Predictive modelling of long-tailed bat distribution in the Hamilton area

Kathy Crewther
Stuart Parsons

Walkingbats Consultancy
www.walkingbats.com

August 2017

Table of Contents

1 Introduction

1.1 Background

A number of acoustic surveys of long-tailed bats (*Chalinolobus tuberculatus*) have been carried out in the Hamilton City area from September 2011 to June 2017 including a comprehensive city-wide survey carried out by Kessels Ecology from September 2011-January 2012 which was the subject of an earlier Project Echo report (Le Roux & Le Roux, 2012). A list of data sources is provided in Appendix I (Table 2). These acoustic surveys have resulted in the availability of presence and absence data, enabling modelling of long-tailed bat distribution.

1.2 Aims

To develop a predictive model of habitats likely to host bats, based on acoustic survey results and environmental variables that influence presence, resulting in:

- the identification of sites likely to host bats;
- a model that can be extended to the Waikato region once acoustic surveys across a range of environments representative of the region have been carried out.

2 Methods

2.1 The model

A species distribution model was created using MaxEnt software. MaxEnt is a statistical learning method that combines species presence data with environmental variables to predict species distribution based on the relationship between where the species occurs and the environments in which it is found (Phillips, Anderson, & Schapire, 2006). It has been used widely since becoming available in 2004 and is considered to have high predictive ability (Elith et al., 2011).

The model calculates the probability distribution across a defined area using the input data to both train and test the model across a number of iterations (Yost, Petersen, Gregg, & Miller, 2008). The contribution of individual variables is measured by jackknife tests which calculate the influence a variable has on the model when used in isolation or when removed. Variables that decrease the predictive ability of the model can be removed to improve the model's performance (Yost et al., 2008). The primary output of the model is a habitat suitability map estimating the probability of presence in each individual cell or pixel on a scale of 0 - 1 (Barnhart & Gillam, 2014).

2.2 Boundary

Initially, the model was trialled using a boundary or "mask" based on a 10 km radius of existing survey data. However, because the surrounding area did not represent the environment typically surveyed, the model did not perform well. The mask was subsequently reduced to a 2.5 km buffer around the existing survey data. The mask was then applied to all environmental variables to ensure that the model made its predictions based only on the area of interest (Figure 1).

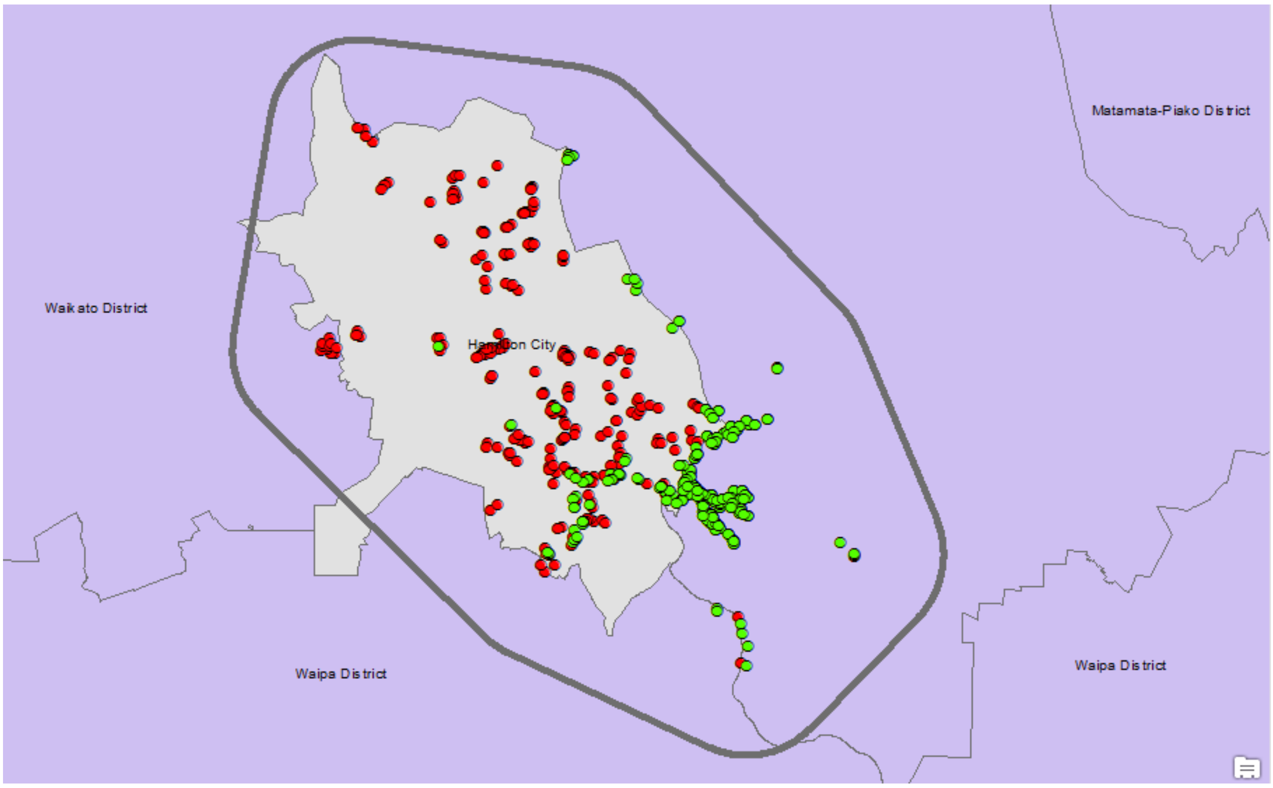


Figure 1. Black line indicates the boundary of the mask created for use by the model based on a 2.5 km buffer around the survey data (green dots = presence data; red dots = absence data). Grey area represents Hamilton City. District boundary layer courtesy of Waikato Regional Council.

2.3 Scale

The cell size (spatial accuracy) selected for the model was 50 m^2, a relatively fine-scale resolution since the area being modelled is comparatively small.

2.4 Variables

A number of potentially predictive environmental variables were trialled in the model (Appendix I). The model takes each variable in turn and evaluates how useful it is on its own, and in conjunction with other variables, in predicting bat presence by using some presence data to train the model and some to test the model. The usefulness of individual variables is then measured by jackknife tests which outline each variable's contribution to training, testing and the overall predictive ability (AUC value) of the model. To establish which variables to use, different permutations of the available variables were trialled 38 times, and the five variables selected were those that were most useful to the model, based on the jackknife tests. The five variables are:

- distance to gullies
- distance to street lighting
- distance to residential areas

- distance to vegetation
- land cover type.

It was decided not to use climatic variables in the model because of the small area being studied. Once mapped in ArcGIS, it became clear that temperature and rainfall across the area did not vary sufficiently to warrant inclusion. In general, climatic variables are only useful if there are major shifts across the relevant area, such as may occur if the area being modelled includes different biomes or major altitudinal changes.

2.5 Data

Survey data included 649 data points, of which 373 recorded bat presence and 276 failed to detect bats. Thirty-five absence records were removed because no geographic coordinates were provided. Two absence data points were removed because their geographic coordinates were shown to be inaccurate when mapped in ArcGIS. Two duplicate presence data points in Hammond Park were removed. In total, this left 610 survey points – 371 presence and 239 absence. However, because a number of presence records had identical coordinates or were located in the same 50 m cell on the map, the model only used 160 presence data points.

Absence data points are not used by MaxEnt, but were mapped in ArcGIS since they can be useful in monitoring any shifts in species distribution over time. Therefore, of the 649 data points provided, it was only the 160 presence data points referred to above that were used to predict bat habitat use within the area of the mask.

Following selection of the variables and input of the presence data, 20 multiple runs (‘replicates’) of the model were carried out using randomly selected presence data points for each run. Replicate runs of the same model potentially allow for it to be more robust as the final model output represents the average findings across all replicates. This means it is less likely to be skewed than if the model is run only once or twice.

The model must reach convergence to maximise its predictive ability and this is done through a series of iterations which take place for every replicate run (Young, Carter, & Evangelista, 2011). The default setting for the maximum number of iterations was therefore changed from 500 to 5000 to allow for this although, in this case, the model only needed to carry out 500-780 iterations over the 20 replicates to reach convergence. Therefore, for each of the 20 replicate runs, the model also ran an average of 597 iterations. The model was run using cross-validation, which randomly splits presence data into equal groups or “folds”. Each iteration of the model is produced leaving out one fold at a time. The left-out fold is then used to evaluate the model. In this way, MaxEnt uses all of the presence data available for validation (Phillips, 2005). Full details of the parameters used to run the model can be found in Appendix I.

3 Results

Habitat suitability maps representing minimum, mean and maximum probability of bat presence were produced based on the average result of the 20 replicate runs of the model. The map representing maximum probability was selected to maximise opportunities for further surveying (Figure 2).

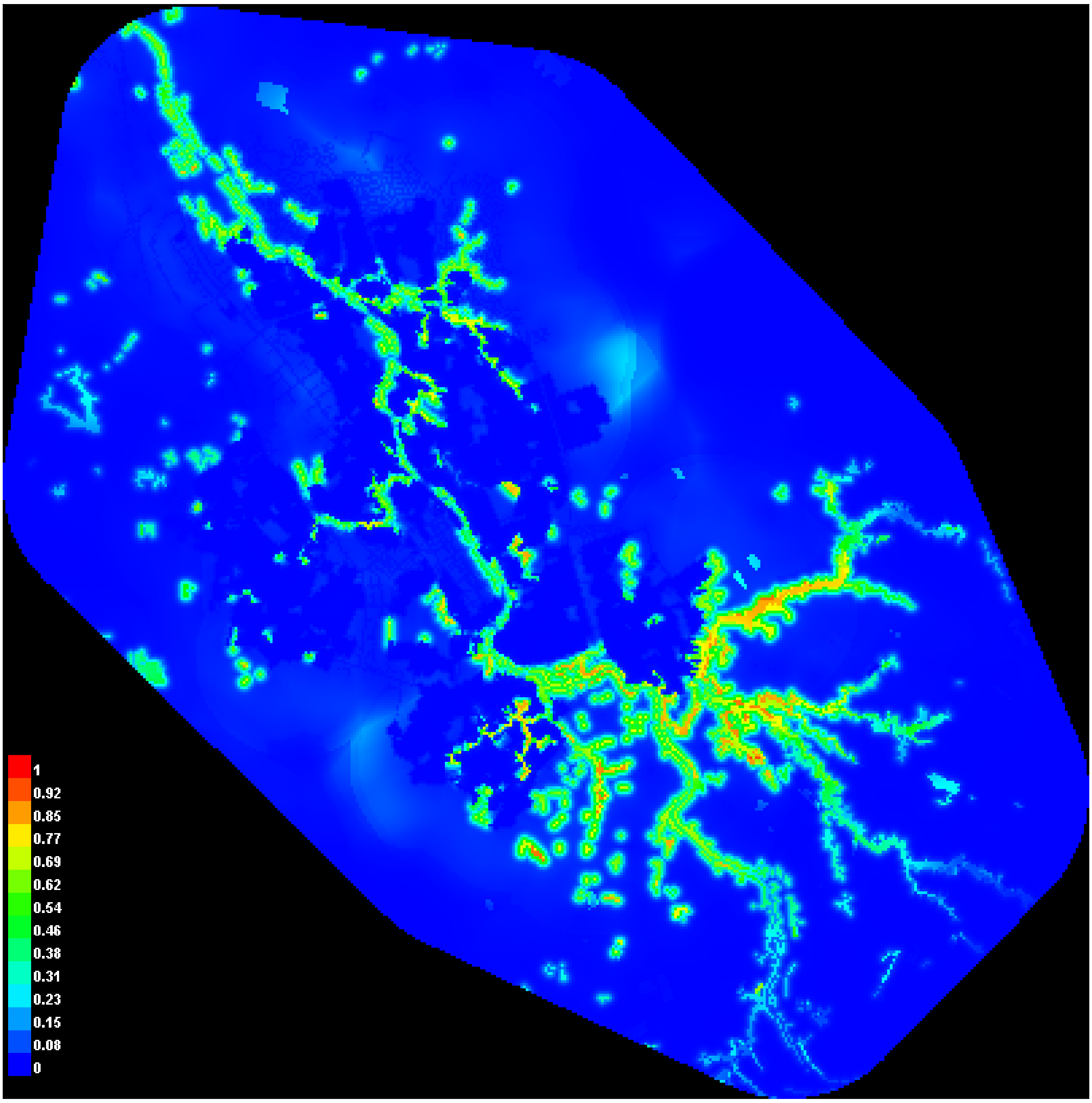


Figure 2: MaxEnt model output based on maximum suitability for *Chalinolobus tuberculatus*. Colours represent a range of pixel values assigned by the model to each 50 m^2 grid cell, on a continuous scale from 0.00-1.00, where 0.00 indicates presence of bats is highly unlikely and 1.00 indicates it is highly likely.

Model output indicated 2.7% of habitat within the mask area as potentially suitable (i.e. having a pixel value of 0.5 of higher) for long-tailed bats, while 97.3% was considered low suitability (Figure 3).

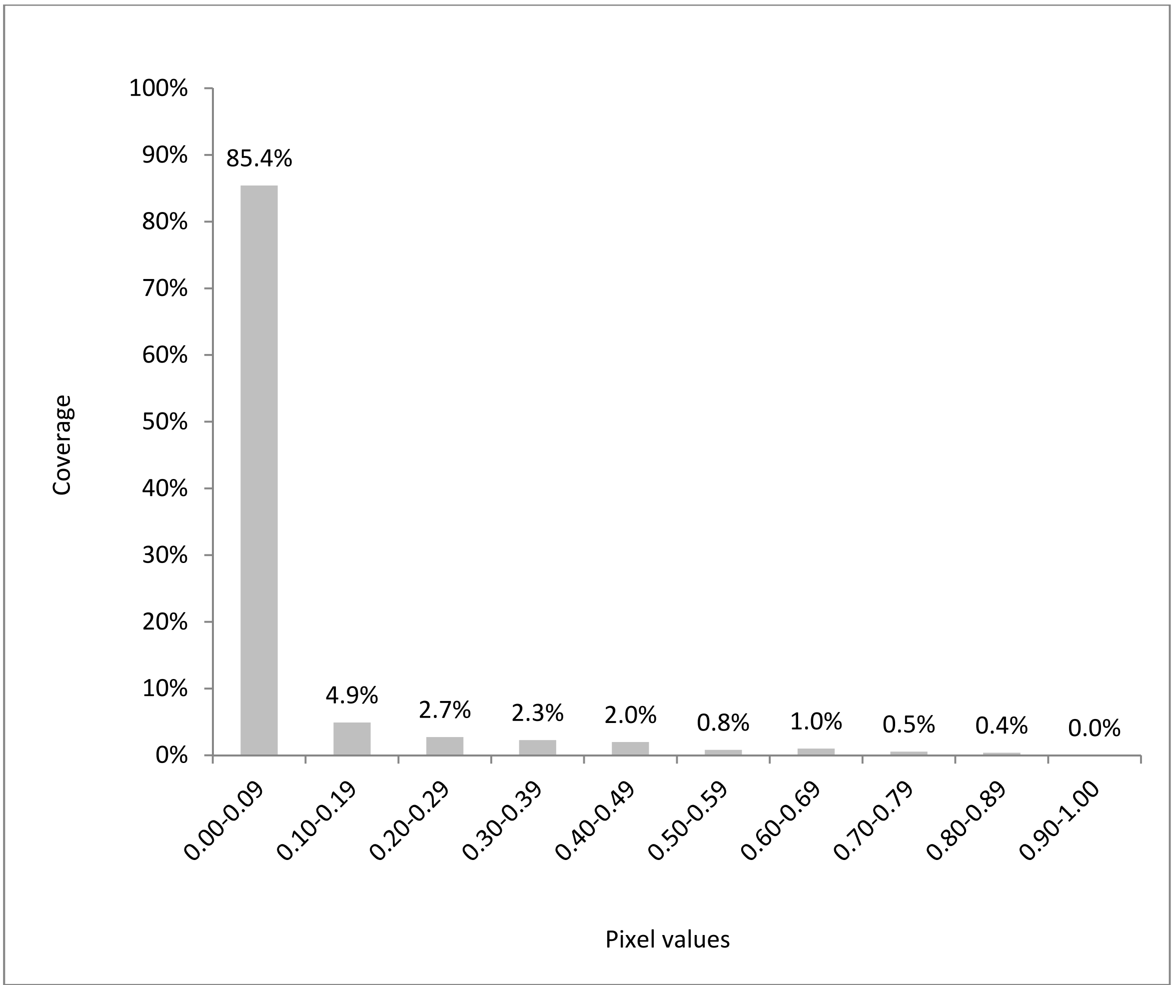


Figure 3. Habitat suitability for *Chalinolobus tuberculatus* in the mask area by percentage of coverage, according to pixel value. Pixel values of 0.00-0.49 indicate low suitability while values of 0.50-1.00 indicate medium-high suitability. The model deemed only 2.7% of the mask area as potentially suitable for long-tailed bats, i.e. having a value of 0.50 and higher.

When the pixel values representing probability of presence are compared to actual survey data locations, it can be seen that the model correctly predicts 60% of absences in the low suitability pixel range (0.00-0.49) and 77% of presences in the medium-high suitability range (0.50-0.99) (Table 1).

Table 1. Percentage of detections based on pixel value allocated by model to habitat suitability map (0 = least suitable habitat; 1 = most suitable habitat)

Pixel value	Presence	Absence	Total	% Presences	% Absences
0.00-0.49	108	162	270	40%	60%
0.50-0.99	263	77	340	77%	23%

Predictive ability increases at the extreme suitability values, with the model correctly predicting presence in 91% of cells with a suitability value of 0.80-0.89, and correctly predicting absence in 82% of cells with a suitability value of 0.00-0.09 (Figure 4).

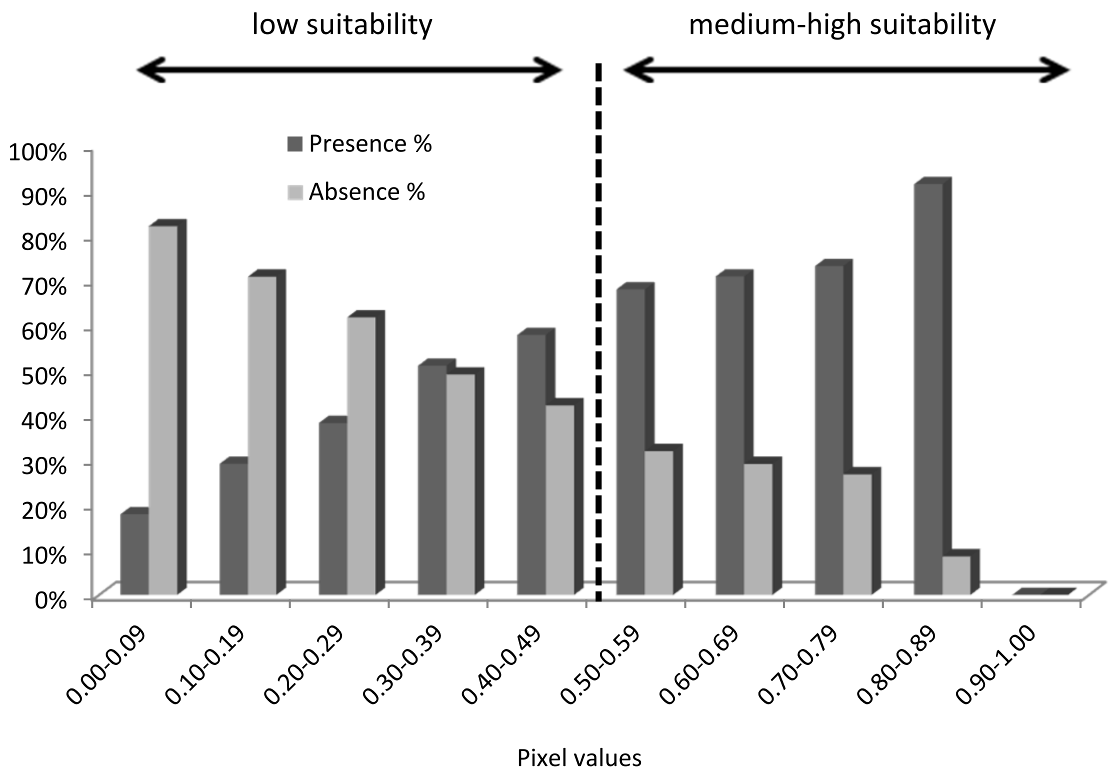


Figure 4. Pixel values allocated by the model compared against cells containing presence and absence data recorded from surveys ($<$0.49 = low habitat suitability; $\geq$0.50 = medium to high habitat suitability).

The model’s predictions included only three pixels within the 0.90-1.00 range and these were in close proximity to presence data points (Figure 5).

Figure 5. Location of cells with pixel value of 0.90-1.00 near Vintners Lane, Tamahere shown against (a) imagery basemap and (b) street name basemap. Inset shows cells in relation to mask boundary (black line) and Hamilton City boundaries (dark grey area).

Jackknife tests of variable importance were produced based on the variables' contributions to training, testing and predictive ability of the model (Figure 6). A probability distribution is generated by the model across all cells or pixels in the area being modelled, starting with a uniform distribution (represented by zero). A number of iterations are carried out using the data provided which increase the probability or gain of sample locations. The resulting figure or "regularised training gain" arrived at by the model is a measure of how well the distribution model fits the data compared to uniform distribution. The exponential of the regularised training gain indicates how many times higher the average sample likelihood is compared to a random pixel (Yost et al., 2008). A good predictive model should perform better than random and, in this case, the regularised training gain was 2.0256 ($e^{2.0256} \approx 7.58$), indicating that the likelihood of the model distribution was 7.58 times higher than random uniform distribution (Figure 6a). The model also evaluates itself by comparing its probability distribution against random distribution using an AUC value, where a value of less than 0.50 indicates that the predictive power is worse than random, 0.50 equals random and a value of greater than 0.50 indicates higher predictive power (Phillips et al., 2006). Therefore, the higher the AUC value, the higher its predictive power is considered to be. In this case, the AUC value was very high at 0.95 (Figure 6c).

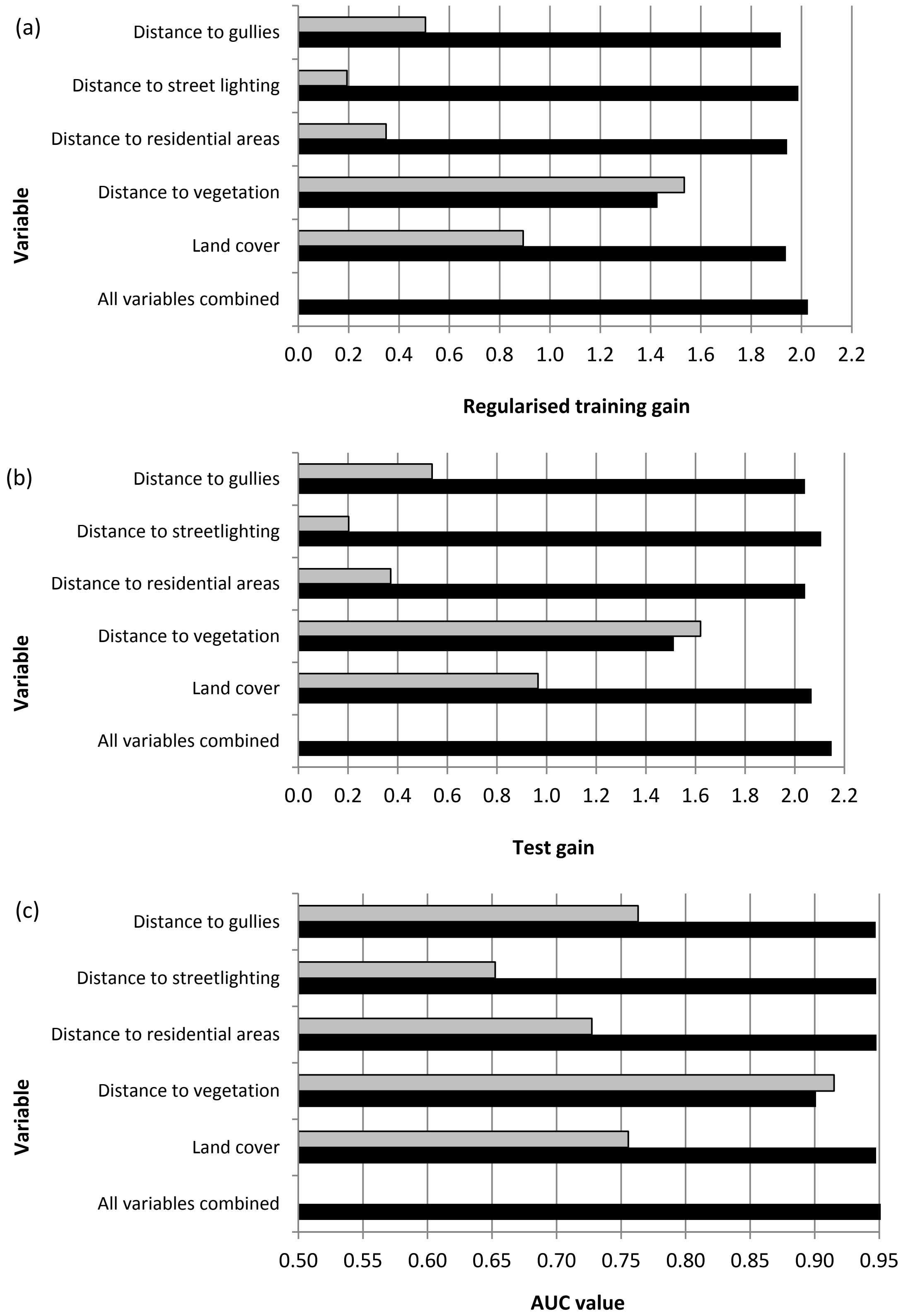


Figure 6. Results of jackknife tests in MaxEnt model identifying the contribution of each variable to (a) training, (b) testing, and (c) AUC value of the model. Values represent the average over twenty replicate runs. Black bars indicate the gain when the variable is used on its own and grey bars indicate the drop in gain when it is removed. Distance to vegetation contributed the most to the training, testing and predictive ability (AUC value) of the model and also resulted in the biggest drop in gain when removed.

The most important variable was distance to vegetation. Jackknife tests show that distance to vegetation resulted in the highest training gain (1.53) and the biggest drop when removed (gain reduced from 2.02 to 1.53) (Figure 6a). The overall test gain was 2.15 with distance to vegetation once again resulting in the highest test gain (1.51), and the largest drop when removed (2.15 down to 1.62) (Figure 6b). The jackknife test on the AUC using test data found that distance to vegetation also contributed the most (0.90) to the predictive ability of the model as measured by the AUC value of 0.95 (Figure 6c). The response curve for distance to vegetation found that probability of bat presence was highest (0.70) with zero distance to vegetation and fell sharply to less than 0.1 at a distance of approximately 125 m (Figure 7).

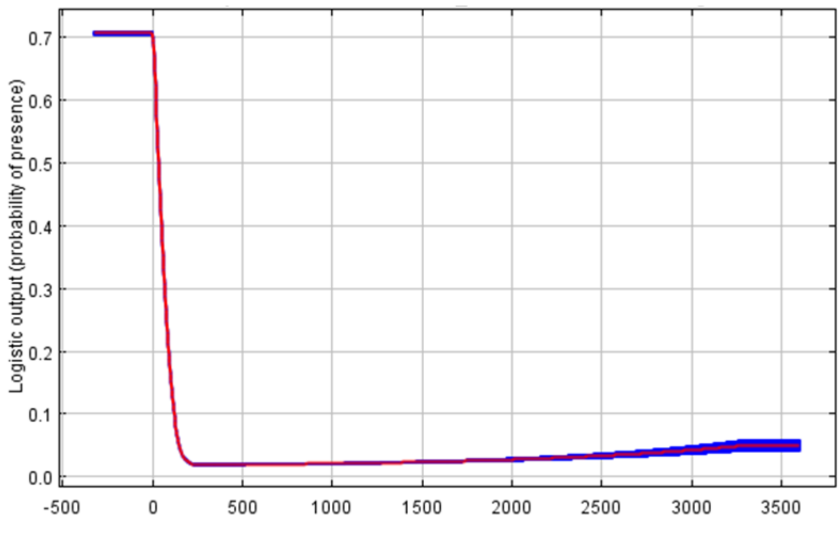
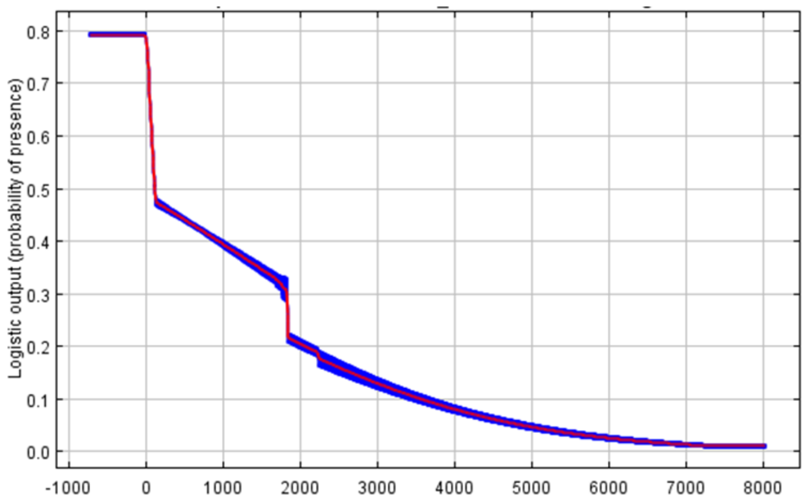


Figure 7. MaxEnt response curve for the distance to vegetation layer. The red line indicates the mean response of the 20 replicate runs carried out by the model while the blue area represents +/- one standard deviation. The Euclidean (straight line) distance shows that the probability of bat presence is highest (0.70) with zero distance to vegetation and drops sharply to less than 0.1 at a distance of approximately 125 m.

Distance to gullies was the second most important variable in terms of its contribution to the AUC value. The gullies layer, which was supplied by Hamilton City Council, only extended to the boundaries of the Hamilton City area and not to the southern boundary of the buffer. Therefore, the effects of gullies in relation to the most southerly presence data points have not been recognised by the model. However, it was considered worthwhile to retain distance to gullies as its contribution to the AUC value, or predictive ability, of the model was second only to the contribution made by distance to vegetation. Although both variables are correlated (vegetation lines the gullies), removing gullies slightly lowered the predictive ability of the model. The response curve for distance to gullies found that probability of bat presence was highest (0.80) with zero distance to gullies and fell sharply to less than 0.5 at a distance of approximately 100 m (Figure 8).



Euclidean (straight line) distance (m)

Figure 8. Response curve for the distance to gullies layer. The red line indicates the mean response of the 20 replicate runs carried out by the model while the blue area represents +/- one standard deviation. The Euclidean (straight line) distance shows that the probability of bat presence is highest with zero distance to gullies.

Land cover type was the third most important variable in terms of the AUC value, but the second most important in terms of training and testing the model. The response curve generated for land cover showed that when this variable was used in isolation, four types predicted greater probability of presence. These were indigenous forest (0.88), broadleaved indigenous hardwoods (0.83), gorse/broom (0.71) and manuka/kanuka (0.62). Other land cover types contributed 0.15-0.53 (Figure 9).

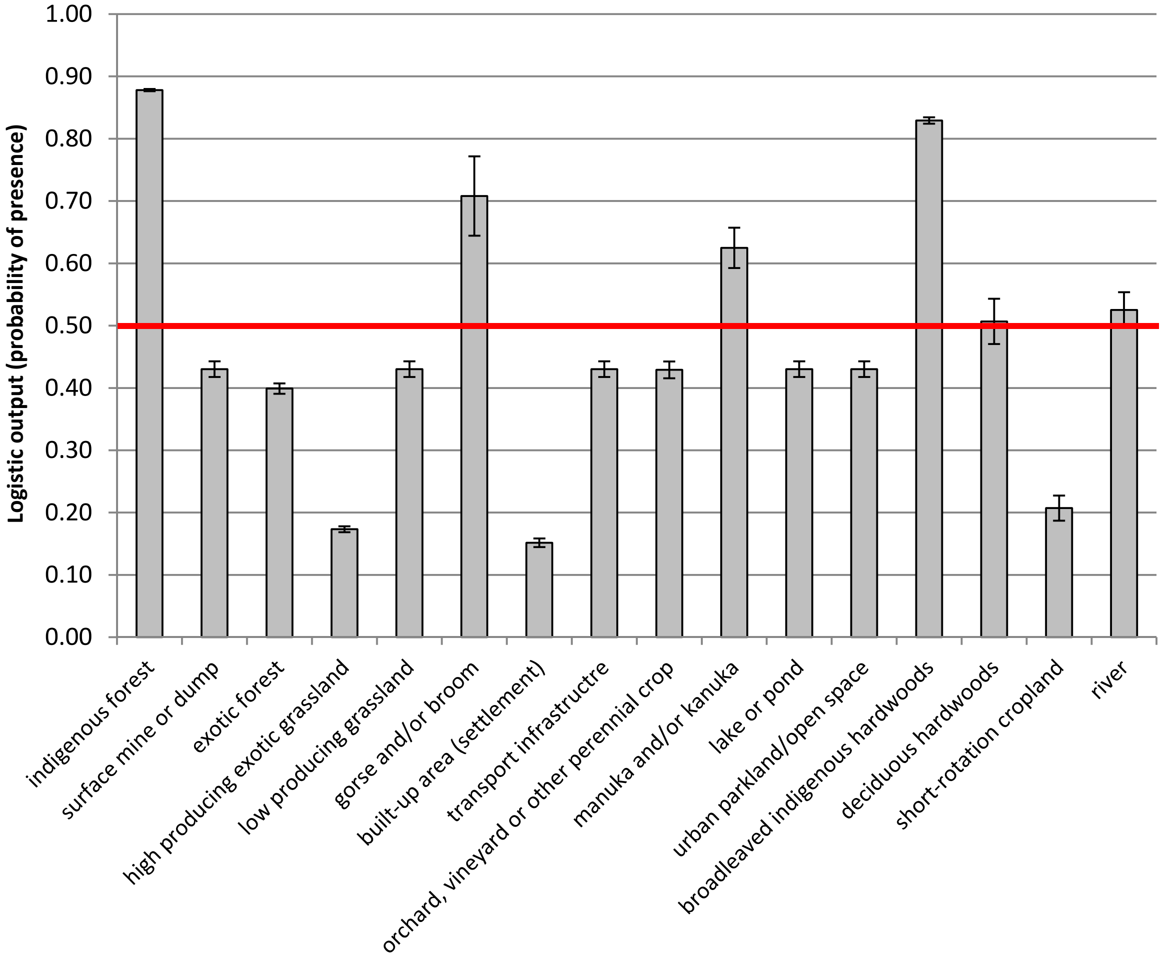


Figure 9. The effect of different land cover categories on a MaxEnt species distribution model's logistic prediction of long-tailed bats (*Chalinolobus tuberculatus*) in the mask area when the land cover variable is the only variable used. Probability of presence is estimated over a scale of 0.00-1.00 with higher numbers indicating higher likelihood of presence and is the mean response over 20 replicate runs +/- one standard deviation. Red line indicates the threshold at which bat presence becomes more likely (0.50+). The model found indigenous forest to be the greatest predictor of bat presence when land cover was considered on its own (i.e. when other variables were excluded).

Distance to residential areas was the fourth most important variable to the model, although its removal resulted in losses to training and testing that were similar to those when distance to gullies was removed. The response curve for distance to residential areas indicated that the probability of bat presence was highest (0.64) at close to zero distance from residential areas, with probability of presence decreasing as distance to residential areas increased (Figure 10).

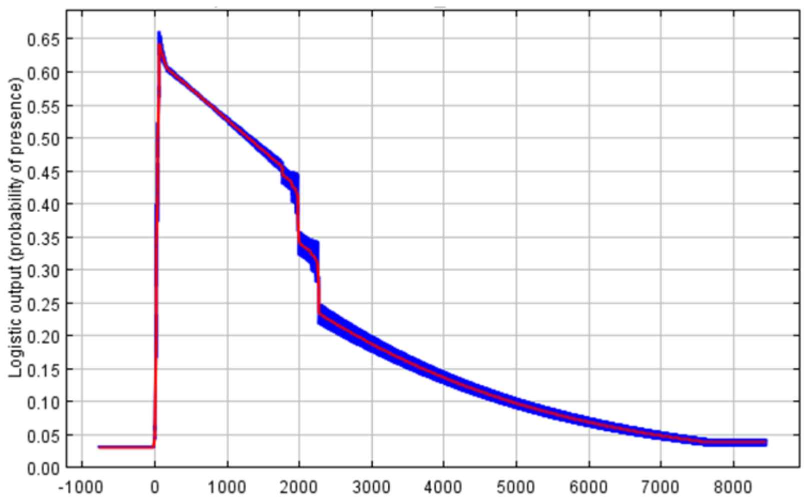


Figure 10. Response curve for the distance to residential areas layer. The red line indicates the mean response of the 20 replicate runs carried out by the model while the blue area represents +/- one standard deviation. The Euclidean (straight line) distance shows that the probability of bat presence is highest close to residential areas and decreases as distance to residential areas increases.

Distance to street lighting was the least important variable, however its removal resulted in some loss to the training and testing of the model. The response curve for distance to street lighting showed the probability of bat presence increasing with a distance of approximately 100 m from street lights to a maximum of 0.57 and then decreasing as distance from street lighting increased (Figure 11).

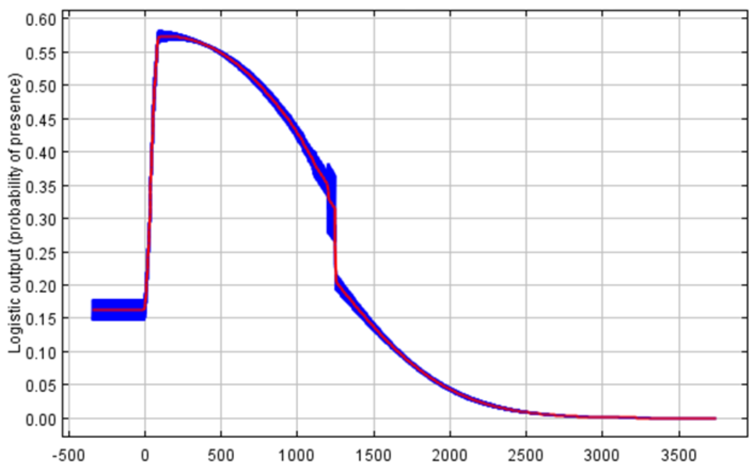


Figure 11. Response curve for the distance to street lighting layer. The red line indicates the mean response of the 20 replicate runs carried out by the model while the blue area represents +/- one standard deviation. The Euclidean (straight line) distance shows that the probability of bat presence is highest approximately 100 m away from street lighting.

4 Discussion

A model's predictive ability may in part be measured by comparing how well the heat map output fits with the data available (Yost et al., 2008). Although this model was better at predicting presence and absence at more extreme values, it nevertheless incorrectly predicted presence for 18% of all cells surveyed in the 0.00-0.09 range and incorrectly predicted absence for 9% of all cells surveyed in the 0.80-0.89 range. However, some of the presence data points were in areas not necessarily considered suitable for bats, such as open pasture. When viewed in ArcGIS using an imagery basemap, it can be seen that, although the area was correctly categorised as high-producing exotic grassland by the land cover variable used in the model, the presence of linear features such as hedges and trees is allowing bats to utilise the area. Since land cover variables rarely include this degree of detail, it is not possible for the model to recognise that these linear features are increasing the possibility of bat presence. Instead, the model has assigned probability of presence values ranging from 0.02-0.19 for the six presence data points shown (Figure 12).

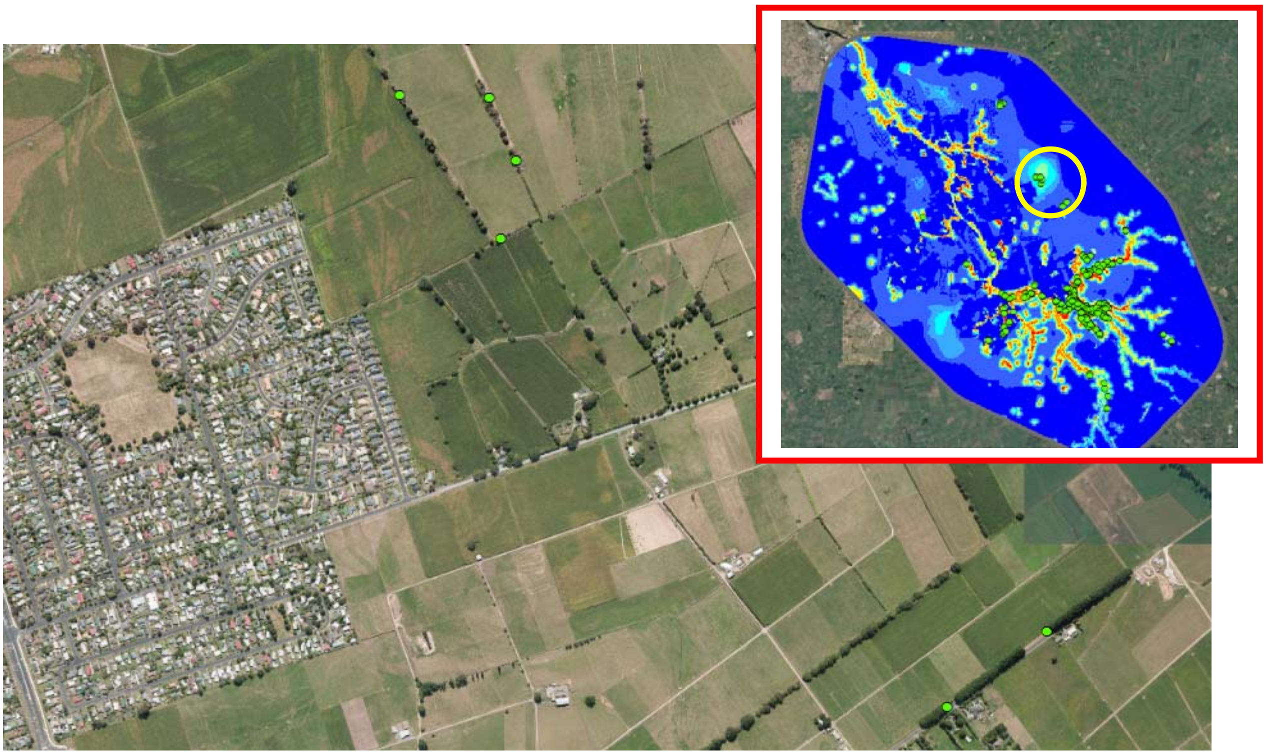


Figure 12. Presence data points (green) of *Chalinolobus tuberculatus* in the mask area show that bats are utilising linear features in open pasture. Inset shows the location of the presence data points (circled in yellow) on the heat map produced by the MaxEnt model where blue indicates that the presence of bats is highly unlikely.

This is one of the limitations of the model; its predictive ability is limited by the extent and accuracy of the detail provided in the variables.

It is less surprising that the model is predicting bat presence in areas where surveys have taken place, and bats have not been detected (Figure 13).

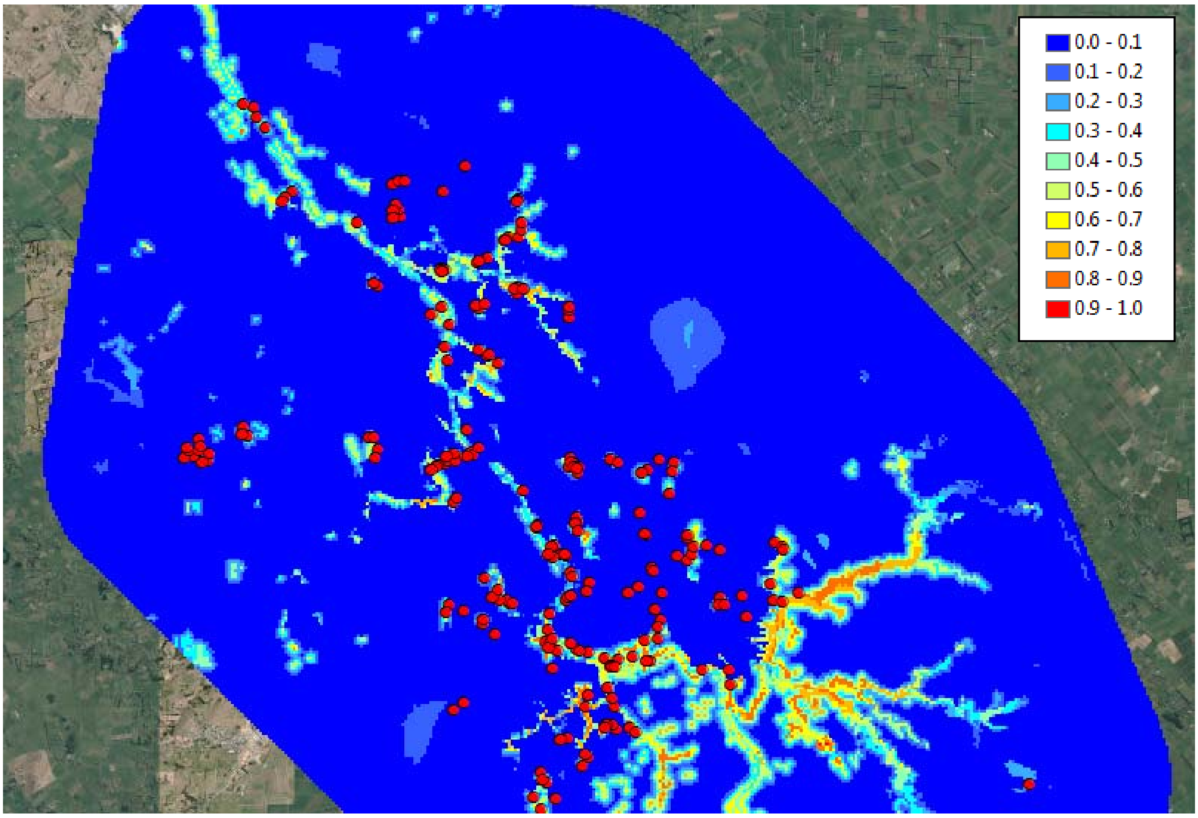


Figure 13. Absence data points (red) mapped against the MaxEnt model output where blue indicates that the presence of bats is highly unlikely and red indicates that it is highly likely.

Although a presence data model should represent the realised niche, since it is informed by where the species is actually found in combination with other variables, in practice it is more likely to represent its potential or fundamental niche. This is due to difficulties incorporating biotic interactions and disturbances into the model, which may have influenced species presence or absence (Guisan & Thuiller, 2005). For bats, disturbance may include light or noise pollution, while biotic interactions may include competition for roost sites and food, or predation. In addition, the surveys themselves may be inaccurate as absence of evidence does not equate to evidence of absence, particularly since bats are nocturnal, cryptic and highly mobile species. Additionally, because the model evaluates each 50 m^2 cell individually, it does not fully reflect the mobility of bats, i.e. the fact that probability of presence is likely to diminish along a gradient away from high probability sites. This can be accounted for in part by the use of “distance to” variables which take proximity to key variables into consideration. However, despite this, the model has assigned low values to some cells which are in close proximity to cells with high probability of presence (Figure 14).

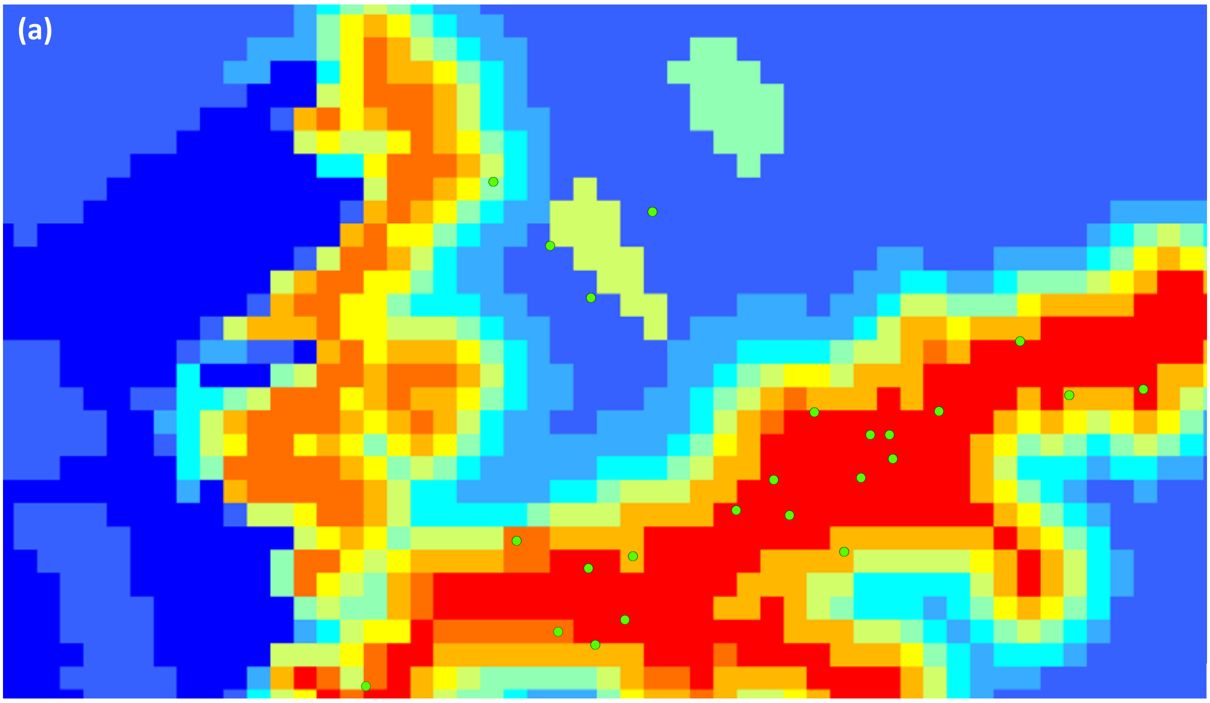


Figure 14. Presence data points (green) shown (a) against the model output where blue indicates that the presence of bats is highly unlikely and red indicates that it is highly likely, and (b) shown against an imagery basemap. The distance between the presence data points deemed high and low probability is 553 m at its shortest point. As can be seen, the model does not fully reflect the mobility of bats, i.e. the fact that probability of presence is likely to diminish along a gradient away from high probability sites. Instead, it evaluates each 50 m^2 cell individually, leading to some cells being deemed low suitability even though they are in close proximity to high suitability cells.

Despite these limitations, the model is recognising the importance of the gully systems and vegetation in the south of the city and the potential of the gully systems in the north of Hamilton. The clearest outcome was in relation to distance to vegetation which consistently contributed highly to the training, testing and predictive ability of the model (Figure 6), and the response curve which showed that probability of presence was highest at zero distance to vegetation (Figure 7). This is unsurprising given the importance of vegetation to bats for navigating, roosting, shelter from predators and edge habitat for foraging. The vegetation to the south of Hamilton covers a larger area and is less fragmented than that further north which could be assisting bats by providing connectivity over a greater area for foraging (Figure 15).

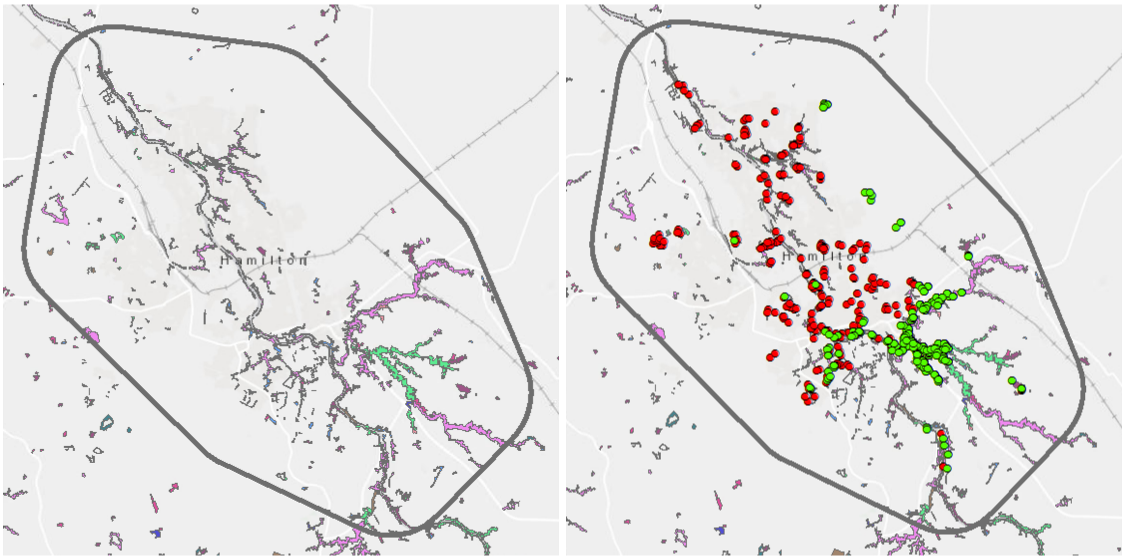


Figure 15. Waikato Regional Council's vegetation layer used as the basis of the distance to vegetation variable. Black line indicates the boundary of the model. Presence data (green dots) are clustered around the larger areas of vegetation while absence data (red dots) are associated with the smaller, more fragmented vegetation further north.

Distance to gullies was the second highest contributor to the AUC value (Figure 6c) with the response curve indicating that the probability of bat presence was highest with zero distance to gullies, falling to less than 0.5 at a distance of approximately 100 m (Figure 8). The importance of the gulley system around Hamilton to long-tailed bats has been noted in other studies. Dekrout et al found bat activity to be correlated to the presence of gullies and noted that gullies were likely to provide connectivity between rivers and forest fragments and also habitat in which to roost and forage (Dekrout, Clarkson, & Parsons, 2014).

When mapped, the influence of both vegetation and gullies on the model's predictions can be seen clearly in the heat map produced (Figure 16).

(a) (b) (c)

Figure 16. Maps produced in ArcMap 10.3 showing (a) vegetation, (b) gullies and (c) the model output or heat map. The area representing higher probability of presence (warmer colours) in the heat map corresponds closely to the area covered by vegetation and gullies, demonstrating the strong influence of both variables on the model.

The distance to rivers layer was not included in the model since its removal did not result in any prediction loss when it was trialled. This was unexpected since other studies have found it to be important, e.g. Bellamy, Scott, & Altringham, 2013; Herkt, Barnikel, Skidmore, & Fahr, 2016; Wordley, Sankaran, Mudappa, & Altringham, 2015. However, the location of rivers was generally correlated with the location of gullies and vegetation and, because of this, did not assist in the model's predictive ability.

The response curve for land cover type indicated that the presence of indigenous forest, broadleaved indigenous hardwoods, gorse/broom (*Ulex europaeus*/*Cytisus scoparius*), manuka/kanuka (*Leptospermum scoparium*/*Kunzea ericoides*), rivers and deciduous hardwoods were associated with a medium to high probability of bat presence (Figure 9). This use by long-tailed bats of a variety of habitats has been seen throughout the country with studies finding bats using not only trees such as beech (*Nothofagus;* Greaves, Mathieu, & Seddon, 2006; Sedgeley & O'Donnell, 1999), kauri (*Agathis australia;* Alexander, 2001), willow (*Salix fragilis;* Sedgeley & O'Donnell, 2004) and pine (*Pinus radiata;* Borkin & Parsons, 2010), but also caves (Guilbert, Walker, Greif, & Parsons, 2007) and limestone outcrops (Griffiths, 2007).

The land cover types with very low probability of bat presence were high-producing exotic grassland and short-rotation cropland, while built-up areas were associated with the lowest probability of presence. Improved pasture was found to indicate low probability of bat presence in the Auckland region also (Crewther, 2016), but surveys in the Hamilton area have detected bats in area classified as pasture due to the presence of linear features such as hedges and trees (Figure 12). In Hamilton, Claudeland's Bush - a kahikatea (*Dacrycarpus dacrydioides*) dominated remnant forest generally considered suitable habitat for bats, no bats have been detected despite being found consistently in Hammond Bush, an urban forest reserve (Dekrout et al., 2014). Dekrout et al noted that Claudeland's Bush was surrounded by a major road network and lacked connectivity with other habitats, while Hammond's Bush was connected both to the Waikato River and a major gulley system. Land cover type on its own, therefore, may not be the best predictor of bat presence since its suitability may be influenced by other factors.

The effect of residential areas was not clear from the model output which suggested that the probability of bat presence peaked at zero distance and then decreased with increasing distance from residential areas (Figure 10). Similarly, the response curve for distance to street lighting suggested that probability of bat presence peaked at a distance of around 100 m, but then decreased with increasing distance from street lighting (Figure 11). It is highly unlikely that bat presence would decrease further away from residential areas and street lighting, especially given that bat presence has been found to be negatively correlated to housing and streetlight density (Dekrout, 2009). The model output is likely to be reflecting survey bias since most surveys were carried out in close proximity to residential areas and street lighting, i.e. the model is seeing the presence of residential areas and street lighting as an indicator of bat presence rather than incidental to it. Certainly, the mapping of presence data points indicates that most are peripheral to areas where housing, street lighting and road density are highest (Figure 17).

(a) (b) (c)

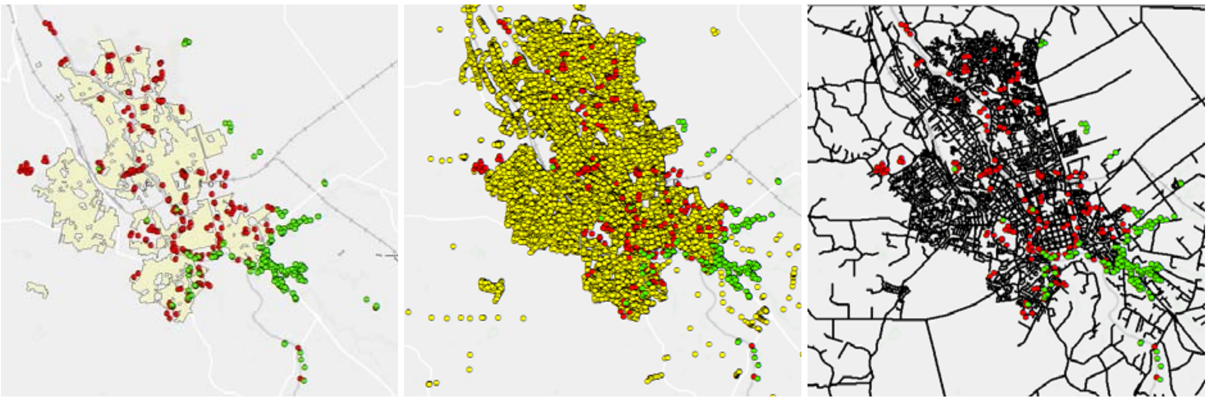


Figure 17. Presence data (green dots) and absence data (red dots) mapped in ArcMap 10.3 against (a) residential areas; (b) street lighting; and (c) roads in the mask area. Most presence data points are peripheral to areas where housing, road and street light density are highest.

It is evident that many of the areas classified as medium to high probability of bat presence by the model have already been recognised as suitable habitat, as they have been surveyed from 2011-2017, with monitoring lasting for 5-23 nights. These areas were also surveyed four times between October 2005 and March 2007 as part of a comprehensive city-wide series of surveys in which the area was divided into 90 x 1 km^2 quadrats, which were systematically surveyed by boat and on foot (Dekrout, 2009). The fact that bats have not been detected in the more northerly areas suggests that there may be behavioural constraints preventing bats from moving up the Waikato River. Dekrout's findings were similar to those of surveys used to inform this model, in that presence data was focused around the Hammond's Bush area to the south of Hamilton. She concluded that bat presence was positively correlated with the topographical complexity associated with gullies and negatively correlated with housing and street light density (Dekrout, 2009). It appears, therefore, that the rivers and gullies which run through the more densely populated parts of Hamilton are not being used by bats and that the majority of activity continues to be confined to the south/southeast region on the edge of Hamilton City (Figure 1).

5 Summary and recommendations

Whilst gullies and vegetation have been shown to be the key predictors of bat presence, the likelihood of presence is also influenced (positively and negatively) by nearby land cover types and reduced in areas where housing and street lighting are most dense. Recorded presence data shows most detections occurring in clusters on the periphery of urban Hamilton, predominantly in the south and southeast, and only extending north to the east of the city in areas of pasture containing linear features such as hedges, and where housing and street lighting density is very low.

Although it appears that the northern gully systems are not being used by bats, it would be worthwhile to survey them from time to time for any change. Certainly, it would be worth surveying areas in the higher suitability range (i.e. 0.70 and over per the maps provided in Appendix II) given that the model's predictive ability generally increased at either extreme, especially in those areas which have not already been surveyed. It would also be worthwhile surveying more of the pastoral areas in the north where there are linear features, as bats may be choosing to forage in these quieter areas on the outskirts rather than use the Waikato River as it enters more populous, noisy and well-lit parts of the city.

It would be useful to maintain a central database of survey results for the Waikato Region to ensure that any changes to bat distribution can be monitored. Although absence data points were not used in the model, a record of these is still important. It is essential that geographic coordinates and monitoring dates are recorded accurately for every ABM placed so that data can be mapped and monitored over time.

Several areas have been identified as suitable for bats, but where bats have not been detected. This may be due to fragmentation of natural areas leading to their isolation. Increasing connectivity of these areas (e.g. Claudeland's Bush) may open them up for use by bats.

6 References

Alexander, J. (2001). *Ecology of long-tailed bats Chalinolobus tuberculatus (Forster, 1844) in the Waitakere Ranges: implications for monitoring.* (Master of Applied Science), Lincoln University, Canterbury, NZ.

Bellamy, C., Scott, C., & Altringham, J. (2013). Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species. *Journal of Applied Ecology, 50*(4), 892-901. doi:10.1111/1365-2664.12117

Borkin, K. M., & Parsons, S. (2010). The importance of exotic plantation forest for the New Zealand long-tailed bat (*Chalinolobus tuberculatus*). *New Zealand Journal of Zoology, 37*(1), 35-51. doi:10.1080/03014221003602190

Crewther, K. (2016). Report on modelled distribution of long-tailed bats in the Auckland region (pp. 42). Report for Auckland Council.

Dekrout, A. S. (2009). *Monitoring New Zealand Long-tailed Bats (Chalinolobus tuberculatus) in Urban Habitats: Ecology, Physiology and Genetics.* (Doctor of Philosophy), The University of Auckland, Auckland.

Dekrout, A. S., Clarkson, B. D., & Parsons, S. (2014). Temporal and spatial distribution and habitat associations of an urban population of New Zealand long-tailed bats (*Chalinolobus tuberculatus*). *New Zealand Journal of Zoology, 41*(4), 285-295. doi:10.1080/03014223.2014.953551

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions, 17*(1), 43-57. doi:10.1111/j.1472-4642.2010.00725.x

Greaves, G. J., Mathieu, R., & Seddon, P. J. (2006). Predictive modelling and ground validation of the spatial distribution of the New Zealand long-tailed bat (*Chalinolobus tuberculatus*). *Biological Conservation, 132*(2), 211-221. doi:http://dx.doi.org/10.1016/j.biocon.2006.04.016

Griffiths, R. W. (2007). Activity patterns of long-tailed bats (*Chalinolobus tuberculatus*) in a rural landscape, South Canterbury, New Zealand. *New Zealand Journal of Zoology, 34*(3), 247-258. doi:10.1080/03014220709510083

Guilbert, J. M., Walker, M. M., Greif, S., & Parsons, S. (2007). Evidence of homing following translocation of long-tailed bats (*Chalinolobus tuberculatus*) at Grand Canyon Cave, New Zealand. *New Zealand Journal of Zoology, 34*(3), 239-246. doi:10.1080/03014220709510082

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters, 8*(9), 993-1009. doi:10.1111/j.1461-0248.2005.00792.x

Herkt, K. M. B., Barnikel, G., Skidmore, A. K., & Fahr, J. (2016). A high-resolution model of bat diversity and endemism for continental Africa. *Ecological Modelling, 320*, 9-28. doi:http://dx.doi.org/10.1016/j.ecolmodel.2015.09.009

Le Roux, D. S., & Le Roux, N. N. (2012). Hamilton city-wide bat survey 2011/2012. Unpublished report for Project Echo prepared by Kessels & Associates, Hamilton.

Phillips, S. J. (2005). A brief tutorial on Maxent. AT&T Research, Florham Park, New Jersey. Retrieved from http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc 22 April 2015

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3–4), 231-259. doi:http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026

Sedgeley, J. A., & O'Donnell, C. F. J. (1999). Roost selection by the long-tailed bat, *Chalinolobus tuberculatus*, in temperate New Zealand rainforest and its implications for the conservation of bats in managed forests. *Biological Conservation, 88*(2), 261-276. doi:http://dx.doi.org/10.1016/S0006-3207(98)00069-X

Sedgeley, J. A., & O'Donnell, C. F. J. (2004). Roost use by long-tailed bats in South Canterbury: examining predictions of roost-site selection in a highly fragmented landscape. *New Zealand journal of ecology, 28*(1), 1-18. Retrieved from http://www.jstor.org.ezproxy.auckland.ac.nz/stable/24058208

Wordley, C. F. R., Sankaran, M., Mudappa, D., & Altringham, J. D. (2015). Landscape scale habitat suitability modelling of bats in the Western Ghats of India: Bats like something in their tea. *Biological Conservation, 191*, 529-536. doi:http://dx.doi.org/10.1016/j.biocon.2015.08.005

Yost, A. C., Petersen, S. L., Gregg, M., & Miller, R. (2008). Predictive modeling and mapping sage grouse (*Centrocercus urophasianus*) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon. *Ecological Informatics, 3*(6), 375-386. doi:http://dx.doi.org/10.1016/j.ecoinf.2008.08.004

Young, N., Carter, L., & Evangelista, P. (2011). A MaxEnt Model v3.3.3e Tutorial (ArcGis v10). Retrieved from http://ibis.colostate.edu/WebContent/WS/ColoradoView/TutorialsDownloads/A_Maxent_Model_v7.pdf 22 April 2015

7 Acknowledgements

We wish to acknowledge and thank the following:

Project lead:

Project Echo

Project coordinator:

Kessels Ecology

Funders:

Department of Conservation
Hamilton City Council
Waikato Regional Council

Survey data provision:

Kessels Ecology
NZTA / Opus

Environmental layers:

Hamilton City Council
NZTA
Waikato Regional Council

Appendix I: Model inputs/output

Data

Presence and absence data were provided by Kessels Ecology and included findings from 14 surveys carried out from 2011-2017 (Table 2).

Table 2. Surveys carried out in the mask area used to inform the model

Data sources	Dates monitored
2011/2012 City-wide Bat Survey	Sept 2011-Jan 2012
2012 Hammond Park Gum Tree Removal - Bat Survey	14-17 Dec 2012
2013 Rugby Park - Bat Survey	May 2013
2014 Cobham Drive - Hamilton Gardens - Gully Assessment	Apr 2014
2014 Sanford Park Tree Removal - Bat Survey	1-5 Oct 2014
2015 NZTA Survey_2	31 March-31 May 2015
2015 Memorial Drive Tree Removal - Bat Survey	June 2015
2015 Sanford Park - Bat Survey	Dec 2015
2016 NZTA Survey_3	26 Jan-4 Apr 2016
2016 Hamilton Gardens Cemetery Tree Removal – Bat Survey	9-12 Feb 2016
2016 NZTA Survey_4	29 March-3 May 2016
2016 Hamilton Gardens Tree Removal – Bat Survey	May-Nov 2016
2016 Wellington Street Tree Removal - Bat Survey	Nov 2016
2017 Community City Survey	Feb-June 2017

Mask

Both presence and absence data were mapped and a minimum convex polygon drawn around the survey points. A 2,500 metre buffer to the minimum convex polygon was then added to create the mask or boundary for the model (Figure 18).

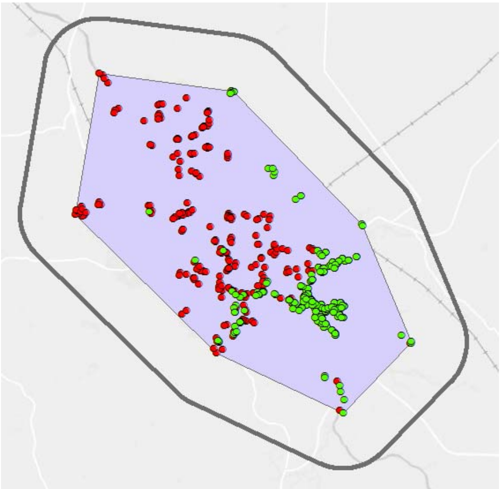


Figure 18. Buffer or mask used for model. Green dots represent presence data, red dots represent absence data, the purple area represents the 100% minimum convex polygon created and the black line represents the 2500 m buffer or mask on which the model is based.

Variables

Variables selected and trialled for use in the model included:

Variable	Name of file	Source	Source
Land cover	LCDB v4.1 - Land Cover Database version 4.1, Mainland New Zealand	Landcare Research NZ LTD	https://lris.scinfo.org.nz/
Vegetation (1)	Biodiversity vegetation 2012	Waikato Regional Council	Requested through personal communication
Gullies	GeotechHazard	Hamilton City Council	Requested through personal communication
Residential areas	NZ Residential Areas	Land Information NZ	https://koordinates.com/
Street lighting (merged files)	Hamilton City Waikato District Waipa District	NZTA with the permission of Hamilton City, Waikato and Waipa Districts	Requested through personal communication
Significant natural areas	Significant Natural Areas	Hamilton City Council	Requested through personal communication
Reserves	Reserves Update 2015	Hamilton City Council	Requested through personal communication
Significant trees	SignifTreesOverlayPt 2012	Hamilton City Council	Requested through personal communication
Vegetation (2)	Vegetative cover map of NZ	Landcare Research NZ LTD	https://lris.scinfo.org.nz/
Elevation	Land Environments New Zealand (LENZ) – Level 4 Polygons (2009)	MfE	https://data.mfe.govt.nz/layer/2358-land-environments-new-zealand-lenz-level-4-polygons-2009/
Roads	NZ Road Centrelines (Topo 1:50k)	Land Information NZ	https://koordinates.com/
Rivers (merged files)	NZ River Polygons (Topo 1:50k) NZ River Centrelines (Topographical 1:50k)	Land Information NZ	https://koordinates.com/
Lakes	NZ Lake Polygons (Topo 1:50k)	Land Information NZ	https://koordinates.com/
Annual mean temperature	BIO_1	BIOCLIM Data for current conditions ~1950-2000 30 arc-seconds ESRI grids	http://www.worldclim.org/current
Mean temperature warmest quarter	BIO_{10}		
Mean temperature coldest quarter	BIO_{11}		
Annual precipitation	BIO_{12}		
Precipitation of wettest quarter	BIO_{16}		
Precipitation of driest quarter	BIO_{17}		

The first five variables listed are those which were selected for the final model. The others were removed as they did not contribute greatly to improving the model's predictive power and, in some cases, were highly correlated to other variables.

All the files used, with the exception of the land cover layer, were converted in ArcGIS to "distance to" files based on Euclidean (straight line) distance from presence data points.

Model parameters

- Replicated run type: Crossvalidate
- Random seed selected (so the model didn't select the same data points for every run)
- Regularization multiplier set to 2 to avoid overfitting of the model
- Replicates: 20
- Maximum iterations changed to 5000
- Other settings kept on the MaxEnt defaults

Appendix II: Proposed areas for surveying

The following maps provide the locations of cells which the model deemed to have a probability of bat presence of 0.70 and higher.

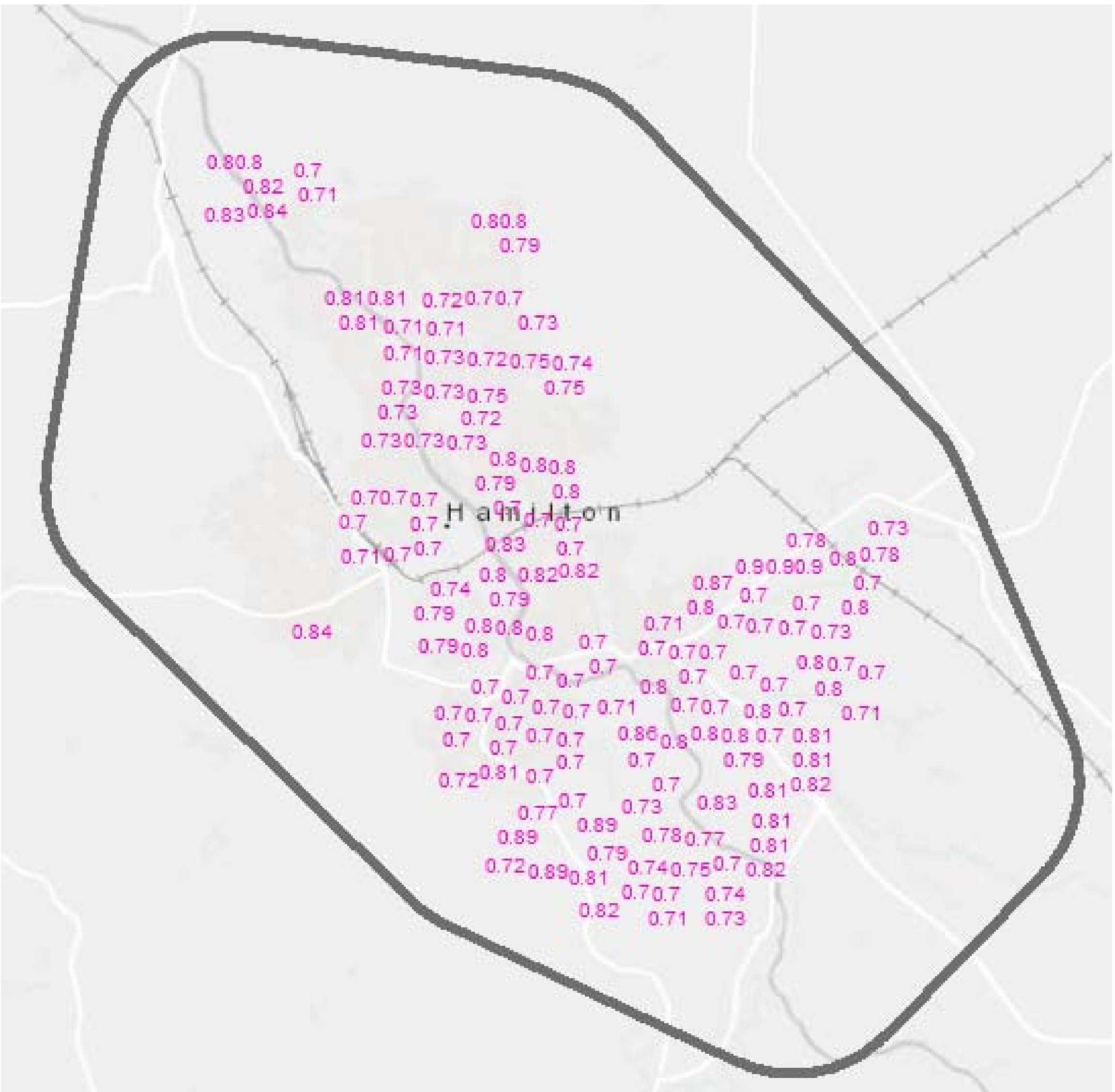


Figure 19. Mask area showing the broad location of cells deemed by the model to have a higher probability of bat presence (0.70+).

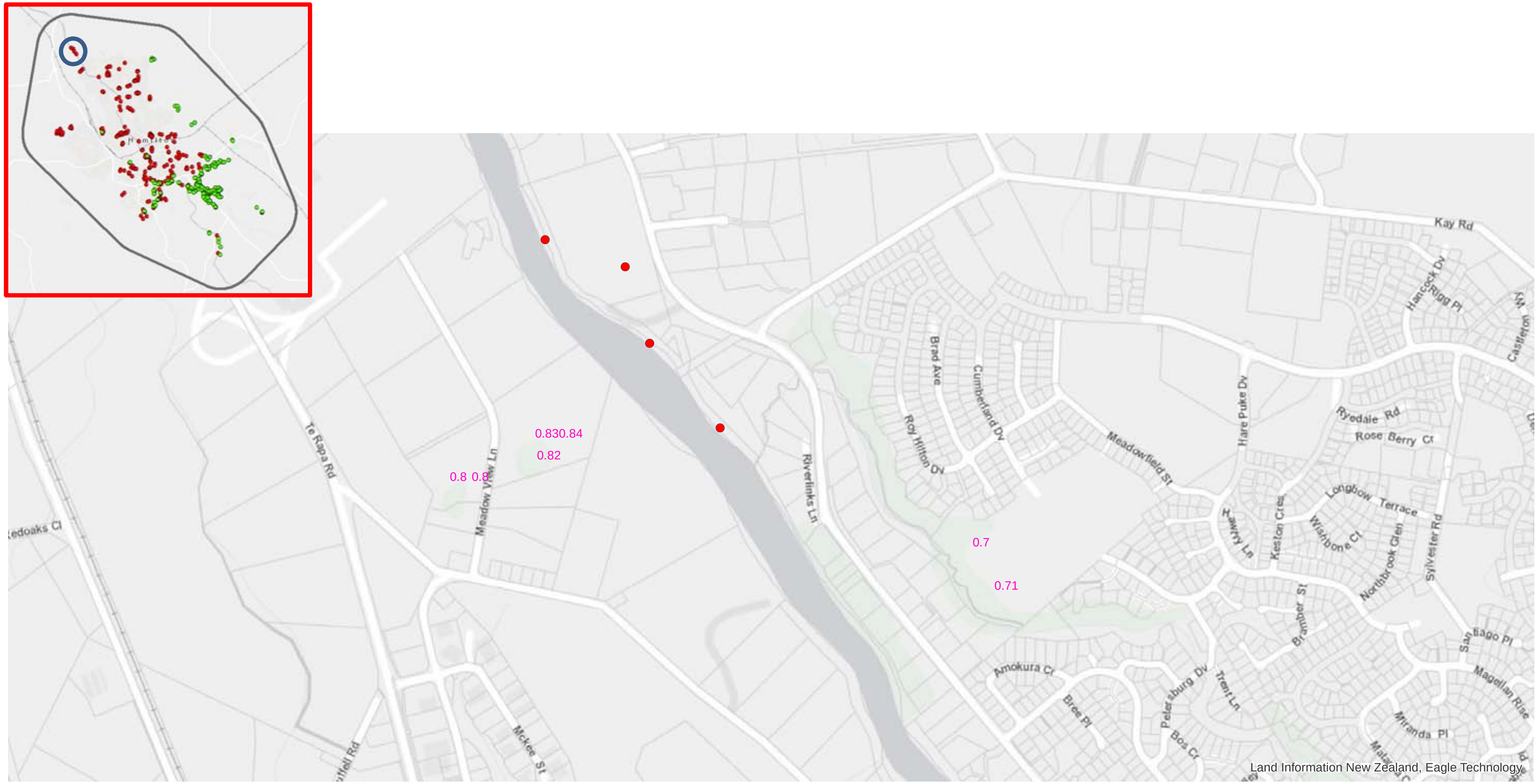


Figure 20. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

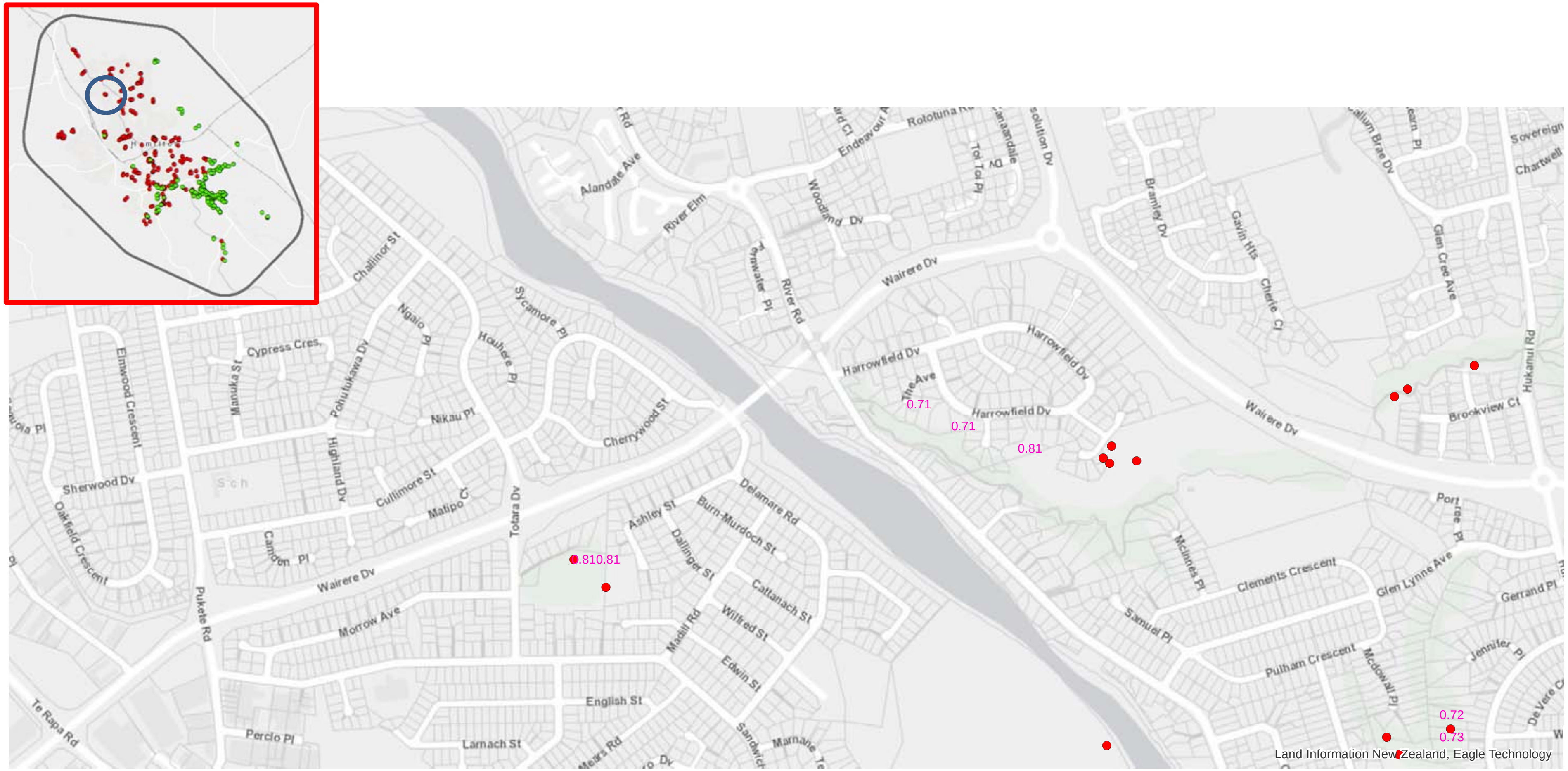


Figure 21. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

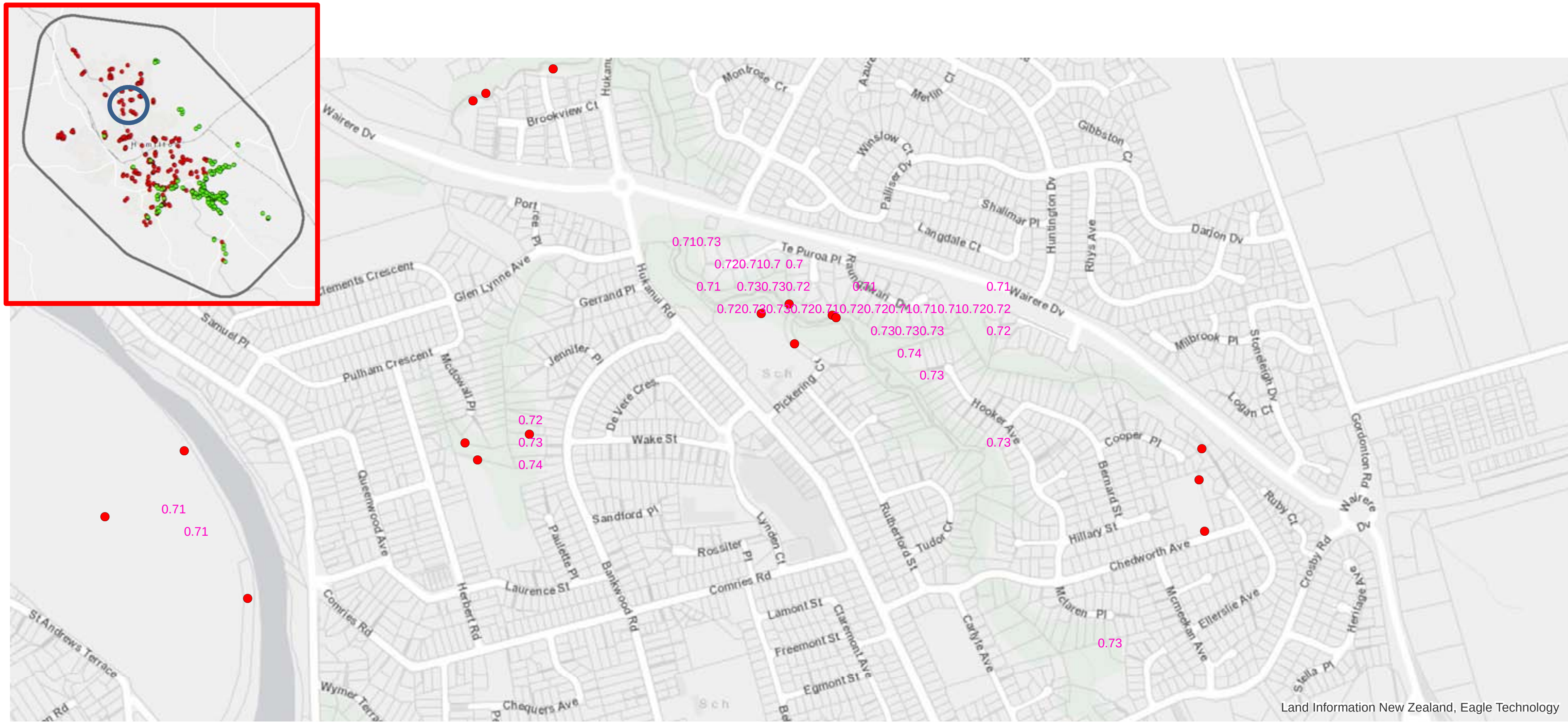


Figure 22. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

Figure 23. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

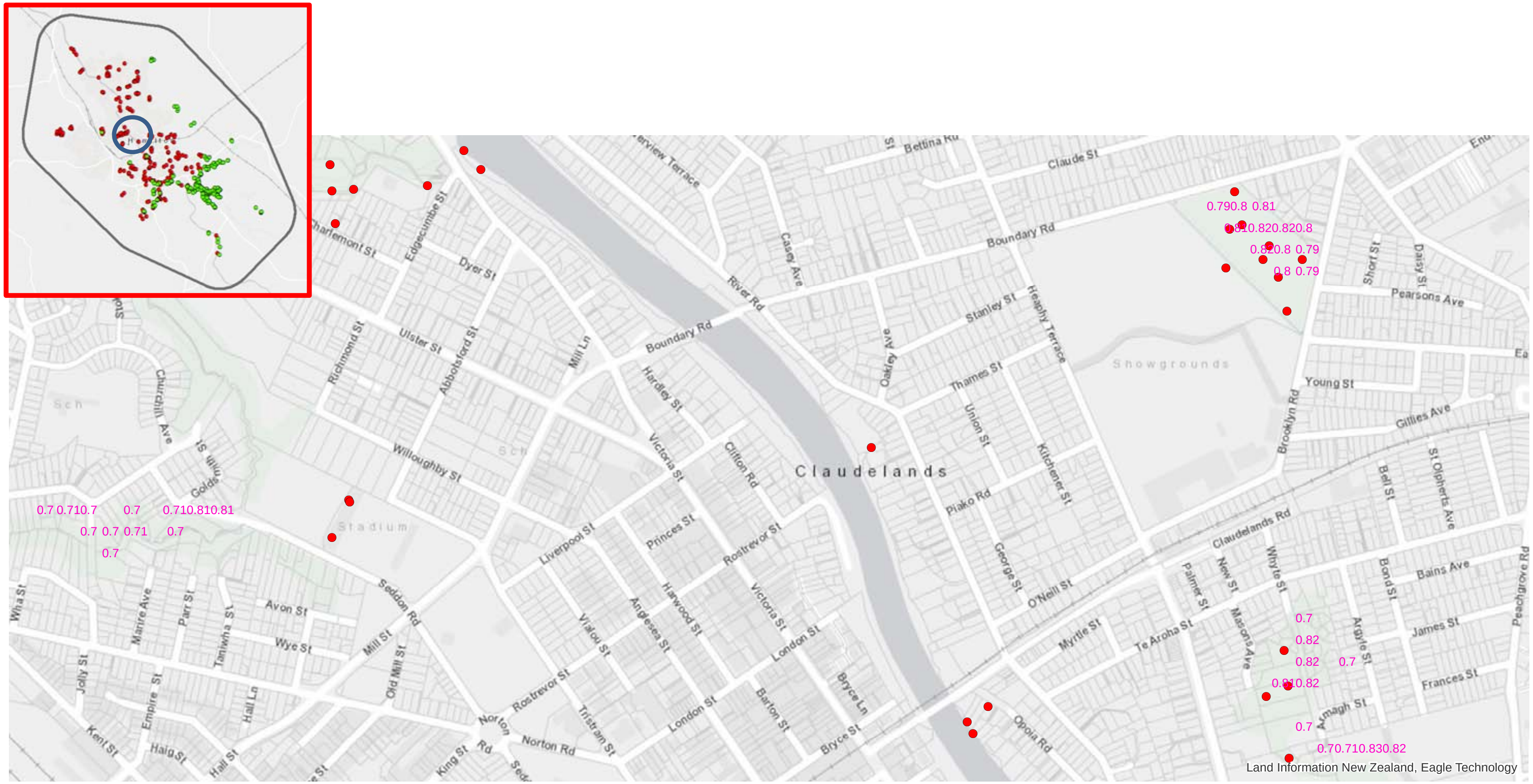


Figure 24. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

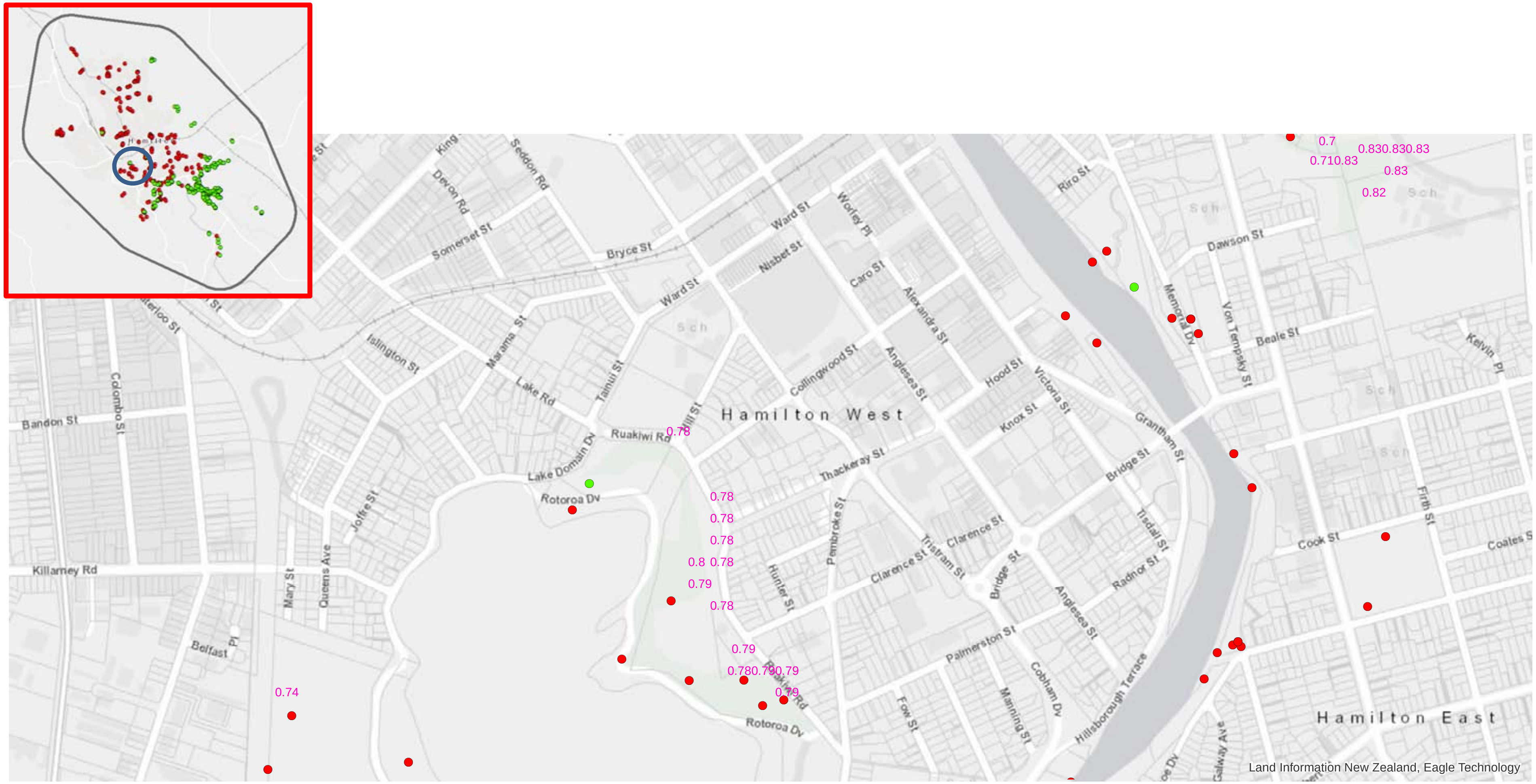


Figure 25. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

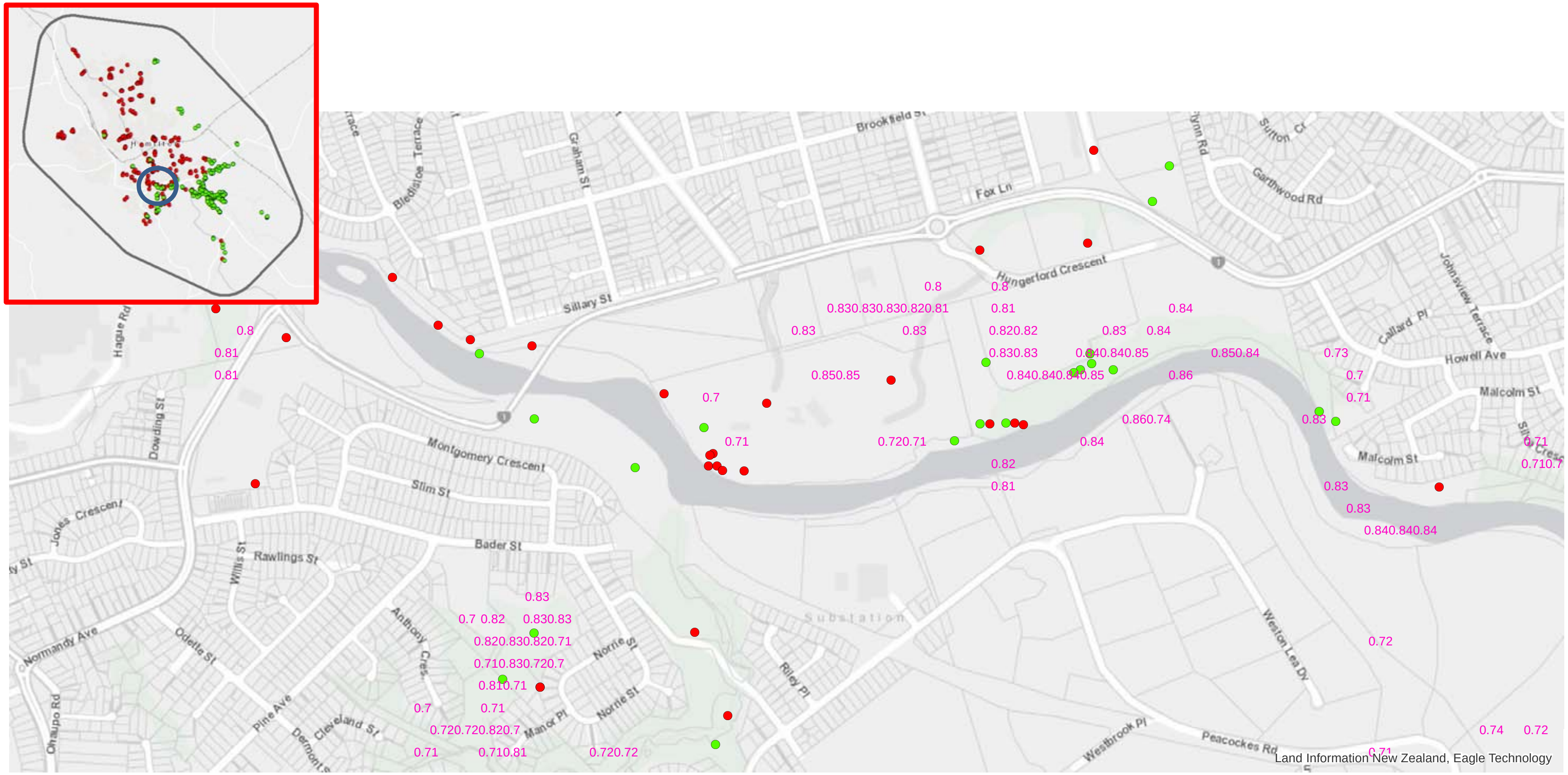


Figure 26. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

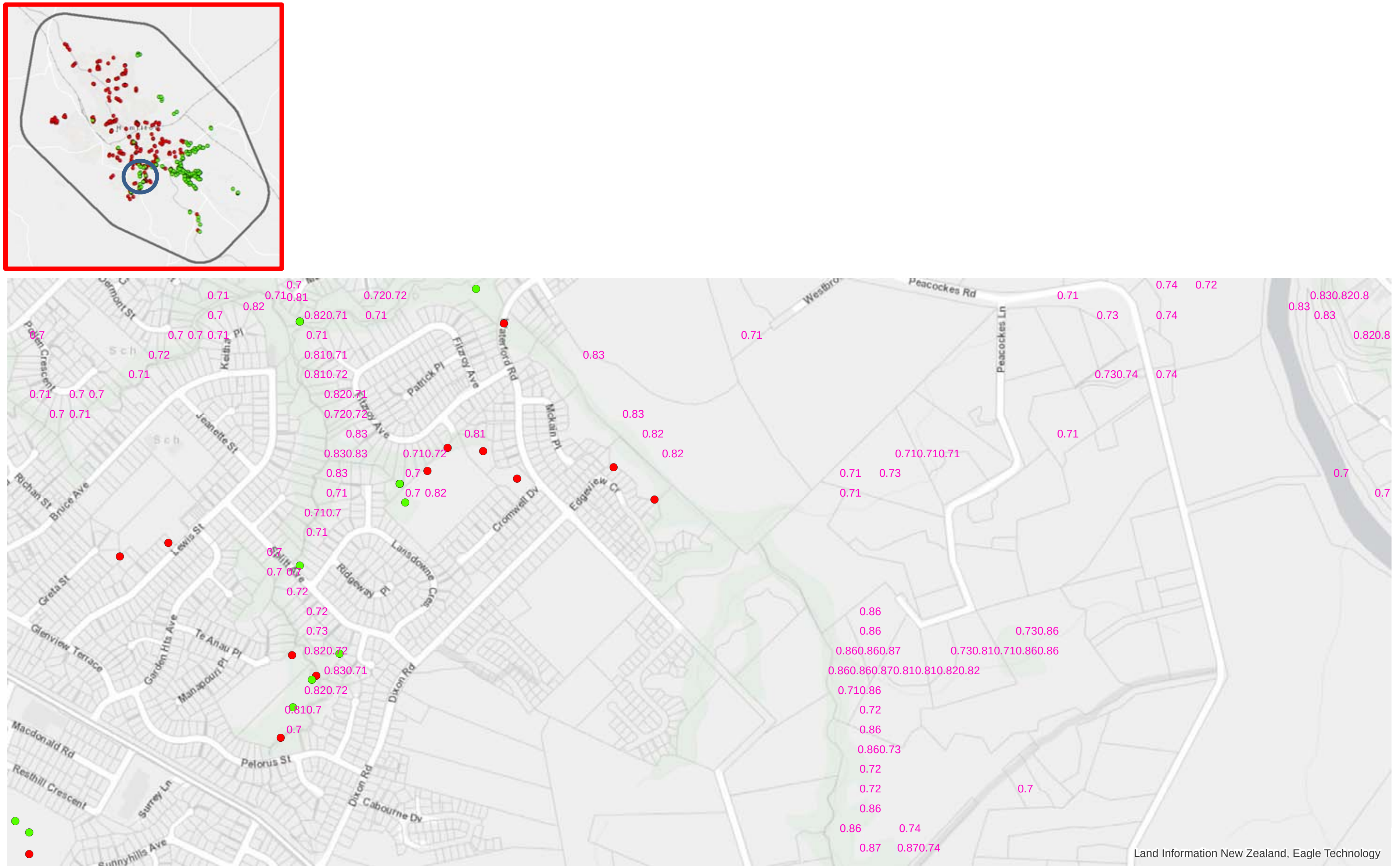
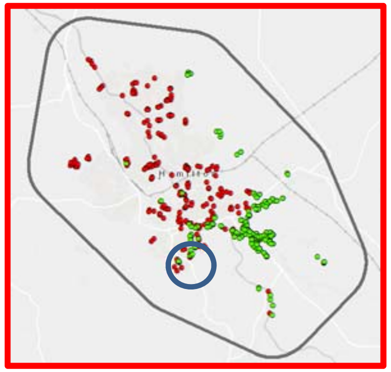


Figure 27. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.



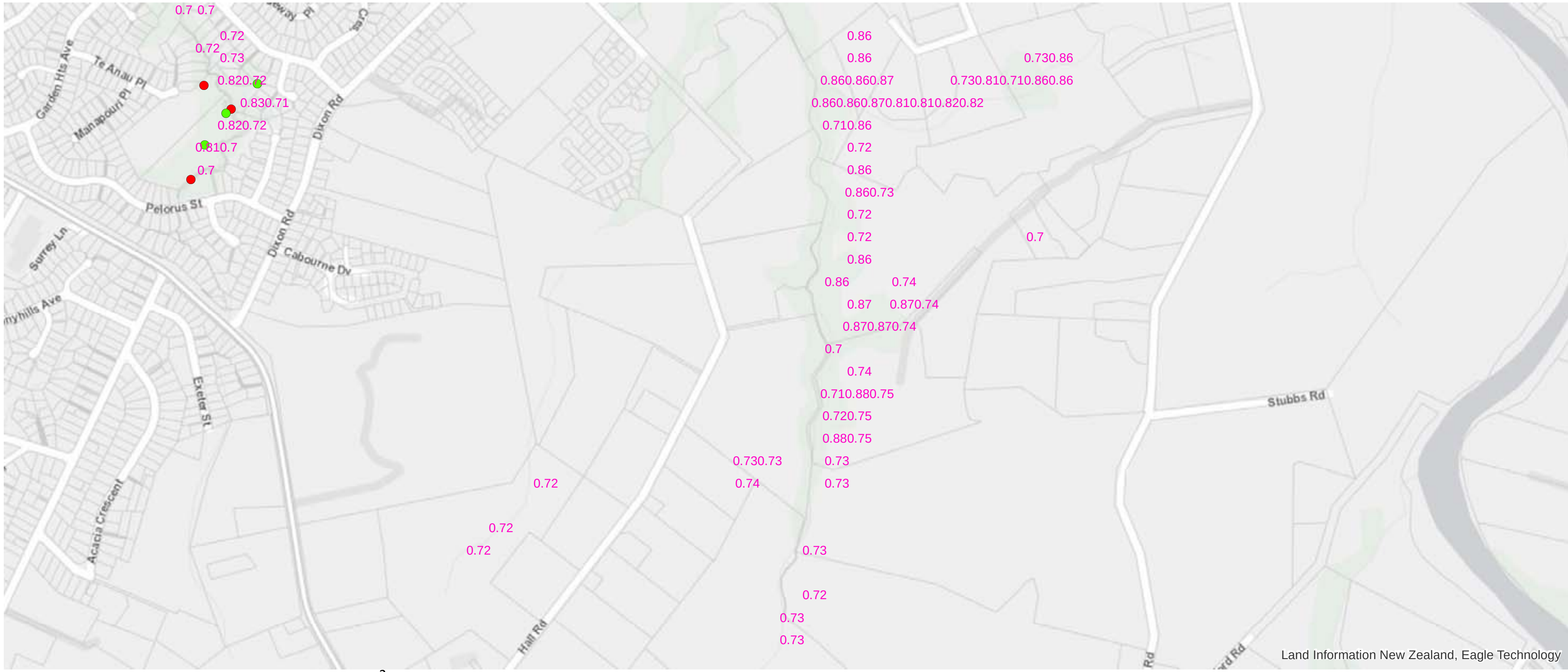


Figure 28. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

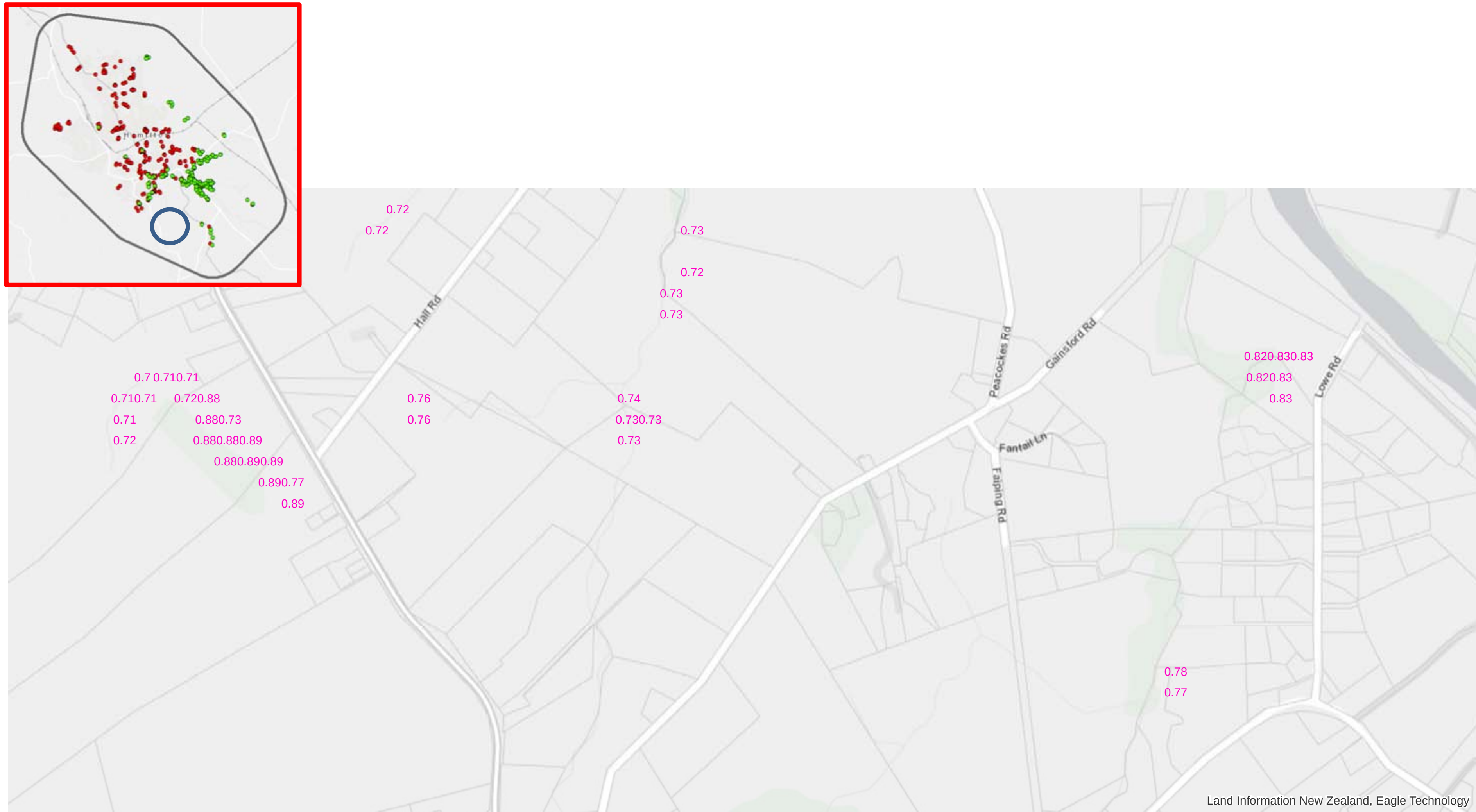


Figure 29. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

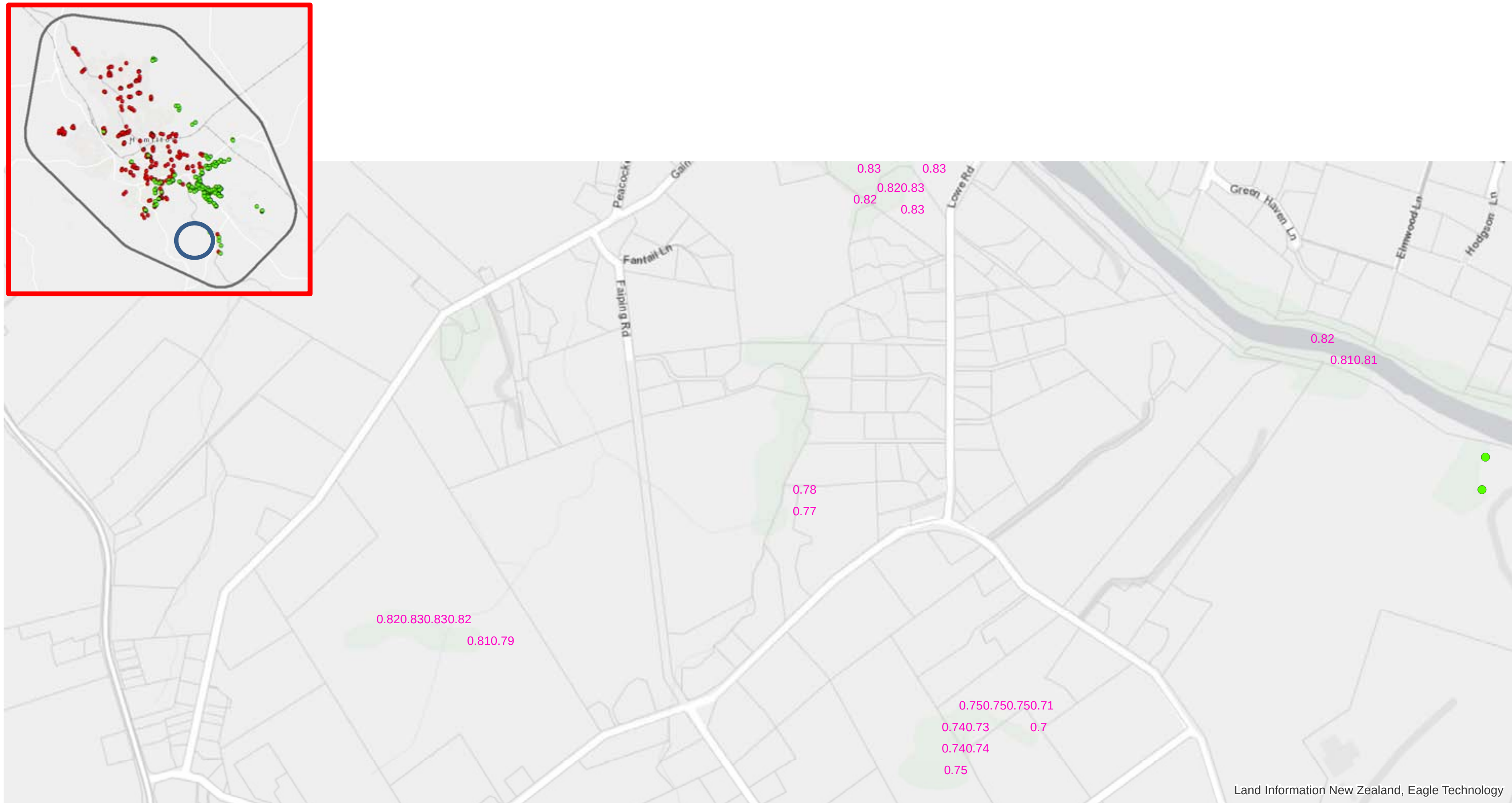


Figure 30. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

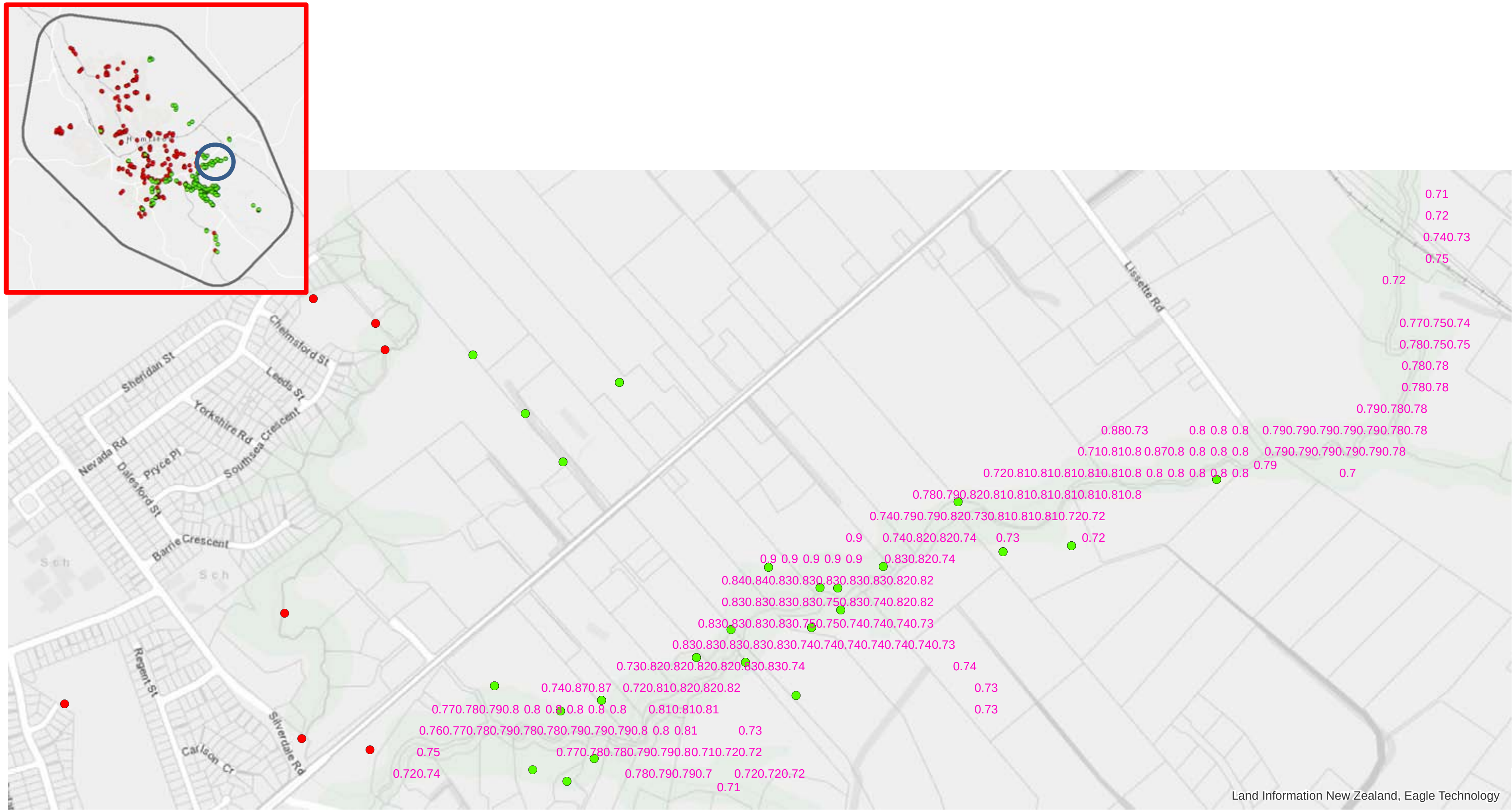


Figure 31. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

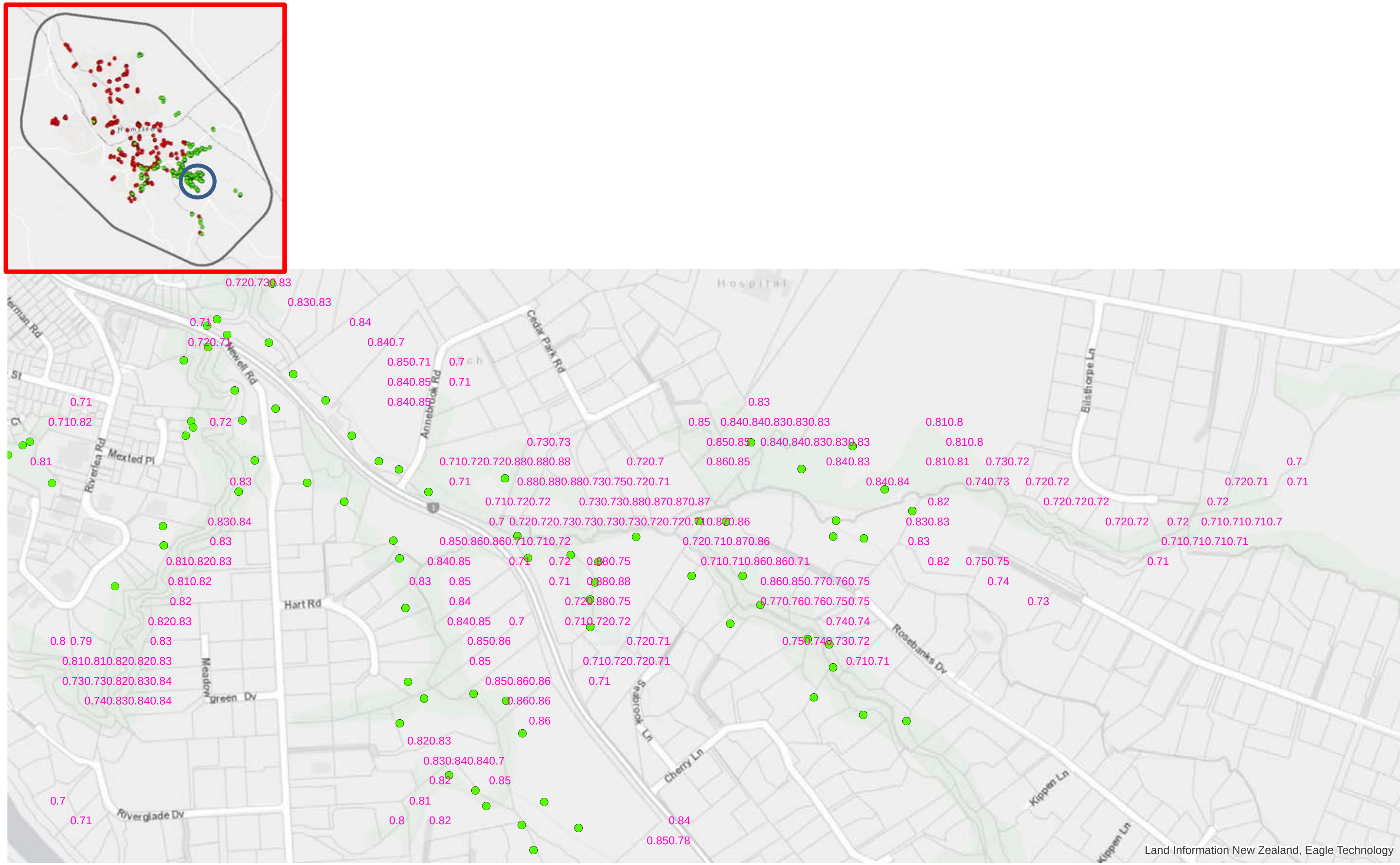


Figure 32. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.

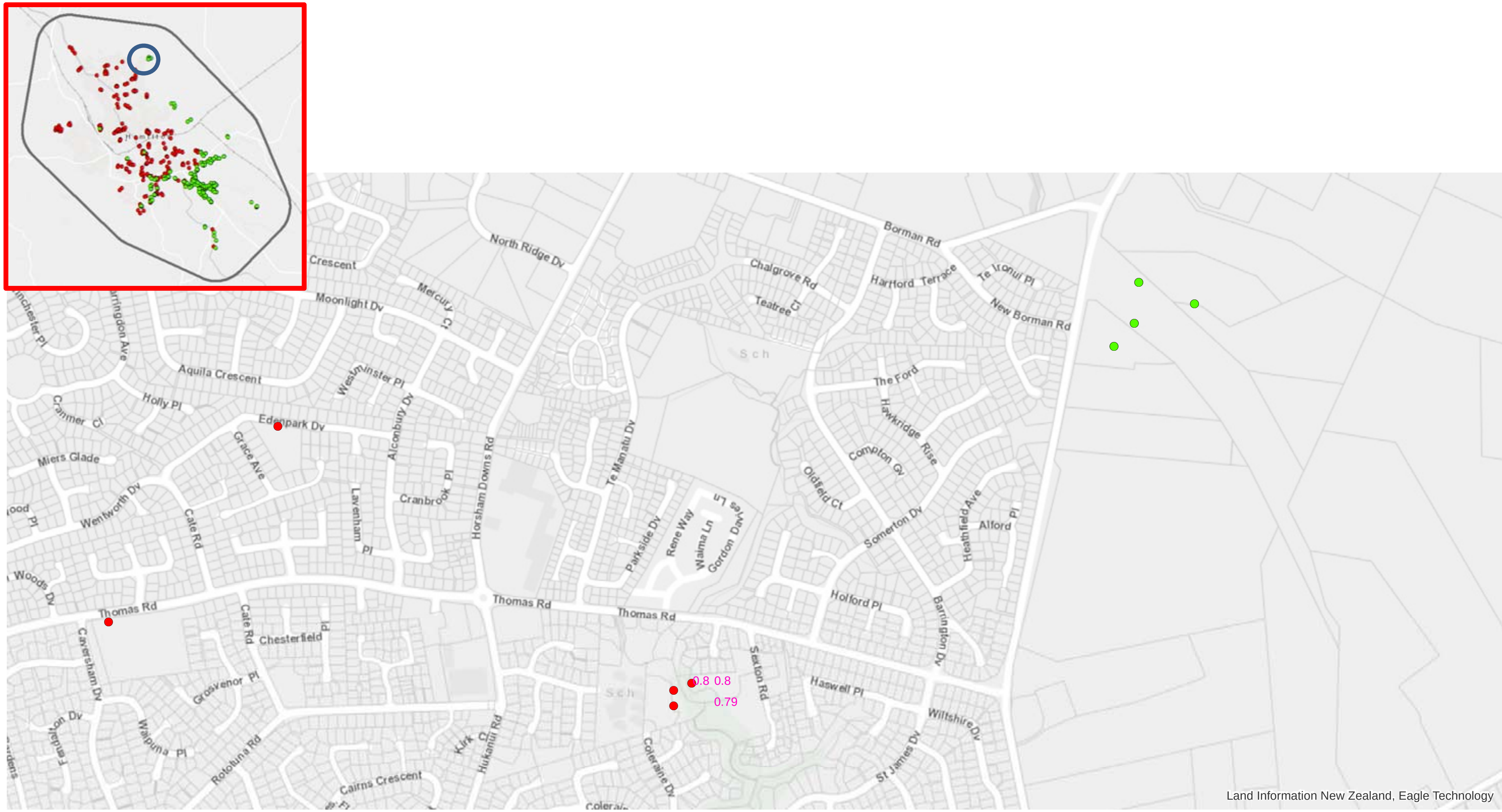


Figure 33. Numbers in pink indicate 50 m^2 cells which the model deemed to be of higher suitability for long-tailed bats (0.70+). Inset shows approximate location of mapped area (circled in blue) in relation to overall mask area. Areas already surveyed are represented by green dots for presence data points and red dots for absence data points.