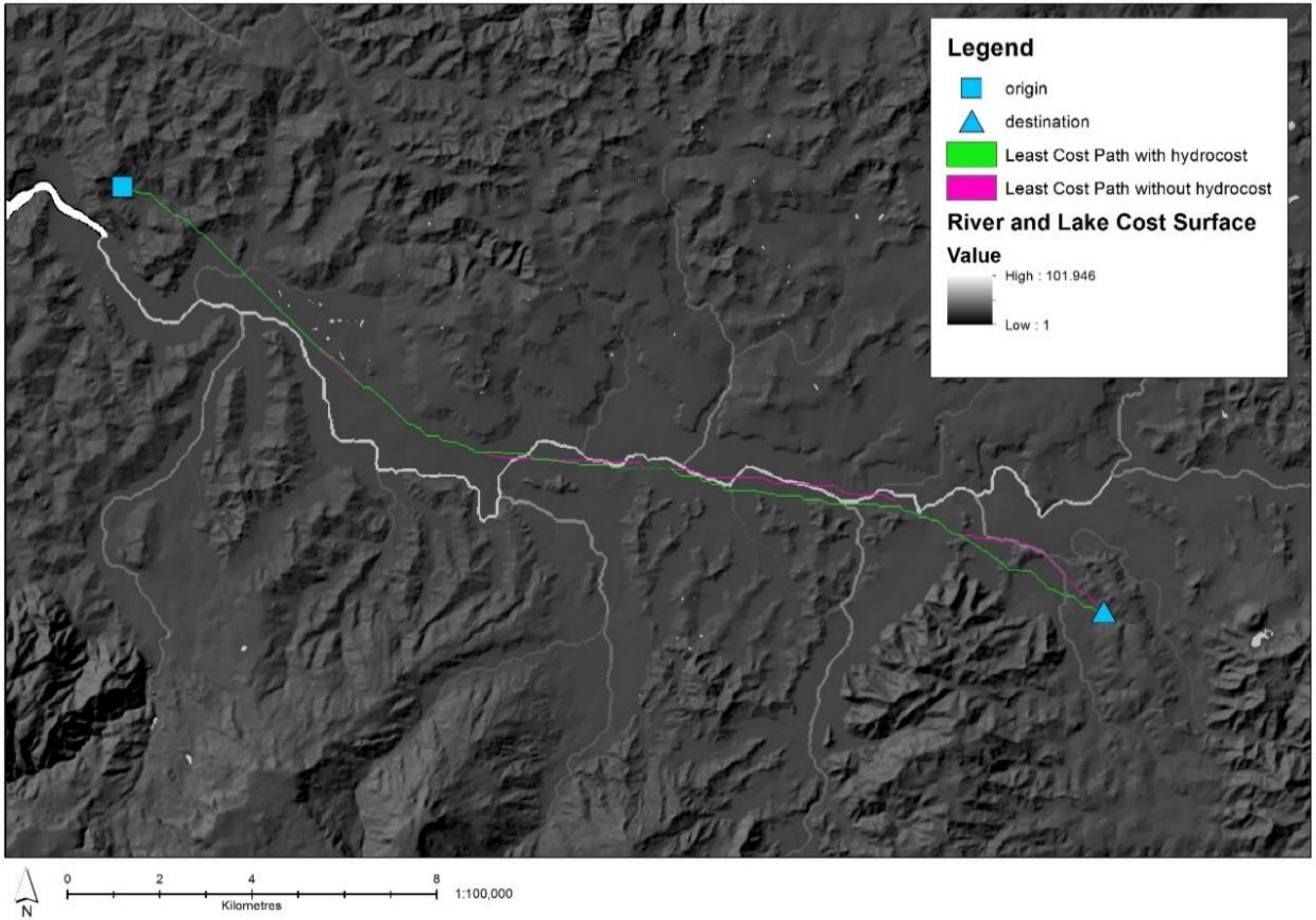


Figure 42: Test 2: Hydro cost effect on least cost paths

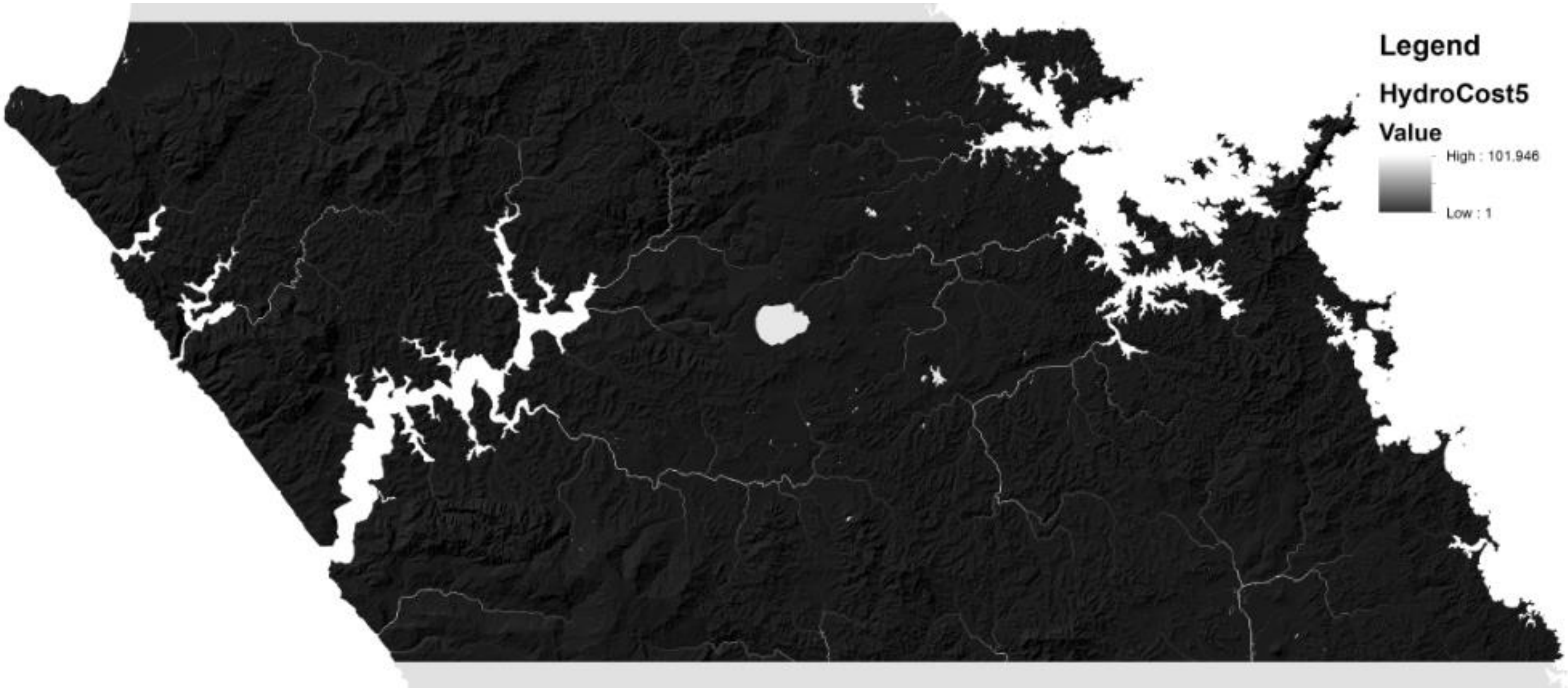


Figure 43: Combined travel cost surface representing lakes and rivers

5.3 PRIMARY CORRIDOR DEFINITION

Methods 1, 2 and 3 described in section 4.8.1 were applied to form three versions of the primary travel path network (figures 46, 47 and 48). In each case, all least cost path raster datasets were reclassified into a binary format and summed using ArcGIS cell statistics. Figure 44 shows the result of the first part of method 1 using the cell statistics sum function to count the number of paths passing directly through each raster cell. This example shows a number of paths running parallel to others that are in close proximity.

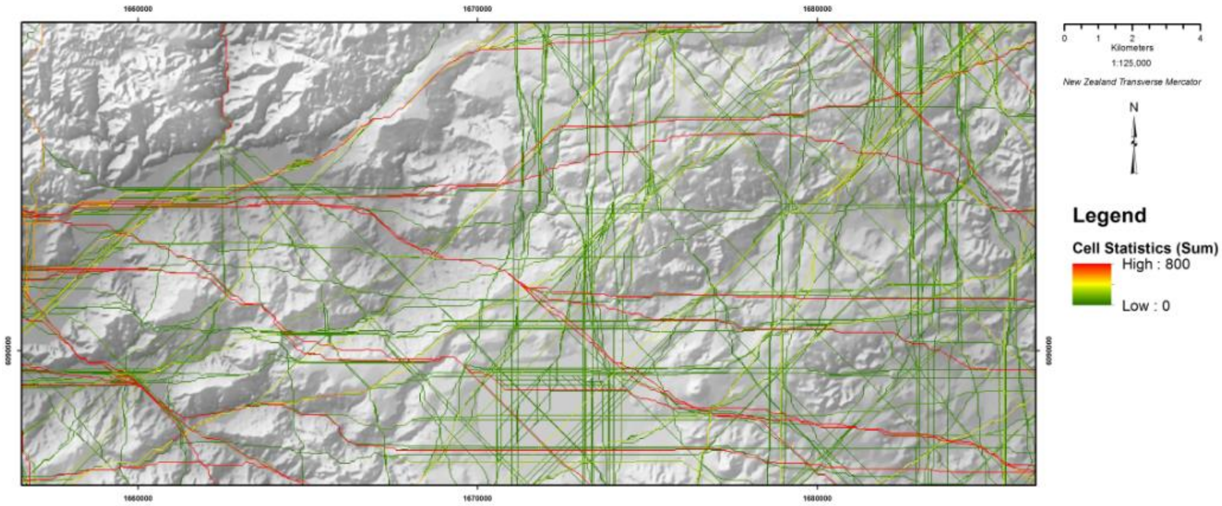


Figure 44: Cells classified by coincident summed paths

Adding the least cost path datasets is limited to being cell based and cannot detect the combined effect of neighbouring paths that form broad travel corridors. This relationship between nearby paths is identified using ArcGIS focal statistics which conducts a path density search around each cell picking up adjacent paths to identify the densest primary corridors (figure 45).

Although each origin-destination method creates a different path pattern in some localised areas, the denser primary routes persist regardless of the number of origin-destinations or their locations. Each method produced clear repeatable least cost corridors in both hilly terrain in the west and across the plains between the Bay of Islands and Hokianga Harbour. Methods 2 and 3 highlight the high level of accessibility throughout the central and north east areas.

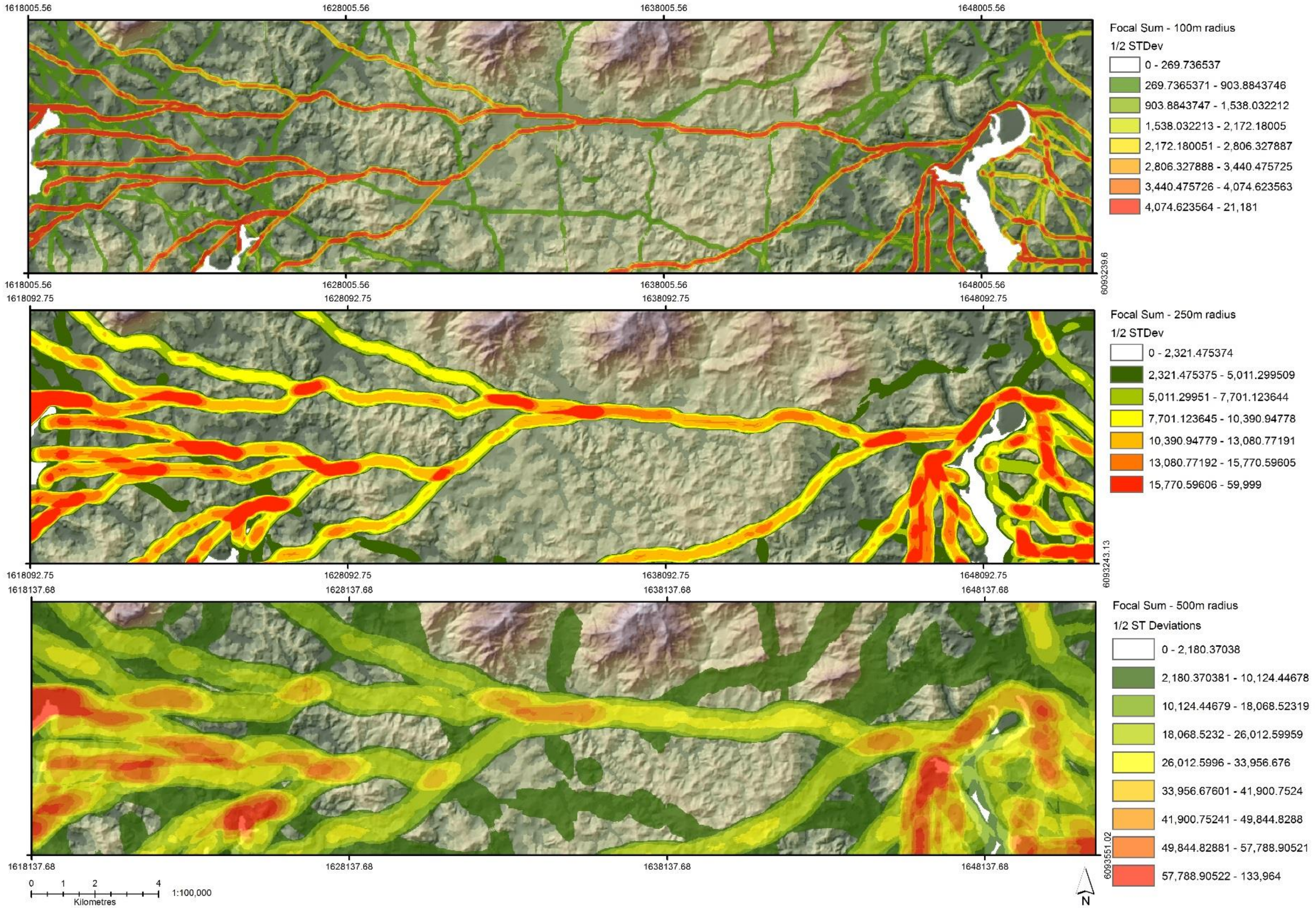


Figure 45: Focal statistics Sum results at different radius settings showing variations in path density

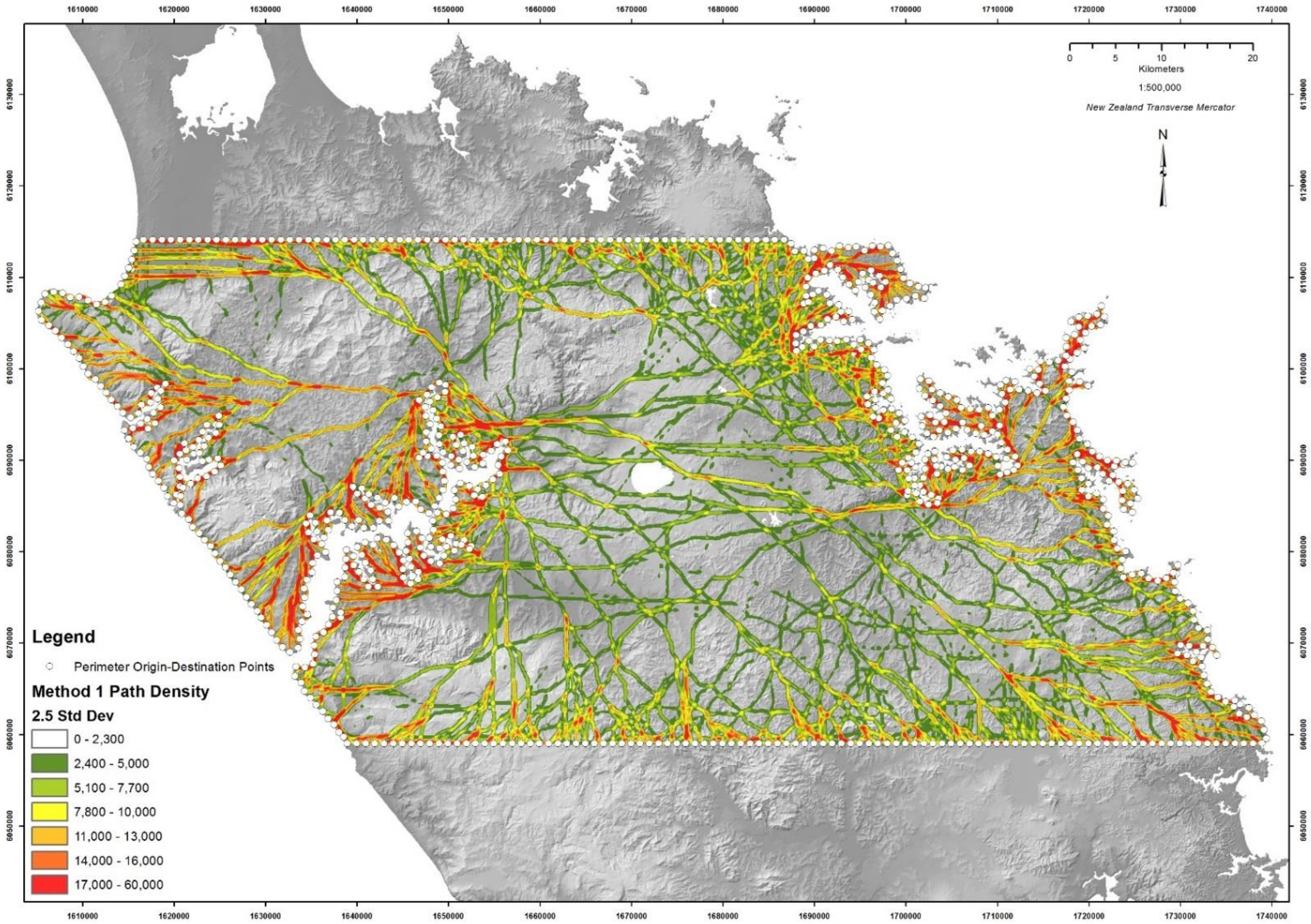


Figure 46: Method 1. 800 Perimeter origin-destination points with 250m radius focal sum analysis

This method was repeated for method 2 and 3 origin-destination patterns.

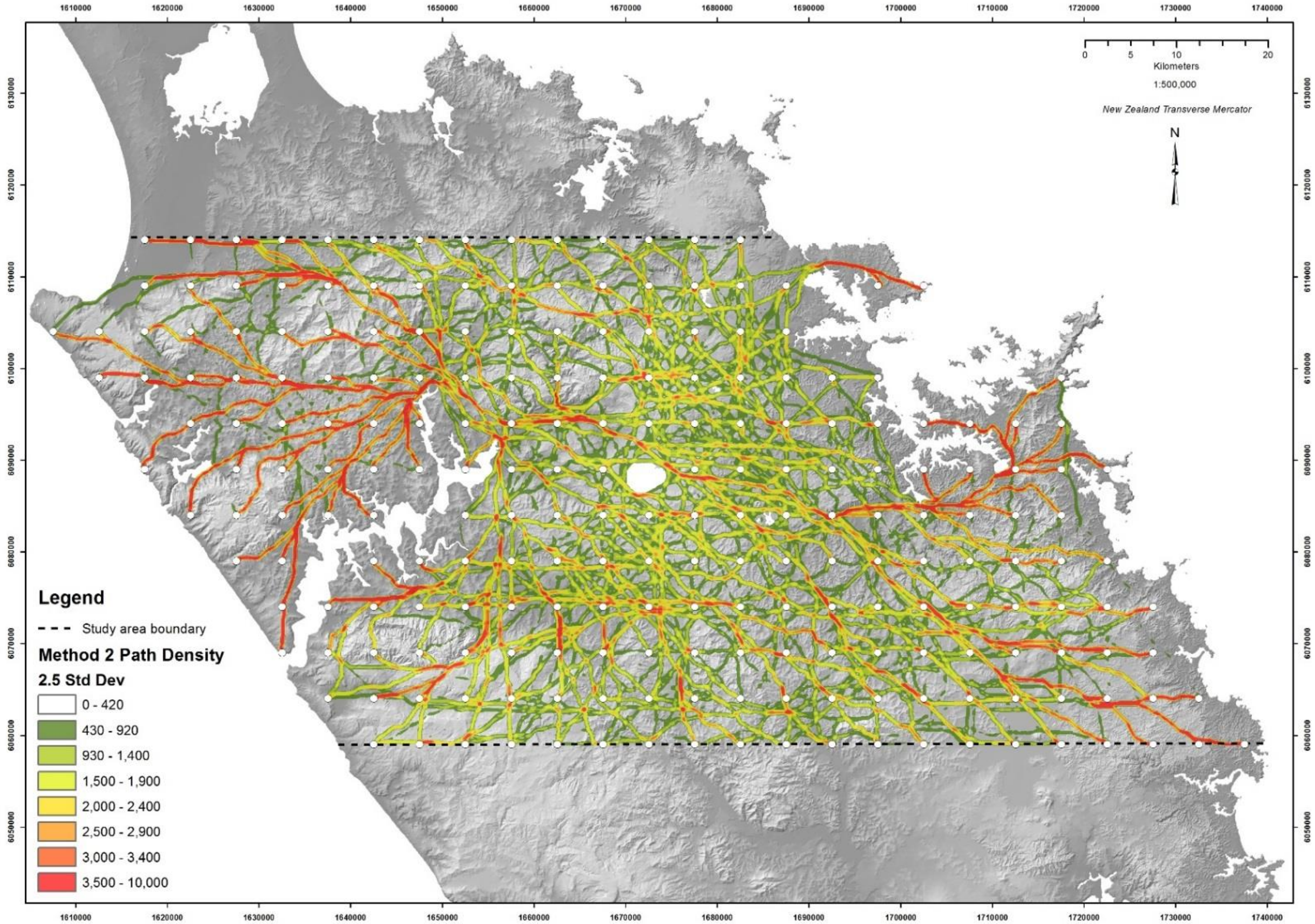


Figure 47: Method 2. 225 origin-destination grid with 250m radius focal sum analysis

Methods 2 and 3 force paths to start within the study area boundary. In places this adds complexity to some areas such as the North West area. However, these paths do not form their own complete corridors. Instead they join the closest main artery formed in Method 1.

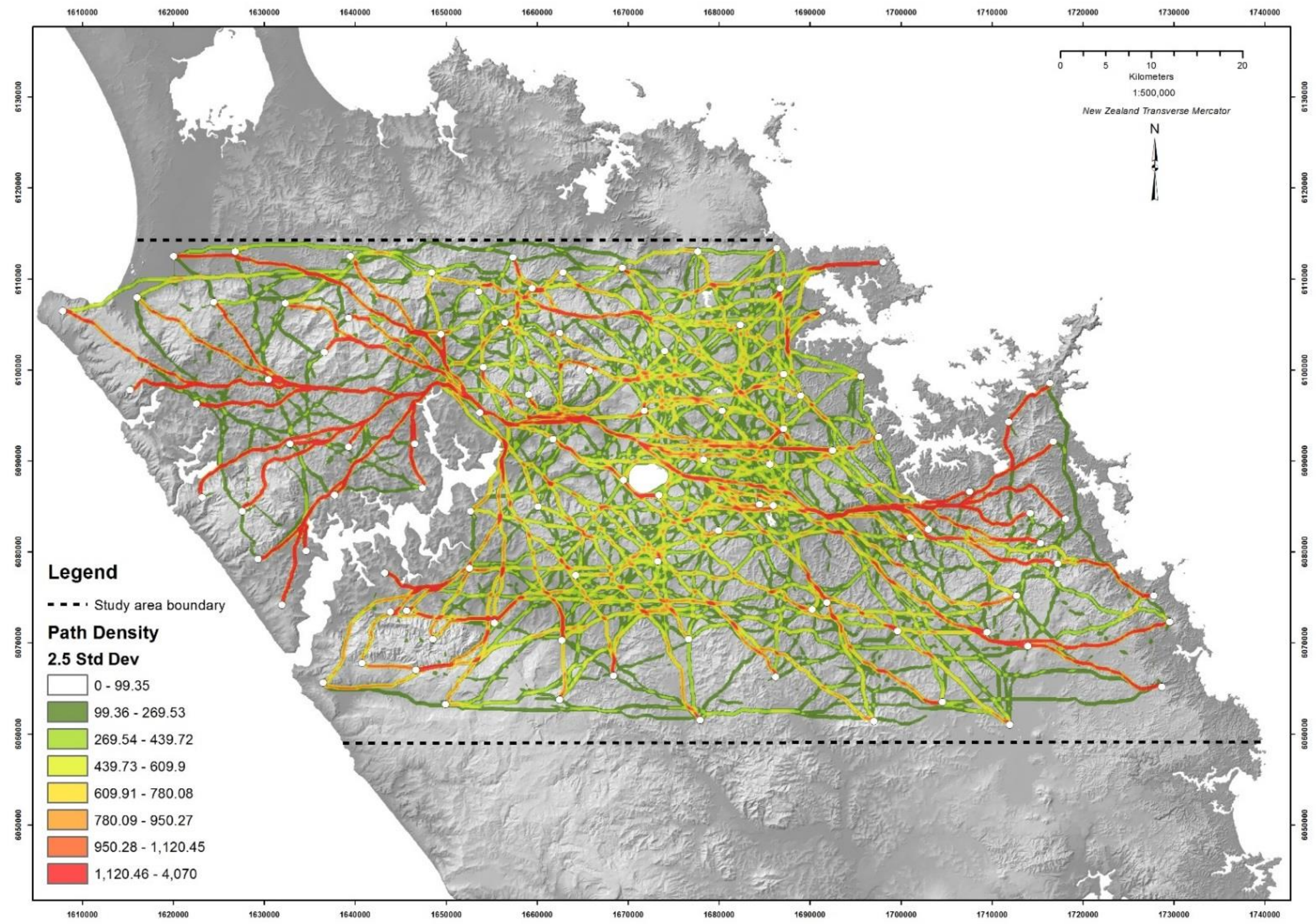


Figure 48: Method 3. 100 random origin-destination points with 250m radius focal sum analysis

Each origin-destination pattern produces what appears to be a similar network of core least cost corridors. The main east-west corridor passing just north of the central Lake Omapere divides a highly accessible central area with coastal corridors being channelled by rougher terrain.

The 100 origin-destination points used in method 3 suggest that the 800 origin-destination points used in method 1 do not appear to change the pattern of the denser travel paths, particularly from east to west coast where the jagged coastline required significantly more points than the northern and southern straight line boundary. Origin points along parts of the southern and northern study area boundary fall within elevated areas resulting in least cost paths being forced over ridges to reach their destinations. Ridge crossings near the boundary should be ignored as a larger study area may provide alternative lower cost routes around these high areas. Within the study area however, there are examples of least cost paths following ridge lines rather than valleys (figures 49 and 50).

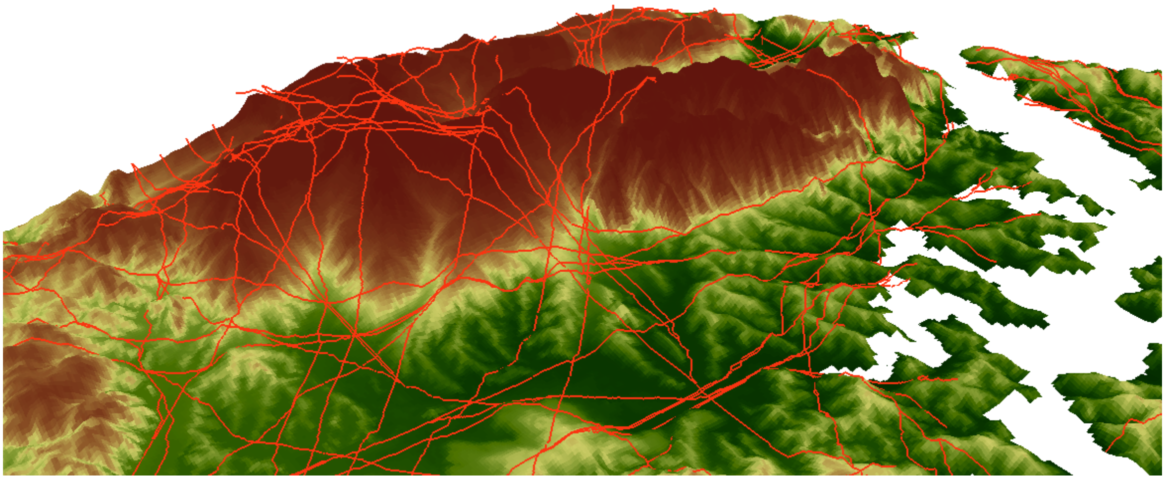


Figure 49: Primary paths originating at high elevation. (ArcScene elevation exaggerated).

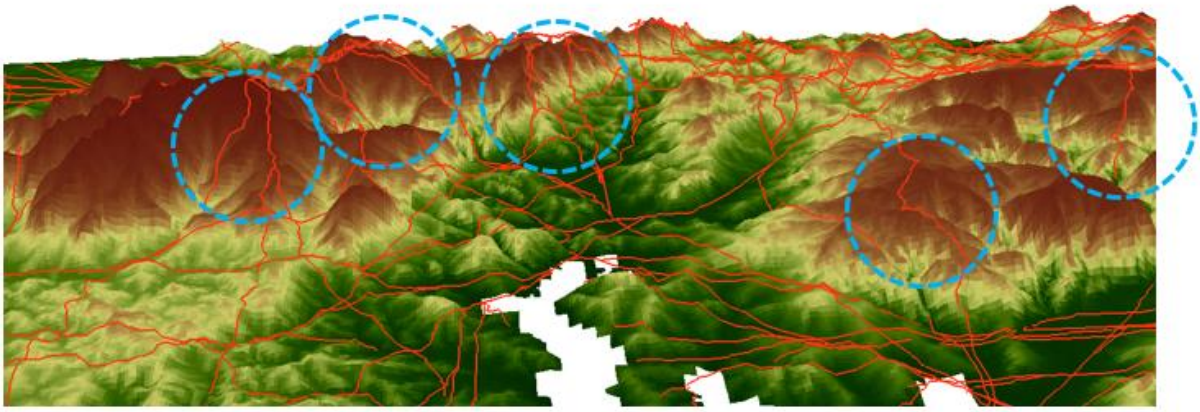


Figure 50: Primary paths shown in ArcScene following ridgelines.

Having established a raster version of the primary path network, a vector version was required as an origin for the secondary path analysis. This was mainly due to the primary raster classification being unable in most cases to produce complete paths as path density varied along each route. The primary path polylines produced at the same time as the raster results were merged into one line dataset and the main routes were selected manually by using the focal density raster dataset as a background.

5.4 SECONDARY PATH DEFINITION

The secondary paths were created using flow accumulation over an anisotropic path distance raster which used the primary path as its origin. Several flow threshold values were attempted to create a secondary path network that wasn't too extensive or too short. Figure 51 shows the results of removing upstream cells with less than 200 cells draining to them. Additional threshold results are shown in figures 65, 66 and 67.

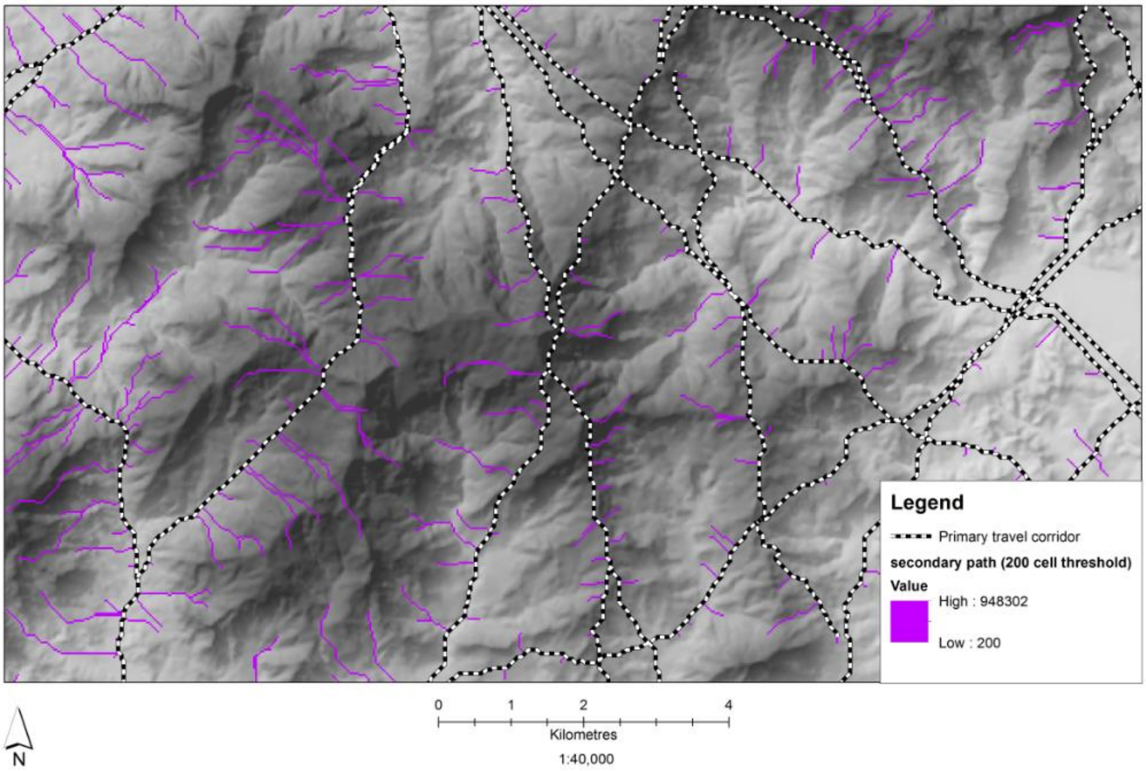


Figure 51: Secondary paths formed using a 200 cell drainage threshold

5.5 HYPOTHESIS TEST RESULTS

This section compares the results of the two methods of classifying travel time intervals leading way from each travel network and the use of statistics to test the research hypotheses.

5.5.1 EQUAL TRAVEL TIME INTERVAL

A Chi Squared test of significance was performed using inland Pa site distribution falling within 4 minute equal travel time intervals leading away from the primary travel path network. This produced a highly significant p value of 0.000011 (table 5) and the null hypothesis was strongly rejected. This is a valid result as only one of the expected site counts is < 5, accounting for just 2% of the expected sites. The Gain Statistic produced a value of 0.43 for the 8 to 12 minute travel time interval which does not indicate a strong relationship. The Chi Squared test was repeated for Pa site distribution within 4 minute walking time intervals from the combined primary and secondary paths (table 6) and resulted in a p value of 0.005627, again rejecting the null hypothesis. However, 4 out of the 7 travel time intervals had expected site counts less than 5, invalidating the result. A multinomial goodness of fit exact test was then used to overcome the low expected site count constraint by using Monte Carlo simulation to achieve an acceptable expected site count per interval (table 7). This produced a p value of 0.054, which meant that the null hypothesis could not be rejected. The strongest Gain Statistic results were 0.69 for the 16-20 minute walking time interval and 0.55 for the 12–16 minute travel time interval were invalid. Valid Gains values were < 0.5 confirming a weak relationship between path and Pa site distribution.

5.5.2 EQUAL AREA INTERVAL

The path distance raster dataset leading away from each network was then reclassified into seven equal area intervals to match the number of classes used in the equal travel time test. The ArcGIS quantile classification method produced a slightly uneven area split with two of the seven intervals having between 1% and 3% less area than others, however this is not considered significant. A Chi Squared test of Pa distribution produced a significant 0.000001 at the 0.1 (99%) confidence level (table 8). The Gain Statistic was 0.44 for the 5.4-8.7 minute travel time interval, representing 26% of Pa sites in 15% of the study area. The equal area Chi Squared test for the combined primary and secondary path network (table 9) also produced a significant *p* value of 0.000214. This test produced a weaker Gain Statistic of 0.37 for the 7+ minute travel time interval.

Figures 52 and 53 show the results of classifying path distance from both travel networks using equal travel time intervals, as applied by Whitley and Hicks 2003.

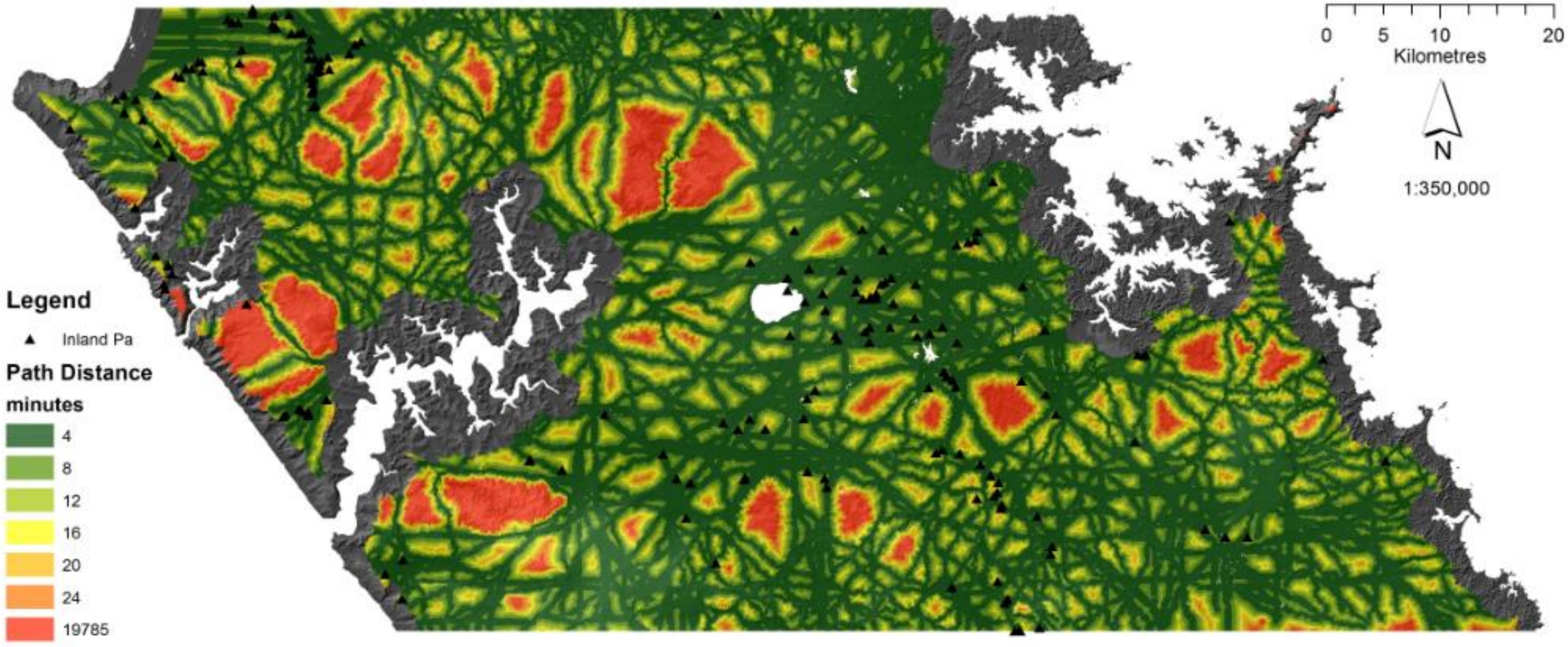


Figure 52: Path distance originating from primary paths by equal travel time intervals.

The introduction of secondary paths significantly reduces the number of travel time intervals between paths. Note that there are very few areas beyond a 16 minute walk of either a primary or secondary path.

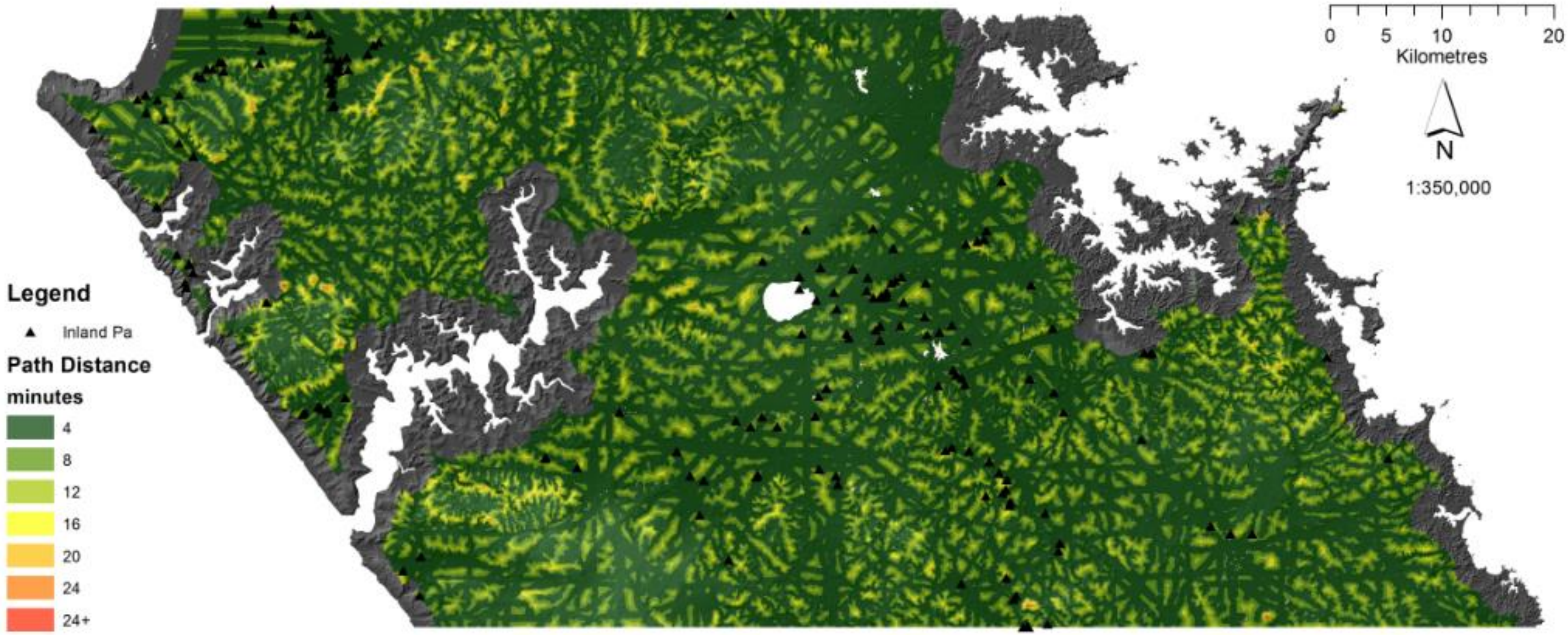


Figure 53: Path distance originating from combined paths by equal travel time intervals.

Figures 54 and 55 show the results of classifying path distance from both travel networks by equal area interval using quantile classification. This method produces narrower intervals near to paths compared with equal travel time intervals.

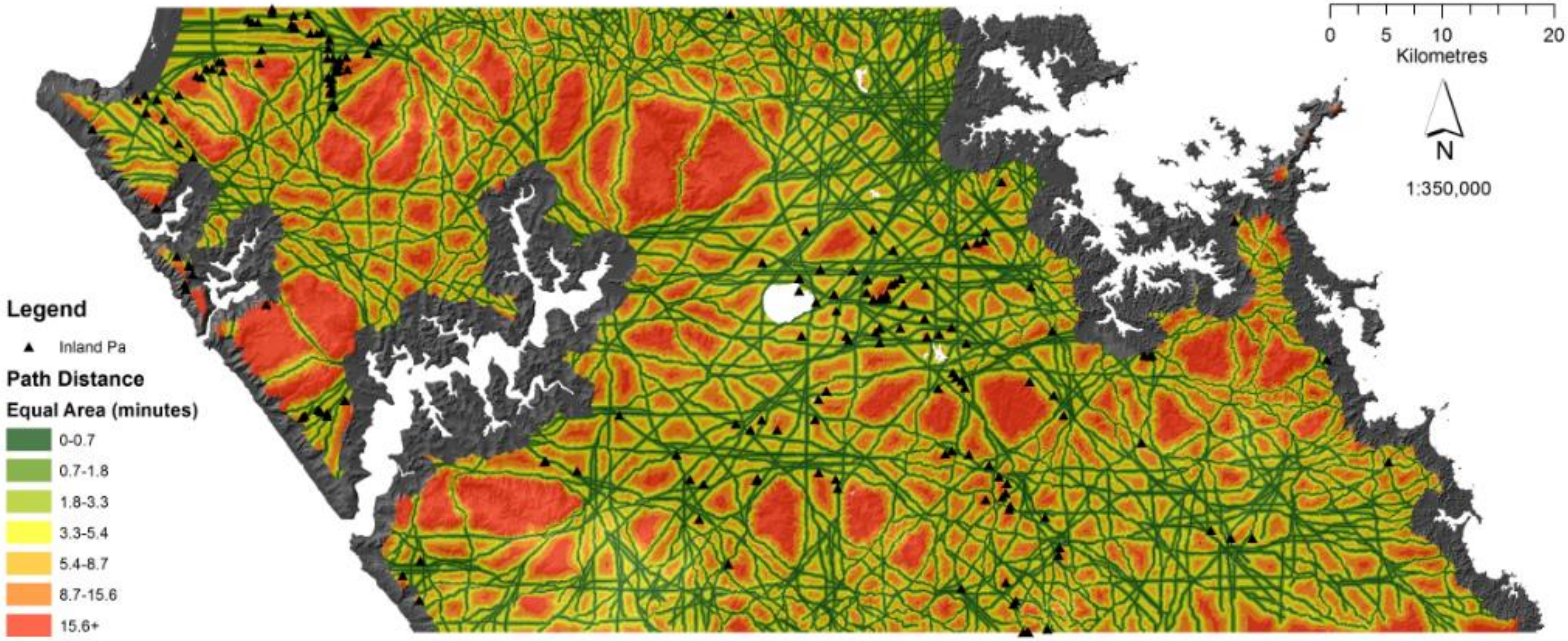


Figure 54: Path distance originating from primary paths by equal area intervals

Equal area intervals provide a clearer division of path distance intervals allowing a valid Chi Squared test to be applied to the Pa site population.

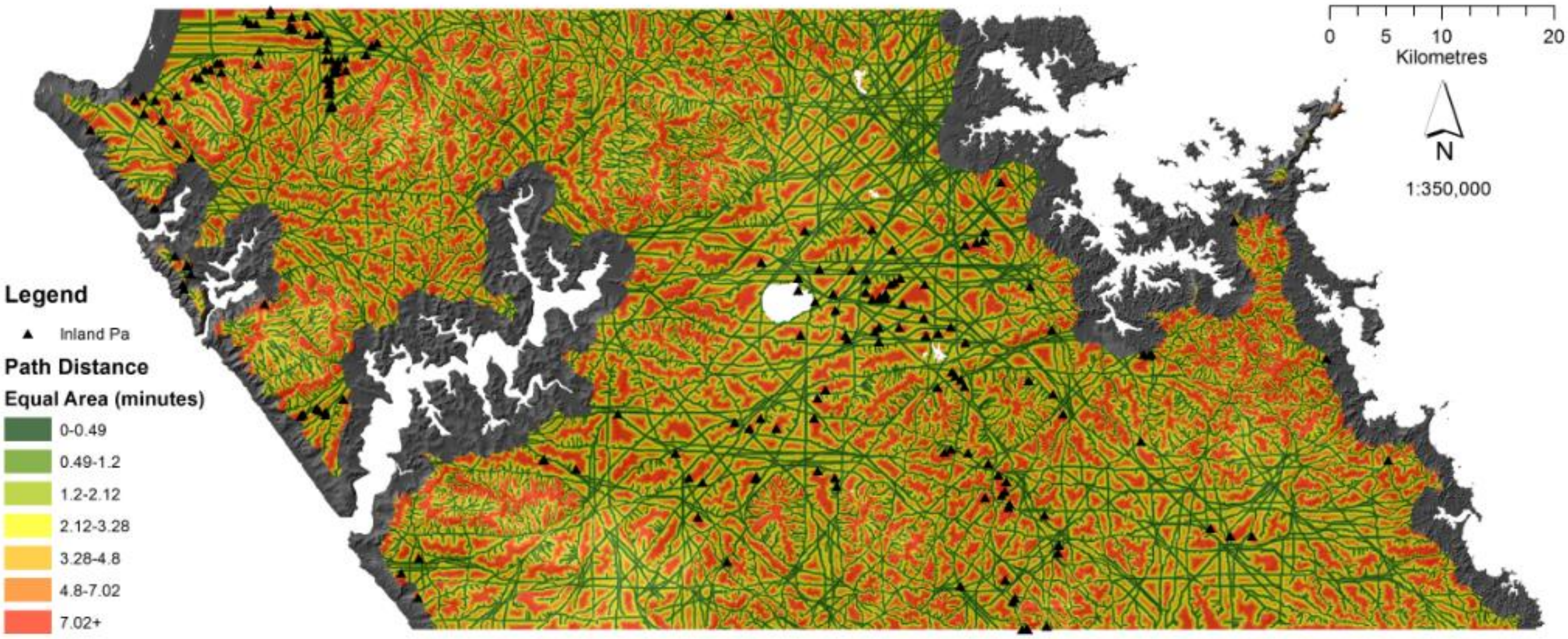


Figure 55: : Path distance originating from combined paths by equal area intervals

5.6 CHI SQUARED TEST OF SIGNIFICANCE - PRIMARY NETWORK (EQUAL TRAVEL TIME INTERVALS)

Tables 5 and 6 show the results of Chi Squared tests of Pa site distribution significance relative to primary and combined path networks respectively using equal travel time intervals.

Travel time (mins)	0 -4	4-8	8-12	12-16	16-20	20-24	24+	Total
Observed sites	74	56	36	6	2	3	4	181
Expected sites	89.79	38.37	20.40	11.13	6.14	3.65	11.52	181
Area (Km2)	2103.63	898.92	477.99	260.71	143.73	85.49	269.89	4240.37

χ2 = Σ(O-E)^2/E	2.78	8.10	11.92	2.36	2.79	0.12	4.91	32.98
% Area	0.50	0.21	0.11	0.06	0.03	0.02	0.06	1.00
% Observed sites	0.41	0.31	0.20	0.03	0.01	0.02	0.02	1.00
α	0.01							
df	6.00							
χ2	32.98							
χ2-crit =CHINV(α,DF)	16.81							
p-value	0.000011							
sig	yes							
Gain Statistic	-0.21	0.31	0.43	-0.85	-2.07	-0.22	-1.88	

Table 5: Chi Squared results for Pa relative to the primary path network

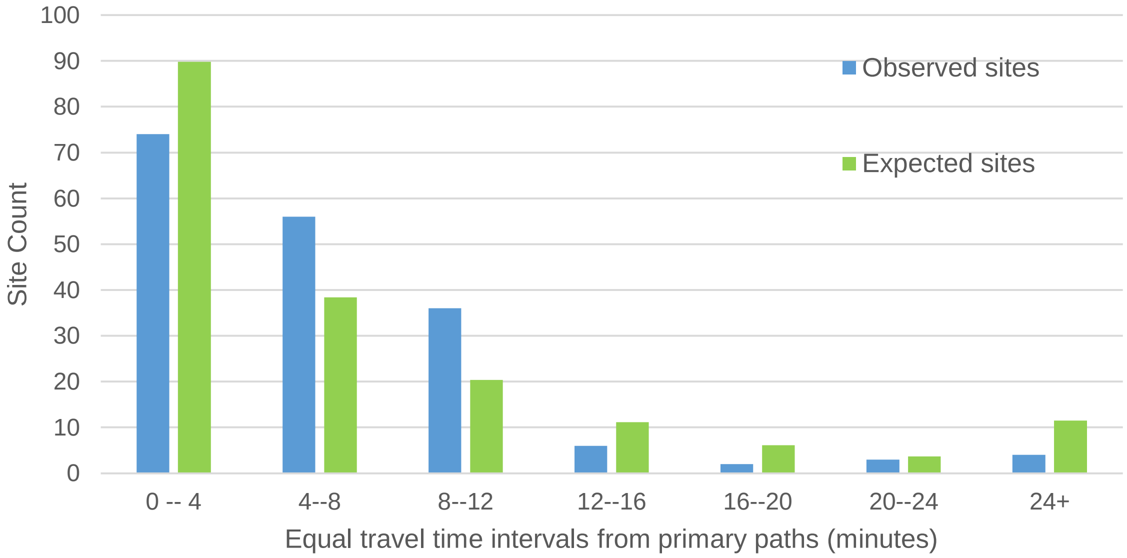


Figure 56: Pa distribution from primary paths by equal travel time intervals

5.7 CHI SQUARED TEST OF SIGNIFICANCE - COMBINED NETWORK (EQUAL TRAVEL TIME INTERVALS)

Travel time (mins)	0 - 4	4-8	8-12	12-16	16-20	20-24	24+	Total
Observed sites	91	61	22	5	1	0	1	181
Expected sites	116.21	46.36	15.07	2.27	0.31	0.04	0.74	181
Area (Km2)	2722.42	1086.13	353.16	53.10	7.22	1.02	17.32	4240.37

χ2 = Σ(O-E)^2/E	5.47	4.62	3.18	3.30	1.55	0.04	0.09	18.25
% Area	0.6420	0.2561	0.0833	0.0125	0.0017	0.0002	0.0041	1
% Observed sites	0.50	0.34	0.12	0.03	0.01	0	0.01	1
α	0.01							
df	6							
χ2	18.25							
χ2-crit =CHINV(α,DF)	16.81							
p-value	0.005627							
sig	yes							
Gain Statistic	-0.28	0.24	0.31	0.55	0.69		0.26	

Table 6: Chi Squared results for Pa relative to the combined path network by equal travel time interval.

Equal travel time interval classification produced low expected site counts which invalidated the Chi Squared test result. An exact test was required to overcome the issue described in section 5.8.

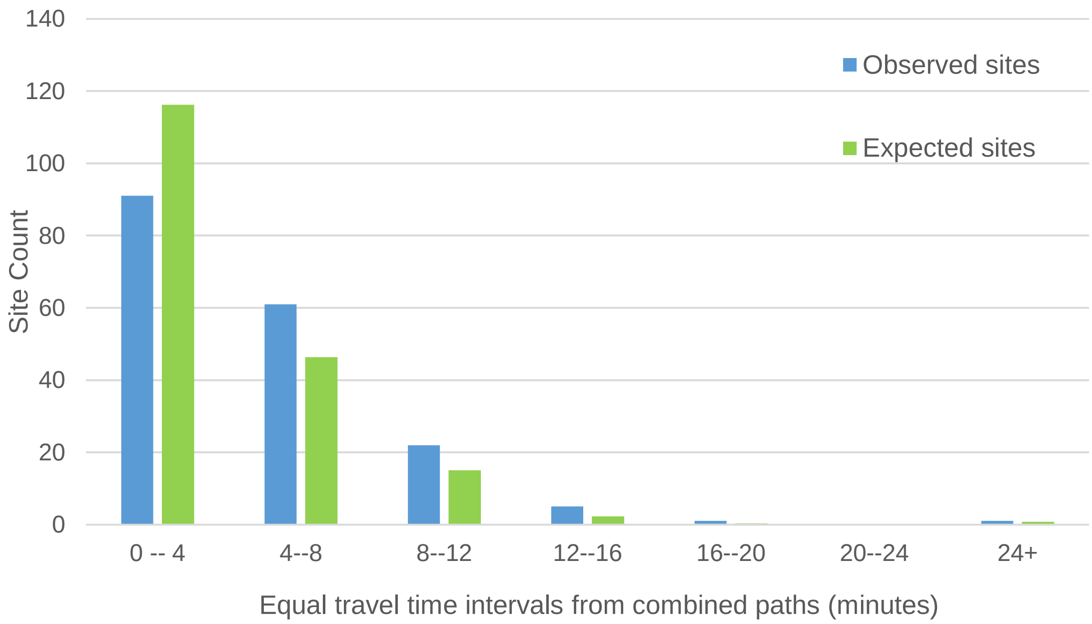


Figure 57: Pa distribution from primary and secondary paths by equal travel time intervals

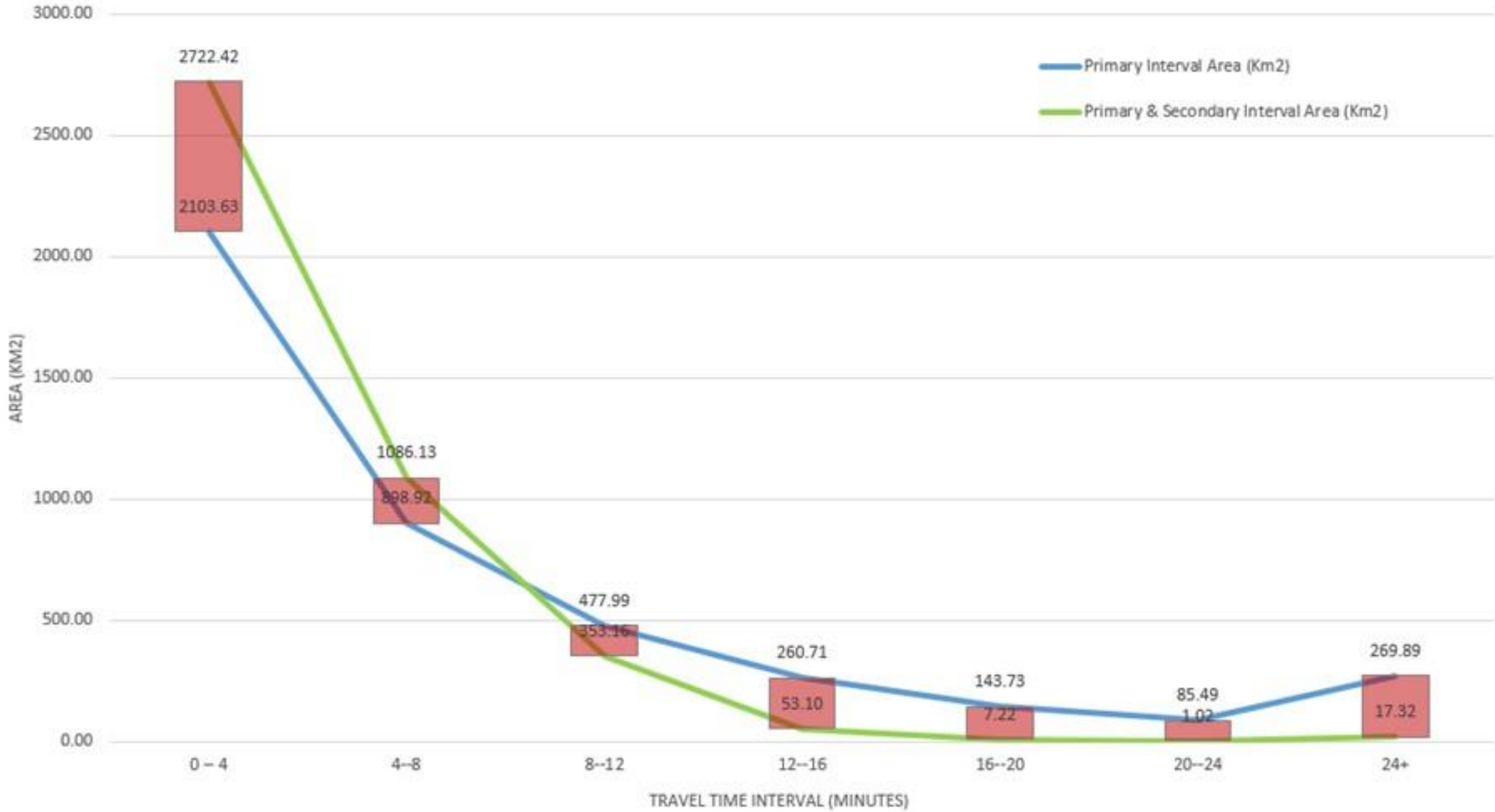


Figure 58: Rate of change between equal travel time interval areas from the primary and combined paths.

5.8 MULTINOMIAL GOODNESS OF FIT TEST – PA RELATIVE TO PRIMARY & SECONDARY NETWORK

The multinomial goodness of fit test accepted the null hypothesis;

Multinomial Goodness of Fit: with Monte Carlo simulation (Number of simulations = 10,000)							
Travel Time Interval (Mins)	Chi-square (Observed value)	Primary Chi-square (Critical value)	Degrees of Freedom	P value	Alpha value	Null hypothesis	Risk to reject null
4	18.254	28.587	6	0.054	0.01	Accepted	5.42%

Table 7: Multinomial goodness of fit test result

5.9 PATH DISTANCE FROM PRIMARY NETWORK (EQUAL AREA INTERVALS)

ArcGIS quantile classification produced six even interval areas, however the final interval was noticeably smaller as shown in table 8 and figure 59.

Equal Area (minutes)	0 - 0.7	0.7 - 1.8	1.8 - 3.3	3.3 - 5.4	5.4 - 8.7	8.7 - 15.6	15.6 +	Total
Observed sites	16	14	30	36	47	29	9	181
Expected sites	26.70	26.67	26.52	26.49	26.48	26.02	22.12	181
Area (Km2)	625.40	624.71	621.31	620.67	620.40	609.62	518.27	4240.37

χ2 = Σ(O-E)^2/E	4.28	6.02	0.46	3.41	15.90	0.34	7.78	38.19
% Area	0.15	0.15	0.15	0.15	0.15	0.14	0.12	1.00
% Observed sites	0.09	0.08	0.17	0.20	0.26	0.16	0.05	1.00
α	0.01							
df	6.00							
χ2	38.19							
χ2-crit =CHINV(α,DF)	16.81							
p-value	0.000001							
sig	yes							
Gain Statistic	-0.67	-0.90	0.12	0.26	0.44	0.10	-1.46	

Table 8: Chi Squared result for Pa relative to primary network

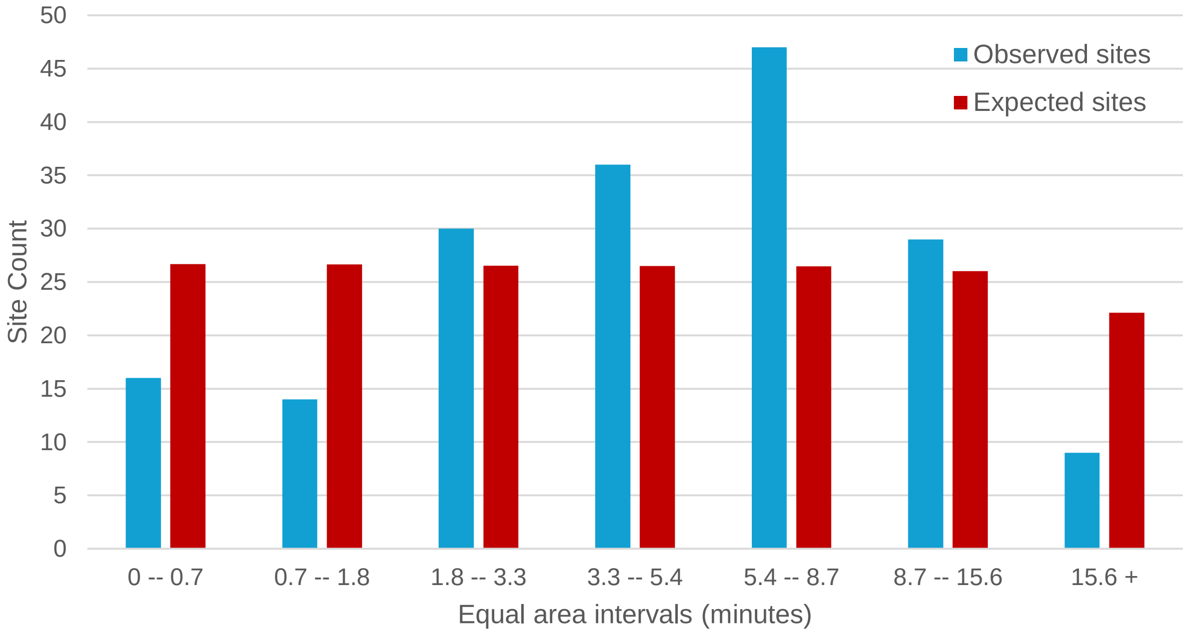


Figure 59: Pa distribution from primary network by equal area interval

5.10 PATH DISTANCE FROM COMBINED NETWORK (EQUAL AREA INTERVALS)

Equal Area (minutes)	0 - 0.49	0.49-1.2	1.2-2.12	2.12-3.28	3.28-4.8	4.8 -7.02	7.02 +	Total
Observed sites	10	21	18	26	29	35	42	181
Expected sites	25.99	25.91	25.75	25.57	25.59	25.82	26.37	181.00
Area (Km2)	608.91	606.98	603.26	599.13	599.42	604.90	617.76	4240.37

χ2 = Σ(O-E)^2/E	9.84	0.93	2.33	0.01	0.46	3.26	9.27	26.09
% Area	0.14	0.14	0.14	0.14	0.14	0.14	0.15	1.00
% Observed sites	0.06	0.12	0.10	0.14	0.16	0.19	0.23	1.00
α	0.01							
df	6.00							
χ2	26.09							
χ2-crit =CHINV(α,DF)	16.81							
p-value	0.000214							
sig	yes							
Gain Statistic	-1.60	-0.23	-0.43	0.02	0.12	0.26	0.37	

Table 9: Chi Squared result for Pa relative to the combined network

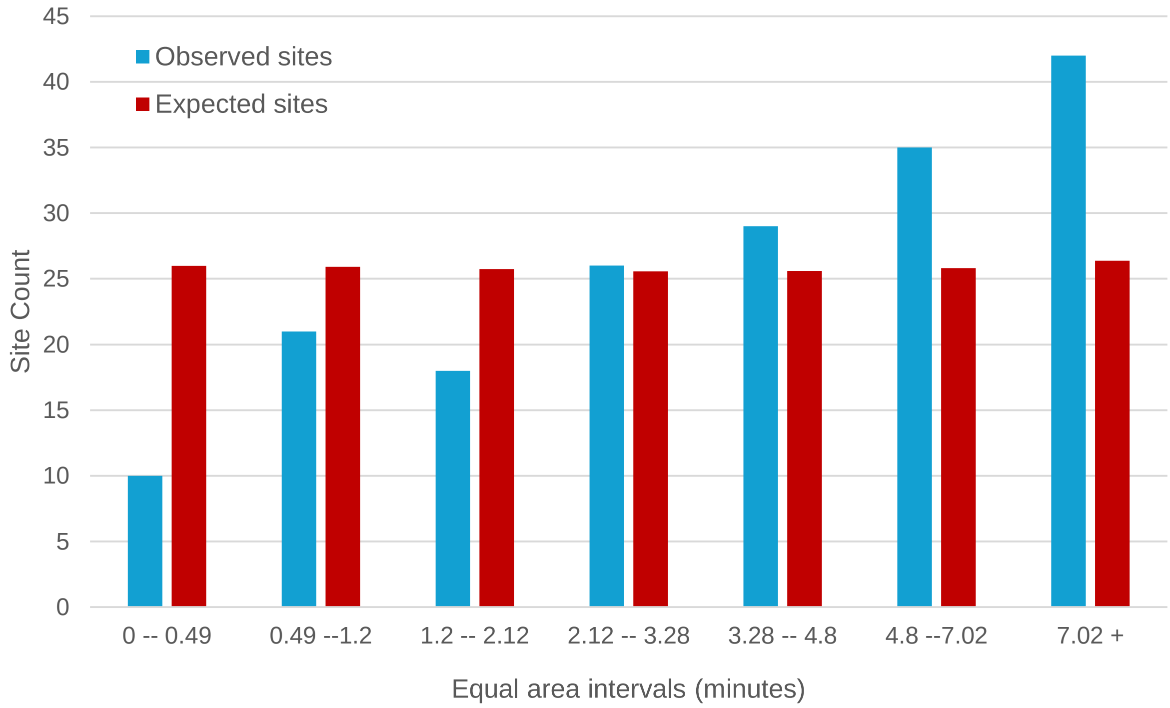


Figure 60: Pa distribution from combined network by equal area interval

CHAPTER 6: DISCUSSION

This research has adapted an existing method of modelling a regional network of least cost paths representing a simulated Prehistoric human travel network. Statistical analysis has confirmed that Pa sites are unevenly distributed across the landscape and significant numbers of Pa are situated in close proximity to the network of regional least cost travel paths. The research objectives set to support the research question and hypotheses are now discussed in sequence;

6.1 TRAVEL COST VARIABLE SUITABILITY

This research used modern lakes, an incrementally increasing river crossing cost, and a vertical slope cost to represent a prehistoric travel cost surface. These are the only variables that are present in current spatial datasets which are assumed to be largely unchanged since European contact. Travel cost variables considered and used in this study are discussed in the following subsections.

6.1.1 RIVERS: AS AN OVERLAND TRAVEL COST

The river travel cost method used in this research is a replication of the method used by Whitley and Hicks. The test described in section 5.2 demonstrates that a simple incremental river cost which is more than one cell wide does have a minor effect on least cost path patterns, for example, forcing overland paths to minimise the number of river crossings. This is a generalised river network used to represent a minor obstacle rather than a complete barrier to overland human movement. This method is considered a reasonable attempt to represent a river crossing travel cost in the absence of a detailed river dataset containing characteristics such as variable depth, width and seasonal flow which could be used to classify the river into levels of crossing difficulty. Until more realistic detailed river crossing data becomes available, rivers cannot be considered a critical human travel cost variable.

The relative cost factor used to convert the river travel cost to a range between zero and one could be enhanced by analysing the relationship between the river and the surrounding land form. If gorges can be identified for example, stretches of river could be considered too treacherous to cross on foot during the wet season or too difficult to access forcing overland paths further away in search of a safer crossing point. The focal statistics method used in this research to determine the primary travel paths could be adapted and used to analyse and classify the river network by each river cell's surrounding average slope. Additional suggestions for modelling river travel cost is provided in section 7.2.

6.1.2 HISTORIC WETLAND REPRESENTATION

Spatial datasets in New Zealand which may be useful as a proxy for overland human travel friction such as current wetlands have been captured for use at 1:50,000 scale. While this appears to be suitable for regional spatial analysis at 25 metre resolution, the level of generalisation within the data means that wet soil classes are not well defined and it cannot be assumed that these mapped areas represent uniform areas of difficult terrain. However, a programme is underway to develop a new soil map series called S-Map, which will have a usable scale of 1:25,000 and should provide an improved dataset for future predictive archaeological modelling use. A revision of the Freshwater Ecosystems of New Zealand (Ausseil et al 2008) project using finer soil classification could provide a useful historic wetland travel cost variable. Due to the poor quality of available wetland spatial data, current lakes were considered the only reliable data source for use as a complete barrier to overland travel.

6.1.3 LAND USE CHANGE

The dramatic change in land use across northern New Zealand pre and post European arrival presents problems when considering modern spatial datasets as a suitable representation of historic land cover. Pre European forest clearance, followed by industrial scale deforestation and swamp drainage post European settlement has resulted in a relatively bare landscape dominated by dairy and sheep farms. Spatial datasets representing prehistoric vegetation are currently not available.

6.2 DEFINING PRIMARY TRAVEL PATHS

This research has successfully replicated an existing method for creating regional natural travel paths (Whitley and Hicks 2003). Primary travel arteries across the study area have been identified based on the assumption that the modern lake and river water cost surface is a reasonable representation of obstacles to overland travel several hundred years ago and acknowledging the lack of spatial data to represent wetland areas which would have presented varying levels of travel cost. Extracting these paths was not straight forward. The irregular path density even along the densest of these routes did not allow a complete path to be extracted by a density classification method. This suggests that while density methods are useful for visually outlining primary paths, such methods are too localised and the path network too complex for the intended purpose of selecting whole path sections. The use of ArcGIS focal statistics to place a focal window around each cell in the combined least cost path raster dataset and count the number of paths within a set radius has helped to define a fuzzy path network to aid manual route selection.

The alternative of vector network analysis looks more promising for being able to isolate whole path segments by retaining attributes representing the number of least cost paths passing through or near these routes. Space syntax and other vector network analysis research (Verhagen et al 2013) potentially offers a more robust method for extracting primary travel paths. Least cost path methods are limited by only being able to represent the optimal path. Whitley and Hicks noted the possibility of regional least cost path analysis failing to identify localised low cost alternative routes. Their solution was to introduce additional nodes at the start and end of sections of the optimal path that coincide with natural alternative passages resulting in new detours being formed. Extending this approach to the Northland area could include the addition of origin-destination points at known settlements, natural convergence points determined from a perimeter only origin-destination test, and known river access locations. The analysis could then be re-run including these intermediate points to introduce intuitive control points, potentially producing additional primary routes.

Moving away from vector methods, in the same way that path distance raster datasets were generated from all origin points in this research project, a Model Builder iteration step could be used to produce path distance raster datasets originating from each origin node to all other nodes, snapped to the underlying DEM, (Lombard and Church 1993). This process could then select path distance raster datasets from pairs of origins and use the ArcGIS Corridor tool to form low to high cost travel corridors. Once this has been repeated for origin pairs from each part of the study area boundary, all corridor raster datasets could then be summed to form a master raster dataset containing cumulative least cost values. Experimental classification such as isolating the lowest 10% cost values (Verhagen et al 2013) could then define a broader primary travel corridor network. This would produce a weighted raster dataset representing a range of low cost paths forming a more complex and realistic travel path network. If the most efficient travel route was flooded or presented a threat to travellers, these alternative routes need to be identified as they could provide an explanation for archaeological evidence sited away from paths generated by the methods presented in this research.

6.3 SECONDARY PATH FINDING LIMITATIONS

The secondary path method used in this study replicated the use of flow accumulation leading away from the primary travel path network over a path distance raster dataset, with a flow threshold limiting the extent of each secondary path (Whitley and Hicks 2003). The use of flow paths rather than least cost paths based on Dijkstra's algorithm means that the flow accumulation method will not necessarily produce the globally optimal path. The flow accumulation path method does however provide a reasonable method of producing these dendritic paths in the absence of destination points. Suggested improvements for modelling secondary paths are provided in section 7.2.

6.4 STATISTICAL SIGNIFICANCE AND STRENGTH OF PA SITE DISTRIBUTION

The null hypothesis for Pa distribution relative to the primary travel path network is strongly rejected by equal travel time and equal area path distance classification. This answers the first hypothesis by confirming that Pa site distribution relative to primary travel paths is not evenly distributed across the landscape. Pa distribution significance relative to primary and secondary paths is less clear. This relationship is described in the following sub sections by the two path distance classification methods used;

6.4.1 EQUAL TRAVEL TIME INTERVALS

The replicated method of using equal travel time (cost distance unit) intervals from paths to measure Pa density (Whitley and Hicks 2003) produced mixed results. Relative to the primary path network, the Chi Squared test resulted in one out of seven classes having an expected site count less than 5, accounting for 14% of the intervals and approximately 2% of Pa site dataset. This low expected site count does not fail the minimum requirements for the Chi Squared test. A maximum Gain statistic of 0.43 falls short of the 0.5 strength threshold confirming a weak relationship between the interval containing the highest Pa site count and least cost path network.

Pa site density measured using equal travel time intervals from primary and secondary paths did produce a moderately high Gain statistic of 0.55 (for the 12-16 minute walking time interval) and 0.69 (for the 16-20 minute walking time interval). However, this classification method produced such small interval areas that the expected site counts were too low to produce a valid result. Although this is a valid workaround to produce a statistical result, the null hypothesis could not be rejected. It was concluded that even though expected site counts were calculated as a ratio of interval area, the reliance on this ratio is flawed as it provides no minimum area limit. When graphed, this method is also not as visually intuitive to interpret as a classification that produces equal interval areas.

6.4.2 EQUAL AREA INTERVALS

Equal area travel time interval classification provides a more intuitive method of measuring Pa site density per and the Chi Squared test requirements are guaranteed to be met if a small number of intervals are used. This interval classification produced a higher Chi Squared statistic (38.19) for Pa relative to the primary path network than the equal travel time method (32.98) suggesting that equal area classification identified a more significant gap between observed and expected site counts. A *p* value of 0.000001 shows a very strong rejection of the null hypothesis.

The maximum gain statistic of 0.44 for the 5.4 - 8.7 minute equal area interval for Pa site proximity to the primary path network is however only marginally stronger than 0.43 Gain produced from the equal travel time classification so the strength of the relation has not improved significantly using equal area site classification. Pa distribution relative to primary and secondary paths is also significant with a *p* value of 0.000214, marginally less significant than the primary network result but confirming that Pa sites are also not evenly distributed relative to the extended path network.

When the equal area intervals are translated to travel time, it is notable that 6 of the 7 intervals are compressed within a 7 minute walk to the combined primary and secondary path network. What appears to be an outlying maximum Gain statistic of 0.37 for the outer 7.02+ minute interval in terms of walking time is still highly accessible. A further interval or two may have helped define the extent of this cluster of Pa sites. Overall, 79% of Pa sites are located within an 8.7 minute walk from a primary path and 77% of Pa are located within a 7 minute walk of a combined primary and secondary path. This decrease in travel time to secondary paths is to be expected as any extension to the primary travel corridors should increase accessibility to Pa sites to varying degrees depending on the surrounding topography.

A minor issue with the use of ArcGIS quantile classification in order to create these intervals of approximate equal area is that the final interval in this case was allocated a slightly lower cell count than the other six intervals. As the difference is only equivalent to 3% of the total area this is not considered critical and should not bias the results. This classification error is also present in the results for Pa proximity to the primary-secondary path network. However in this example the final interval has a slightly higher cell count than the other intervals, so quantile classification effectively provides a best fit of equal cell proportions.

6.5 ARCHAEOLOGICAL SITE CLASSIFICATION

The New Zealand Archaeological site dataset requires improvement in the following areas in order to become more useful for predictive archaeological modelling.

- Greater spatial accuracy
- Alternatives to point representation of sites
- Categorical classification accuracy
- Temporal site classification

The archaeological literature in New Zealand is extensive, however this is not reflected in the national spatial dataset content. The spatial accuracy of the point dataset is gradually being improved through a systematic re-survey of sites, however the inconsistency between high numbers of coastal sites vs. sparse inland site numbers remains and introduces a bias towards formally surveyed zones.

Point representation of sites could be improved by simply using polygons for larger sites such as battle fields and large Pa sites. Archaeological site types would benefit from a reclassification, upgrading early amateur descriptions of site types to more formal site terminology. Dating of sites is a tougher challenge, however broad period classification into pre and post European settlement based on site design and carbon dating would provide significant value to future research regarding archaeological site distribution relative to low cost travel paths.

6.6 INTRODUCTION OF SOCIAL AND CULTURAL VARIABLES

Introducing known prehistoric settlements, culturally significant sites and natural resource locations as origin-destination points for anisotropic least cost path modelling will start to provide a more realistic network of low cost travel routes. The New Zealand Archaeological site database only records formally identified sites, and many culturally sensitive sites have remained unrecorded in databases or maps. Recently this situation has started to change via the negotiations between iwi groups and the Crown, settling land claims dating back to early European settlement and forced land take by the government. This process has required claimant groups to engage a Historian and mapping contractor to document sites of significance, to tell their story to support claims on rivers, coastlines and land.

Local Councils are trying to encourage iwi to share the locations of sensitive areas so they can form part of operative district plans where they should receive some form of protection. If there is a willingness from indigenous groups to contribute their knowledge of significant sites and trails, this could provide further meaningful origin-destinations and potentially act as verification of least cost paths.

7.0 CONCLUSION

This concluding chapter links the original research question, hypotheses and objectives with the findings and interpretations. The overall objective of this research is to answer the question; *Are inland Pa sites positioned in close proximity to least cost overland travel paths*?, and to test the supporting hypotheses;

- *There is a difference between observed and expected Pa site distribution.*
- *Inland Pa site density measured relative to regional least cost paths will result in a Gain statistic > 0.5.*

In order to answer the question this research was designed to replicate and enhance an existing GIS based method of creating a low cost regional prehistoric human travel network scenario applied to a new geographic setting within a quantitative and instrumental theoretical framework. The replicated methods from the original research and adaptations included in this study has found a significant pattern in the inland Pa site distribution when measured using travel cost proximity to least cost paths. Although the current archaeological evidence does not provide a strong association between Pa sites and paths, the original contribution to knowledge presented by this research is evidence that this GIS method demonstrates that inland Pa sites are unevenly distributed and clustered in close proximity to least cost travel paths. This provides quantitative analytical support for theoretical claims made in the archaeological literature.

7.1 RESEARCH FINDINGS

1. This research found that currently available spatial terrain data can be used to create a bare earth prehistoric landscape with lakes providing the only realistic obstacle to overland movement. Other travel cost elements discussed in Chapter 2, such as wetlands cannot be identified in sufficient detail from associated wet soil types due to the generalisation of soil maps. Rivers are also problematic as a travel cost variable as described in section 2.4.3 and demonstrated in section 4.7.1, as they do not present a complete barrier to overland travel and current data lacks variation in the physical attributes of rivers which might identify higher cross cost locations.

2. The lack of temporal Pa site data mentioned in section 4.4 used in this study compared to the rich period information described in the original research presents the possibility of the Pa site dataset containing sites developed post European contact where the siting criteria may have been different with less or no focus on close proximity to the most efficient travel routes.

3. The method used to define primary travel routes is limited to a manual selection of vector lines using a focal density raster for reference. This is a weakness of the replicated method and provides a challenge for future research.

4. The flow accumulation method used to create secondary paths was recreated. However, as it can't guarantee the lowest cost path, this method needs to be developed to overcome this limitation. This relates to secondary path limitations described in section 6.3.

5. When testing archaeological site distribution relative to path networks, this research found that the equal travel time interval method used by the original author is inappropriate and introduces the risk of invalidating statistical tests by failing to manage the requirements associated with minimum expected site counts which are directly related to the proportion of path distance interval areas. Equal area classification avoids this risk.

6. (1st hypothesis): There is significant statistical evidence demonstrating that Pa sites are not evenly distributed relative to the most efficient regional travel routes. This knowledge provides a significant contribution to the body of archaeological knowledge in New Zealand and supports international use of similar methods of deriving regional travel networks. The Pa distribution test result also supports the original author's statement concerning the universal and transposable nature of the technique.

7. Although Pa sites are not sited directly on paths, 79% of Pa sites are located within an 8.7 minute walk from a primary path and 77% of Pa are located within a 7 minute walk of a combined primary and secondary path suggesting a high level of accessibility.

8. (2nd hypothesis): The hypothesis stating an expected high Gain statistic for Pa site density relative to regional travel paths is not supported by the current recorded Pa site locations.

This research supports the findings of other applications of regional least cost travel network models. Anisotropic methods are now commonly used, with few examples of isotropic methods as used by Whitley and Hicks. The application of focal statistics in this research to isolate major travel arteries by measuring path density within adjacent raster cells is similar to recent research using vector line density and natural neighbor interpolation described in section 2.6. Introducing the concept of modelling theoretical travel networks with a GIS to New Zealand opens up opportunities for Archaeologists to take this approach further and attempt to offer archaeological explanations for significant spatial correlations discovered using anisotropic least cost path methods.

This research prompts further questions that may help develop the use of least cost path analysis into a mainstream archaeological exploratory tool.

- *Can Archaeologists use their expert knowledge to produce a more realistic prehistoric human travel cost surface dataset?*
- *Could qualitative archaeological research on human land use in prehistoric swamp areas be cross referenced with mapped soil classes to confirm probable extents of historic wetlands and potentially weight these areas by relative travel cost?*
- *Have archaeological surveys already been conducted along the primary and secondary modelled paths presented in this study to validate or challenge the method used?*

The strength of this research is that the method can be used anywhere where there is a suitable terrain surface dataset, digitised water bodies and other features considered to have been a barrier or hindrance to human travel and a dependent variable found near to travel routes. The limitations of the research are that the method can only model the most efficient paths across a basic travel cost surface and these paths may not correspond directly to actual historic paths but can provide an indication of route direction and avoidance of steep terrain and other high cost travel areas.

Potential applications for this research method include;

- Prospective archaeological field survey planning. Once refined, this method may challenge the continued focus on coastal areas and redirect future surveys to inland areas which may come under threat from infrastructure development.

- Longer term, if prehistoric routes can be verified by ground truth survey, this research may help road and utility asset owners to manage the risk of encountering archaeological sites where modelled travel paths intersect modern asset networks.

This research follows earlier national scale research into least cost travel routes and provides a modest contribution to the body of knowledge concerning Prehistoric human movement patterns within New Zealand.

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Several recommendations are now provided for future research which may improve different aspects of modelling regional low cost travel paths.

1. The development of a more realistic travel friction surface representing prehistoric vegetation would be a valuable enhancement of the process presented by this research. The focus should be on identifying historic wetlands not associated with post European land use change which could be achieved through more detailed soil and surface geology mapping. Archaeological input to this would be vital to understand which areas would have posed partial or complete barriers to human movement.

2. The addition of river width and depth attributes to the national river network dataset would be helpful for future research even though some rivers have changed alignment and flow rates over time. Without a more detailed river network, the value of rivers as an overland human travel cost variable is questionable. Archaeological qualitative research would be essential to guide river and lake travel cost improvements.

3. Further analysis of least cost path methods to determine whether travel time or energy consumption produces the most realistic least cost routes would be an important area of research if combined with a realistic local travel cost dataset. Although Tobler's hiking algorithm has been applied extensively, it may be appropriate to ground truth this method for a particular research area to reflect differences in regional terrain characteristics and walking time.

4. Although river travel was not considered in this research, canoe travel is well documented by early European explorers. Future research could combine river and overland travel to try to model a more comprehensive transportation network. An important objective of this research subject would be to determine whether the river network would have been fully or partially navigable by canoe. If this question could be answered, it may help to distinguish between travel mode catchments. If rivers were considered partially navigable, this would narrow the area being researched and support the need for overland paths for long distance travel.

5. Additional validation of least cost path results using historic accounts and site survey may provide archaeological evidence to support statistical test results.

6. Improvements to GIS software are required to address known limitations related to neighbourhood search extents when calculating rates of change in slope and inconsistent application of Dijkstra's algorithm in least cost path analysis. The ability to directly compare ArcGIS, IDRISI and GRASS GIS model results using a common raster format would provide a more robust method by allowing researchers to replicate and challenge previous research using different GIS software.

7. An alternative to path distance based statistical distribution tests is to consider the use of statistical tools provided by ArcGIS and other leading GIS software. In ArcGIS for example, the Incremental Spatial Autocorrelation statistics tool allows for autocorrelation of distance and Z scores, in much the same way as the a-spatial Chi Squared test shows cluster categories. Other ArcGIS statistical tools include Ripley's K function for multi-distance spatial clustering and the Anselin Moran's I tool, which looks for hot and cold clustering with weighted values. Currently however, ArcGIS statistical tools are based on Euclidean distance, and are not suitable for local analysis where slope is a critical factor without further development.

8. Instead of relying on stream modelling techniques to identify potential secondary travel paths, a GIS could be used to identify areas that might be considered attractive for siting defensive Pa sites such as mountain ridges and elevated plateaus. A site suitability approach could then form destinations for secondary travel paths, linking primary travel corridors to attractive defensive positions nearby. Figure 61 provides a simple method of defining these locations using ArcGIS 10. If ridge area cells are grouped into sections of ridge or isolated plateaus and vectorised, this allows a GIS to use Dijkstra's guaranteed optimal least cost path to generate a more robust and meaningful network of secondary paths.

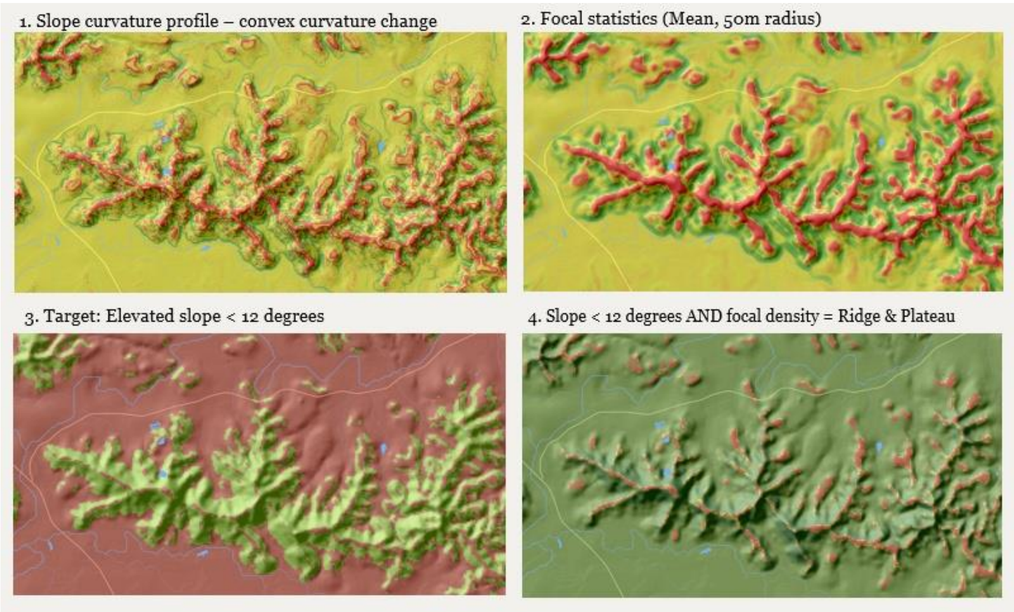


Figure 61: Example of ridges defined for use as a least cost path destination.

9. The study area presented in this research contains contrasting landforms with the north east and central areas having little relief and therefore much more accessible than the southern and western areas. These flatter areas also force least cost paths to be linear which may be realistic or simply a reflection of the lack of true travel cost variables. More detailed accessibility based research may benefit by sub dividing the study area into landscape character zones to determine whether Pa site to natural travel path accessibility is higher in certain topographic zones.

REFERENCES

Alexanderson, G. 2006. Euler and Konigsberg's Bridges: A Historical View. Bulletin of the American Mathematical Society 43, #4: 567-573.

Anderson, A.1991. The chronology of colonization in New Zealand. Antiquity 65.

Arnold, G., Newsome, P., Heke, H. 2004: Predicting archaeological sites in New Zealand. DOC. Science Internal Series 180. Department of Conservation, Wellington.

Aubry, T, Luís, L. Mangado Llach, J, M. Matias, H. 2012. We will be known by the tracks we leave behind: Exotic lithic raw materials, mobility and social networking among the Côa Valley foragers (Portugal). Journal of Anthropological Archaeology, Vol. 31. Elsevier.

Ausseil, A.G., Gerbeaux, P., Lindsay Chadderton, W., Stephens, T, Brown, D., and Leathwick, J. 2008. Wetland ecosystems of national importance for biodiversity: Criteria, methods and candidate list of nationally important inland wetlands. Landcare Research Contract Report: LC0708/158.

Balstrøm, T., 2002. On identifying the most time-saving walking route in a trackless mountainous terrain. Geografisk Tidsskrift, Danish Journal of Geography, 102: pp.51-58.

Barton, P. L., 1980. Maori geographical knowledge and mapping: a synopsis. The Turnbull Library Record, 13:5–25. Bateman Ltd, Auckland, 2000. Plate from New Zealand 1815-1870: the making of a colony, Students' Bateman Historical Atlas.

Bell, T, & Lock, G., 2000. Topographic and Cultural Influences on Walking the Ridgeway in Later Prehistoric Times. In: Lock, G (ed), Beyond the Map. Archaeology and Spatial Technologies, pp. 85-100. Amsterdam: IOS Press.

Bell, T., Wilson, A. and Wickham, A. 2002. Tracking the Samnites: Landscape and Communications Routes in the Sangro Valley, Italy. American Journal of Archaeology, Vol. 106, No. 2, pp. 169-186. Archaeological Institute of America.

Bellavia, G. 2002. Extracting natural pathways from a digital elevation model: applications to landscape archaeological studies. In Burenhult, G., Arvidsson, J. (ed.), Archaeological Informatics: Pushing the Envelope. Oxford, Archaeo-press.

Brassey, R. 1998. Archaeology in New Zealand. Pre-European use of dacite obsidian in northern New Zealand. Archaeology in New Zealand 41.

Bridge, Cyprian (Lieutenant-Colonel), 1807-1885. Bridge, Cyprian, 1807-1885: Native village with swing and pa in the distance. [Engraved by] J. Whymper, from a sketch by Col. Bridge. [London, 1859]. Thomson, Arthur Saunders, 1817-1860. National Library of New Zealand.

Carballo, D. M., and Pluckhahn, T. 2007. Transportation Corridors and Political Evolution in Highland Mesoamerica: Settlement Analyses Incorporating GIS for Northern Tlaxcala, Mexico. Journal of Anthropological Archaeology 26 (4): pp.607-629.

Cochran, W. G. 1952. The χ2 test of goodness of fit. Annals of Mathematical Statistics, 25, pp. 315-345.

Conolly, J. & Lake, M., 2006. Geographical Information Systems in Archaeology.

Cambridge Manuals in Archaeology. Cambridge: Cambridge University Press

Cumberland, Kenneth B., 1949. Aotearoa Maori: New Zealand About 1780. The Geographical Review, 39: pp. 401-24.

Dalla Bona, L.1994. Archaeological Predictive Modelling Project, Ontario Ministry of Natural Resources. Centre for Archaeological Resource Prediction, Lakehead University, Thunder Bay.

De Silva, M. and G. Pizziolo 2001. Setting up a "Human Calibrated" Anisotropic Cost Surface for Archaeological Landscape Investigation, in: Stančič, Z. and T. Veljanovski (eds.), Computing Archaeology for Understanding the Past. CAA 2000. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 28th Conference, Ljubljana, April 2000 (BAR International Series 931). Archaeopress, Oxford, pp. 279-288.

Dijkstra, E. W., 1959. A note on two problems in connection with graphs. Numerische Mathmatik.

Drennan, R,D. 2009. Statistics for Archaeologists, Interdisciplinary Contributions to Archaeology.

Encyclopedia of New Zealand (Online) http://www.teara.govt.nz/

Ericson, J.E. & Goldstein, R., 1980. Work Space: A New Approach to the Analysis of

Energy Expenditure within Site Catchments. Anthropology UCLA, 10.

Fabrega Alvarez, P., Parcero Oubina, C., 2007. Proposals for an archaeological analysis of pathways and movement. Archeologia e Calcolatori 18.

Gaffney, V. & Stančič, Z., 1992. Diodorus Siculus and the island of Hvar, Dalmatia: testing the text with GIS. In: G. Lock & J. Moffet eds. 1992. Computer Applications and Quantitative Methods in Archaeology 1991. *BAR International Series*, 577.

Gaffney, V.L. and P.M. van Leusen, 1995. GIS and environmental determinism. In Lock, G and Stancic, Z. (eds.), GIS and Archaeology: a European Perspective. Francis and Taylor.

Gietl, R., Doneus, M., and Fera, M., 2008. Cost Distance Analysis in an Alpine Environment. Comparison of Different Cost-surface Modules. In: Posluschny, A., Lambers, K., Herzog, I. (ed.) Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2-6, 2007.

Golson, J, 1957. Field Archaeology in New Zealand. Journal of the Polynesian Society 66 (1): pp.64–109.

Gorbey, K C., 1970. Pa Distribution in New Zealand. Unpublished M.A. thesis, Anthropology, University of Auckland.

Gorenflo, L.J Gale, N. 1990. Mapping Regional Settlement in Information Space, Journal of Anthropological Archaeology, 9.

Gorenflo, L.J Gale,N, 1996. Regional efficiency in prehispanic central Mexico: insights from geographical studies of archaeological settlement patterns, in: A.G. Mastache, J.R. Parsons, M.C. Serra Puche (Eds.), Arqueologı`a Mesoamericana. Homenaje a William T. Sanders, Instituto Nacional de Anthropologı`a e Historia, Mexico City, 1996, pp. 135–159.

Green, R.C. 1975. Adaptation and change in Maori culture. In: Kuschel, G. (Editor). Biogeography and Ecology in New Zealand. pp. 591-641. Junk, The Hague

Harris, T.M. and Lock, G. 1990. The diffusion of a new technology: a perspective on the adoption of Geographic Information Systems within United Kingdom archaeology, in Allen, K.M., Green, S., and Zubrow, E, (eds.), Interpreting Space: GIS and archaeology, 33-53, Taylor and Francis.

Hare, T. S., 2004. Using measures of cost distance in the estimation of polity boundaries in the Post classic Yautepec valley, Mexico. Journal of Archaeological Science, 31.

Hart, P., Nilsson, N., and Raphael, B. 1968. A formal basis for the heuristic determination of minimum cost path. IEEE Transactions on Systems Science and Cybernetics, 4 (2): pp.100-107.

Herzog, I., 2010. Theory and Practice of Cost Functions. In: Melero, et al., 2010, pp.431-434.

Herzog, I. & Posluschny, A., (2011). Tilt – Slope-Dependent Least Cost Path Calculations Revisited. In: E. Jerem, F. Redo & V. Szeverényi (eds). On the Road to Reconstructing the Past. Proceedings of the 36th CAA conference, Budapest, 2008, pp. 212-218.

Herzog, I. 2012. Least-cost networks. Proceedings of the 40th Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2012, Southampton.

Herzog, I. 2013. Testing Models for Medieval Settlement Location. Springer International Publishing, Switzerland.

Houle, J.L, Broderick, L.G. 2014. GIS approaches to past mobility and accessibility: An example from Bronze Age Khanuy Valley, Mongolia. In Past Mobilities: Archaeological Approaches to Movement and Mobility, Ashgate Publishing.

Howey, M.C.L., 2007. Using multi-criteria cost surface analysis to explore past regional landscapes: a case study of ritual activity and social interaction in Michigan, AD 1200-1600. Journal of Archaeological Science 34.

Imhof, E., 1950. Gelande und Karte. Zurich: Eugen Rentsch Verlag.

Judge, J.W. and Sebastian, L. 1988 (eds.). Quantifying the Present and Predicting the Past. Theory, Method and Application of Archaeological Predictive Modeling. U.S. Department of the Interior, Bureau of Land Management, Denver.

Kamermans, H., J. Deeben, D. Hallewas, P. Zoetbrood, M. van Leusen and Ph. Verhagen 2005. Project proposal. In M. van Leusen and H. Kamermans (eds), *Predictive Modelling for Archaeological Heritage Management: A research agenda*. Nederlandse Archeologische Rapporten 29, 13-23. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Kenny, J.A., and Hayward, B.W., 1993. Geological Society of New Zealand Miscellaneous Publication No. 78, Geological Society of New Zealand, Lower Hutt.

Kinaston RL, Walter RK, Jacomb C, Brooks E, Tayles N, et al. (2013) Correction: The First New Zealanders: Patterns of Diet and Mobility Revealed through Isotope Analysis. PLoS ONE 8(9): 10.1371/annotation/05b5437f-cea1-43f4-88c1-53f1377683ef. doi:10.1371/annotation/05b5437f-cea1-43f4-88c1-53f1377683ef

K.L. Jones/Department of Conservation.

Kohler, T.A, and Parker, S.C. 1986. Predictive models for archaeological resource location. In: Schiffer, M.B. (ed), Advances in Archaeological Method and Theory 9, Academic Press, New York.

Kohler TA, 1988 Predictive locational modeling: history and current practice. In Judge W.J., & Sebastian, L. (eds.) Quantifying the present, predicting the past. Denver: US Department of the Interior Bureau of Land Management. pp.19-59.

Kvamme, K., 1988, Development and testing of quantitative models. In W.J. Judge and L. Sebastian (eds.), Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling, U.S. Department of the Interior, Bureau of Land Management, Denver, pp. 325 - 428.

Land Air Water Aotearoa (Online) http://www.lawa.org.nz/

Land Information New Zealand (Online). https://data.linz.govt.nz/

Langmuir E. Mountain craft and Leadership. The Scottish Sport Council/MLTB; Cordee, Leicester: 1984.

Leahy, A.L., and Walsh, W.D., 1976. Archaeological Survey Report, Bay of Islands and Kerikeri / Paihia Area. The New Zealand Historic Places Trust.

Leathwick, J.R. 2000: Predictive models of archaeological site prediction in New Zealand. Science and Research Internal Report 181. Department of Conservation, Wellington.

Leathwick, J.R. et al, Freshwater Ecosystems of New Zealand (FENZ Geodatabase). Version One, August 2010 User Guide. Department of Conservation.

Leusen, P. M. van, 1993. Cartographic modelling in a cell-based GIS. Computing the Past. CAA92: Computer Applications and Quantitative Methods in Archaeology, Aarhus University Press, Aarhus.

Leusen, P. M. van, 2002. Pattern to Process. Methodological Investigations into the Formation and Interpretation of Spatial Patterns in Archaeological Landscapes. PhD thesis. Rijksuniversiteit Groningen. Groningen.

Leusen, P.M van, J. Deeben, D. Hallewas, P. Zoetbrood, H. Kamermans and Ph. Verhagen 2002. Predictive Modelling for Archaeological Heritage Management in the Netherlands. Baseline Report. Den Haag: NWO

Leusen, M. van and H. Kamermans (eds) 2005. Predictive Modelling for Archaeological Heritage Management: *A research agenda*. Nederlandse Archeologische Rapporten 29. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.

Limp, W.F., 1990. Intersite Analysis: Aboriginal use of the Rush locality, in Sabo, GI (ed), Archaeological Investigations at 3MR80-Area D in the Rush Development Area, Buffalo National River, Arkansas, Professional Paper 38 vol. 1: 295-345. Santa Fe: US National Park Service.

Llobera, M., 1996. Exploring the Topography of Mind: GIS, Social Space and Archaeology. Antiquity, 70.

Llobera, M., 2000. Understanding movement: a pilot model towards the sociology of movement. In Lock, G. Beyond the Map. Archaeology and Spatial Technologies.

Llobera, M. Fabrega-Alvarez, P. Parcelero-Oubina,C. 2010. Order in movement: a GIS approach to accessibility. Journal of Archaeological Science 37, 2011.

Llobera, M. & Sluckin, T.J., 2007. Zigzagging: Theoretical insights on climbing strategies. Journal of Theoretical Biology, 249.

Lock,G. and Stancic,Z. 1995, Archaeology and Geographical Information Systems: A European Perspective. Taylor & Francis.

Lock, G., and Harris, T.M., 1996. Danesbury revisited: an Iron Age hillfort in a digital landscape, Anthropology, Space, and Geographic information and Analysis. Oxford University Press.

Lock, G. (ed.), 2000. Beyond the Map. Archaeology and Spatial Technologies. NATO Science Series, Series A: Life Sciences -Vol. 321. IOS Press, Amsterdam.

Lock, G., and Pouncett, J. 2010. Walking the Ridgeway Revisited: The Methodological and Theoretical Implications of Scale Dependency for the Derivation of Slope and the Calculation of Least-Cost Pathways, in B. Frischer, J. Webb Crawford and D. Koller (eds.), *Making History Interactive*, Proceedings of the Computer Applications and Quantitative Methods Conference, Williamsburg 2009,Oxford: Archaeopress, 191-202.

Lombard, K., and Church, R.L., 1993. The gateway shortest path problem: generating alternative routes for a corridor routing problem, Geographical Systems, 1.

Malczewski, J., 2004. GIS-based land-use suitability analysis: a critical overview. Progr. Plann. 62 (1), pp. 3-65.

McCoy, M.D., Ladefoged, T.N., Blanshard, A., Jorgensen, A. 2010. Reconstructing lithic supply zones and procurement areas: an example from the Bay of Islands, Northland, New Zealand. Journal of Pacific Archaeology 1. pp.174-183.

McCoy et al 2011, A cost surface model of volcanic glass quarrying and exchange in Hawai'i. Journal of Archaeological Science, 38. 2547-2560. Elsevier Ltd.

McCoy MD, Carpenter J (2014) Strategies for Obtaining Obsidian in Pre-European Contact Era New Zealand. PLoS ONE 9(1): e84302. doi:10.1371/journal.pone.0084302

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd edition.). Sparky House Publishing, Baltimore, Maryland. (Online) http://www.biostathandbook.com/

McFadgen, B G 2007. Hostile Shores: Catastrophic Events in Prehistoric New Zealand and their Impact on Māori Coastal Communities. Auckland University Press.

McGlone, M.S. 1983b. Polynesian deforestation of New Zealand: a preliminary synthesis. Archaeology in Oceania 18: 11-25.

McGlone, M.S. 1989. The Polynesian Settlement of New Zealand in relation to environmental and biotic changes. New Zealand Journal of Ecology, Vol 12.

Minetti, A.E., 1995. Optimum gradient of mountain paths. Journal of Applied Physiology, 79.

Minetti, A.E., Moia, Ch., Roi, G.S., Susta, D. & G. Ferretti., 2002. Energy cost of Walking and running at extreme uphill and downhill slopes. Journal of Applied Physiology 93.

Monteleone, K., E. J. Dixon, and A. D. Wickert (2013), Lost Worlds: A predictive model to locate submerged archaeological sites in SE Alaska, USA, in Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012, edited by G. Earl, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska, and D. Wheatley, Amsterdam University Press.

Moore P.R 1982. Obsidian deposits at Huruiki, Northland. New Zealand Archaeological Association newsletter 25: 240_246.

Moore, P.R. 2012. The obsidian sources of Northland, New Zealand. Journal of the Royal Society of New Zealand.

Murrieta-Flores, P. (2012): Understanding movement during Late Prehistory through spatial technologies: The role of natural areas of transit in South-Western Iberia. Trabajos de Prehistoria, CSIC, 69 (1): 114-133.

NIWA, Bay of Islands Sedimentation Report 2010 for the Northland Regional Council collected as part of the 2010 LINZ Oceans 20/20 Bay of Islands (BOI) survey.

Pandolf, K.B.B., Givoni, R.F.Goldman. 1977. Predicting Energy Expenditure with Loads while standing or walking very lowly. Journal of Applied Physiology, 43 (4). Pp.577-581.

Rapaport, E, and Snickars, F., GIS-Based Road Location in Sweden: A Case Study to Minimize Environmental Damage, Building Costs and Travel Time. (pp. 135–153), in John Stillwell, Stan Geertman, and Stan Openshaw (Eds.), Geographical Information and Planning (New York: Springer, 1999).

Reeler, C. 1997, Fuzzy logic, neural networks and the analysis of pa sites, in M. NORTH, I. JOHNSON (eds.), Archaeological Applications of GIS: Proceedings of Colloquium II, UISPP XIIIth Congress (Forlì, Italy, September 1996), Sydney University Archaeological Methods Series, n. 5.

Santee, W.R., Allison, W.F., Blanchard, L.A., Small, M.G., 2001. A proposed model for Load carriage on sloped terrain, Aviation, Space and Environmental Medicine 72.

Scott, C. 2007. A Cost-surface analysis of the Exchange of Obsidian in Pre-Historic New Zealand. New Zealand Journal of Archaeology 29: 57–87

Shennan, S. 1997. Quantifying Archaeology. Edinburgh University Press. ISBN: 0-7486-0791-9.

Sheskin, D.J 2011. Handbook of parametric and nonparametric statistical procedures. Taylor and Francis Group. ISBN: 978-1-4398-5801-1.

Stancic, Z., and Kvamme, K.L. 1999. Settlement Pattern Modelling Through Boolean Overlays of Social and Environmental Variables. In New Techniques for Old Times, CAA98, J.A.Barcelo, I.Briz, and A.Vila, eds., BAR International Series 757, Tempus Reparatum, Oxford, pp. 231-237.

Strahler, A.N. 2001. Quantitative Analysis of Watershed Geomorphology. In: Transactions of the American Geophysical Union 8, 913-920.

Sutton, D.G., 1987: A paradigmatic shift in Polynesian prehistory: implications for New Zealand. New Zealand Journal of Archaeology, 9, pp.135-155.

Sutton, D. G. 1990. Organisation and Ontology: The Origins of the Northern Maori Chiefdom, New Zealand." Man 25(4): pp. 667-692.

Sutton, D.G., Furey, L., and Marshall, Y.M. 2003. The Archaeology of Pouerua. Auckland University Press, Auckland.

Surface-Evans, S.L., 2012. Cost catchments: a least cost application for modeling Hunter-Gatherer land use. In: White, D.A., Surface-Evans, S.L. (eds.), Least Cost Analysis of Social Landscapes: Archaeological Case Studies. University of Utah Press, Salt Lake City.

Te Area, The Encyclopedia of New Zealand, http://www.teara.govt.nz/ . The Ministry for Culture and Heritage, New Zealand.

The Land Resource Information Systems Portal.(Online) https://lris.scinfo.org.nz/

The National Institute of Water and Atmospheric Research (Online) http://www.niwa.co.nz/

The New Zealand Archaeological Association.(Online) http://archsite.org.nz/

Thomas, D. H., 1978. The Awful Truth about Statistics in Archaeology. American Antiquity 43(2): pp.231-244.

Tobler, W. 1993. Three Presentations on Geographical Analysis and Modeling. Technical Report 93-1. National Center for Geographic Information and Analysis. (Online) http://www.ncgia.ucsb.edu/Publications/Tech_Reports/93/93-1.PDF.

Tripcevich, N. 2009. Workshop 2009, No1. Viewshed and Cost Distance. (Online) http://mapaspects.org/book/export/html/3743

Verhagen, P., and Whitley, T.G. 2011. Integrating Archaeological Theory and Predictive Modeling: A Live Report from the Scene. Journal of Archaeological Method and Theory 19(1): 49-100.

Verhagen, P. & Jeneson, K., 2012. A Roman Puzzle. Trying to Find the Via Belgica with GIS. In: A. Chrysanthi, P.M. Flores & C. Papadopoulos eds. 2012. Thinking Beyond the Tool. Archaeological computing and the interpretive process. *BAR International Series*,2344.

Verhagen, P., Brughmans, T., Nuninger, L., and Bertoncello, F. 2013. The Long and Winding Road: Combining Least Cost Paths and Network Analysis Techniques for Settlement Location Analysis and Predictive Modelling. In G. Earl, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska and D. Wheatley (eds). Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012, pp. 357-366.

Verhagen, P. 2013. On the Road to Nowhere? Least Cost Paths, Accessibility and the Predictive Modelling Perspective. In F. Contreras, M. Farjas and J.F. Melero (eds.) Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Granada, Spain, April 2010. Archaeopress, Oxford, pp. 383-390.

Walton, T. 2002. Mapping regional variation in settlement in late prehistoric New Zealand. Archaeology in New Zealand 45(4):325-330.

Warren, R.E., 1990. Predictive Modelling in Archaeology: A Primer. In Allen, K.M., Green, S.W., and Zubrow, B.W., (eds.), Interpreting Space: GIS and Archaeology. London. Taylor and Francis.

Westcott, K.L., and Brandon, R.J.(eds.). 2000. Practical Applications of GIS for Archaeologists: A Predictive Modeling Toolkit. Taylor & Francis, London.

Wheatley, D. 1995. Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application, in Lock, G & Z Stancic (eds.), GIS and Archaeology: a European Perspective: pp.171-186.

Wheatley, D. 2003. Making space for an archaeology of space. (Online): D. Wheatley 2004 'Making space for an archaeology of place', Internet Archaeology 15. http://dx.doi.org/10.11141/ia.15.10

Wheatley, D. & Gillings, M., 2002. Spatial Technology and Archaeology. The Archaeological Applications of GIS. London, New York: Taylor & Francis.

White, D.A., 2012. Prehistoric trail networks of the Western Papagueria: a multifaceted least cost graph theory analysis. In: White, D.A., Surface-Evans, S.L. (eds.), Least Cost Analysis of Social Landscapes: Archaeological Case Studies. University of Utah Press, Salt Lake City.

White, D.A., and Barber, S.B. 2012. Geospatial modeling pf pedestrian transportation networks: A case study from pre-Columbian Oaxaca, Mexico.

White, D.A & Surface-Evans, S.L. (eds). 2012. Least Cost Analysis of Social Landscapes. Archaeological Case Studies. The University of Utah Press, Salt Lake City.

Whitley, T.G. and I. Burns, 2007. An Explanatory Framework for Predictive Modelling Using An Example from Marion, Horry, Dillon and Marlboro Counties, South Carolina. Paper prepared for Computer Applications in Archaeology Conference, Fargo, North Dakota, April 18-21, 2006.

Whitley, T.G., and Hicks, L.M., 2003. A geographic information systems approach to understanding potential prehistoric and historic travel corridors. Southeastern Archaeology 22.

Whitley, T.G. and Burns, G., 2008. Conditional GIS Surfaces and their Potential for Archaeological Predictive Modelling. In: Posluschny, et al., 2008.

Yates, D., Moore, Moore, D., McCabe, G. 1999. The Practice of Statistics (1st Ed.). New York: W.H. Freeman.

Zadeh, L.A., 1965. Fuzzy Sets, Information and Control, Volume 8.

Zaiontz, C. (2014) Real Statistics Using Excel (Online) http://www.real-statistics.com/free-download/

Zakšek, K., Fovet, E,. Nuninger, L., Pobodnikar, T., 2008. Path Modelling and settlement pattern. In Posluschny, A., Lambers, K., Herzog, I. (eds.). Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Berlin, Germany, April 2-6, 2007, pp: 309-315.

Zubrow, E.B.W, 1990. Contemplating space: a commentary on theory. In Allen et al. 1990, pp. 67-72.

APPENDICES

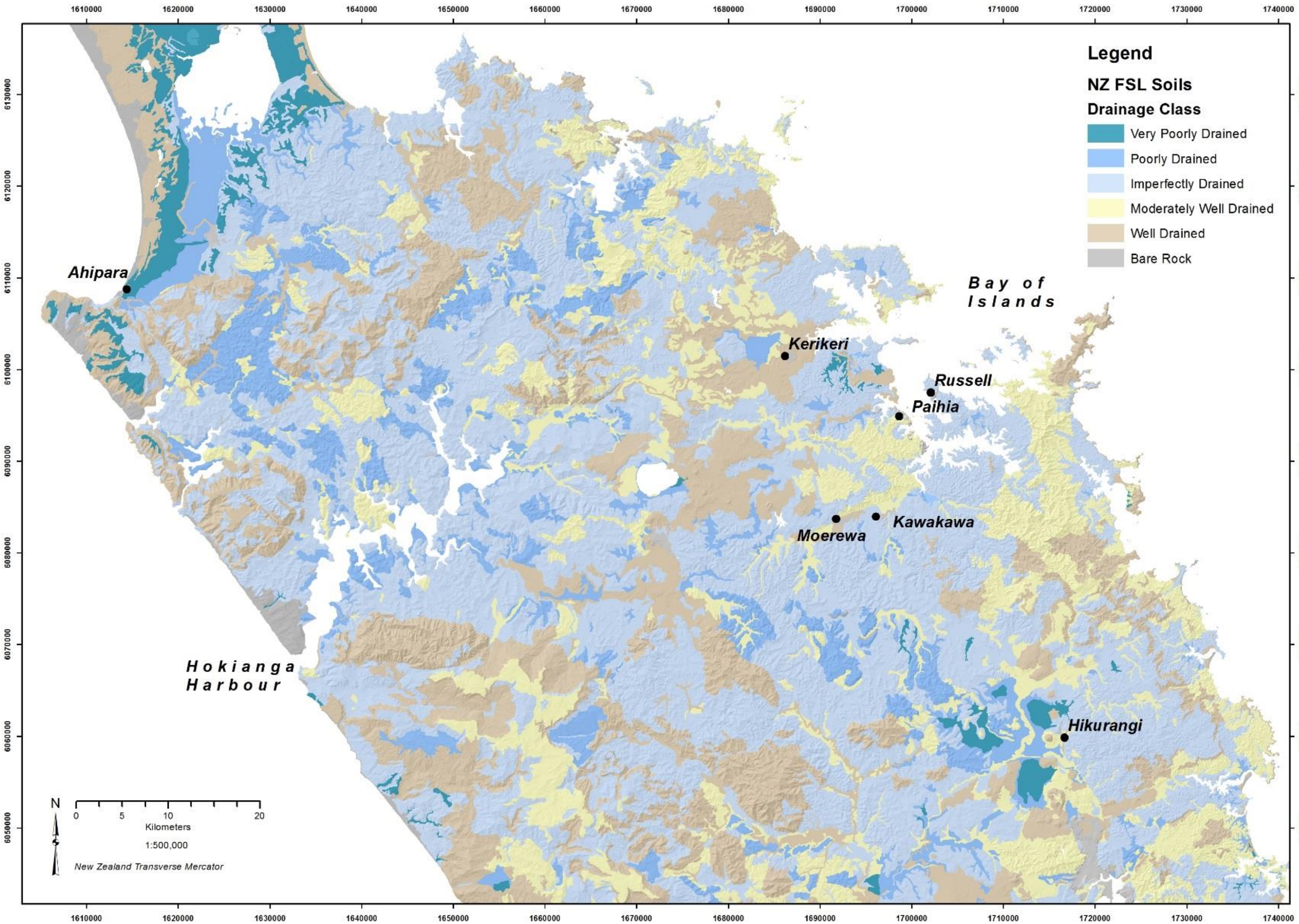


Figure 62: The New Zealand Fundamental Soils dataset

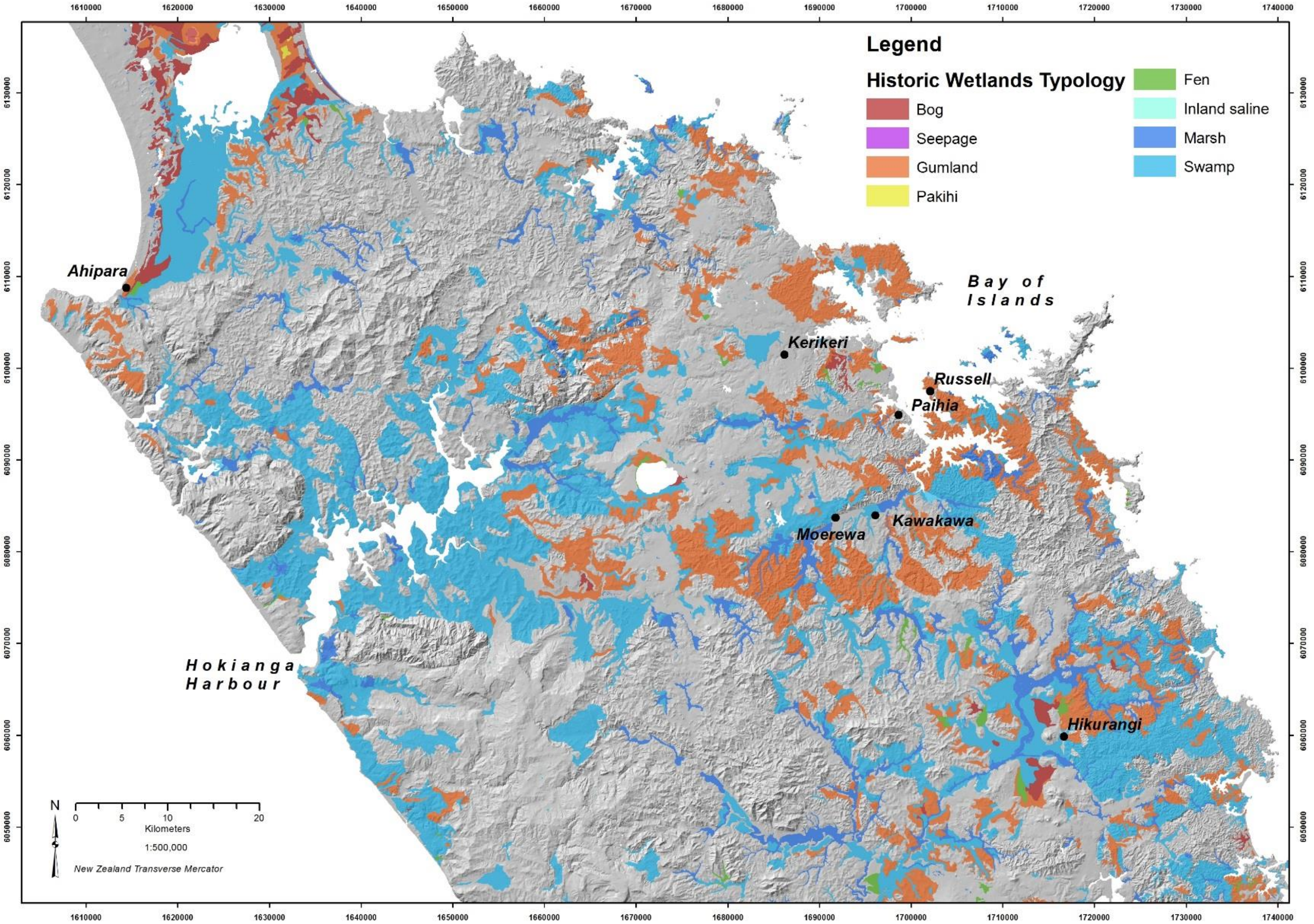


Figure 63: FENZ Historic wetland dataset (Ausseil et al 2008)

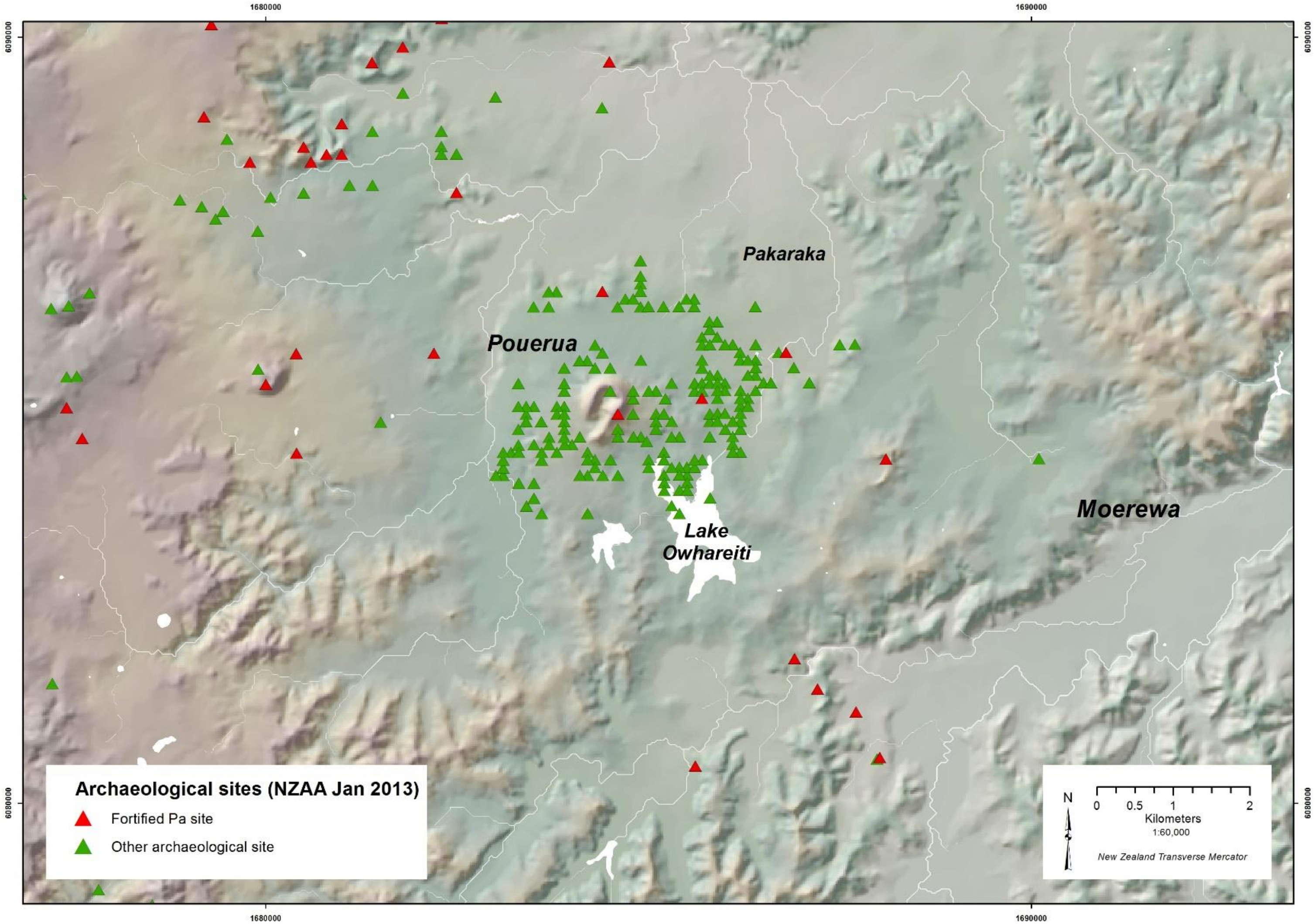


Figure 64: Clustered archaeological sites around Pouerua

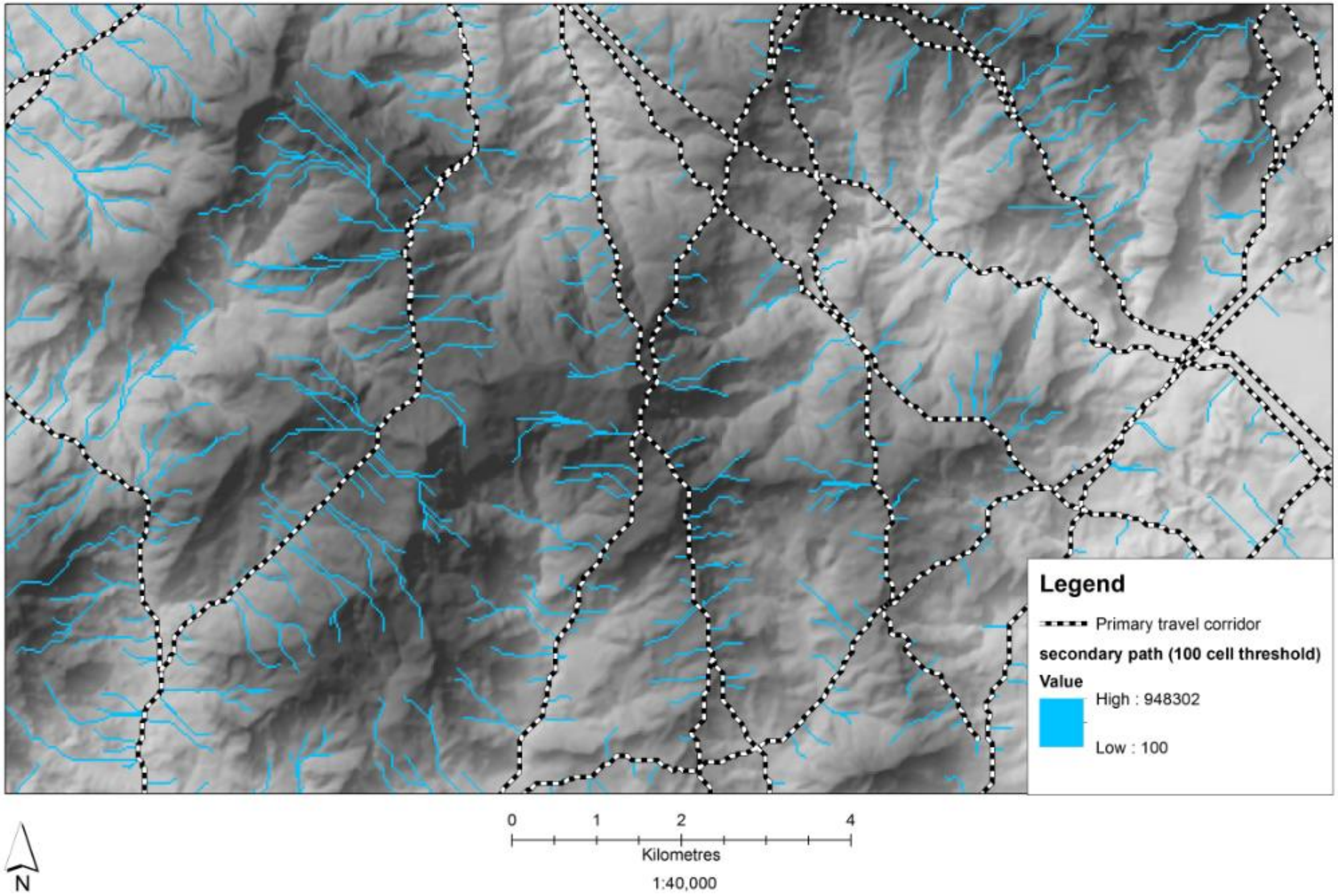


Figure 65: Secondary flow paths with 100 cell threshold applied

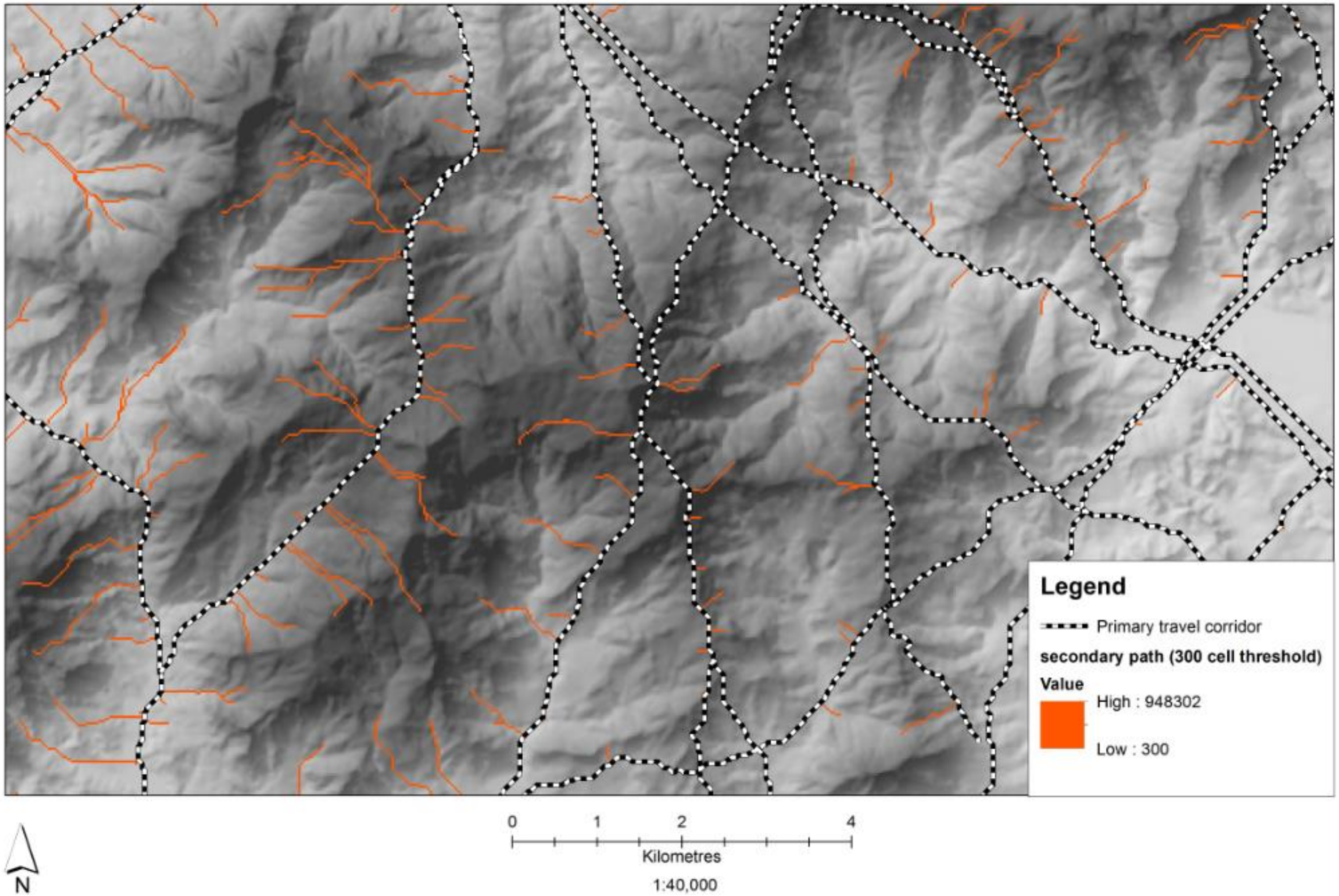


Figure 66: Secondary flow paths with 300 cell threshold applied

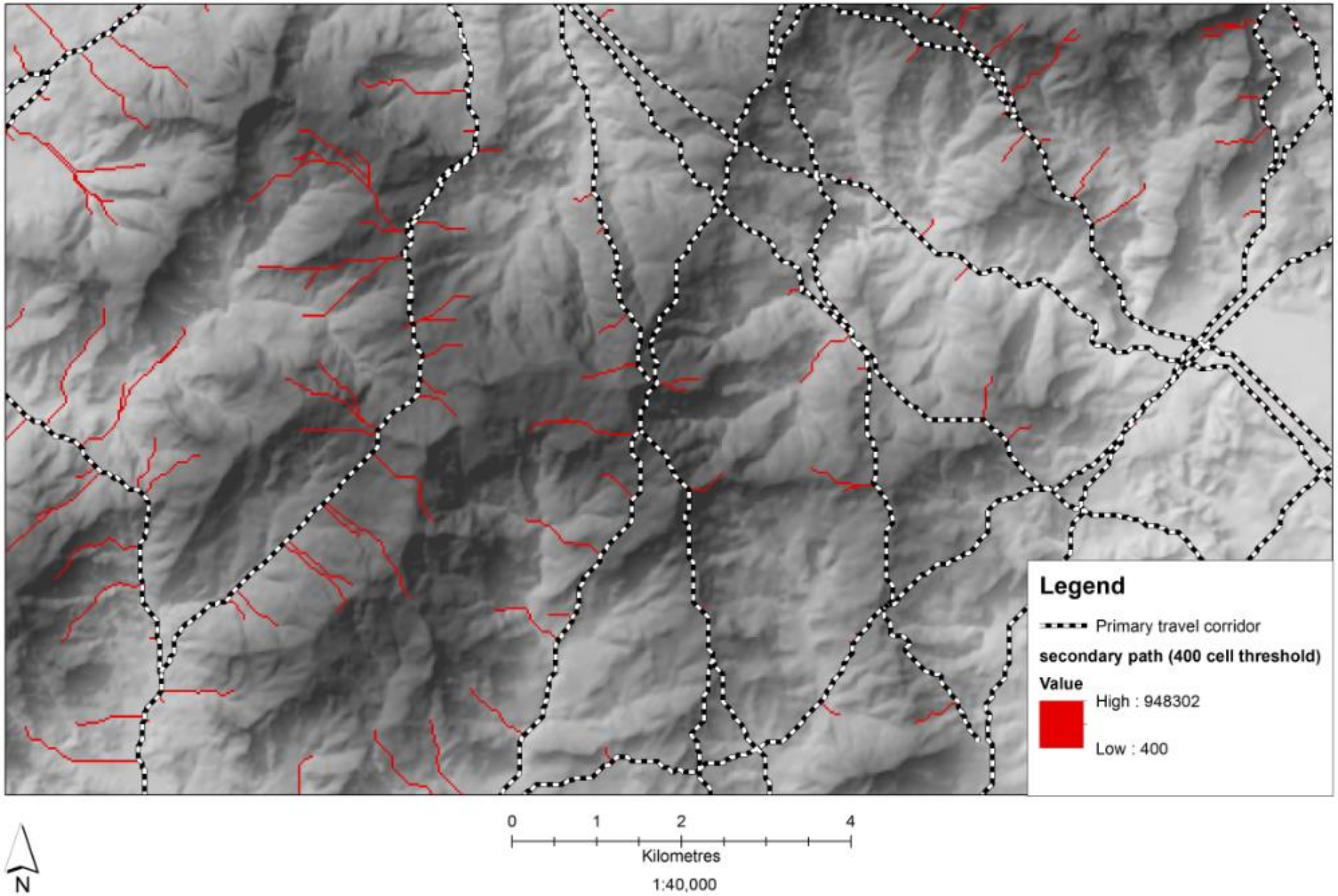


Figure 67: Secondary flow paths with 400 cell threshold applied

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table 10: A Chi Squared distribution table

(Pennsylvania State University, Department of Statistics).