Te Rangahau Aroturuki i ngā Rākau Rangatira o Te Wao Nui ā Tiriwa

2021 Waitākere Ranges Kauri Population Health Monitoring Survey

June 2022, Technical Report 2022/8

aucklandcouncil.govt.nz

Chapter 3

Multivariable analysis of risk factors associated with symptomatic kauri and detection of *P. agathidicida* in the Waitākere Ranges

Te Mātatini o te tātari i ngā whakaputanga tūraru e hāngai ana ki kauri e whai tohumate ana, i te kitenga hoki o te puruheka patu kauri i Te Wao Nui ā Tiriwa

Authors

Karyn Froud[1], Jun-Hee Han[2], Chris Compton[2], Yue Chin Chew[3], Hugo Geddes[3], Georgia Edwards[3], Bruce Burns[4], Stuart Leighton[3], Sarah Killick[3], Edward Ashby[5], Alastair Jamieson[3], Lisa Tolich[3]

[1]Biosecurity Research

[2]Massey University

[3]Auckland Council

[4]Auckland University

[5]Te Kawerau ā Maki

3.1 Abstract

Te whakatūporotanga

The aims of this study were to generate and test hypotheses about the associations of environmental, host and pathogen-related risk factors with i) symptoms in kauri consistent with kauri dieback and ii) the presence of *Phytophthora agathidicida*, the causal agent of kauri dieback.

Multivariable logistic regression models and spatial modelling were used to investigate symptomatic kauri and detection of *P. agathidicida* in separate models from data collected from a cross-sectional survey and GIS-generated landscape variables. Data from 2140 randomly selected kauri were used to investigate the risk factors associated with the binary outcome of symptomatic vs non-symptomatic kauri, based on the symptomatic criteria of the case definition for kauri dieback disease (Chapter 2). Data from a subset of 761 kauri with soil samples analysed for *P. agathidicida* using a soil bioassay were used to investigate the risk factors associated with a *P. agathidicida* detection vs not detected.

This study identified three factors that were significantly associated with presence of symptomatic kauri and four factors that were significantly associated with presence of *P. agathidicida* in spatial models.

For the symptomatic kauri model, the strongest association was between symptomatic kauri and proximity to *P. agathidicida* sites (point locations of *P. agathidicida* detections). Prevalence was highest close to *P. agathidicida* sites and reduced with distance away from *P. agathidicida* sites. Symptomatic kauri prevalence was also higher closer to historic timber sites (timber mills and saw pits) (reducing with distance away from them) and increased with increasing tree size (DBH).

For the *P. agathidicida* model, pathogen prevalence was higher with decreasing elevation, and with decreasing distance from historic timber sites and from the coastline. It was also higher as the distance to the closest neighbouring tree decreased. The results generated hypotheses for further investigation into understanding or managing these relationships, such as managing the distribution of *P. agathidicida.* In addition, our results found several associations of note (where the associations had wider credible intervals) between symptomatic kauri prevalence and distance to the coast, neighbouring tree distance, and distance to the closest uphill track; and *P. agathidicida* prevalence and distance to the closest track and presence of tanekaha (*Phyllocladus trichomanoides).* These require further investigation.

Both modelled outcomes had potential misclassification bias, in that effect sizes for risk factors may have been pushed towards no effect (towards the null hypothesis). Misclassification bias may have been present due to the low sensitivity of the diagnostic test for *P. agathidicida*, missing true positives, and the potential misclassification of symptomatic trees as non-symptomatic, using a conservative symptom-based cut-point.

These results can be used to prioritise future surveillance and research, as well as inform potential management interventions to reduce the spread of *P. agathidicida* and development of disease through appropriate biosecurity and ecosystem protection measures.

3.2 Introduction

Te whakataki

There is a strong relationship between *P. agathidicida* and kauri dieback disease, with both pathogenicity and Koch's postulates having been demonstrated (Bellgard et al., 2016, Gadgil, 1974). The presence of *P. agathidicida* is necessary to cause kauri dieback but a pathogen is rarely sufficient to cause disease in the absence of other factors, in that other component causes such as a vulnerable host and particular environmental conditions (e.g., drought, rainfall, disturbances) are required for disease to develop (Rothman and Greenland, 2005, Martin, 2008). In addition, it is uncertain how many kauri with symptoms that look like kauri dieback observed in the forest are caused by *P. agathidicida* compared to other abiotic or biotic causes. All potential causes of disease and tree death are important when the aim is a healthy forest.

An observational study design was used to identify and collect risk factors for symptomatic vs non-symptomatic kauri and for *P. agathidicida* detection vs non-detection as separate outcomes as described in Chapter 2. These potential risk factors will be assessed using an analytical cross-sectional study. The cross-sectional study design is a type of observational study, which is a commonly applied in human and animal health investigations, with only recent application in plant health (Rothman et al., 2008, Dohoo et al., 2009, Froud and Cogger, 2015). This is a novel approach for investigating kauri dieback, which has previously followed a pathogen-centric approach (Bradshaw et al., 2020). A key difference between observational and experimental studies is that extraneous factors, called confounders, are not able to be managed through randomisation. These are therefore typically controlled for during the analysis stage of an investigation using multivariable statistical models (Dohoo et al. 2009e). Cross-sectional studies have robust guidelines for their application (Sargeant et al., 2016, O'Connor et al., 2016, Vandenbroucke et al., 2007).

The type of observational study design selected depends on the research question. Ideally, a longitudinal study such as a cohort study would be used to obtain the strongest evidence for a causal link between risk factors and disease. However, when disease is already widely distributed, as in the New Zealand kauri dieback outbreak (Hill et al., 2017), a cross-sectional study is a more appropriate approach, because it collects outcome and risk factor data at a single point in time with the aim of identifying factors that are associated with an increased or decreased prevalence of the outcome. In this case symptomatic kauri or *P. agathidicida* detection. The risk factors identified in a well-designed cross-sectional study may not be causal, however, as long as results are interpreted with caution around temporality (in that a cause precedes an outcome) and potential confounding, they should be interpreted as factors that contribute significantly to an increased or decreased prevalence of disease (Maes et al. 2001). Results can be used to prioritise which factors should be investigated further, using either experimental studies or more comprehensive observational studies (e.g., a cohort study or case-control study) to determine causal relationships (Mann 2003).

This study investigated a range of environmental, anthropogenic, host and pathogen-related risk factors to generate and test hypotheses on associations with i) symptoms in kauri consistent with kauri dieback and ii) *P. agathidicida* detected in soil beneath kauri in Te Wao Nui ā Tiriwa / the Waitākere Ranges parkland. The intended outcome of this study is to inform kauri dieback control measures to reduce the presence of *P. agathidicida* and the development of disease symptoms in kauri to enhance kauri health.

3.3 Methods
Ngā tikanga

3.3.1 Dataset
Trees were randomly selected from a sample frame of trees classified as kauri using remote sensing, based on the Meiforth et al. (2020) methodology and detailed in Chapter 2. A total of 2140 randomly selected trees were surveyed and a subset of 761 trees were soil sampled for *P. agathidicida*.

3.3.2 Outcome variables
Each surveyed tree was visually assessed and classified as symptomatic or non-symptomatic (which included healthy and ill-thrift trees) as described in Chapter 2. Dead trees were excluded from the study.

Soil samples were collected around the base of pre-selected trees at the time of visual assessment and tested using the soil sampling bioassay as described in Chapter 2 and classified as *P. agathidicida* detected or not detected.

3.3.3 Initial risk factor variable selection
Individual kauri tree health factors were identified through two hui involving kauri ecosystem health experts from mana whenua and research organisations.

For each tree, potential risk factor variables were either collected during the ground-based survey (Chapter 2, Appendix A) or derived by later Geographic Information System (GIS) analyses based on existing Auckland Council or national datasets (Chapter 2, Appendix A, Appendix G). Among the aggregated data, over 100 variables (Appendix C) were collected which were potentially associated with the presence of symptomatic kauri or detection of *P. agathidicida*, the outcome variables of this study.

Using the variables identified as potential risk factors, a univariable screening test (simple logistic regression) for each binary (yes/no) outcome (e.g., symptomatic kauri vs non-symptomatic kauri and *P. agathidicida* detected vs *P. agathidicida* not detected) was conducted. Based on the results of the univariable screening test (Appendix C), all variables with a P-value < 0.2 were identified for either outcome for further consideration. Among these, any variables that either (1) contained a large number of missing values (except the variable of the distance to the closest uphill track, which was a variable of interest), or (2) was an (in)direct result of the outcome variables were discarded as they were not on the causal pathway for symptomatic kauri or *P. agathidicida*. Once the variables were identified, any plausible correlations between the variables were manually assessed in turn to select the most biologically meaningful variable among a group of highly correlated ones (e.g., correlated groups of common species) to be included in the multivariable models. A Bayesian network analysis was further conducted to investigate any additional correlations that were missed during the manual examination (Lewis and McCormick, 2012). Based on the correlation between variables, causal path models were constructed for each outcome to aid in variable selection for modelling (Figure 3-2 and Figure 3-3). Finally, the correlations between the variables in the path models were differentiated as either a potential biological confounding effect or simple correlation. The univariable screening and Bayesian network analysis were conducted in R using "glm" and "bnlearn" packages (R Core Team, 2020). The casual path models were developed using the "DAGitty" programme (Textor et al., 2016).

3.3.4 Non-spatial multivariable models

The variables from the screening and initial selection process were investigated using frequentist-based, non-spatial multivariable logistic regression models for symptomatic kauri or *P. agathidicida* detection. As part of the model building process, three key variables of interest that were highly correlated with each other, namely the distance to the closest track, road, or uphill track, were checked separately to identify the variable that best explained the data. Therefore, for each outcome, three models were established with the model building process of each model starting with a full model containing either one of these key variables of interest (i.e., the distance to the closest track, road, or uphill track) and other variables from the initial selection process. From each full model, any non-significant variables with P-values > 0.05 were removed from the model in a stepwise manner with the variable in the order of the largest P-value being removed first. However, regardless of P-value, the distance to the closest track, road, and uphill track for symptomatic kauri and *P. agathidicida* models and the distance to the closest *P. agathidicida* site for the symptomatic kauri models were retained in each model because they were key interest factors and to allow comparison between the three models for each outcome. Also, any biological confounders identified during the discussion with experts remained in the model regardless of the P-value to account for potential confounding when using observational data (refer to glossary). The models were examined for any statistical confounders identified as causing > 20% change in any of the coefficients of remaining variables when they were removed. If identified, they were retained in the final model.

However, there was an exception in the management of statistical confounders in the case of the diameter at breast height (DBH) in the *P. agathidicida* model. This was because (1) DBH was kept in the model even though it had a P-value > 0.50 since it was a biological confounder of the association between the distance to the closest tree and *P. agathidicida* detection, and (2) there were strong correlations between DBH and other risk factors. As the model coefficient for DBH was highly variable following the removal of insignificant risk factors from the *P. agathidicida* model, it needed to be retained.

Once a final non-spatial model was established, potential interactions (refer to glossary) between variables were examined. An interaction term between the distance to the closest timber site and the number of archaeological sites within 500 m significantly decreased the variability of the model, however, the interaction term was not statistically significant in any of the models.

In this study, the final three non-spatial models for each outcome shared the same risk factors except the three different road/track variables. However, due to differences in calculation of the three variables of interest, the final models were based on different numbers of observations. The difference in numbers of observations was due to how the uphill track variable was calculated; in that if a tree had no track above it, no value could be calculated. Therefore, the comparison between the final models for each outcome was based on a reduced dataset without any missing values. The models were compared using standard statistical criteria of the Akaike information criteria (AIC) and area under the ROC (receiver operating characteristic) curve (AUC) with lower AICs and higher AUCs indicating a better model. Once the final multivariable non-spatial model for each outcome was chosen, between the three options, it was re-run using the full observations available depending on the track/road variable that best suited the data. The linearity assumption of any continuous variables for the final multivariable non-spatial model for each outcome was evaluated by converting the variable into an ordinal variable of four groups (based on its quartile values) and visually examining the linearity of the coefficients of the ordinal variable. Also, a Hosmer-Lemeshow test was conducted to examine the goodness-of-fit of the final multivariable non-spatial model for each outcome by splitting the data into eight groups based on percentiles of predicted probability. After confirming the lack of any violation of linearity assumptions or goodness-of-model fitness, standardised residuals (the difference between the observed values and value predicted by the model) were calculated to investigate any remaining spatial dependence in the data that the multivariable models had not adjusted for. The spatial correlation (i.e., the values for trees close to each other may be more similar than the values of those further apart) was examined by assessing covariance in the residual values as a function of distance via computing omnidirectional variograms to a distance of 100 metres.

3.3.5 Spatial multivariable models

Due to evidence of spatial correlation in the standardised residuals from the non-spatial (frequentist) multivariable models, separate Bayesian spatial models were developed for each outcome variable. For a kauri *i*, the presence of outcome (presence of symptomatic kauri or detection of *P. agathidicida*), *Yi*, can be mathematically described as

$$Y_i = Bernouilli(P_i)$$

$$logit(P_i) = \beta_0 + BC + W_i$$

where P_i is the probability of a kauri *i* showing the outcome, β_0 is the intercept, *C* is a matrix with rows corresponding to the covariate pattern from the non-spatial multivariable model for each sampled location, B is a vector of the covariate coefficients, and W_i is a zero-mean Gaussian spatial random effect term with a Matérn covariance function (Matérn, 2013). By using the formula above, the remaining spatial correlation in the data (i.e., W_i) was expected to be adjusted after considering the result of the final non-spatial multivariable models (i.e., $\beta_0 + BC$).

The covariate coefficients and spatial correlations were inferred based on a stochastic partial differential equation via integrated nested Laplace approximations. In brief, the inferring process relied on a very fine mesh consisting of small triangles, and the value of W_i is determined depending on the location of i within a triangle. In this study, the Waitākere Ranges parkland study area was converted into a fine mesh that consisted of small triangles for where kauri were sampled and large triangles for where the trees were not sampled or outside of the study area boundary (Figure 3-1). For the small and large triangles, the maximum length of triangle edge was set as 1/15 and 1/5, respectively, of the diameter of the study area. All the parameter values for generating the mesh were based on recommendations provided by Moraga et al. (2021). The diameter was calculated as the distance of easting difference between the east-most and west-most kauri sampled. Cut-off values were set as 1/5 of the maximum length of the small triangle. The use of cut-off values was to avoid generating too many small triangles where kauri were closely located to each other to decrease the computational burden. A coefficient of the Matérn covariance function was set as 0.5, which is identical to the exponential covariance function. The modelling was developed in R using the contributed INLA package (R Core Team, 2020).

Once the model was established, the standardised residuals of the models were calculated, and the covariance was examined by variogram to investigate whether the use of a spatial model properly adjusted for the remaining spatial correlation. Also, the standardised residuals were plotted over the study area to visually examine whether there was any distinctive spatial pattern in the residuals. Variables were retained in the final models if the 95% credible intervals (Bayesian equivalent of confidence intervals) for their coefficients did not overlap the null value, if they were significant in the non-spatial model, or if they were considered a biological confounder. Although the measure of association calculated for this study was the prevalence odds ratio (POR), it was presented and interpreted as the prevalence ratio (PR) and assumes that the POR is a good approximation of PR in this study to aid interpretation.

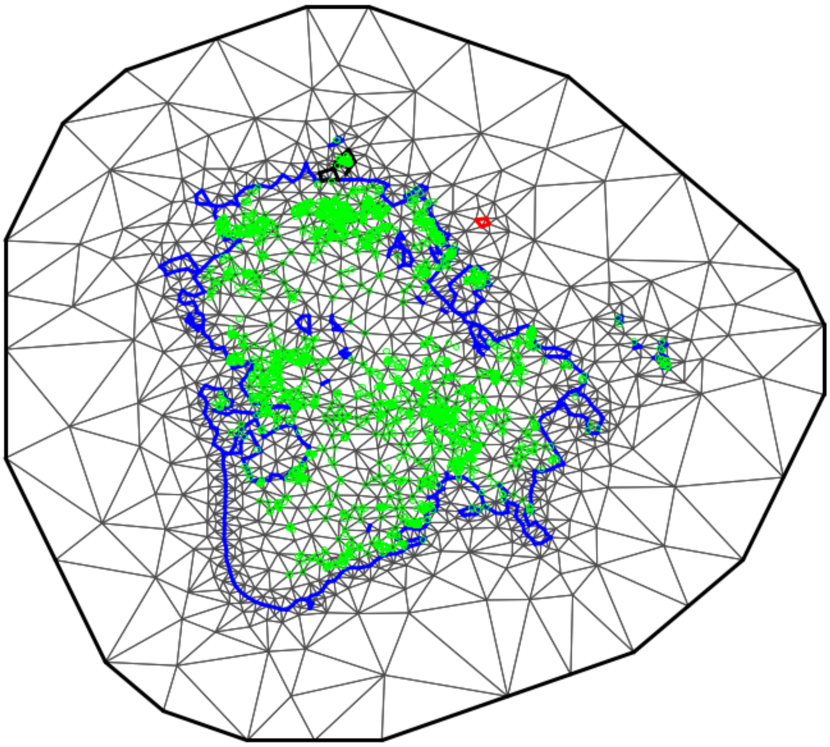

Figure 3-1. A mesh generated for a stochastic partial differential equation via integrated nested Laplace approximations for spatial multivariable models. Blue line indicates the boundary of Waitākere Ranges Regional Park and green dots are the location where kauri were sampled. Red line indicates a disjunct area of Waitākere Ranges Regional Park where no kauri were sampled. The black line denotes areas outside the study area.

3.4 Results

Ngā hua

3.4.1 Initially selected variables

Among 101 potential risk factors for each outcome variable, 39 and 29 variables showed a P-value < 0.2 for the presence of symptomatic kauri and detection of *P. agathidicida*, respectively. The result of the univariable screening tests for the variables with P-value < 0.2 is presented in Appendix C, and the association between the variables are illustrated as a causal path diagram in Figure 3-2 (for presence of symptomatic kauri) and Figure 3-3 (for detection of *P. agathidicida*). In the figures, variables in green with a black triangle are potential risk factors selected for the multivariable model. Variables in white are those omitted from the model due to being highly correlated with the selected potential risk factors, whereas variables in grey are discarded for reasons such as containing too many missing values or being an (in)direct result of the outcome variable. Green lines between any two selected risk factors indicate a potential confounding effect based on discussion with experts.

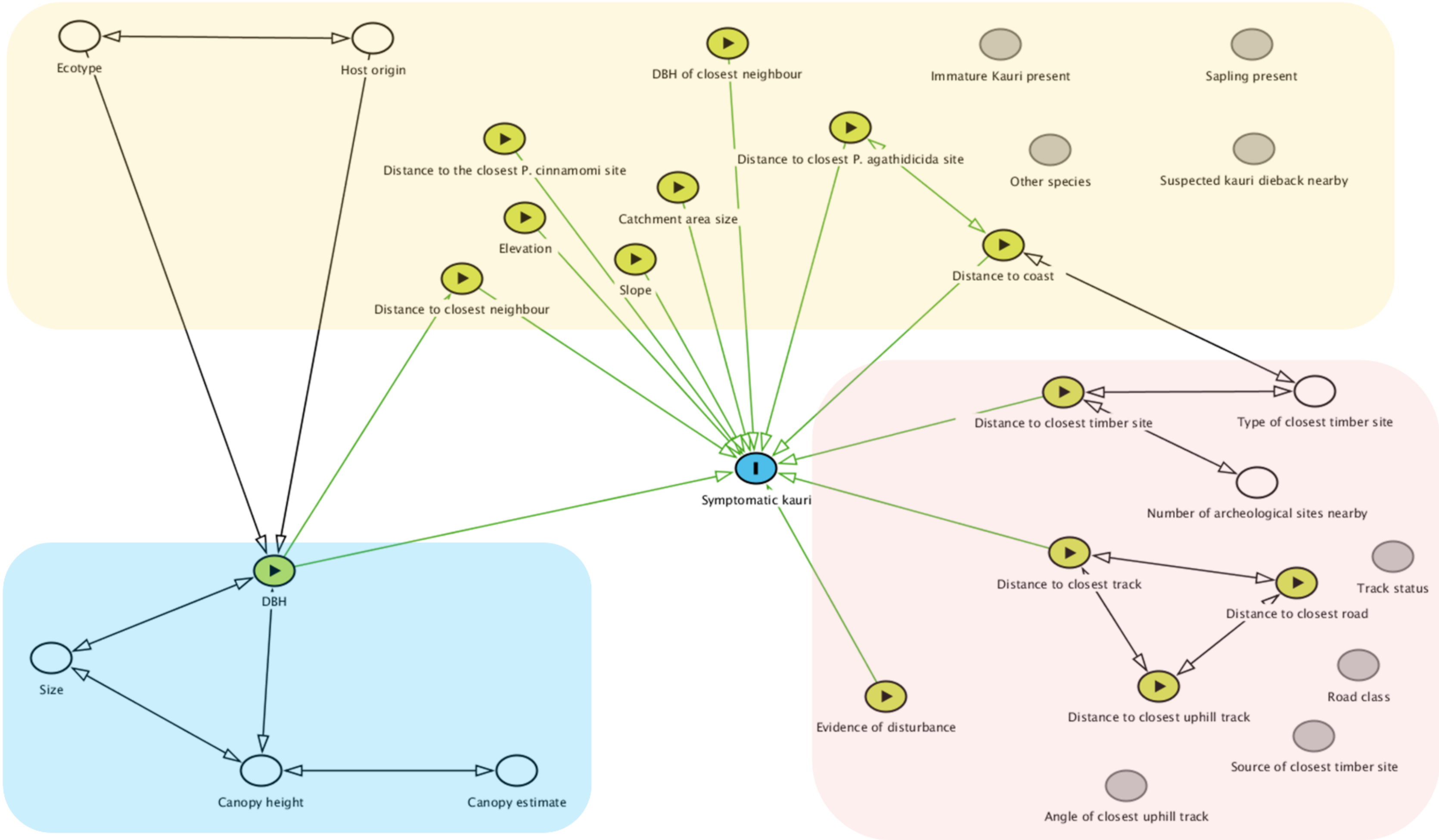

Figure 3-2. A path diagram of potential risk factors for the presence of symptomatic kauri in the Waitākere Ranges Regional Park, Auckland. The variables are grouped in three categories: (1) individual tree factors (blue square), (2) environmental factors (yellow square), and (3) anthropogenic factors (red square). Please note that not all the correlations between variables are shown to enhance readability.

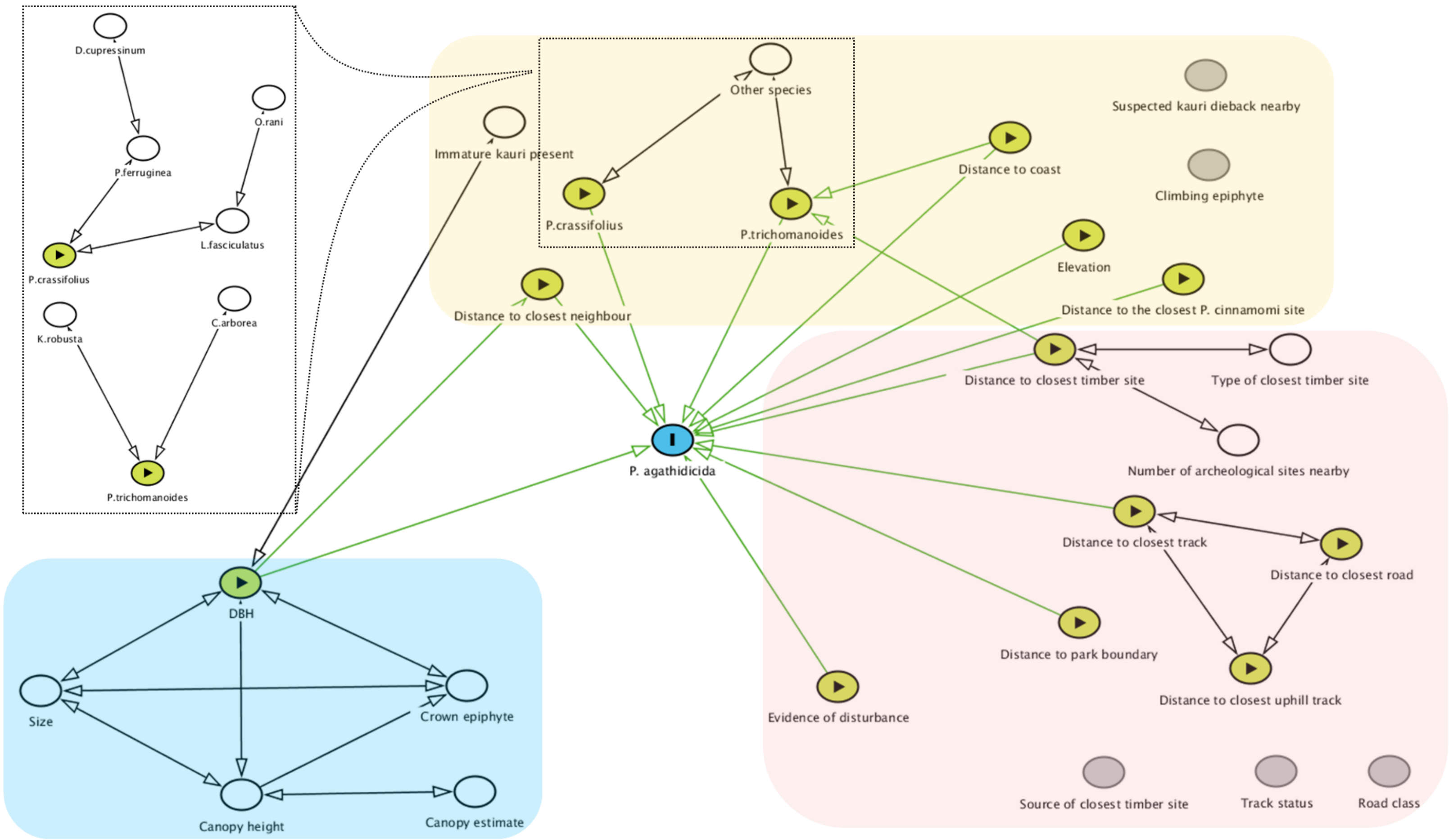

Figure 3-3. A path diagram of potential risk factors for the detection of *Phytophthora agathidicida* in kauri of Waitākere Ranges Regional Park, Auckland. The variables are grouped in three categories: (1) individual tree factors (blue square), (2) environmental factor (yellow square), and (3) anthropogenic factor (red square). Please note not all the correlations between variables are shown to enhance readability. Where *P. crassifolius* is lancewood (*Pseudopanax crassifolius*) and *P. trichomanoides* is tanekaha (*Phyllocladus trichomanoides*).

3.4.2 Results of non-spatial models

For the presence of symptomatic kauri, three models (one for each of the road/track variables) were built. The variables: diameter at breast height (DBH); distance to the closest neighbouring tree; distance to the closest *P. agathidicida* site; distance to the closest coast; distance to the closest timber site and the relevant road/track variable remained across the three final models due to either biological or statistical significance after the variable selection process. The number of observations for the three final models for symptomatic kauri presence with either the distance to the closest track, road, and uphill track was 2094, 2094, and 1856, respectively.

For the detection of *P. agathidicida*, three models (one for each of the road/track variables) were built. The same variables as the disease model (except the distance to the closest *P. agathidicida* site) remained in the final models after variable selection, along with distance to closest *P. cinnamomi* site and elevation. The three final models for the detection of *P. agathidicida* with the distance to the closest track, road, and uphill track were based on 729, 729, and 644 observations, respectively. The results of the final non-spatial multivariable models are presented in Appendix D.

To compare the three models with different key variables, the same dataset for each outcome was used (based on the uphill track variable). It reduced the size of complete datasets to 1862 and 644 observations for the presence of symptomatic kauri and detection of *P. agathidicida*, respectively. Based on these datasets, non-spatial multivariable models were reconstructed and compared. The AIC and AUC values of the reconstructed final models depending on the inclusion of different key variables of interest (i.e., the distance to the closest track, closest road, and closest uphill track) are presented in

Table 3-1. The results indicate that, although small differences in the measure of model fitness occurred, the final models including the distance to the closest uphill track for symptomatic kauri presence and the distance to the closest track for *P. agathidicida* detection best explained the data. Based on this, a final non-spatial multivariable model for each outcome variable was chosen.

Table 3-1. A comparison of final non-spatial multivariable logistic regression models incorporating either the distance to the closest track, distance to the closest road, or distance to the closest uphill track. Values are the Akaike information criteria or the area under the ROC curve for each model. The model with its value underlined indicates the model that best explained the data.

Model outcome	**Distance to the closest**		
	Track	Road	Uphill track
Presence of kauri dieback disease			
Akaike information criteria (AIC)	1721.5	1722.4	1720.4
Area under the ROC curve (AUC)	0.693	0.692	0.695
Detection of *Phytophthora agathidicida*			
Akaike information criteria (AIC)	353.8	356.1	354.6
Area under the ROC curve (AUC)	0.836	0.832	0.836

The variograms of the standardised residuals from the multivariable models for symptomatic kauri presence (**Figure 3-4**) and *P. agathidicida* detection (**Figure 3-5**) indicated a weak remaining spatial correlation at close distance (up to approximately 35 metres), suggesting a need to use a spatial model (Bayesian geostatistical multivariable logistic regression) for symptomatic kauri presence (and potentially *P. agathidicida* detection as well) to account for the remaining spatial correlation. Although the variogram for the detection of *P. agathidicida* did not provide strong evidence of remaining spatial correlation, this may have been due to the smaller sample size compared with the symptomatic kauri outcome.

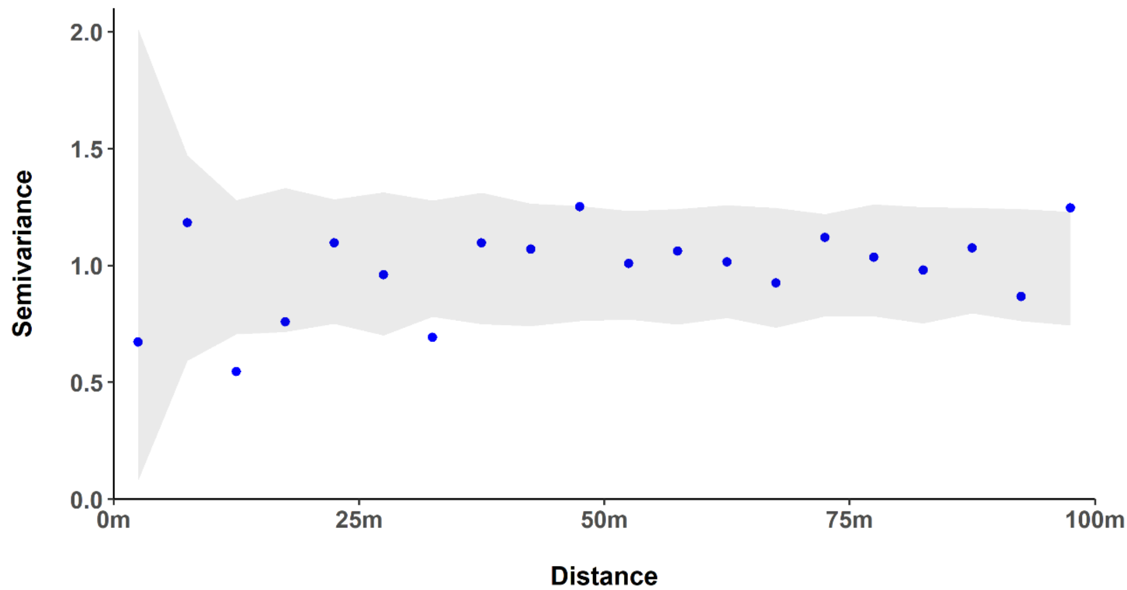

Figure 3-4. A variogram of standardised residuals of a non-spatial multivariable logistic regression model for the presence of symptomatic kauri in the Waitākere Ranges parkland, Auckland (blue points). Any blue points outside of the grey area indicate a spatial correlation, where the grey area was computed by permutation of the standardised residual 500 times.

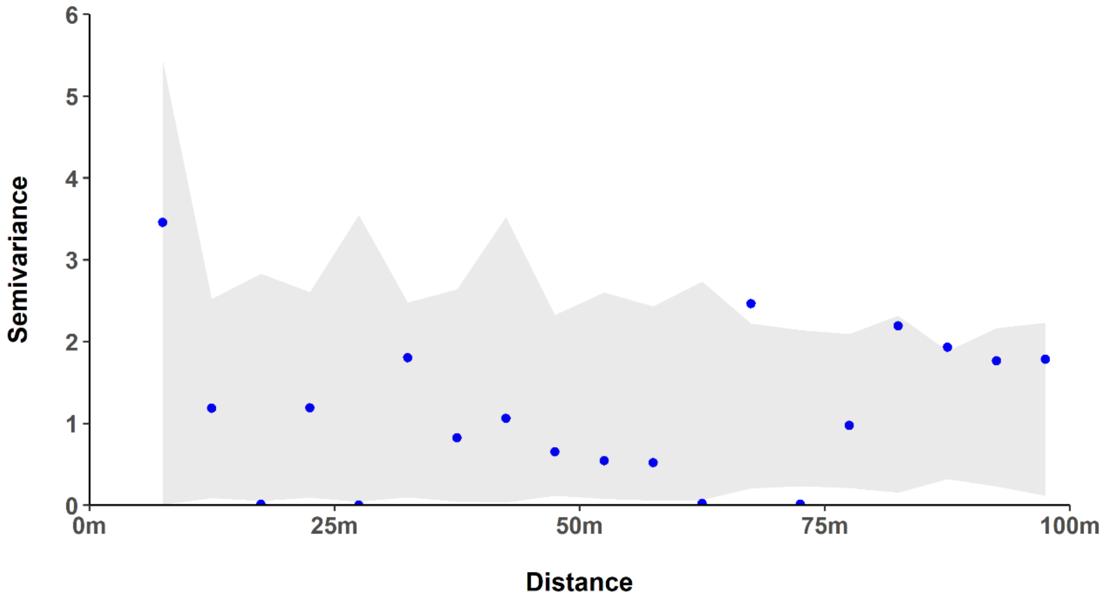

Figure 3-5. A variogram of standardised residuals of a non-spatial multivariable logistic regression model for the detection of *Phytophthora agathidicida* in kauri of the Waitākere Ranges parkland, Auckland (blue points). Any blue point outside of the grey area indicates a spatial correlation, where the grey area was computed by permutation of the standardised residual 500 times. However, this variogram has a low sample size so it could be an impractical indicator of spatial correlation.

3.4.3 Results of spatial models

The results of the spatial multivariable models are presented in Table 2 (for symptomatic kauri presence) and Table 3 (for *P. agathidicida* detection). Note that there is a transition from talking about significance and p-values with the frequentist based non-spatial models to association and credible intervals with the Bayesian spatial models (refer to Kruschke and Liddell (2018) for further reading on how these differ).

There was a small difference of coefficient values between non-spatial and spatial models for both outcomes. This is because only a weak spatial correlation was indicated from the variograms. However, the coefficient of the distance to the closest coast was greatly affected by adjusting the spatial correlation for the *P. agathidicida* model. After adjusting for spatial autocorrelation, the strength of the association between some of the other explanatory variables and either symptomatic kauri or *P. agathidicida* was both reduced (the point estimates were closer to 1) and became more uncertain (i.e., the magnitude of the credible intervals around the association measure increased and included one). For example, in the model for the detection of *P. agathidicida*, after accounting for spatial autocorrelation in the data, the upper band of the 95% credible interval of the prevalence odds ratio for the presence of tanekaha (*Phyllocladus trichomanoides*) nearby, included the value of one. This indicates an association between *P. agathidicida* and the presence of tanekaha, with a small probability (<5%) that the association is either less than or equal to one (i.e., no association). These have been referred to as associations of note in the discussion.

The prevalence of symptomatic kauri decreased in trees with increasing distance from *P. agathidicida* sites and increasing distance from a timber site and increased in trees with increasing DBH of kauri. Examples are provided in **Table 3-2** and **Figure 3-6**. In addition, associations of note were detected with a reduction in prevalence odds of symptomatic kauri with increased distance from the closest neighbouring tree and closest uphill track. A smaller association with distance from coast was observed after adjusting for spatial autocorrelation.

The prevalence odds of kauri detected with *P. agathidicida* reduced with increasing elevation, greater distance to a neighbouring tree, historic timber site or the closest coast (**Table 3-3** and **Figure 3-7**). In addition, associations of note were detected with an increase in prevalence odds of *P. agathidicida* with the presence of tanekaha and a reduction in prevalence odds of *P. agathidicida* with increased distance from the closest track. There was a low probability of an association with *P. cinnamomi* after adjusting for spatial autocorrelation. No association was found with DBH, however it remained in the model as a potential confounder for the closest neighbouring tree relationship.

Table 3-2. A result of spatial multivariable logistic regression model for the presence of symptomatic kauri, consistent with kauri dieback in the Waitākere Ranges Regional Park, Auckland. The median (95% credible interval (CI)) of the coefficients and prevalence odds ratio of the potential risk factors are presented, in order of the strength of association.

Variables	**Coefficient (95% CI)**	**Prevalence odds ratio (95% CI)**
Intercept	-0.805 (-1.317 ~ -0.331)	Reference
Distance to the closest *P. agathidicida* site (100 m)	-0.055 (-0.077 ~ -0.034)	0.947 (0.926 ~ 0.967)*
Distance to the closest timber site (100 m)	-0.027 (-0.046 ~ -0.009)	0.973 (0.955 ~ 0.991)*
Diameter at breast height (10 cm)	0.076 (0.047 ~ 0.106)	1.079 (1.048 ~ 1.112)*
Distance to the closest neighbouring tree (m)	-0.091 (-0.189 ~ 0.005)	0.913 (0.828 ~ 1.005)
Distance to the closest uphill track (100 m)	-0.055 (-0.122 ~ 0.011)	0.947 (0.885 ~ 1.011)
Distance to the closest coast (100 m)	-0.006 (-0.014 ~ 0.003)	0.994 (0.986 ~ 1.003)

Interpretation of factors with the strongest associations (*) after accounting for other variables in the model, demonstrating the effect of one unit difference from the average value of the variable:

- Distance to the closest *P. agathidicida* site: The prevalence odds of symptomatic kauri was 0.95 times (5% less) for each 100 m increase in distance from the closest *P. agathidicida* site. i.e., symptomatic kauri prevalence was higher closer to *P. agathidicida* sites.
- Distance to the closest timber site: The prevalence odds of symptomatic kauri was 0.97 times (3% less) for each 100 m increase in distance to the closest timber site. i.e., symptomatic kauri prevalence was higher closer to historical timber sites.
- Diameter at breast height (DBH): The prevalence odds of symptomatic kauri for trees with a DBH of 70 cm was 1.08 times (8% greater) than that of kauri with a DBH of 60 cm i.e., symptomatic kauri prevalence increased with tree size.

Table 3-3. A result of spatial multivariable logistic regression model for the detection of *Phytophthora agathidicida* in kauri soil samples in the Waitākere Ranges Regional Park, Auckland. The median (95% credible interval (CI)) of the coefficients and prevalence odds ratio of the potential risk factors are presented, in order of the strength of association.

Variables	**Coefficient (95% CI)**	**Prevalence odds ratio (95% CI)**
Intercept	1.150 (-1.806 ~ 4.403)	Reference
Elevation (100 m)	-0.906 (-1.907 ~ -0.046)	0.404 (0.149 ~ 0.955)*
Distance to the closest neighbouring tree (m)	-0.456 (-0.777 ~ -0.178)	0.634 (0.460 ~ 0.837)*
Distance to the closest timber site (100 m)	-0.132 (-0.259 ~ -0.034)	0.877 (0.772 ~ 0.966)*
Distance to the closest coast (100 m)	-0.060 (-0.164 ~ -0.005)	0.942 (0.848 ~ 0.995)*
Presence of *P. trichomanoides* (tanekaha)	0.664 (-0.161 ~ 1.566)	1.942 (0.851 ~ 4.787)
Distance to the closest track (100 m)	-0.140 (-0.437 ~ 0.129)	0.870 (0.646 ~ 1.138)
Distance to the closest *P. cinnamomi* site (100 m)	-0.024 (-0.060 ~ 0.007)	0.977 (0.942 ~ 1.007)
Diameter at breast height (10 cm)	0.038 (-0.047 ~ 0.119)	1.038 (0.954 ~ 1.126)

Interpretation of factors with the strongest associations (*) after accounting for other variables in the model, demonstrating the effect of one unit difference from the average value of the variable:

- Elevation: The prevalence odds of *P. agathidicida* was 0.41 times (59% less) for each 100 m increase in elevation. i.e., *P. agathidicida* prevalence was higher at lower elevations.
- Distance to the closest neighbouring tree: The prevalence odds of *P. agathidicida* was 0.64 times (36% less) for each 1 m increase in distance away, i.e., the wider the gap between the kauri tree and its closest neighbour, the lower the *P. agathidicida* prevalence.
- Distance to the closest timber site: The prevalence odds of *P. agathidicida* was 0.88 times (12% less) for each 100 m increase in distance away, i.e., *P. agathidicida* prevalence was higher closer to historic timber sites.
- Distance to the closest coast: The prevalence odds of *P. agathidicida* was 0.94 times (6% less) for each 100 m increase in distance away, i.e., *P. agathidicida* prevalence was higher closer to the coast.

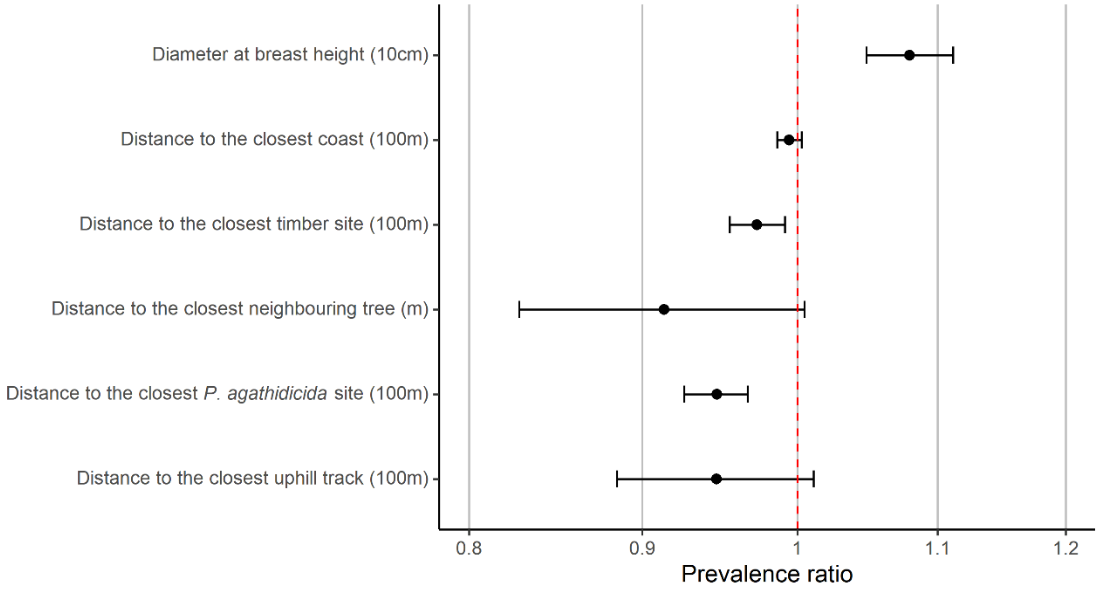

Figure 3-6. A forest plot depicting the prevalence odds ratio (PR) of potential risk factors for the presence of symptomatic kauri in the Waitākere Ranges parkland, Auckland. The black dot and horizontal bars respectively indicate the PR and its 95% credible interval (CI). Risk factors with their PR and 95% credible intervals fully to the left or right of the red dashed vertical line are associated with the outcome, where most of the PR and 95% credible intervals are to the left or right of the red line the association is protective or increases the prevalence odds of symptomatic kauri respectively, and where the black dot and credible intervals are centred on the red dashed line, the strength of the association is low (e.g., distance to coast).

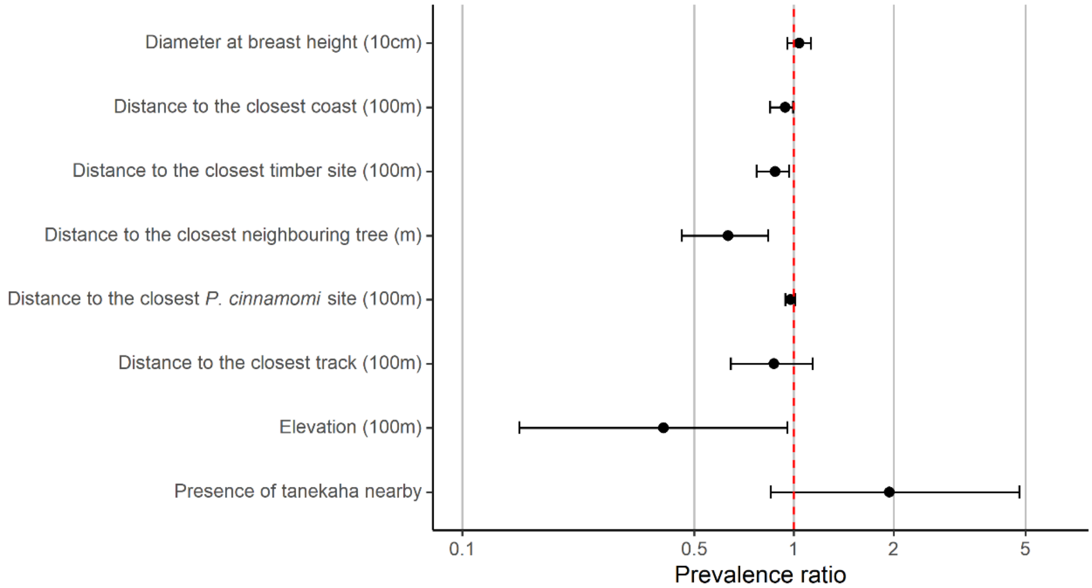

Figure 3-7. A forest plot depicting the prevalence odds ratio (PR) of potential risk factors for the detection of *Phytophthora agathidicida* in kauri in the Waitākere Ranges Regional Park, Auckland. The black dot and horizontal bars respectively indicate the PR and its 95% credible interval. Risk factors with their PR and 95% credible intervals fully to the left or right of the red dashed vertical line are associated with the outcome, where most of the PR and 95% credible intervals are to the left or right of the red line the association is protective or increases the prevalence odds of *P. agathidicida* respectively, and where the black dot and credible intervals are centred on the red dashed line, the strength of the association is low (e.g., diameter at breast height). Note that the x axis is illustrated in a log scale and has a wider range than the symptomatic kauri plot.

3.5 Discussion

Te matapaki

The aim of this study was to identify which environmental, host, anthropogenic and pathogen-related risk factors were associated with either symptomatic kauri or presence of *P. agathidicida*. It also aimed to identify factors much less likely to be causally related to symptomatic kauri or *P. agathidicida* presence. For those that were associated the aim was to generate hypotheses on the possible nature of the relationships. This will inform new studies designed to answer questions about these relationships and identify management interventions to enhance kauri health.

Proximity to *P. agathidicida* sites was strongly associated with symptomatic kauri in the symptomatic kauri model, so discussing the *P. agathidicida* model first will provide insight into the symptomatic kauri model. Below we present the associated risk factors found through the spatial models and discuss potential causal or non-causal hypotheses for these relationships. The strongest associations are discussed first, followed by the associations of note.

3.5.1 *P. agathidicida* model

There were four risk factors that were strongly associated with *P. agathidicida* detection in soil, three of which were environmental factors and one anthropogenic factor. In addition there were three associations of note, two were environmental and one was anthropogenic. It is easier to intervene with anthropogenic factors than environmental factors which tend not to be modifiable; however, they can inform management such as placement of amenities or replanting areas.

3.5.1.1 Elevation

The prevalence of *P. agathidicida* in kauri was higher at lower elevations, after accounting for all other factors. This was an interesting finding, especially as it remained highly associated after coastal proximity was controlled for. Previous reports of a negative relationship between *P. cinnamomi* prevalence and elevation in Southeast Australia support this finding (Wilson et al., 2003). The association may be due to environmental constraints on pathogen survival, such as the warming that occurs with increased solar radiation, or changes in soil pH and moisture. It may also be related to opportunities for vectored or natural spread. As a soil-borne water-mould, it is more likely that prevalence due to natural spread would be greater at lower elevations as water is carried downhill. This is consistent with the direction of effect in the model and with research on other *Phytophthora* species showing that propagules are washed down catchments (Redondo et al., 2018). However, other unmeasured factors such as soil type and chemistry may also affect the presence of *P. agathidicida* in soil and differ with elevation, especially in areas where significant disturbance has occurred. When the soil samples for this study were collected, additional volumes of soil were taken for distribution to a range of collaborating researchers, and soil chemistry or microbiota relationships may become clearer when their research is completed. Elevation is not a modifiable variable, but this result provides information about potentially higher risk areas for future surveillance or replanting.

3.5.1.2 Distance to historic timber sites

The prevalence of *P. agathidicida* was higher closer to historic timber sites, after accounting for other factors. This association could be related to other unmeasured confounding factors but suggests a hypothesis of introduction and spread through increased soil disturbance near these sites. This association was also observed for the disease model, potentially suggesting that inoculum load is greater in these areas, increasing disease risk. It is also reasonable to assume that *P. agathidicida* is easier to detect in soils with a high inoculum load. An increased pathogen prevalence near historic logging has also been observed in other *Phytophthora* diseases (Socorro Serrano et al., 2015, Homet et al., 2019).

3.5.1.3 Distance to coast

The prevalence of *P. agathidicida* was higher closer to the coast. It is possible that the association observed in this study may relate to other unmeasured confounding factors such as higher human habitation and disturbance or climatic differences between coastal areas and the inland forest. Coastal areas are where most modification has happened over time in the Waitākere Ranges (S. Leighton, Auckland Council, pers. comm.) and this association could be related to historic introduction and spread pathways of *P. agathidicida*, a hypothesis supported by the association with historic timber sites. It is also consistent with mātauranga Māori (indigenous knowledge) that when the moana (ocean) is depleted, so too is the whenua (land), making the trees near the coast more vulnerable from this exploitation. Another possible explanation is that rainfall amounts are up to 3 times higher in the centre of the Park compared with the coastal fringe. For example, the range in rainfall is approximately 1 m in Piha through to just over 3 m in the upper Nihotupu Basin (S. Leighton, Auckland Council, pers. comm.). This raises the hypothesis that *P. agathidicida* may be more prevalent in dryer areas or where the host is under increased pressure from dry conditions; future investigation into the relationships between rainfall and other climatic factors on *P. agathidicida* presence would be useful. Depth to water was not associated with an increase or decrease in *P. agathidicida* prevalence (or disease) and typically *Phytophthora* species are more associated with wet soils (e.g., Gyeltshen et al. (2021), Donald et al. (2020), Weste and Ruppin (1975), Weste and Vithanage (1979), Venette and Cohen (2006)), although Sena et al. (2019) found *P. cinnamomi* was more prevalent in drier areas in Kentucky, United States. Another potential hypothesis is that dry areas may have a higher presence of the oospore life stage, which is longer lived and may be easier to detect in the soil bioassay.

The higher prevalence of *P. agathidicida* detection near both the coast and historic timber sites being associated with an introduction pathway is supported by research by Weir et al. (2015) and Winkworth et al. (2021). *Phytophthora agathidicida* is likely an introduced species into New Zealand as Weir et al. (2015) indicate that the centre for diversity of Clade 5 Phytophthora species which includes *P. agathidicida* is East Asia/Pacific. Winkworth et al. (2021) provided some evidence that the limited number of *P. agathidicida* isolates from the Waitākere Ranges they examined (Huia (3) and Piha (1)) were diversifying from the late 1700s onwards, although the authors acknowledge the research requires further sampling. This study raises the hypothesis of historical introduction from the coast and human assisted movement of *P. agathidicida* through timber and other disturbances. This is also supported by the limited distribution of *P. agathidicida* around the periphery of the study area found in the Chapter 2.

3.5.1.4 Distance to closest neighbouring tree

The lesser the gap between the monitored tree and its closest neighbouring tree, the higher the prevalence of *P. agathidicida*. It is postulated that with 20% of neighbouring trees also being kauri, this is likely to indicate enhanced localised spread of *P. agathidicida* between kauri within a stand. In addition, soil samples may be collecting root material from several kauri and maximising the opportunity for *P. agathidicida* detection.

3.5.2 Symptomatic kauri model

There were three risk factors that were strongly associated with symptomatic kauri, one anthropogenic factor, one host related factor and one pathogen related factor. Two other environmental risk factors and one anthropogenic factor were associations of note.

3.5.2.1 Distance to closest *P. agathidicida* site

Trees that were closer to a *P. agathidicida* site had a higher probability of being a symptomatic kauri than trees that were further away from *P. agathidicida* sites, indicating localised tree to tree spread. This finding was not unexpected and is supported by extensive research showing a strong association between kauri dieback disease and *P. agathidicida* (Bradshaw et al., 2020). Both pathogenicity and Koch's postulates have been demonstrated between *P. agathidicida* and kauri dieback (Bellgard et al., 2016, Gadgil, 1974) and the case definition for symptomatic kauri in our model was based on expert agreement on the symptoms of kauri dieback caused by *P. agathidicida*. Not all symptomatic trees were near *P. agathidicida* detected sites, which indicates that while *P. agathidicida* management will be important in reducing disease, some other factors are also contributing to a decline in kauri health and should be investigated.

3.5.2.2 Distance to historic timber site

Symptomatic kauri prevalence was higher the closer the tree was to historic timber sites, after accounting for proximity to *P. agathidicida* and other risk factors.

This indicates that the relationship is beyond that of an introduction pathway of the pathogen. It is hypothesised that proximity to historical timber sites is an indication of soil disturbance and tree damage. Historical logging was extremely destructive to surrounding forest from not only the felling of kauri but the entire process, including the creation of the timber mills, digging of saw pits and then radiating out from these areas, the chutes, bullocks and tramways to move kauri logs to site for processing (Figure 3-8). It is also possible that this association is a proxy for wider disturbance of sites after logging. Often farming was attempted in the wake of logging, leading to full clearance of remaining forest and loss of topsoil. The Manukau, Waitematā and Kaipara harbours are full of silt that would have once been rich soils that were washed away following forest clearance by early Europeans (Hayward et al., 2006).

There is potential to investigate in finer detail the strength of the relationship between timber mills, saw pits and other sites associated with kauri logging and potentially other large soil disturbance activities, such as dam building, using this data and historical records.

It may also be relevant to query and isolate other archaeological features from available datasets (i.e., the cultural heritage inventory, historic tracks and tramlines) to determine the significance of additional archaeological classes (e.g., historic access and transport, historic land use, European and pre-European settlement and activity) in relation to symptomatic kauri and *P. agathidicida* distribution.

Figure 3-8. Historic images of i) a kauri log on a cutting table inside the Piha timber mill (photographer A.P. Godber, Auckland Libraries Heritage Collections JTD-04L-00124) and ii) a felled kauri crown showing surrounding forest devastation after the sawn log has been removed (photographer A.P Godber, Auckland Libraries Heritage Collections JTD-04D-03327).

3.5.2.3 Kauri diameter at breast height (DBH)

The prevalence of symptomatic kauri increased with the size (DBH) of the kauri host. The results were surprising from a physiological viewpoint as *P. agathidicida* infection reduces water uptake in kauri roots, decreasing the infected tree's ability to replace water lost through evaporation at the leaf surface (Killick, 2022). Infected trees are also less conservative of water, operating at a narrower hydraulic safety margin overall (Killick, 2022). While this is true independent of kauri size, larger trees have greater water storage capacitance than smaller trees (Kaplick et al., 2017); therefore, larger kauri should decline slower or later than smaller kauri. On the other hand, increasing tree size affects the availability of soil water, which may also be a factor (Ruess et al., 2021). Bradshaw et al. (2020) state that smaller trees generally decline at a faster rate than larger trees, although it is difficult to measure the rate of decline in individual trees without knowing when they became infected. In a cross-sectional prevalence study, subjects are observed at a single point in time and prevalence can be influenced by the duration of disease (Grimes and Schulz, 2002). If larger trees survive with disease longer than smaller trees, then they are likely to make up a larger proportion of the prevalent population as smaller trees with disease are removed when they die. This survey provides the baseline measure of symptomatic kauri prevalence and repeated surveys on the same cohort of trees will provide more evidence of this relationship by measuring the incidence of new symptomatic kauri developing over time.

It is also biologically plausible that the high proportion of trees that are regenerating from logging that occurred in the late 1800s and early 1900s (i.e., 100-120 year old trees transiting from ricker to intermediate size classes (Bergin and Steward, 2004)) are facing increased competition with higher vulnerability to disease which could be driving this association. The distribution of DBH in trees included in this study was shown to be left skewed towards smaller (average 60 cm) trees, with few very large mature trees (Chapter 2). The association was strongly linear when tested, but this relationship requires more investigation. In addition, large trees within the Waitākere Ranges, especially around the Cascade area where symptomatic kauri risk was high (Chapter 2) were extensively bled for kauri gum in the same period as logging occurred increasing root disturbance and affecting tree health. There may also be a physiological reason for some protection from symptoms in younger or smaller trees, such as greater root growth rates in some younger trees (Rosenvald et al., 2013). The strength of the association between tree size and symptoms was strong and this could be an important finding for the long-term management of kauri. The size classes of kauri cannot be manipulated for management, however trees at greater risk could be prioritised for protection and enhanced monitoring to inform early treatment.

It is possible that the association between symptomatic kauri prevalence and DBH was an unmeasured confounding factor, for example, trees with a DBH of less than 10 cm were deliberately excluded from the study because symptoms are hard to detect on very small trees. It is also possible that symptoms, in particular basal lesions are more obvious on larger trees, which may have contributed to the observed association.

3.5.3 Associations of note

3.5.3.1 Distance to tracks

The distance to tracks (closest or uphill) was significantly associated with *P. agathidicida* detection and disease in the non-spatial models. However, the association reduced (the point estimates were closer to 1) and became more uncertain (i.e., the magnitude of the credible intervals around the association measure increased and included one) after adjusting for spatial autocorrelation. It is biologically plausible that an association exists and additional analysis of different track types, historic tracks, and whether there is a similar association between ridgelines and *P. agathidicida* and symptomatic kauri prevalence will provide a more complete picture of the relationships with track and transport networks. It would also be possible to undertake quantitative bias analysis on the non-spatial model results to investigate if misclassification of the outcome variables is masking a greater effect.

3.5.3.2 Distance to closest neighbouring tree

The association towards a lower prevalence of disease as the distance between monitored trees and their closest neighbour tree increases contrasts with the relationship between an increase in symptomatic kauri prevalence as tree size increases. As mean tree size increases, it would be expected to see a decline in density suggesting greater distances between trees. It is possible that these relationships are confounded by whether the nearest neighbour is a kauri or not, which was the case in 20% of trees (Chapter 2). Further investigation of the data to understand size classes in relation to closest neighbouring tree species and the importance of this relationship is possible with the data collected during this study using different outcome variables.

3.5.3.3 Presence of tanekaha

An interesting association between *P. agathidicida* and the presence of tanekaha (*Phyllocladus trichomanoides*) nearby (within 10 m of the monitored tree) was found. During screening, 8 of the 15 common plant species showed an initial association and formed into two distinct groupings (**Figure 3-3**) when inter-variable correlations were investigated. One was represented best by lancewood (*Pseudopanax crassifolius*) and the second was best represented by tanekaha. The groupings are well aligned with the developmental phases of kauri forest, i.e., mature, old-growth forest and newer regenerating forest respectively (Ahmed and Ogden, 1991, Ogden and Stewart, 1995). Presence of tanekaha could be a proxy for forest characteristics differentiating these two forest types that may favour *P. agathidicida* or be related to increased disturbance and spread. Tanekaha are also more common on drier ridges and in areas with extreme conditions (Kaplick et al., 2018). Another potential biological association could be related to the possibility of tanekaha acting as an alternative host for *P. agathidicida*. To date there have been some laboratory indications that tanekaha may be an alternative host for *P. agathidicida* (Ryder et al., 2016), however no field evidence exists as yet. As with the other factors of note, the relationship remains uncertain and further investigation is warranted. The data collected in this study will aid researchers to locate kauri sites with tanekaha where *P. agathidicida* has been detected for future studies.

3.5.3.4 Distance to closest *P. cinnamomi* site

There was no association between symptomatic kauri and *P. cinnamomi,* however, there was a weak initial association between *P. agathidicida* and distance to the closest *P. cinnamomi* site in the non-spatial model, with a very small decrease in *P. agathidicida* prevalence with increasing distance from *P. cinnamomi* sites. However, this relationship became very weak in the spatial model. It does raise an interesting hypothesis that the introduction pathways of *P. agathidicida* and *P. cinnamomi* may have been similar, however from the *P. cinnamomi* distribution results in Chapter 2, historically in New Zealand (Podger and Newhook, 1971) and internationally (Sena et al., 2019) it is clear that *P. cinnamomi* is much more efficient at spreading within the landscape, most likely due to a much wider host range.

3.5.4 Variables of interest with no association found

There were several variables of note that were found to have no association to symptomatic kauri and/or detection of *P. agathidicida* in our models.

P. cinnamomi was not associated with symptomatic kauri in this study, a factor that has been uncertain in the past (Podger and Newhook, 1971, Bellgard et al., 2013, Beever et al., 2010), although Beever et al. (2009) also found no association with disease in kauri within the Waipoua Forest in 2003. Podger and Newhook (1971) concluded *P. cinnamomi* was important in disease observed in older 80-100-year-old regenerating stands (now 120-150 years old), however when the site was revisited in 2006, remaining trees appeared healthy (Beever et al., 2009).

It was also surprising that the depth to surface water index which gave an indication of areas more prone to being moist or dry was not associated with increased symptomatic kauri prevalence or *P. agathidicida.* It may be that the depth to water index used was not a good model for wet or waterlogged sites (Davison, 2018) which are postulated to enhance infection through weakened roots, higher sporulation and mobility of the motile zoospores as has been observed in other native tree-*Phytophthora* pathosystems (Donald et al., 2020, Jung et al., 2018).

Similarly, it was postulated that the distance to hydrological features would be an associated factor. However, distance to overland flow path (watercourses) did not indicate a relationship with symptomatic kauri or with *P. agathidicida* detection. Despite this, it is considered important to investigate this relationship which could consider stream order or detailed watershed analysis to determine whether a tree's location in the sub-catchment influences *P. agathidicida* or symptomatic kauri prevalence.

It is also important to note that disturbance at the tree base by pigs and other hoofed animals was included in the initial model building but was not significant in the non-spatial model. However, the study design was not optimal to collect data on pig and other potential soil-disturbing and pathogen vectoring pest animal species and no existing geospatial datasets were suitable for investigation. It may be useful to obtain pig surveillance data similar to that used for Bovine TB (*Mycobacterium bovis*) in New Zealand (Nugent et al., 2015). Further research to understand pig density and pest animal relationships with *P. agathidicida* and symptomatic kauri would be helpful.

3.5.5 Study limitations

The symptomatic criteria of the case definition (Chapter 2) used to classify symptomatic and non-symptomatic trees relies on set cut-points for canopy scores (greater or equal to 3 out of 5) and more yellow than green canopy colours, along with the presence of trunk or lateral root basal lesions, which can be caused by physical damage or biological factors. The Stevenson and Froud (2020) case definition we applied states that the symptoms need to be consistent with kauri dieback, as assessed by approved observers. The survey was undertaken by experienced and well-trained observers that were familiar with kauri dieback to reduce the level of misclassification. However, the non-symptomatic class contains trees that can be either healthy or showing a level of ill-thrift below the case definition cut-points. Therefore, the ill-thrift trees will contain both stressed trees from other causes which might recover, and trees that may transition into the prevalent (symptomatic) population. Misclassified ill-thrift trees into the non-symptomatic class are most likely to push prevalence odds ratios towards 1 (the null) and may have reduced effect sizes. Further research looking at modelling specific symptoms with *P. agathidicida* detection may inform an improved case definition to explore risk factors and improve effect size estimates.

For the *P. agathidicida* model, the diagnostic test sensitivity for the soil bioassay is relatively low (details in Chapter 4). That means that we may have missed over a third of the true positives and misclassified them as not detected. As with the symptomatic kauri outcome, this misclassification would most likely lead to an underestimation of the true effect and pushed effect sizes towards the null. Therefore, risk factors that were associated in the final model, but partly crossed the null value, remain likely to be biologically important and have been considered for hypothesis generation. The sample size for the *P. agathidicida* model was lower than the symptomatic kauri model and this was evident with higher spatial variability and wider credible intervals. Sample sizes for soil sampling in future risk factor studies may need to be increased.

3.6 Conclusion

Te whakatau

For the symptomatic kauri model, the strongest association was between symptomatic kauri and proximity to *P. agathidicida* sites (point locations of *P. agathidicida* detections) which reinforces the need to manage *P. agathidicida* to reduce tree-to-tree spread and symptom development. Symptomatic kauri prevalence was also higher closer to historic timber sites (reducing with distance away from them) and increased with increasing tree size (DBH).

For the *P. agathidicida* model, associations were found showing *P. agathidicida* prevalence was higher with decreasing elevation, and with decreasing distance from historic timber sites and the coastline. It was also higher as the distance to the closest neighbouring tree decreased. In addition, our results found associations of note that are potentially biologically important between symptomatic kauri prevalence and distance to the coast, neighbouring tree distance, and distance to the closest uphill track; and *P. agathidicida* prevalence and distance to the closest track and presence of tanekaha. These require further investigation, particularly around effect size impacts from misclassification bias.

The results generated hypotheses for further investigation into understanding or managing these relationships, such as managing the distribution of *P. agathidicida* and development of disease through appropriate biosecurity and ecosystem protection measures.

Find out more: kauri@aucklandcouncil.govt.nz
or visit knowledgeauckland.org.nz and
aucklandcouncil.govt.nz

Auckland Council
Te Kaunihera o Tāmaki Makaurau