Predicting archaeological sites in New Zealand

DOC SCIENCE INTERNAL SERIES 180

Greg Arnold, Peter Newsome, and Hamish Heke

Published by
Department of Conservation
PO Box 10-420
Wellington, New Zealand

DOC Science Internal Series is a published record of scientific research carried out, or advice given, by Department of Conservation staff or external contractors funded by DOC. It comprises reports and short communications that are peer-reviewed.

Individual contributions to the series are first released on the departmental website in pdf form. Hardcopy is printed, bound, and distributed at regular intervals. Titles are also listed in the DOC Science Publishing catalogue on the website, refer http://www.doc.govt.nz under Publications, then Science and Research.

ISSN 1175-6519

ISBN 0-478-22585-7

In the interest of forest conservation, DOC Science Publishing supports paperless electronic publishing. When printing, recycled paper is used wherever possible.

This report was prepared for publication by DOC Science Publishing, Science & Research Unit; editing by Helen O'Leary and layout by Geoff Gregory. Publication was approved by the Manager, Science & Research Unit, Science Technology and Information Services, Department of Conservation, Wellington.

CONTENTS

Predicting archaeological sites in New Zealand

Greg Arnold, Peter Newsome, and Hamish Heke

Landcare Research, Private Bag 11-052, Palmerston North, New Zealand

A B S T R A C T

Statistical modelling of the differences between the environments where archaeological sites are known to occur and random sites was used to create maps and GIS coverages depicting the relative likelihood of any 100 × 100 m area of New Zealand containing an archaeological site. Coverages were developed separately for pa, middens, and pits as well as for all sites. The coverage for all sites shows very good discrimination over the whole country, but discrimination was less satisfactory for the separate categories in the South Island, where sites are rarer. The strongest indicator of a site was that there is a known site nearby, but a wide variety of other measures had an influence, particularly a site's altitude and its distance from the sea.

Keywords: Spatial modelling, prediction, archaeological sites, environment, pa, pits, middens, New Zealand

Arnold, G.; Newsome, P.; Heke, H. 2004: Predicting archaeological sites in New Zealand. *DOC Science Internal Series 180*. Department of Conservation, Wellington. 24 p.

1. Introduction

This work builds on a preliminary analysis (Leathwick 2000), which showed that it was feasible to predict the likelihood of occurrence of archaeological sites from a range of environmental variables. The aim of this study is to construct a working tool for predicting the likelihood of a site at any point in New Zealand. The earlier study has been refined to give separate predictions for pa, middens and pits as well as for all sites. The spatial resolution has also been enhanced from 1000 × 1000 m to 100 × 100 m, some additional variables have been tried, and some old ones refined.

The results are presented as maps with contours showing the likelihood of a site. The data used to construct these maps are available from DOC Science & Research Unit.

Apart from being of general archaeological interest, these maps will be of value for warning when a development is at risk of infringing on archaeological remains.

2. Data

2.1 ARCHAEOLOGICAL SITES

A list of 44 300 sites, 39 100 in the North Island and 5200 in the South Island, categorised as pa, pits, or middens, was supplied by DOC from the Central Index of New Zealand Archaeological Sites (CINZAS) database. CINZAS provides an electronic index to the more than 53 000 records in the New Zealand Archaeological Association Site Recording Scheme. Locations were specified as New Zealand Map Grid (NZMG) easting and northings.

A 100 × 100 m grid was generated covering the North and South Islands separately. Each site was mapped into a cell in this grid. Where multiple sites occurred in the same cell that cell was repeated in the database, but downweighted in the analysis to contribute as though it were a single cell. For example two middens in a cell would be recorded separately but each given a weight of 0.5.

2.2 ENVIRONMENTAL VARIABLES

The environmental variables tested are listed below. Not all of these were ultimately found to be useful.

Climate, as used in Leathwick (2000) and described in Leathwick et al. (2003):

Mean annual temperature (*MAT*)

Temperature seasonality (*TSeas*)

Mean annual solar radiation (*MAS*)

Solar radiation seasonality (*SSeas*)

Ratio of rainfall to potential evapotranspiration in the driest month (*RtoPET*)

Spring air saturation deficit (*VPD*)

Soil, where the soil parent material categories of Leathwick (2000) were simplified to three variables, each giving ordered categories. These, too, are now described in Leathwick (2003):

Calcium (*Ca*)

Acid P (*AcidP*)

Particle size (*PSize*)

Proximity to water, which was refined to resolve three classes of water and, in the case of seawater, the nature of the coastline. Distances were calculated from the hydrological layers in the 1:50 000 Land Information New Zealand (LINZ) Topo50 database. All are Euclidean distances from the centroids of grid cells.

Distance to navigable water. This was defined as any river with both banks marked separately, lakes or lagoons (*DistNavW*).

Distance to any other fresh water. This was any river or stream marked by a single line (*DistW*).

Distance to sea (*DistSea*)

Type of coastline: shingle, mud, rock, sand, later summarised as two presence/absence variables, *Rock* or *Sand.* Occasionally both were present.

Topography, as used by Leathwick (2000), with the addition of topographical position. All topographical themes were derived from a 100 × 100 m Digital Elevation Model (DEM) generated from the 20 m contour set contained in the LINZ Topo50 database:

Elevation (*Elevation*)

Topographical position, measured as the number of the eight adjacent cells higher than the given cell (*TopoPos*)

Aspect, fitted as *cos* (*Aspect* + *k*) to avoid the discontinuity between 360° and 0° (*Aspect*)

Slope, which gave a better fit as *cos* (*Slope* + *k*) (*Slope*)

Finally, the distance to the nearest known archaeological site was introduced, to utilise any spatial correlation between sites (*DistSite*).

2.3 DATA ACCURACY

All distance data (i.e. distance to streams, rivers, and beaches), and data based on elevation (i.e. aspect, slope, prominence), should be of such accuracy that changes over 100 m will be of the same order of magnitude as their errors. Climate variables are interpolated from climate station data using elevation, described in Leathwick et al. (2003). The elevation variable will, therefore, contain all the short-distance information in the climate data, so the climate data will mainly contribute to large-scale prediction. Soil data were interpolated from the New Zealand Land Resource Inventory as also described in Leathwick et al. (2003).

An unexpected inaccuracy lay in the location of the sites themselves. These were read from 1:50 000 topographical maps and were expected to have an error of within ± 100 m. However, 4300 sites were located beyond the shoreline, which is over half the number of sites recorded within 100 m from the shore. The source of the error is more likely to be an incorrect scale conversion at some time in the past than a map reading error. Presumably an error of similar magnitude exists in the location of all sites.

For the purposes of this study, sites recorded as being in the sea were moved to the nearest cell on land. However, errors of a few hundred metres in the location of sites mean predictions can only be expected to be accurate at this scale.

3. Statistical methods

As in Leathwick (2000), a sample of cells (39 184 in the North Island and 6281 in the South Island, roughly equal to the number of known sites), containing no known archaeological site, were selected randomly, and the environment of these cells was compared with the environment of cells containing sites. The statistical procedures look for those variables that distinguish random cells lacking sites from cells known to contain sites, defining a continuum of environments from those strongly represented in the random cells to those strongly represented in the archaeological site cells. The proportion of archaeological site cells in each environment is then a measure of how probable it is that a site might be found in that environment. This does not give an absolute estimate of the probability of a cell containing a site, a value that can be estimated only by the impractical process of physically searching a large sample of 100 × 100 m cells for archaeological sites, randomly located throughout the country. The measure does, however, give a measure of the relative likelihood of a site that is broadly comparable over the whole country.

Leathwick (2000) used generalised additive models to uncover those variables associated with the presence of sites. Being additive, these models assume that a factor has the same effect regardless of the level of other factors. Here, this assumption was checked by using a classification tree (Venables & Ripley 1999, p. 310). This looks for a simple question like:

Is variable > x?

the answer to which depends on the presence of a site. The question divides the cells into two categories, those likely to contain sites and those not, and the process is then repeated for each category. The result is a tree of questions. Interactions are detected by different variables becoming involved in different branches of the tree. For example, for the South Island, the question:

Is distance to beach > 1800 m?

was the best single question for predicting the presence of a site. If the answer was 'Yes', the next question was about mean temperature, but if 'No', it was about altitude. This preliminary analysis showed that variables that were important were different for the North and South Island, for coastal or inland cells, and for cells near known sites. Separate models were estimated within these categories, with presence/absence of a site measured by the log-odds ratio as in Leathwick (2000), and Venables & Ripley (1999, p. 349). Predictor variables were first fitted as smooth curves, but generally straight lines gave as good predictions. On occasion, a quadratic term was added, or an upper or lower limit, where a the curve showed that the variable's effect was bounded. This simplified the calculation of estimates for the extremely large number of 100×100 m^2 cells required to cover the whole country.

4. The fitted models

Since the classification tree showed that different models were needed in the North and South Islands, and that within each island different variables were important near existing sites or a beach, three separate models were fitted for each island. Table 1 lists the variables used in each. For pa, pits and middens, a single model was fitted for all distances from the beach, and the *MAT* effect estimated using an interaction with *DistSea*. This is denoted by *MAT*DistSea* in Table 1.

Detailed descriptions of the predictive models are given in Appendix 1.

4.1 CRITERIA USED

Three criteria were used in determining which variables to include in the models.

Statistical

With these very large samples, variables can be statistically significant without adding a great deal to the prediction. Nevertheless, significance (based on the fall in model deviance) provided a useful screening device to eliminate variables that added no predictive information.

Effect

The amount the variable contributed to the predicted *log(odds)* over its range. Typically, a variable that changed the *log(odds)* by more than 0.5 would be included.

Consistency

Some variables are sufficiently highly correlated that one can be substituted for the other with little change in the predictions. Where this happened, variables

TABLE 1. VARIABLES USED FOR EACH OF THE FITTED MODELS FOR PREDICTING ARCHAEOLOGICAL SITES.

ALL SITES					
North Island					
In vicinity of known site	*DistSite Elevation*	*MAT TSeas*	*AcidP*	*DistSea Sand*	
Within 4 km of beach	*DistSite Elevation*	*MAT VPD*	*AcidP*	*DistSea DistW DistNavW*	*TopoPos*
Beyond 4 km of beach	*DistSite Elevation Slope*	*MAT TSeas*	*AcidP*	*DistSea DistW*	*TopoPos*
South Island					
In vicinity of known site	*DistSite Elevation Slope*	*MAT TSeas*	*AcidP*	*DistSea DistNavW Rock*	*TopoPos*
Within 4 km of beach	*DistSite Elevation Slope*	*MAT RtoPET*		*DistSea DistNavW*	
Beyond 4 km of beach	*DistSite Elevation Slope Aspect*	*MAT TSeas*			
PA					
North Island					
In vicinity of known site	*DistSite Slope*	*MAT VPD*	*AcidP*	*Sand*	*TopoPos*
No known site near	*DistSite Elevation Slope*	*VPD MAT*DistSea*	*AcidP*	*DistNavW*	*TopoPos*
South Island					
In vicinity of known site	*DistSite DistSea Slope*	*MAT TSeas VPD*		*Sand*	*TopoPos*
No known site near	*DistSite DistSea Elevation*	*MAT VPD*		*DistNavW Sand*	*TopoPos*
MIDDENS					
North Island					
In vicinity of known site	*DistSite * MAT*	*VPD*	*AcidP*	*DistSea*	*TopoPos*
No known site near	*DistSite Elevation Slope*	*VPD*	*MAT*DistSea*	*DistNavW*	*TopoPos*
South Island					
In vicinity of known site	*DistSite Slope*	*VPD*	*AcidP*	*MAT*DistSea*	*TopoPos*
No known site near	*DistSite Elevation Slope*	*VPD*	*MAT*DistSea*	*DistNavW*	*TopoPos*
PITS					
North Island					
In vicinity of known site	*DistSite Slope*	*MAT RtoPET*		*DistSea DistNavW*	
No known site near	*DistSite Slope*	*MAT RtoPET*	*AcidP*	*DistSea DistNavW*	
South Island					
In vicinity of known site	*DistSite Slope*	*MAT TSeas MAS*		*DistW Sand*	
No known site near	*DistSite Slope*	*MAT RtoPET*	*AcidP*	*DistW DistNavW*	

were chosen to provide a consistent set of variables over a set of models. *MAS* and *MAT* are an example, where *MAT* was chosen unless *MAS* was clearly better according to the effect criteria.

5. Interpreting the maps

Applying the model to each 100×100 m^2 cell generates a number, which is the relative likelihood of a site occurring in that cell. Each of the five categories of site requires one complete coverage of the country with these cells. The full detail of a likelihood for each cell can be accessed using ArcView software with the shape files prepared as an adjunct to this report and available from the authors. A broader view of the coverage is provided by the colour maps,

Appendix 2 (Figs A2.1–A2.5). Here, the colours display five categories within which sites are of broadly of equal likelihood. In the absence of any absolute measure of probability, these categories have to be somewhat arbitrary, but for convenience they have been given the following descriptive labels: Existing, Likely, Possible, Unlikely, and Negligible. Table 2 shows the numerical relative likelihoods corresponding to these categories.

TABLE 2. RELATIONSHIP BETWEEN RELATIVE PROBABILITIES AND MAP CATEGORIES FOR ARCHAEOLOGICAL SITES.

	EXISTING	LIKELY	POSSIBLE	UNLIKELY	NEGLIGIBLE
North Island sites	0.90–1	0.5–0.90	0.15–0.5	0.03–0.15	0–0.030
South Island sites	0.75–1	0.5–0.75	0.07–0.5	0.008–0.07	0–0.008
North Island middens	0.80–1	0.5–0.80	0.16–0.5	0.01–0.16	0–0.010
South Island middens	0.75–1	0.1–0.75	0.07–0.1	0.008–0.07	0–0.008
North Island pa	0.85–1	0.5–0.85	0.04–0.5	0.002–0.04	0–0.002
South Island pa	0.85–1	0.5–0.85	0.06–0.5	0.003–0.06	0–0.003
North Island pits	0.85–1	0.5–0.85	0.05–0.5	0.002–0.05	0–0.002
South Island pits	0.75–1	0.5–0.75	0.07–0.5	0.008–0.07	0–0.008

The *Negligible* category was chosen to give a category within which there were very few known sites. The *Existing* category generally indicates known sites are nearby, because this is the only environmental factor that raises the relative probability over 0.7. This effect explains the most prominent feature of the maps, a patch around a red dot.

In areas free of known sites, prediction is much less certain because the absolute probability of a site lying in a randomly selected 100×100 m^2 cell must be very small. There are 40 million 100×100 m^2 cells covering the North Island, but only 39 000 (less than 0.1%) contain known sites. Even if 10% of all known sites have so far been discovered (a low value to illustrate) there is still only a 1% chance of a random cell containing a site. As a broad estimate, the *Possible* category indicates absolute probabilities of this order of magnitude.

Where the predictive model works well, a greater proportion of the area is given a definite classification, *Negligible* or *Likely*. Where the model works less well, more area is classified as *Possible* or *Unlikely*. For example, Fig. A2.3 has much less dark blue and red than others because the model was not able to predict pa sites in the North Island. Pa in the South Island are generally so rare that the same level of unpredictability maps to the *Negligibl*e category. An estimate of the absolute probabilities corresponding to these categories could be built up over time by recording whether or not sites are found each time a map prediction is used.

6. Strength of predictions

Figure 1 compares the distribution of predicted probabilities between random sites and known sites. The plot for all sites in the North Island illustrates a very clear discrimination, with very high probabilities for the known sites and very small probabilities for the random sites. This plot exaggerates how well the discrimination would work on an unknown site, because the model was inevitably tuned to the known sites, and being unknown might itself give a site different properties. Nevertheless, the plots show that the model does perform well for all sites, particularly in the North Island, but also in the South Island.

Many of the plots for the known sites show a small increase in frequency at low probability. This is because the low probability environment is very much more common than the high probability environment. What matters for prediction are the relative frequencies, and low probability is always very much more frequent for random sites than known sites. Prediction is inherently more difficult for rarer objects, which explains why the North Island predictions are better than those for the South Island. Pa are also hard to predict for this reason, but as they tended to be more isolated from one another the strongest predictor, *DistSite*, was less powerful.

An intended application of this study was to provide a tool for warning when land development might encroach on a archaeologicalally important site. The study shows that the strongest predictor is a site already in the area. This is not a particularly surprising discovery, but at least the information is incorporated in the model. More usefully, the plots show large areas of the country where a site would be most unlikely, particularly in the North Island. However, for the South Island, where sites are rarer, a high relative probability will be a useful warning of the presence of a site, but a low probability will not provide quite as strong an indication of absence, particularly for pa and pits.

7. Conclusion

This analysis has produced maps (Figs A2.1–A2.5) which show the relative likelihood of a 100 × 100 m area of New Zealand containing a site of archaeological interest. The most likely areas tend to be those close to existing sites, but environmental variables do provide useful discrimination in areas where nothing of archaeological interest has yet been found. They also specify large tracts of country where sites are exceedingly unlikely to be found.

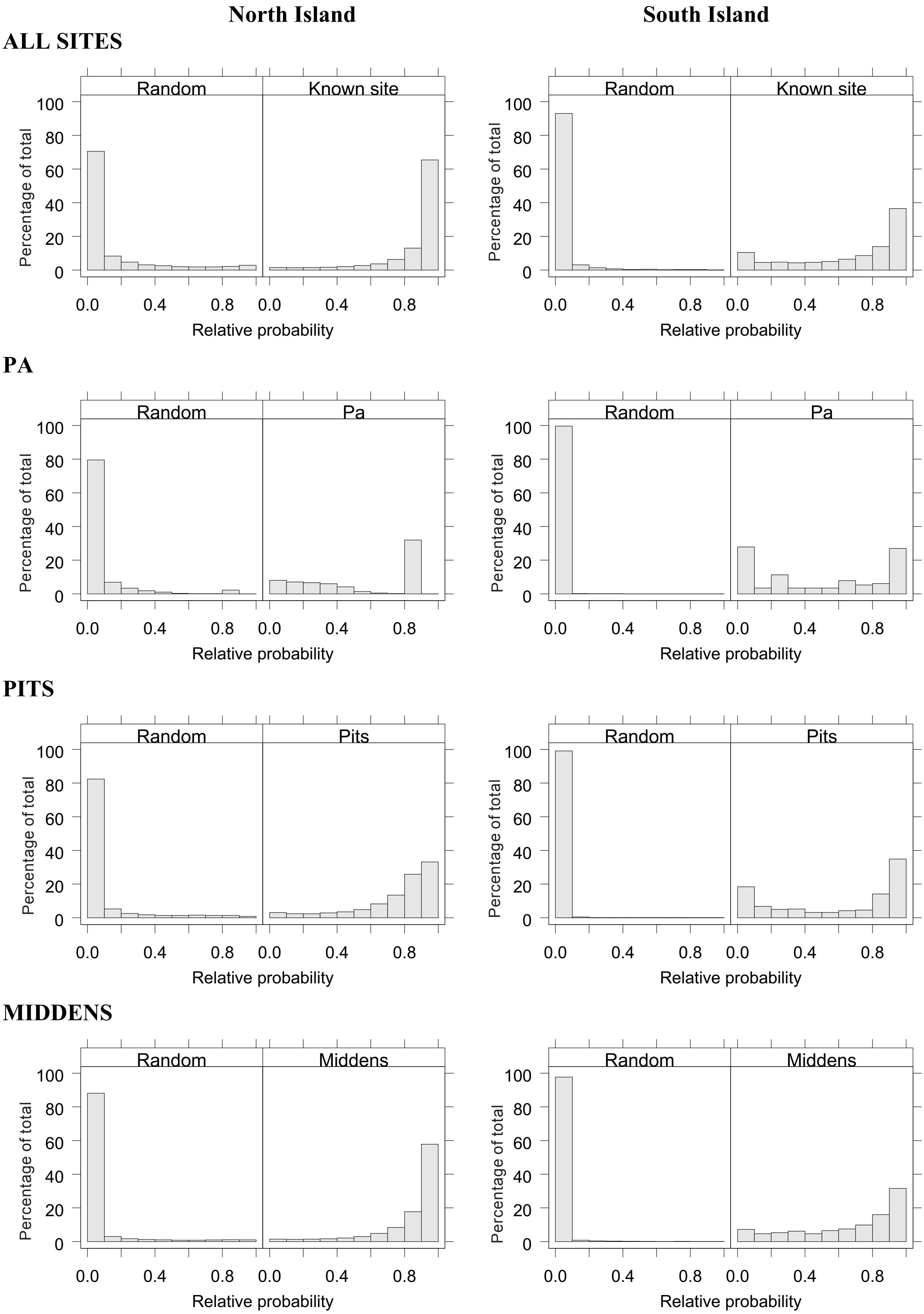


Figure 1. Distribution of predicted relative probabilities of occurrence of sites, for each site type.

8. References

Leathwick, J.R. 2000: Predictive models of archaeological site prediction in New Zealand. *Science & Research Internal Report 181*. Department of Conservation, Wellington. 12 p.

Leathwick, J.R.; Morgan, F.; Wilson, G.; Rutledge, D.; McLeod, M.; Johnston, K. 2003: Land environments of New Zealand: technical guide. Ministry for the Environment, Wellington. 237 p.

Venables, W.N. ; Ripley, B.D. 1999: Modern applied statistics with S-PLUS. (3rd edn). Springer-Verlag, New York. 501 p.

9. Acknowledgements

This report resulted from work done under DOC science investigation no. 3271.

Appendix 1

Details of the predictive models used for occurrence of archaeological sites in New Zealand

The variables are defined in Section 2.2. A few variables in this list were tried, but not found useful. Frequently variables were used in formulae, which are reproduced here as computer code for accuracy and brevity. These code terms are defined as:

NA: Variable not used for this prediction.

min(X, N): The minimum of X and N, which puts an upper bound on X.

max(X, N): The maximum of X and N, which puts a lower bound on X.

ifelse(L, X, Y): If L is true use X, otherwise use Y.

cos(X): The cosine of X, where X is in radians.

^ : Exponential. So X^2 is X squared.

All sites, North Island

VARIABLE	NEAR KNOWN SITE		WITHIN 4 km OF SEA		BEYOND 4 km OF SEA	
	Coefficient	*t* ratio	Coefficient	*t* ratio	Coefficient	*t* ratio
Constant	−0.6259	0.61	−0.6540	−0.27	6.704	8.04
DistSite	−0.1831	−3.36	−0.9503	−55.84	−0.9402	−78.3
min(DistSea, 4000)	−0.000114	−4.21	NA	NA	NA	NA
ifelse(DistSea > 4000, DistSea, 0)	−0.000002	−0.48	NA	NA	NA	NA
max(DistSea, 20000)	NA	NA	NA	NA	0.000002	1.72
max(1000, DistSea)	NA	NA	−0.001012	−16.41	NA	NA
Elevation	0.001428	2.17	NA	NA	−0.000991	−3.66
max(Elevation, 400)	NA	NA	0.000094	0.24	NA	NA
cos(Slope*pi/180)	NA	NA	NA	NA	5.065	7.724
min(130, MAT)	NA	NA	0.106588	5.83	NA	NA
MAT	0.04085	7.13	NA	NA	0.003094	0.83
TSeas	0.4384	3.08	NA	NA	0.2263	2.88
VPD	NA	NA	−0.006879	−1.60	NA	NA
AcidP	−0.001315	−0.04	0.004838	0.26	0.1259	6.53
(AcidP – 2.5)^2	−0.08493	−3.38	0.03295	2.05	0.004939	0.40
max(DistNavW, 20000)	NA	NA	0.000032	1.77	NA	NA
DistW	NA	NA	0.000039	1.75	−0.000031	−0.43
Sand	−0.02331	−0.63	NA	NA	NA	NA
TopoPos	NA	NA	−0.08476	−6.69	−0.06378	−6.00

All sites, South Island

VARIABLE	NEAR KNOWN SITE		WITHIN 4 km OF SEA		BEYOND 4 km OF SEA	
	Coefficient	*t* ratio	Coefficient	*t* ratio	Coefficient	*t* ratio
Constant	−51.193	−4.85	−4.993	−1.32	1.456	0.47
DistSite	NA	NA	−0.6110	−25.80	−0.7799	−44.60
DistSite > 10	−0.7275	−4.91	NA	NA	NA	NA
DistSite > 9.5	0.02820	0.18	NA	NA	NA	NA
DistSea	NA	NA	−0.00074	−16.77	NA	NA
max(DistSea, 40000)	−0.000067	−3.47	NA	NA	NA	NA
Elevation	NA	NA	NA	NA	−0.002034	−7.06
max(Elevation, 400)	0.04217	2.69	NA	NA	NA	NA
min(Elevation, 100)	NA	NA	−0.020032	−16.54	NA	NA
cos(Slope * pi / 180)	32.49	4.17	12.70	3.40	10.07	3.37
cos((Aspect + 28) * pi)/180)	NA	NA	NA	NA	0.2279	4.70
MAT	NA	NA	0.017652	4.88	−0.01882	−3.55
min(MAT, 120)	0.075319	6.38	NA	NA	NA	NA
TSeas	1.300	3.93	NA	NA	0.9666	5.76
TSeas^2	3.6967	6.31	NA	NA	−0.5691	−3.06
RtoPET	NA	NA	−0.005031	−3.40	NA	NA
AcidP	0.3497	3.18	NA	NA	NA	NA
DistNavW	NA	NA	−0.000003	−0.67	NA	NA
max(1000, DistNavW)	−0.000051	−2.97	NA	NA	NA	NA
(TopoPos – 4)^2	0.4458	4.30	NA	NA	NA	NA
TopoPos	0.1543	0.93	NA	NA	NA	NA
Rock	3.587	0.98	NA	NA	NA	NA

Pa sites, North Island

VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
Constant	−27.896	−7.40	−21.65	−5.80
DistSite	NA	NA	−0.8896	−44.04
min(DistSite, 9)	1.841	9.54	NA	NA
max(DistSite, 9)	−0.4788	−16.30	NA	NA
min(DistSea, 40000)	NA	NA	0.000023	8.14
min(Elevation, 300)	NA	NA	−0.002132	−5.36
cosSlope	15.961	4.73	25.38	6.90
ifelse(DistSea<4000,MAT,100)	NA	NA	0.03813	7.96
ifelse(DistSea^{3}4000, MAT,100)	NA	NA	0.02880	7.29
MAT	0.01702	6.20	NA	NA
VPD	−0.0229	−4.83	−0.03229	−6.50
max(AcidP, 3)	0.1749	3.43	0.4758	10.44
min(DistNavW, 5000)	NA	NA	−0.000044	−3.07
max(TopoPos, 5)	−0.1681	−2.98	−0.2049	−4.00
ifelse(Sand & DistSea< 4000, 1, 0)	−0.1545	−2.72	0.07147	0.91

Pa sites, South Island

VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
Constant	−1415	−3.33	4.5818	0.46
min(DistSite, 9)	31.620	0.81	NA	NA
DistSite	NA	NA	−0.6969	−5.24
max(DistSite, 9)	−1.315	−3.84	NA	NA
min(DistSea, 40000)	−0.000269	−2.46	0.000175	2.89
(min(DistSea,20000)−10000)* min(MAT,140)	NA	NA	−0.000003	−3.14
min(Elevation, 800)	NA	NA	−0.01132	−3.19
max(0, Slope − 10)	−0.07171	−1.34	NA	NA
cos((min(Slope − 5, 5) * pi)/180)	1120	4.69	NA	NA
max(MAT, 100)	NA	NA	0.04840	1.59
min(MAT, 100)	NA	NA	−0.03582	−0.41
TSeas	1.742	1.81	NA	NA
TSeas^2	3.284	2.79	NA	NA
MAT	0.1658	4.62	NA	NA
VPD	NA	NA	0.05186	2.24
min(50, VPD)	0.2089	4.71	NA	NA
min(DistNavW, 25000)	NA	NA	0.00005	1.88
max(TopoPos, 4)	NA	NA	−1.466	−1.81
min(TopoPos, 3)	−0.9843	−3.31	NA	NA
ifelse(Sand & DistSea< 4000, 1, 0)	−0.5952	−1.21	−1.163	−2.76

Pit sites, North Island

VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
(Intercept)	−77.57	−14.96	−71.65	−11.95
DistSite	NA	NA	−1.057	−44.24
min(DistSite, 9)	1.1021	11.57	NA	NA
max(Dist, 9)	−0.8299	−31.36	NA	NA
min(DistSea, 40000)	NA	NA	−0.000022	−8.25
min(DistSea, 80000)	−0.000018	−10.70	NA	NA
max(Slope, 15)	NA	NA	−0.04811	−3.56
cos((min(Slope – 10, 5) * pi)/180)	85.42	16.62	95.83	16.30
max(MAT, 140)	−0.03062	−6.39	−0.06809	−9.38
min(MAT, 140)	NA	NA	0.01638	3.62
min(60, RtoPET)	−0.02771	−6.14	−0.0443	−9.32
max(AcidP, 1)	NA	NA	0.1399	5.86
max(DistNavW, 25000)	−0.000070	−2.47	−0.000057	−1.55

Pit sites, South Island

VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
(Intercept)	−549.8	−8.28	−122.5	−6.55
DistSite	NA	NA	−0.7452	−12.52
max(DistSite, 9.6)	−0.9658	−6.09	NA	NA
cos((min(Slope – 10, 5) *pi)/180)	NA	NA	118.8	6.37
cos((min(Slope – 7, 7)* pi)/180)	567.4	8.48	NA	NA
AcidP	NA	NA	0.1098	1.32
MAT	NA	NA	0.1272	12.74
max(MAT, 100)	0.1615	3.97	NA	NA
max(MAT, 120)	−0.5263	−6.60	NA	NA
min(MAS, 140)	0.2873	5.82	NA	NA
min(−0.1, TSeas)	−3.458	−7.61	NA	NA
RtoPET	NA	NA	−0.006955	−1.39
DistW	0.000777	3.05	NA	NA
min(DistW, 1000)	NA	NA	0.00076	2.47
max(DistNavW, 25000)	NA	NA	−0.000138	−2.90
ifelse(Sand & DistBch <4000, 1, 0)	0.5760	1.91	NA	NA

Midden sites, North Island

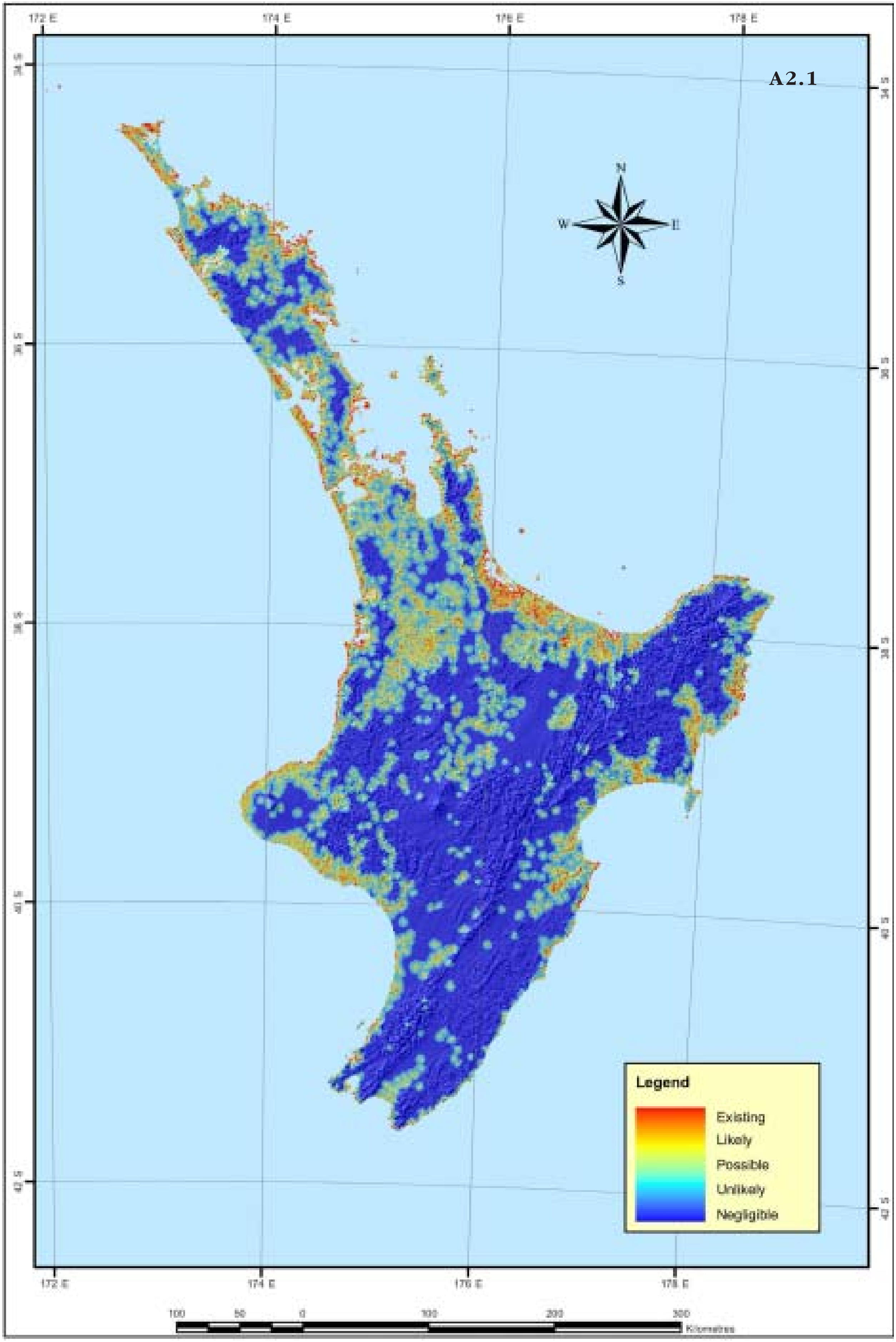
VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
Constant	−1.399	−1.01	0.1301	0.03
DistSite	NA	NA	−1.105	−29.34
max(DistSite, 9)	−0.927969	−32.83	NA	NA
min(DistSite, 9)	0.592326	7.92	NA	NA
min(DistSea, 80000)	0.000334	8.69	0.000368	13.97
(min(DistSea,80000)−20000)* min(MAT,140)	−0.000003	−11.15	−0.000003	−17.95
Elevation	NA	NA	−0.006796	−9.22
cos(((Slope − 5) * pi)/180)	NA	NA	11.96	2.64
min(MAT, 140)	0.03152	3.84	−0.02508	−2.77
min(−0.1, TSeas)	−3.458	−7.61	NA	NA
min(MAS, 140)	0.2873	5.82	NA	NA
max(40, VPD)	−0.0885	−10.32	−0.08828	−6.52
max(AcidP, 3)	−0.5810	−9.19	NA	NA
min(DistNavW, 40000)	NA	NA	−0.000008	−1.09
min(DistW, 1000)	NA	NA	0.00076	2.47
TopoPos	−0.1142	−7.45	−0.1169	−5.94

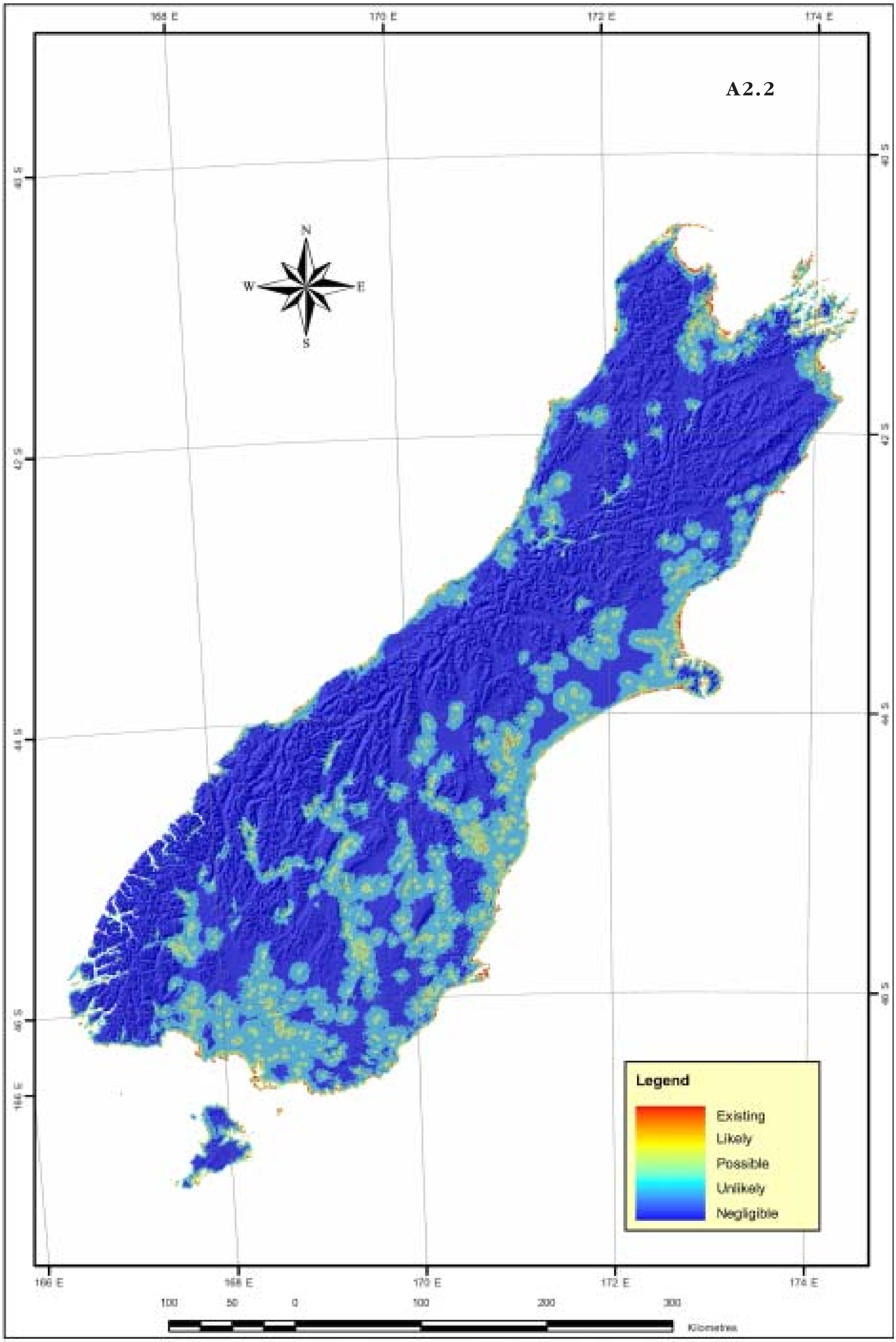
Midden sites, South Island

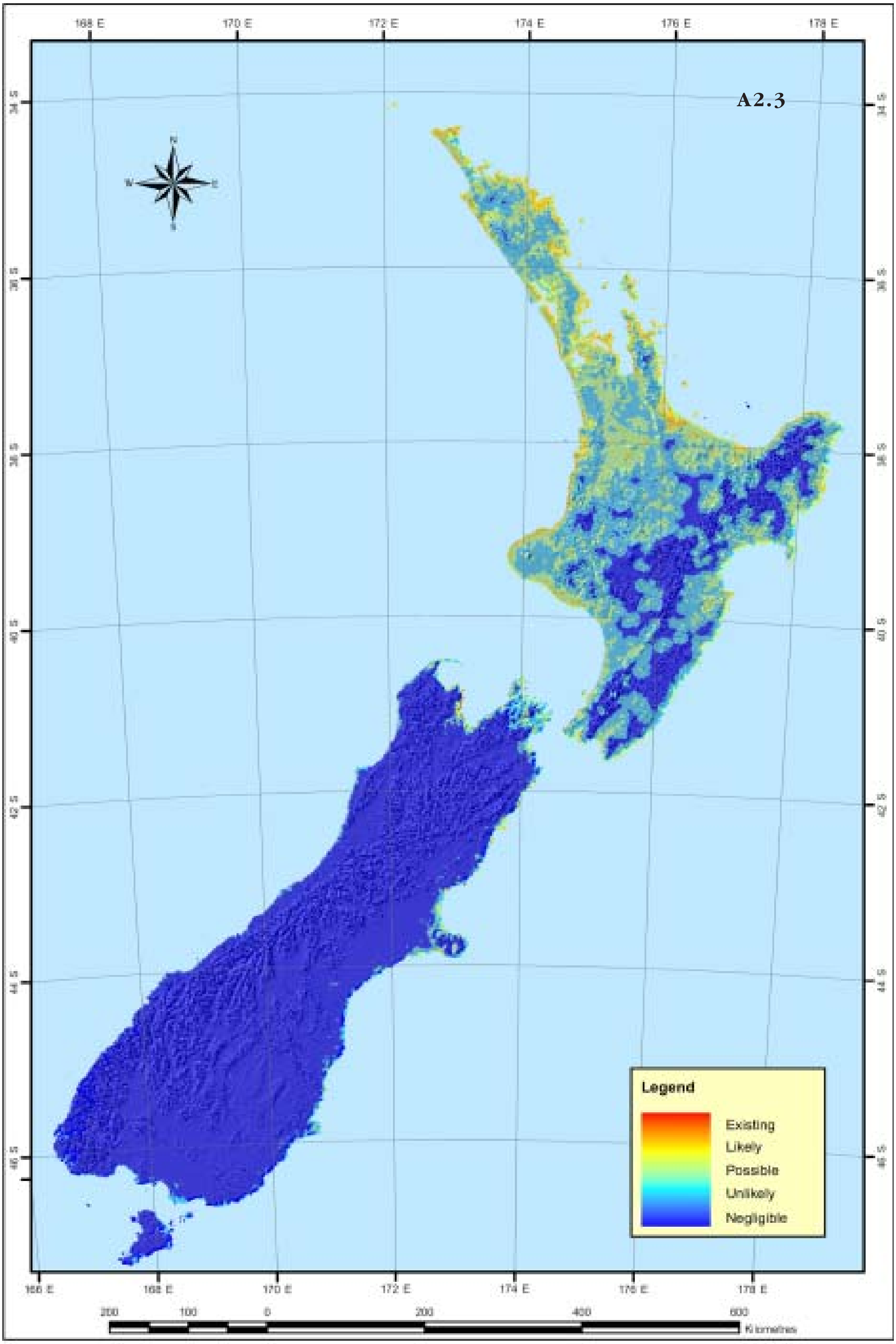
VARIABLE	NEAR KNOWN SITE		DISTANT FROM KNOWN SITE	
	Coefficient	*t* ratio	Coefficient	*t* ratio
(Intercept)	−414.049	−1.88	−16.037	−2.62
min(DistSite, 9)	46.615368	1.90	NA	NA
max(DistSite, 9)	−0.900220	−10.50	NA	NA
DistSite	NA	NA	−0.723707	−18.97
min(DistSea, 60000)	0.000335	4.60	NA	NA
min(DistSea, 40000)	NA	NA	0.000797	13.62
min(DistSea,60000)−20000)* min(MAT,120)	−0.000004	−5.25	NA	NA
min(DistSea,40000)−20000)* min(MAT,140)	NA	NA	−0.000009	−15.31
Elevation	NA	NA	−0.013863	−11.69
Slope	−0.101470	−8.70	NA	NA
cos(((Slope − 5) * pi)/180)	NA	NA	22.309353	3.71
min(MAT, 120)	−0.012675	−0.74	−0.164360	−11.37
max(30, VPD)	−0.020053	−1.91	NA	NA
max(40, VPD)	NA	NA	0.034333	2.57
max(AcidP, 3)	−0.347447	−1.86	NA	NA
min(DistNavW, 40000)	NA	NA	0.000011	1.56
TopoPos	−0.000734	−0.02	0.155437	5.16

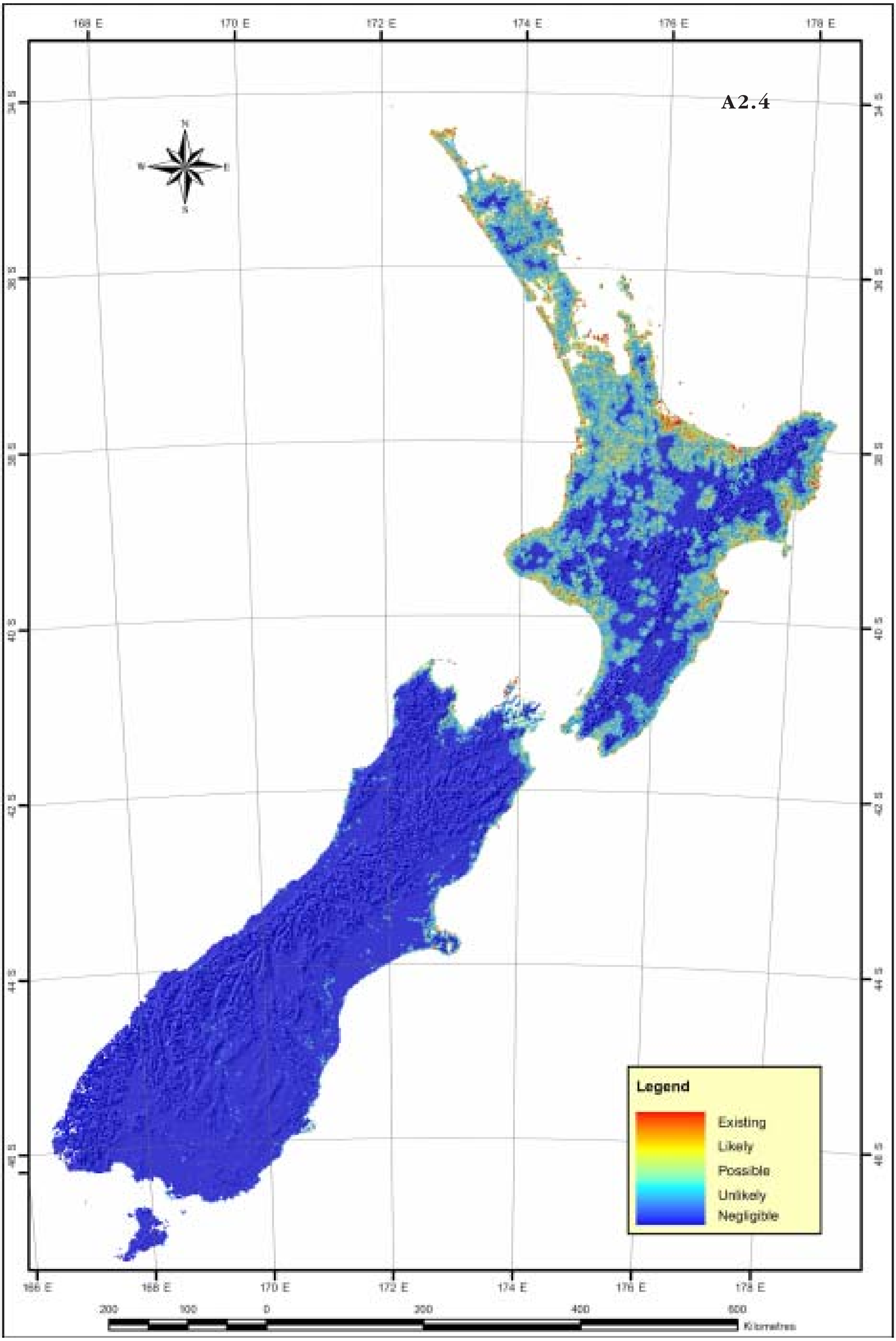
Appendix 2

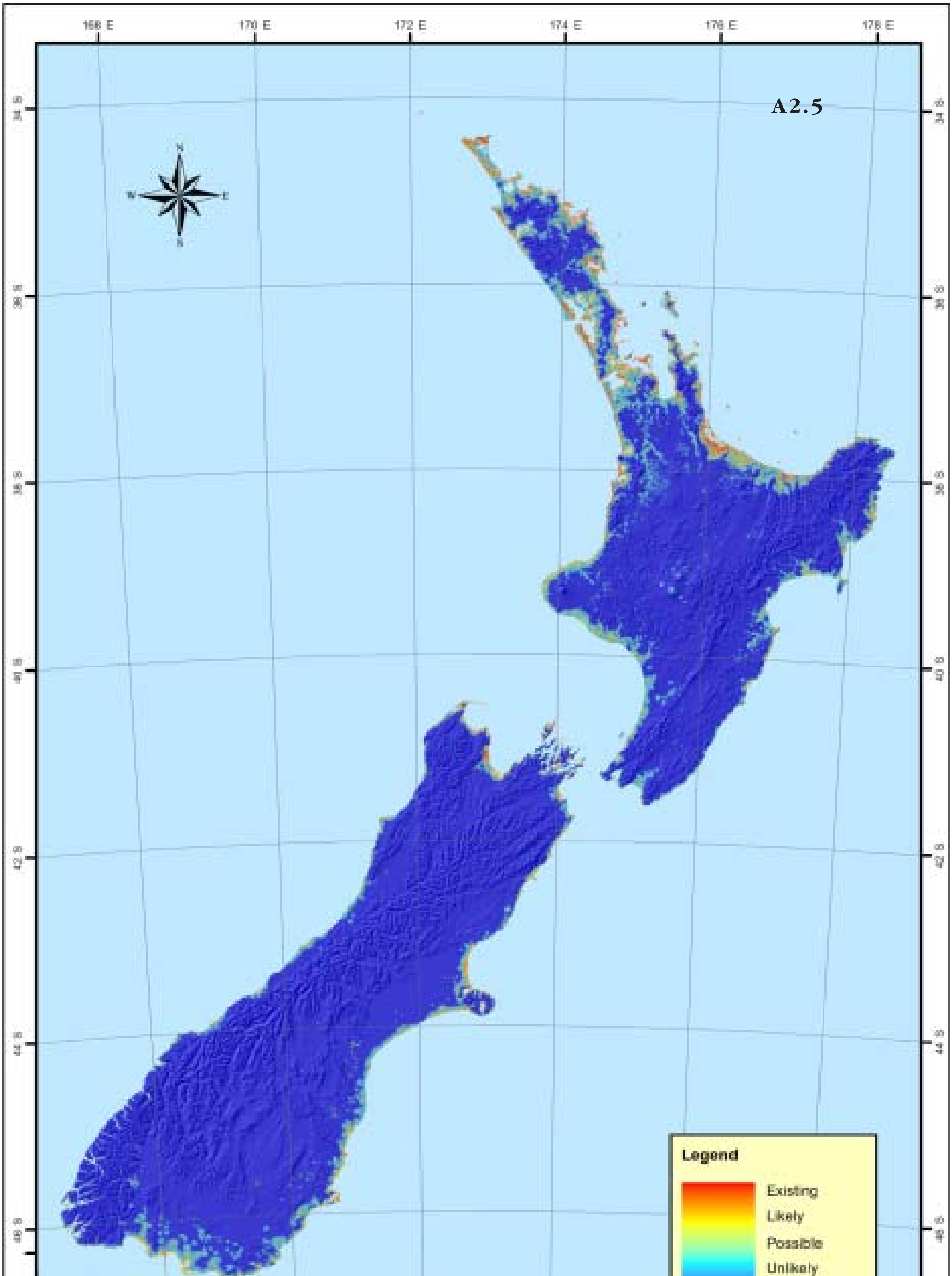
Maps of relative probabilities of occurrence of archaeological sites in New Zealand


A2.1
172 E
174 E
176 E
178 E
34 S
36 S
38 S
40 S
42 S
N
W
E
S
Legend
Existing
Likely
Possible
Unlikely
Negligible
100
50
0
100
200
300
Kilometres


A2.2
N
S
E
W
Legend
Existing
Likely
Possible
Unlikely
Negligible
Kilometres


A2.3
Legend
Existing
Likely
Possible
Unlikely
Negligible
Kilometres


A2.4
Legend
Existing
Likely
Possible
Unlikely
Negligible
Kilometres


A2.5
168 E
170 E
172 E
174 E
176 E
178 E
166 E
Legend
Existing
Likely
Possible
Unlikely
Negligible
200
100
0
200
400
600
Kilometres