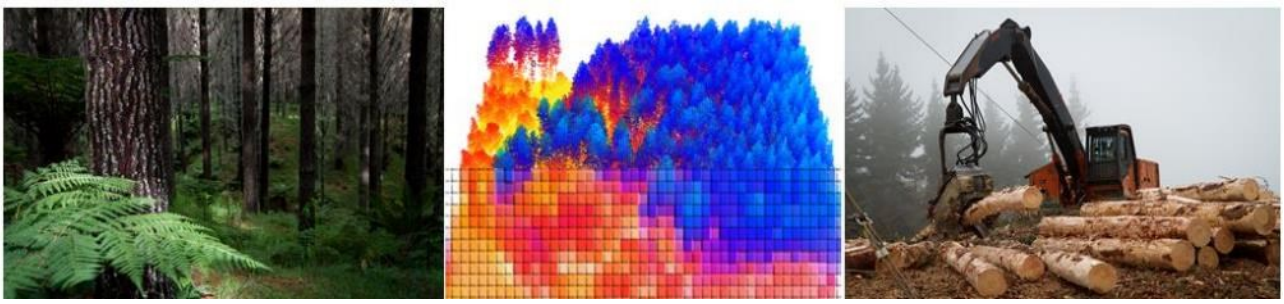
Resilient Forests Research Programme

SCION FORESTS PRODUCTS INNOVATION

Identifying Key Parameter for RNC Severity Modelling and Prediction

Alan Yu Shyang Tan, Stuart Fraser, Nari Williams, Chanatda Somchit



Date: April 2020

Report No: RFP-T005

FOREST GROWERS RESEARCH

TABLE OF CONTENTS

supported by forestgrowers commodity levy

EXECUTIVE SUMMARY

This Technical Note aims to capture the research required for the development of a predictive model for Red Needle Cast (RNC) severity. First, we outline previous work undertaken in this area. Following which, we identify the objectives for our RNC modelling work planned for the Resilient Forest programme. We will use the RNC Susceptible-Infectious model as our basis for modelling RNC severity and identified four key parameters for focusing our efforts on understanding influence of environmental, genetics and management factors on RNC.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

INTRODUCTION

Red Needle Cast (RNC) is a type of foliar disease recognised to be occurring in New Zealand since 2008. Although RNC has been established in New Zealand for several years, we still lack sufficient understanding to predict and efficiently control RNC in the field. Research aim 3 (RA3) of the Resilient Forests programme aims to address this knowledge gap and to develop capabilities, such as predictive models, that can aid disease management.

A key output from the prior *Healthy Trees Healthy Future* (HTHF) research programme is the RNC Susceptible-Infectious (SI) model [1, 2]. The model provides a theoretical basis to model and quantify the dynamics of RNC at the spore production level. From the model and related studies in the HTHF, new questions such as how environmental factors drive RNC and how can the model be used to develop new approaches for estimating disease severity in the field, were raised.

This Technical Note aims to provide an overview of our strategies for modelling RNC severity and development of tools by addressing the following:

1) Identifying key parameters of the SI model and plans for understanding influence of environmental factors on these parameters from a modelling perspective.
2) Providing an overview of the framework and requirements for the development of a tool for RNC severity prediction.

The modelling work described in this document contributes towards advancing our understanding of how environmental factors drive RNC and the development of new knowledge on epidemiology and impacts of foliage diseases in radiata pine.

This document is structured such that in Section 2, we provide a brief overview of prior RNC modelling work and list some of the identified gaps in previous work. In Section 3, we first list the objectives for the Resilient Forests RNC modelling work. We then identify key SI model parameters to focus on and discuss our strategies for modelling these parameters and associated data requirements.

Overview of prior RNC modelling

Research carried out in the HTHF programme has resulted in models for understanding and predicting RNC. We provide a brief overview of these models in the following subsection as context for understanding the current progress in RNC modelling.

RNC SI and Needle Assay Model

A system model describing the Susceptible-Infectious behaviour of RNC [1, 2] was formulated by Wake et al. as part of the effort to understand the dynamics of RNC. The RNC SI model attempts to describe the transition of healthy needles to infected needles from a *Phytophthora* spores production perspective. The model was formulated as follows:

$$S = \frac{dS}{dt} = bS(1 - S) - kSP \tag{1}$$

$$P = \frac{dP}{dt} = gIP\left(1 - \frac{P}{K1}\right) - mP \tag{2}$$

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

where:

- S: amount of healthy needles.
- P: amount of *Phytophthora* spores.
- I: amount of infected needles.
- b: per-capita birth rate of needles.
- k: transmission parameter for spores.
- g: per-capita birth rate of spores.
- m: per-capita death rate of spores.
- $K1$: carrying capacity of spores for a needle.

Focusing on the production and mortality of spores of *Phytophthora pluvialis* and the infection and death of needles allowed us to understand the RNC infection process. This process can be illustrated as a flow diagram, as shown in Figure 1.

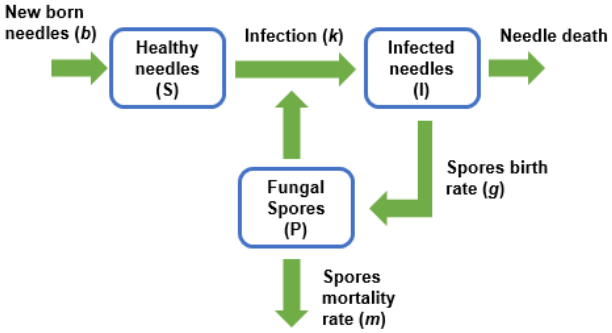


Figure 1. **RNC infection process at needle level, as modelled by the RNC SI model.**

The ability to model various components of the infection process enabled further studies for understanding RNC. For example, Gomez-Gallego et al. [3] extended the model to further understanding of key epidemiological dynamics such as incubation time, infection rates and pathogen mortality rates for susceptible and resistant genotypes. The extended model (referred to as "*needle assay model"*) was fitted to data from the detached needle assays and artificial inoculations experiments.

The needle assay model was formulated as follows:

$$\frac{dS}{dt} = -kSP \tag{3}$$

$$\boldsymbol{\frac{dQ}{dt} = -\frac{dS}{dt} - nQ} \tag{4}$$

where:
$\boldsymbol{Q}$: fraction of infected needles
$\boldsymbol{n}$: per-capita spore death rate.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Experimental observables for isolation rate $\boldsymbol{R_{isol}}$, presence of mycelium $\boldsymbol{R_{myc}}$ and presence of sporangia $\boldsymbol{R_{spor}}$, derived from the detached needle assay study, were defined as follows:

$$\boldsymbol{R_{isol} = Q} \tag{5}$$

$$\boldsymbol{R_{myc} = c_m Q(t - \tau_m)} \tag{6}$$

$$R_{spor} = R_{myc} \tag{7}$$

The needle assay model was formulated based on the following assumptions:
No new needles were born, hence $b = 0$.
Infected needles stay infected, but live pathogen may die.
Time delay for detection by isolation was small (can be ignored).
Time delay for production of mycelium and sporangia were almost the same.
The time delay was treated as a distribution of incubation periods.

These assumptions were reflected in the modifications to the RNC SI model, resulting in Eqn. 3 to 7.

Findings from the needle assay model fitting and analysis study showed that infection rate and incubation period remained the same for resistant and susceptible genotypes. However, higher pathogen death rate and lower detection of mycelium and sporangia were observed in resistant genotypes. Smaller lesion lengths, suspected to be due to lower sporangia production, were also observed on needles from resistant genotypes. These findings have highlighted the need to factor in genotype information into any RNC severity predictions. We will discuss this further in Section 3.

Through the extension of the RNC SI model to the needle assay model, we have gained insights to the incubation period and infection rate of RNC. However, the influence of environmental factors such as climate and terrain in driving RNC remains a knowledge gap. We address this gap in our modelling plans for RA3, which we will discuss further in Section 3.

Spatio-Climate RNC Model

The HTHF programme had also made initial efforts in building an RNC predictive model in its efforts to develop tools for RNC disease management. A machine learning-based model (referred from here on as *"spatio-climate RNC model"*) was trained using data covering topological, climate and disease field observations.

The spatio-climate RNC model was built using the eXtreme Gradient Boosting machine learning model (XGBoost) [4], a state-of-the-art Gradient Boosting Machine approach. XGBoost is well-suited for learning on structured data, even with mixed data types, due to its use of decision trees in its modelling. Climate, topological and disease field observation information were extracted from different data sources, massaged and amalgamated into a dataset for training the spatio-climate RNC model.

Disease field observation information were extracted from two data sources, the Forest Health Database (FHDB) and data from long term monitoring of 50 plantation sites (referred to from here as *long-term monitoring dataset*). Information on field sites with RNC observed (see *Appendix – Forest Health Database* for fields extracted), including a binary value that indicated the presence of RNC, were extracted from the FHDB. These data points were merged with the long-term monitoring dataset (see *Appendix – Long-term monitoring dataset* for field information), which captures severity of RNC for 50 monitored sites on five severity levels.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

A clustering approach was used to resolve disparity along the spatial and temporal dimensions of the merged dataset[1]. Data points from both datasets were clustered along the "*compact location name*" and "*bioregion code*" variables. This created clusters with unique spatial representation, illustrated in Figure 2. Data points within each cluster were then rolled up according to NZ seasons for seasonal analysis (i.e. data points grouped using a three-month window and statistical values such as *mean*, and *quantiles* were calculated for each window). This ensures consistent data representation temporally for each cluster (i.e. there are data points for each season in each cluster).

To address differences in disease expression representation between the two datasets (binary value in FHDB compared to a five-level discrete representation in long-term monitoring dataset), a probability of infection was calculated for each cluster-season. The infection probability was derived using a beta distribution as prior, modelled using collected field observations, and calculated as a posterior infection rate. The posterior infection rate indicates the likelihood of observing RNC infection for a given season-cluster based on known field observations.

Figure 2. **Regional clusters formed from grouping sites along compact location name and bioregion code**

Quick analysis on the clustered data showed that RNC seasonal patterns can be detected. Figure 3 captures the graphs showing the posterior infection rate for all clusters aggregated at season-level. A three-year cycle, with peaks at Spring (Oct) can be observed. This observation is consistent with RNC disease expression having seasonally driven poly-cyclic population growth which is variable across years [5].

[1] Data points do not match up spatially and difference in inspection patterns generated large temporal gaps for observations in each site.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

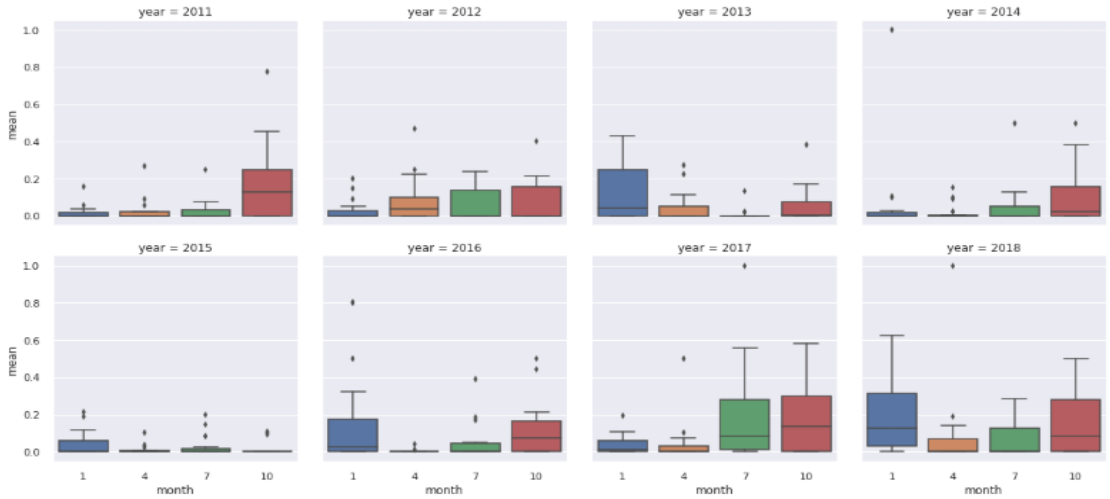


Figure 3. **Graphs showing aggregated posterior infection rate across seasons from 2011 to 2018 for all clusters. Inter-year pattern can be observed by looking at Apr (2nd bar) and Oct (4th bar) in each plot. Intra-year pattern can be observed through comparing posterior infection rate for each Oct.**

[1] Data points do not match up spatially and difference in inspection patterns generated large temporal gaps for observations in each site.

Climate information was extracted from the NIWA Virtual Climate Station Network [6]. Climate stations overlapping with the clusters were identified using geolocation information. Daily weather conditions for each identified station were extracted from 1972 till 2018. Details of the climate information extracted are shown in *Appendix – NIWA Virtual Climate Station Network.*

Topological information was gathered by extracting terrain information for each cluster from Land Information New Zealand's (LINZ) geo-spatial databases. Spatial attributes (i.e. shape files) were used to extract information on elevation, aspect, slope and steepness for each cluster. Using the topological information and each cluster's location, other information such as distance and orientation to water bodies were derived.

Extracted climate and topological information were concatenated with the disease expression data and rolled up for calculating cluster-season representation. Initial testing of the trained spatio-climate RNC model showed good prediction capabilities for clusters where historical trend of RNC infection was known. However, the model was still unable to generalised predictions to clusters where historical RNC infection history is missing, with model performance ranging between 0.255 to 0.304 R^2.

Besides prediction capabilities, we also looked at how input information was used by the model for prediction. Using the trained model, feature ranking lists that showed improvements in accuracy/contributions (entropy) when using each feature were generated. We generated two features ranking list using the average entropy gain and total entropy gain across all trees metrics. Both lists showed spatial and climate features were ranked highly, with the following features appearing in both lists' top 15 features:

- Season
- Soil moisture 12 weeks
- Relative humidity 24 weeks
- Wind speed 12 weeks
- Max temperature 24 weeks
- Precipitation 12 weeks
- Direction to coastline (mean)
- Precipitation 24 weeks

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Other features such as *elevation*, *slope* and *distance to coastline* were also highly ranked but were not consistent (i.e. only appeared in one list or appeared outside the top 15 ranks).

The spatio-climate RNC model demonstrated the importance of including information such as climate, spatial and disease information into RNC models. This observation aligns with studies that found climate conditions to be significant drivers for RNC [7, 8].

The model's inability to generalise predictions to new study areas showed more understanding and data on RNC are needed. Although we were able to gain insights to importance of different information through the ranking list, the abstraction of the machine learning model meant we still do not understand how exactly climate and spatial factors affect RNC.

RESILIENT FOREST RNC MODELLING

Modelling Objectives

The Resilient Forest RNC modelling work aims to address some of the gaps identified in previous works. Specifically, our plans for the modelling will focus on the following objectives:

1. Understand climate, topological, genetic and management factors' influence on RNC.
2. Model and quantify impacts of RNC on productivity on *Pinus Radiata*.
3. Develop a model for predicting severity of RNC for a given location as part of an effort to develop tools for RNC management.

Success in the planned modelling work will advance our understanding in how environmental factors drive RNC. Availability of a model that can predict severity of RNC given information, such as the climate, topological and genotype, surrounding a location will enable end user tools for RNC management to be developed.

Key Modelling Parameter

Modelling RNC is complex due to the dynamics in interactions between different factors in the host-pathogen-environment relationship. To manage this complexity, we explore a hybrid modelling approach that leverages both machine learning to discern complex inter-variables relationships for parameter estimation and established knowledge on RNC for an informed modelling approach.

At the core of our modelling approach is the RNC SI model. Our strategy is to identify key parameters that drive RNC using the RNC SI model and focus our studies on these parameters. The following list shows the identified key parameters:

- New born needles, b
- Infection, k
- Spores birth rate, g
- Spores mortality rate, m

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

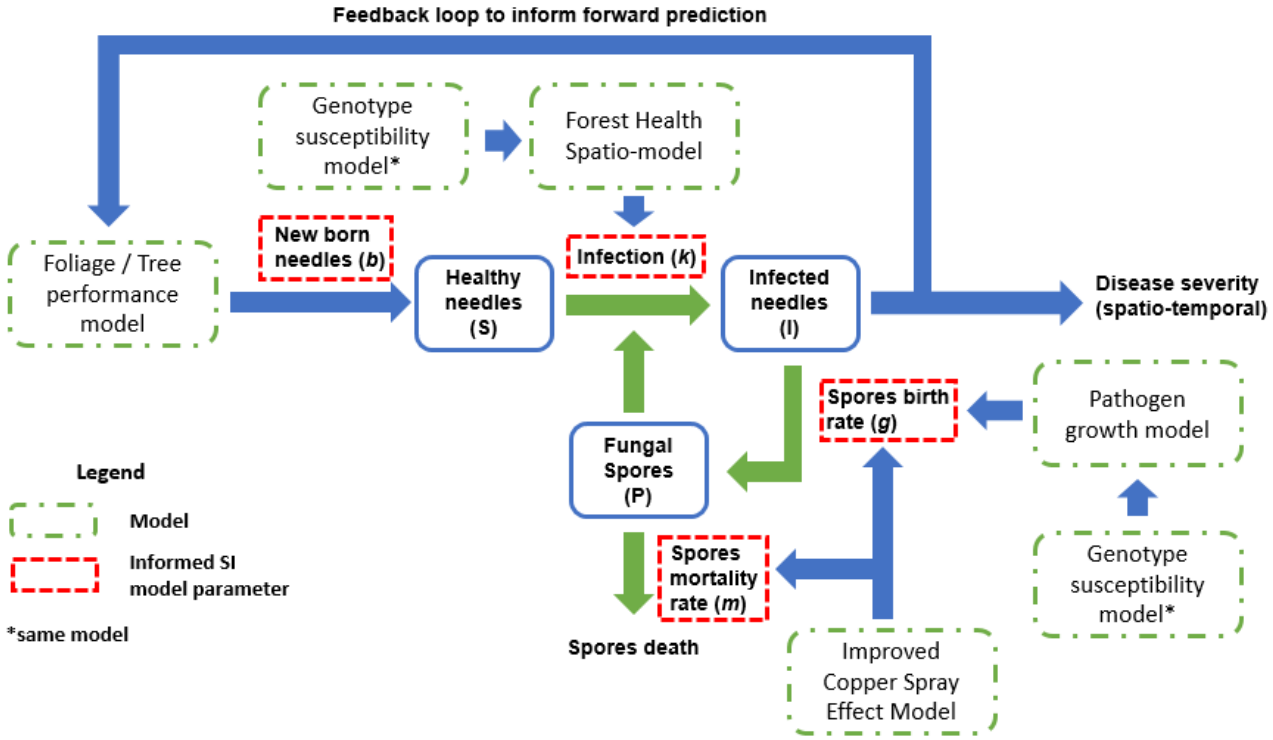


Figure 4. **Overview of our RNC severity modelling framework planned for the RA3 of the Resilient Forest Programme.**

Table 1. Table matrix showing model to dataset linkage

Model / Dataset	Foliage Productivity model	Forest Health Spatial model	Pathogen growth model	Genotype Susceptibility model	Copper treatment model
NIWA VCSN	✓	✓	✓		
Terrain info	✓	✓	✓		
Large-scale Remote sensing foliage health monitoring	✓	✓			
Field Spores Trap			✓		
Elite Clones trial				✓	
Industry Clones trial				✓	
Genotype variation		✓		✓	
Foliage sensor network data	✓				
Copper spray field trial					✓

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

For each key parameter, a data-driven approach will be used to develop models that can be used to estimate values for each parameter. Through these models, we will also be able to understand, in detail, how various external factors influence RNC. The developed parameter models will then collectively inform the RNC SI model, which we will use for predicting RNC severity for a given geolocation.

Figure 4 depicts the key parameters in relation to the RNC SI model and our overall modelling framework's schematic. Data requirements for the modelling of each key parameter is illustrated in Table 1. In the following subsections, we briefly discuss our modelling plans and associated data requirements for each key parameter.

New Born Needles - b

Prior studies on RNC have identified a poly-cyclic pattern across multiple season-years in disease expression. Current hypothesis is that the availability of healthy needles for new infection is one of the main factors for the observed poly-cyclic behaviour in disease expression. As such, it is important that we can forecast quantity of healthy needles produced spatially and temporally to forecast RNC severity.

Efforts to understand how tree health and environmental factors impact the availability of new needles will be partly covered under Resilient Forests RA3.1.3. A workplan for designing and implementing a framework for foliage health and productivity monitoring will be produced under RA3.1.3, providing data for the parameter estimation of b.
Data, which includes microclimate at individual trees collected from autonomous sensor network, aerial LiDAR and field observations, will be collected. They will provide insights about foliage performance, climate conditions and foliage health. Disease impact on foliage performance is factored in by including foliage health related parameters in the model. Doing so allows us to realistically model foliage dynamics under field conditions.

As RNC severity has a physiological feedback on needle production, this can be fed back to the model for estimating needle emergence for forward prediction. This will enable the model to iteratively forecast foliage productivity while considering disease severity and environmental conditions. The feedback loop is illustrated in Figure 4.

Infection – k

Key findings derived from studies in the HTHF programme have highlighted the importance of infection history and climate conditions for predicting RNC severity for a given forest plot. We hypothesise the infection rate of Phytophthora pluvialis, the cause of RNC, is heavily influenced by climate conditions, host genetics and topographical conditions. By analysing information from these factors with a plot's historical RNC severity information, will enable us to derive an understanding on how RNC expression varies with respect to its environmental conditions.
We adopt a machine learning approach for discerning complex interactions between variables in the terrain, climate, forest health and genotype dimensions. We then employ statistical analysis methods to investigate discovered interactions for developing new understanding on drivers for RNC expression.

Large-scale historical and current forest health data will be gathered using a remote sensing approach, designed and implemented in RA3.1 of this programme. This dataset is key for the machine model to learn from past disease expression information. The goal of the machine model will focus on estimating the infection parameter, k, used for determining the percentage of infected needles we can expect. Once we can quantify both healthy and infected needles, we will be able to estimate RNC disease severity for a given location at different level of detail (e.g. plot, stand, tree).

Spores birth rate – g

Prior studies have shown different radiata pine genotypes exhibit varying levels of RNC susceptibility-resistance. Gomez-Gallego et al.[3] have also reported on observing different levels of sporangia and mycelium production between susceptible and resistant genotypes. As such, we

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

hypothesis genotype information, together with environmental information, plays an important role in determining Phytophthora pluvialis spores' reproduction rate.
Datasets from trials with different radiata pine genotypes (e.g. Elite and Industry clone trials) and future studies, such as those intended under the Resilience Forest RA3.3, can inform a model for inferring genotype RNC susceptibility.

Building on the genotype susceptibility ranking model developed in the HTHF programme, we can extend the model to scale susceptibility for a wider range of genotypes. This ranking model will help inform a pathogen growth model that factors in climate, topographic and field spore level (from RA3.1.3 framework) information. The objective of the pathogen growth model will focus on deriving an estimate for the spore birth rate parameter, g.

The parameters g and m will be further elucidated with for chamber trials investigating the upper and lower environmental limits of pathogen growth on inoculated plants with the intent to inform the dynamics of pathogen growth and mortality. These studies will specifically explore the conditions and interactions between relative humidity, needle wetness and temperature (degree-hours) in relation to spore release and infection.

Spores mortality rate – m

Ongoing studies on treatments for RNC have identified copper as a likely candidate for disease control. Data gathered from copper field trials will allow analysis and the development of a copper spray effect model for simulating the effects of copper treatments on RNC in the field. Such a model can then be applied to the RNC SI model to allow incorporating management factors for RNC severity prediction.

The copper spray effect model will work on top of the pathogen growth model in estimating the spore's mortality rate parameter, *m*. Estimation of the birth rate, *g* and mortality rate, *m*, will have to be done in conjunction to simulate the immediate and short-term effects of spray treatment and subsequent regeneration rate of spores after treatment.

Overall RNC severity model

With the ability to estimate each key parameter based on collected field conditions and information, we will be able to use the RNC SI model to estimate disease severity for the given location. The RNC severity model (overall framework) can be used as basis for tools that can be delivered to hands of end users for RNC management and treatment planning.
Such a tool will allow end users to forecast severity of RNC for their forest plots, plan and optimise treatment plans for reducing impacts of RNC on forest gains.

Expansion of RNC severity model to incorporate other needle diseases

The framework presented here offers an opportunity to better elucidate the complex spatial and temporal dynamics of the key pathogens impacting *Pinus radiata* in New Zealand. Careful differentiation and monitoring disease severity (S-I) and pathogen generation (g) and mortality (m) for other key needle pathogens (*Dothistroma septosporum, Phytophthora kernoviae* and *Cyclaneusma minus*) with this framework will allow the disease severity, host genotype and spatio-temporal variation across the landscape for each to be better forecast and managed.

Baseline models for each pathogen will be informed by prior epidemiological studies and historic records from the NZFS Forest Health database and updated with comparative field assessments in the upcoming programme (RA3.1.3).

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

ACKNOWLEDGEMENTS

Funding for this research came from Scion and the Forest Growers Levy Trust, with the support of the NZ Forest Owners Association (FOA) and the NZ Farm Forestry Association (FFA).

REFERENCES

[1] G. Wake, N. Williams, and T. Pleasants, "A dynamical systems model for poly-cyclic foliar forest pathogens," *2018,* Tree disease, epidemiology, systems model vol. 59, p. 14, 2018-03-12 2018.

[2] G. Wake and F. Zaidi, "A proposed systems biology model to determine the spread of red needle cast in pine trees and its control: Formulation and preliminary analysis," Scion, Contracted Report Contract Reference J02006, 2017.

[3] M. Gomez-Gallego, R. Gommers, M. K.-F. Bader, and N. M. Williams, "Modelling the key drivers of an aerial Phytophthora foliar disease epidemic, from the needles to the whole plant," *PLOS ONE,* vol. 14, no. 5, p. e0216161, 2019.

[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA, 2016, pp. 785-794, chen16: ACM.

[5] M. A. Dick, N. M. Williams, M. K.-F. Bader, J. F. Gardner, and L. S. Bulman, "Pathogenicity of Phytophthora pluvialis to Pinus radiata and its relation with red needle cast disease in New Zealand," *New Zealand Journal of Forestry Science,* vol. 44, no. 1, p. 6, 2014/05/23 2014.

[6] A. Tait and R. Turner, "Generating Multiyear Gridded Daily Rainfall over New Zealand," *Journal of Applied Meteorology,* vol. 44, no. 9, pp. 1315-1323, 2005/09/01 2005.

[7] M. Gómez-Gallego, J. M. LeBoldus, M. K.-F. Bader, E. Hansen, L. Donaldson, and N. M. Williams, "Contrasting the Pathogen Loads in Co-Existing Populations of Phytophthora pluvialis and Nothophaeocryptopus gaeumannii in Douglas Fir Plantations in New Zealand and the Pacific Northwest United States," *Phytopathology®,* vol. 109, no. 11, pp. 1908-1921, 2019.

[8] S. Fraser, M. Gomez-Gallego, J. Gardner, L. S. Bulman, S. Denman, and N. M. Williams, "Impact of weather variables and season on sporulation of Phytophthora pluvialis and Phytophthora kernoviae," *Forest Pathology,* vol. 50, no. 2, p. e12588, 2020.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

APPENDICES

Appendix: Dataset descriptions

NIWA Virtual Climate Station Network

Following information can be retrieved from the NIWA VCSN dataset. The dataset spans from Jan 1972 till Jul 2018. The data is also available for NZ wide for both North and South islands.

Variable	Description
Agent No.	VCSN Station Identifier
Date	Local day date with hour/minute (note: for the subsequent variables, when/duration the measurements are recorded)
MSLP	Mean Sea Level Pressure in hPA at 9am local day
PET	Penman Potential Evapotranspiration total in mm from 9am local day
RAIN	Total amount of rain in mm from 9am local day
SMI	Soil moisture in mm from 9am local day calculated from rainfall and evapotranspiration. Base value is -150mm (permanent wilting point) based on soil store capacity. Value of 0mm indicates soil is at field capacity. >0mm indicates runoff.
ETMP	10cm earth temperature in degC at 9am local day
RAD	Accumulated global solar radiation in MJ/m^2 from midnight local day
TMAX	Max temperature in degC from 9am local day
TMIN	Min temperature in degC to 9am local day
VP	Vapour pressure in hPa at 9am local day
WIND	Mean wind speed in m/s at 10m above ground over 24hrs from midnight local day
Lat	Latitude of station. Note, this information is stored in a separate file and can be retrieved using Agent No.
Lon	Longitude of station. Note, this information is stored in a separate file and can be retrieved using Agent No.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Forest Health Database

Field observations captured from Jan 1960 till Jun 2017. Observations capture information of presence of various types of visible tree diseases.

Variable	Description
Bioregion	Name of region for site observed
Inspected	Date of inspection
Identification name	Type of pathogen identified
Identification comments	Comments from inspector
Serial no.	ID assigned to observation
Location	Road/area of the observed site
Stand	Name/ID of stand where observed site is
Species	Tree species name
Estab	Date the stand was established. 0 for unknown
Inspector name	Name of inspector
Inspection comments	Comments on stand/site by inspectors
Easting	Coordinates of site
Northing	Coordinates of site
Latitude	Coordinates of site
Longitude	Coordinates of site
NZMG N	Coordinates of site
NZMG E	Coordinates of site
Disorder ID	Assigned ID for observation
Disorder name	Brief description of observation
Disorder agent	Type of pathogen
Disorder comment	Comments of observation

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Long-term Monitoring Dataset

The survey captures long term RNC monitoring for about 50 sites. Each site is monitored several times a year. Data collection started in Oct 2015. All sites are located on central North Island. Infection levels have been quantized to five levels. The 50 sites are shown in the following map figure.

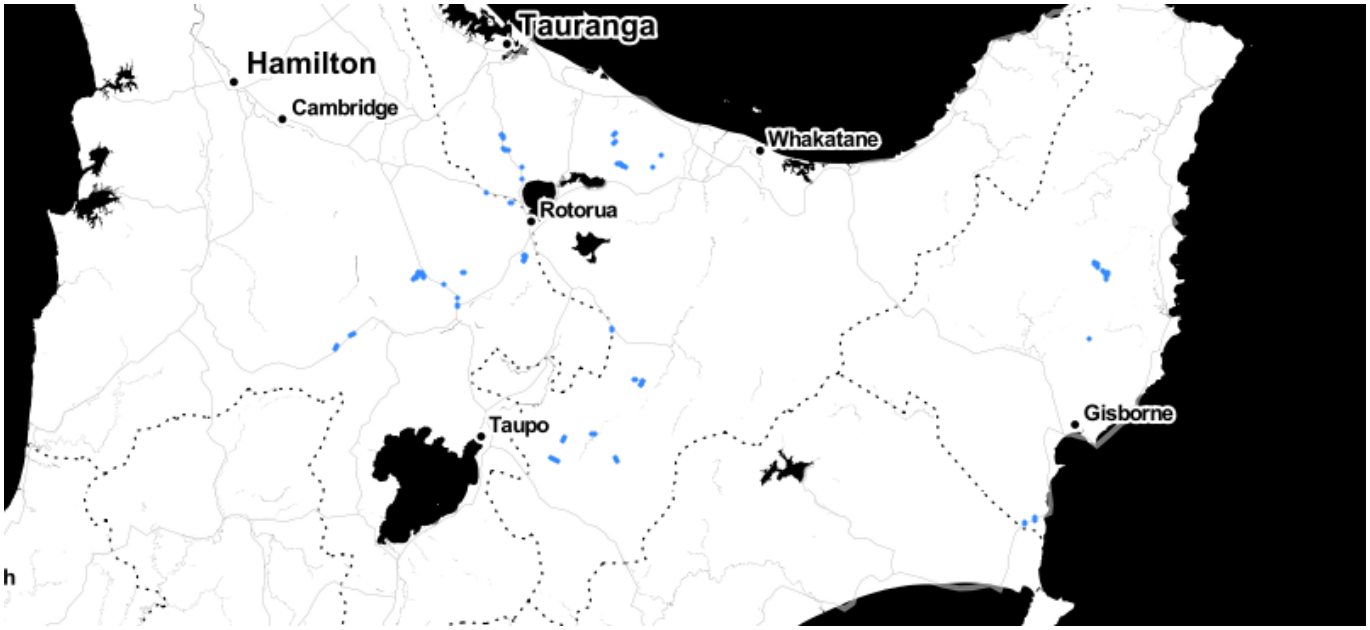


Figure: Location of the 50 sites being monitored for RNC.

Variable	Description
site no	Site id
site name	site name
start easting	easting of a transect head
start northing	northing of a transect head
end easting	easting of a transect tail. If not available, it is a point-transect.
end northing	northing of a transect tail. If not available, it is a point-transect.
age	age categories, (`Pre-Mid`, `Post-Mid`, `Mature`)
length	transect length in km.
side of road	which side/facing of the road where observations are made, (`N`, `S`, etc)
month	month and year for an observation.
date	date for an observation.
transect score	[0,100], percentage of infection in a transect.
worst tree	[0,100], percentage of infection for the worst tree.

GIS Topology Data (Spatial layers)

This dataset comprises of GIS data extracted from the LINZ database. Terrain info for North and South were processed and stored separately. For each island, “aspect”, “elevation”, “slope”, “TPI1000”, “TPI200” and “TPI100” were computed and stored in TIF format along with the shape files required for GIS data extraction.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Elite Clone Trial Data

This data captures full visual assessment of 45 Elite clone genotypes' susceptibility to Red Needle Cast. Data captures assessment of detached needle assay exposed to RNC at 70 days post inoculation. Symptom observation was noted on a scale of 0-4 being 0, 1-25%, 25-50%, 51-75% and 76-100% for a given symptom (can have more than 1 symptom on the same needle).

Variable	Description
Index (no name.	A unique identifier with no specific physical meaning. Example is 1059285T70.
Genotype	Type of gene, e.g., 09_10 (Note: The first part of the label (09) refers to the family ID. Genotypes with the same family ID will have the same parents. The second part of the label (10) is simply a number to identify the genotype in the family.)
Pathogen	Type of pathogen, e.g., PLU and H2O. H20 is the control.
Block	An experiment unit. Every block has a tree of each genotype in it.
Row	Positional variable in a grid layout.
Position	Positional variable in a grid layout.
Treatment	Type of treatment, e.g, P. pluvialis
Chamber	Name of chamber
Genotype	Same as `Genotype` above.
qPRC vial#	Can be ignored.
Time point	Number of days after starting the experiment. For example, T70 means 70 days after the start of the experiment.
Dark green	A phenotype of RNC.
Light green	A phenotype of RNC.
Yellow	A phenotype of RNC.
Brown	A phenotype of RNC.
Casting	A phenotype of RNC.
RNC bands	A phenotype for an early-stage RNC.
Comments	Text description.
No name	"Plant health 1=1-25%, 2=26-50%, 3=51-75%, 4=76-100%" is a note on how the percentage of symptom for a fascicle to be mapped to a numeric category. Category 4 means that >76% of a fascicle is having the relevant symptom.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Elite Clones Fog Chamber Data

Isolation data from the Elite Clones Fog Chamber work. For each sample and time point, 6 pieces were plated onto selective agar and the number of pieces from which P. pluvialis was isolated was recorded. Time points 7, 14 and 70 were done in full (all genotypes). Time points 28, 42 and 56 only had 20 genotypes analysed.

Variable	Description
Genotype	Type of gene, e.g., 09_10 (Note: The first part of the label (09) refers to the family ID. Genotypes with the same family ID will have the same parents. The second part of the label (10) is simply a number to identify the genotype in the family.)
Pathogen	Type of pathogen, e.g., PLU and H2O. H20 is the control.
Block	An experiment unit. Every block has a tree of each genotype in it.
Row	Positional variable in a grid layout.
Position	Positional variable in a grid layout.
Treatment	Type of treatment, e.g, P. pluvialis
Chamber	Name of chamber
Time point	Number of days after starting the experiment. For example, T70 means 70 days after the start of the experiment.
Fascicle	A bundle of leaves as a sample unit for experiment.
Plate assessment	Binary variable indicating the absence or presence of P. pluvialis. Positive for presence of P. pluvialis.
Pieces plated	Needle index within a fascicle.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx

Industry Relevant Clone Trial Data

44 genotypes were assessed over three time points (14, 27 and 53 days after exposure in the field) in the field. The trial was established in 9 fully randomized blocks. Trees assessed were all two-year-old trees. Whole plant assessment was done. Bottom whorl is 1 year old while top whorl is current year. Top whorl was assessed close to branching point, i.e. not where the current growth takes place.

Variable	Description
Date	Days after inoculation.
Whorl	Whorl position
Rep	Replication index
Assessed by	Initials of the assessor.
Genotype	Sample genotype, e.g., 268007.
Black Bands	A phenotype of RNC.
Olive lesions base	A phenotype of RNC.
Olive lesions other	A phenotype of RNC.
yellow/green mottle	A phenotype of RNC.
Casting	A phenotype of RNC.
bbcast	Derived.
olcast	Derived.
ygcast	Derived.
Dothi	A phenotype of other disease.
Cyclaneusma	A phenotype of other disease.
Yellow Speck	A phenotype of other disease.
Yellow	A phenotype of other disease.

RFP-T005 Identifying Key Parameter for RNC Severity Modelling and Prediction_G11.docx