PFR SPTS No. 19875

Aotearoa myrtle rust surveillance library

Campbell R, Teasdale S, Bradshaw P
Plant & Food Research Motueka[1]; Department of Conservation: Wellington[2], Nelson[3]

September 2020

Confidential report for:

Ministry for the Environment

DISCLAIMER

LIMITED PROTECTION

CONFIDENTIALITY

PUBLICATION DATA

Campbell R, Teasdale S, Bradshaw P. September 2020. Aotearoa myrtle rust surveillance library. A Plant & Food Research report prepared for: Ministry for the Environment. Milestone No. NA. Contract No. NA. Job code: P/346029/01. PFR SPTS No. 19875.

Report prepared by:

Rebecca Campbell
Scientist/Researcher, Plant Pathology
September 2020

Report approved by:

Beccy Ganley
Science Group Leader, Plant Pathology – Bioprotection
September 2020

For further information please contact:

Rebecca Campbell
Plant & Food Research Riwaka
Old Mill Road
RD3
Motueka 7198
NEW ZEALAND
Tel: +64 3 907 3595
DDI: +64 3 907 3649
Email: Rebecca.Campbell@plantandfood.co.nz

Contents

1 Introduction

Myrtle rust (MR), causal agent *Austropuccinia psidii,* was reported for the first time on mainland, New Zealand, in Northland, in May 2017. The pathogen has a wide host range on Myrtaceae plant species, of which New Zealand has many culturally, environmentally and economically important native and introduced species. After the initial detection in Northland, presence and absence data were collected by various stakeholders, primarily the Ministry for Primary Industries (MPI) and the Department of Conservation (DOC), with suspected MR infection reported through MPI. In April 2018, the incursion response was transitioned to long-term management of the pathogen. It was therefore no longer a requirement to report all MR detections to MPI, and also no longer a priority for DOC and MPI to continue the level of surveillance undertaken in the incursion response. Citizen surveillance was launched in November 2017 by the means of the Myrtle Rust Reporter (MRR) app, with data hosted through iNaturalist. Surveillance and incidence data from all known sources (including those listed above, as well as other individualised stakeholder data collections) were collated and mapped at the national scale for distribution to stakeholders and the public. Maps were distributed approximately monthly, to a range of stakeholders as an effective communication tool to help understand the response to the incursion and the extent of surveillance.

There is a need for these multiple data sets to be accessible to inform future disease incursions, MR research, public engagement in biosecurity, transparency, disease risk and other data uses. A combined, single database for the data which were collected throughout the incursion and to date, along with the metadata, is a valuable resource for researchers in the future, who were not necessarily part of the original response and data collection. The collated datasets have been further quality controlled, complete metadata compiled and data provided for the purpose of the Ministry for the Environment (MfE) State of the Environment reporting, in a single combined database. This completes the historical MR data resource for future MR research and surveillance. Expert users and collectors of the original datasets have been consulted to complete the metadata, and methods of compiling the disparate data sources have been thoroughly documented.

In addition to the surveillance data, there have been associated tools developed that have helped inform risk and priority areas for surveillance and management of the disease. A climatic pathogen process risk model developed (Myrtle Rust Process Model) allowed prediction of the national spatial risk of infection, latent period and sporulation risk with respect to weather patterns providing a near real-time risk prediction to help inform areas which should be focused on for surveillance during the period of the incursion response. As part of a Ministry of Business, Innovation and Employment (MBIE) Endeavour funded project (Beyond Myrtle Rust), there has also been a focus on taonga locations and trees as areas of importance to understand the presence/absence, management or prevention of the impacts of the rust. As an initial dataset of these locations, we have created a set of locations of iconic Myrtaceae plants.

A surveillance library has been prepared and is supplied with this report. In this report we describe the following components of the surveillance library:

1. Myrtle rust surveillance data compiled from multiple data sources
2. Iconic trees layer of important myrtaceous plant hosts
3. Infection risk and latent period rasters created using the Myrtle Rust Process Model/climatic risk model.

1.1 Surveillance library

1.1.1 Myrtle rust surveillance data compiled from multiple data sources

An extensive surveillance programme was deployed at the start of the MR incursion response. This was able to track the spread of the pathogen in different parts of the country on a variety of host plants (Toome-Heller et al 2020). Knowing where and when the disease has been recorded is important information for land managers, plant growers, risk assessors, conservation land and anyone with interest in our myrtaceous vegetation, including plants of national importance such as mānuka and pōhutukawa. A timeline outlining relevant events for the data collection processes is shown in Table 1.

Table 1. Some key points in the timeline since myrtle rust incursion in New Zealand, including dates first reported, media campaigns, and operational changes. Dates of first report for regions up until April 2018 are from Toome-Heller et al (2020).

Date	Key event
23 March 2017	Myrtle rust first reported in New Zealand on Raoul Island
3 May 2017	first reported on mainland New Zealand in Northland
May 2017	MPI media campaign
17 May 2017	first reported in Taranaki
22 May 2017	first reported in Waikato
11 June 2017	first reported in Bay of Plenty
September 2017	MPI media campaign
September 2017	Myrtle Rust Process Model operational
22 November 2017	first reported in Auckland
November 2017	Myrtle Rust Reporter app. goes live
30 November 2017	first reported in Wellington
15 March 2018	first reported in Manawatu
26 March 2018	first reported in Taupo
27 March 2018	first reported in Tasman
10 April 2018	first reported in Gisborne
April 2018	Incursion response transitioned to long-term management
19 April 2018	first reported in Coromandel
7 June 2018	Myrtle rust first reported in Marlborough
9 May 2019	first reported in West Coast at a nursery in Greymouth, only confirmed West Coast report (MPI investigation determined natural infection and incursion not contained)
4 March 2020	Christchurch nursery, only confirmed Canterbury report (MPI investigation determined infected plant transport and incursion contained)

The incursion response was initiated immediately after the first detection in New Zealand by MPI and continued as an incursion response until it transitioned into long-term management (April 2018) (Toome-Heller et al 2020). Strict plant movement controls were deployed when the disease was initially detected in Kerikeri, Northland in May 2019. Toome-Heller et al (2020) describe the incursion response, data collection and diagnostic methods and summaries for the MPI data collected. The MPI data sets included notification from members of public, government, non-government organisations and targeted surveys by designated surveillance teams. Plants were examined from nurseries, designated high risk areas, conservation land, home gardens and public areas. Media campaigns asked the public to report suspected infection to the MPI exotic pest and disease hotline and this did result in peaks of calls (but was not necessarily correlated with an increase in MR detection) to the hotlines correlated with May and September media releases (Toome-Heller et al 2020). The first seven new area detections were reported to MPI through the public and subsequent new area detections were found by targeted surveys based on the MR climatic risk models. Trained surveillance teams mainly targeted known infected areas and also areas predicted to be high risk by the climatic risk models (Beresford et al. 2018; Toome-Heller et al. 2020). Limited surveillance was conducted in low risk areas. Samples and photos were examined and processed by the diagnostic team in the MPI Plant Health and Environment Laboratory (PHEL). Initially, all suspected infected plant material was processed through PHEL. Subsequently, after MR was confirmed in a number of geographical regions samples were only taken if they were from new areas/regions or suspected new hosts, or images did not allow conclusive identification. All other later inspections were confirmed by trained staff from images (Toome-Heller et al. 2020). The majority of MR detections throughout the MPI response period were in Taranaki and the western Bay of Plenty, followed by Auckland and Waikato (Toome-Heller et al. 2020).

1.1.2 Iconic trees layer of important myrtaceous plant hosts

Infection risk to trees of national importance is informative for targeting surveillance, prevention and management. This layer provides a list, identity and location for priority nationally iconic myrtaceous trees and can be combined with the locations of and proximity to existing known infections in conjunction with the modelled climatic risk, to make informed predictions, assessments and precautions. Relating to trees that are well known and have significance to people also helps publicise the need for community surveillance. The iconic trees data were compiled as part of the Beyond Myrtle Rust programme (MBIE Endeavour) to aid knowledge integration with mana whenua.

1.1.3 Infection risk and latent period rasters created using the Myrtle Rust Process Model/climatic risk model

The Myrtle Rust Process Model was developed in collaboration between The New Zealand Institute for Plant and Food Research Limited (PFR) and the National Institute of Water and Atmospheric Research (NIWA) to assist MPI during the incursion response. Development of the weather based risk prediction model started in June 2017 and was operational from September 2017 (Beresford et al. 2018). The Myrtle Rust Process Model predicts three epidemiological processes and respective risk indices: infection risk, latent period and sporulation risk in relation to weather variables (relative humidity (RH), temperature and solar radiation) specifically for New Zealand (Beresford et al 2018). The process model captured diurnal cycling of infection processes (rather than averaged (e.g. monthly) climatic data) which enabled the production of near real-time updates. The infection risk index proved to be most useful for surveillance planning during the incursion response and was used for tactical planning of incursion surveillance, using information to predict how weather conditions would affect regional and seasonal risk.

The model was developed from published international information on *A. psidii* because at the time of initial development the pathogen could not be studied experimentally in New Zealand (Beresford et al. 2018); it had a “controlled organism” status as a new to New Zealand species. Since then it has been updated with field and controlled environment data from Australia and New Zealand and continues to be updated as new data are obtained. The latent period represents the time delay between an infection event and the appearance of disease symptoms. The period of time between the infection event and the production of uredinia (new inoculum) is defined as the latent period. When uredinia have appeared, MR symptoms can also readily be confirmed visually. The latent period was not definitive when the model was first developed (Beresford et al. 2018), but has since been updated with field and controlled environment data from Australia and New Zealand, resulting in the minimum latent period reducing from 12 to 5-7 days (Beresford et al. 2020). The latent period is generally shorter on more susceptible hosts and at higher temperatures, and is not dependent on external wetness or high RH (Beresford et al. 2020). The shorter the latent period, the higher the disease risk and this is strongly related to mean air temperature (Beresford et al 2020). Small increases in temperature can cause substantial decreases in latent period (increased risk). The sporulation sub-model, a product of temperature and latent period, is not included in this data set but indicates that spore availability is also greater when the latent period is shorter. Potential risk predicted by the Myrtle Rust Process Model is only based on climatic factors and does not use host susceptibility as an input, consider changes in susceptibility of host tissues with age, nor consider pathogen inoculum availability (this information is not available at this time scale). The risk predictions are based on susceptible hosts and spores being present in the area of interest and assumes that known susceptible common host species will dictate the patterns of spread and severity in New Zealand (Beresford et al. 2018). The infection risk sub-model is a function of wetness (hours of high (>=85%) RH, including a minimum wet period incorporated with a Gompertz growth function), hourly solar radiation and temperature (mean temperature during moist period). Spore germination is favored by low light intensity, optimal temperatures and periods of moisture availability.

Spatial patterns were predicted using the Myrtle Rust Process Model and the virtual weather grid produced by the New Zealand convective scale model (NZCSM). Daily values of the risk indices form the sub-models summarised as average daily risk/latent period over the previous 7 days (Sunday - Saturday). NIWA and PFR provided colour coded maps as a weekly operational service to MPI. Calibrations between the forecast weather data and weather stations were conducted. Risk predictions were checked for bias between NZCSM predictions and 100 NIWA climate stations, and time series of NZCSM were also validated against a number of nearest climate stations. Validation is difficult from MR surveillance data as MR has not been in New Zealand long enough.

2 Methods

2.1 Myrtle rust surveillance data compiled from multiple data sources

Much of the methods detail for the surveillance data is documented in Toome-Heller et al (2020) which covers the MPI data collection and incursion response. The large majority of the data were collected during the formal incursion response by MPI, and are therefore recorded in the MPI datasets. See the Table 1 for key dates that may have influenced the frequency and location of surveillance records.

Frequency and spatial coverage depended strongly on the organisation/individual collecting the data. MPI incursion data targeted high risk areas and areas proximal to known infection to delimit extent of spread (as documented in the "Surveillance Library" and Toome-Heller et al. 2020). Once the incursion response transitioned to long-term management (April 2018), data collection appeared to be opportunistic, convenience driven, heavily reliant on the public, but without much publicity or funding. This is similar for the Department of Conservation (DOC) surveillance data in the transition to long-term management, with the addition of asymptomatic observations recorded as part of the scoping process for DOCs Myrtaceae seed collection project (i.e. semi-targeted surveillance biased to areas predicted to be free from MR infections or with populations of Myrtaceae identified as targets for seed preservation).

The composite surveillance data were comprised of 10 separately collected data sets (Table 2). Detailed methods for the dataset combination methods, including SQL, are in the appendix document (Appendix 1. myrtle-rust-combined-surveillance-dataset-methodology). Appendix 2 (methodology-field-matching) lists the fields from the original data sets and how they informed and were included in the combined data set. Some key assumptions and processes are summarised below for each key data set.

Arguably the most important and difficult fields to translate correctly into the combined data set were the fields defining whether myrtle rust symptoms were visible or not. This was a combination from a number of different fields, depending on the data source. Appendix 2 (methodology-field-matching) lists the fields these details were derived from to inform the 'Myrtle rust symptoms' fields in the combined dataset. Which subsequently informed the 'Myrtle Rust Present' field to simplify the attribute to asymptomatic, symptomatic or suspected.

'Myrtle rust present' was summarised as a 'suspected' record in the final combined dataset to capture where two or more fields relating to whether MR symptoms was observed disagreed (one indicating symptoms present, the other indicating asymptomatic). For example, 'canopy impact score', 'manager comments' and 'survey completed' fields. For example, where the 'CanopyImpactScore' indicated rust present, whereas the 'TreeBranchScore' suggested no rust present.

Table 2. List of individual surveillance datasets compiled into the composite surveillance dataset after data cleaning. MPI = Ministry of Primary Industries, PFR = The New Zealand Institute for Plant and Food Research Limited, MRR= Myrtle Rust Reporter, MRRI = Myrtle Rust Raoul Island, DOC = Department of Conservation.

Original data source name	Data source name in the combined data set	First record	Last record	Total number of records	Number of records positive for MR
MPIPlantSurvey	MPI_Revisits	2017/05/16	2018/08/30	176,904	1,427 (95 suspected)
MPIHosts	MPI_MyrtleRustHosts	2017/04/07	2018/09/03	144,543	2,606
RestrictedPlaces (MPI)	MPI_Current_Restricted_Places_Dec19	2017/05/03	2019/06/25	1,185	1,158
DocHistoric	DOCPCLSurveillanceHistoric	2017/07/18	2018/12/12	7,592	259
DocCurrent	DOCPCLSurveillance	2019/02/01	2019/07/23	115	10
Plant and Food Research	Plant_and_Food_Research	2017/07/03	2020/06/08	3,016	30
Auckland Botanic Garden	Auckland Botanic Garden	2018/08/01	2020/06/01	869	In iNaturalist
Dunedin Botanic Garden	Dunedin Botanic Garden	2017/12/20	2017/12/20	15	In iNaturalist
Horizons Regional Council	Horizons Regional Council	2018/03/01	2018/10/01	2	2
iNaturalist-MRR	iNaturalist_Myrtle_Rust_Reporter	2017/09/23	2020/06/27	534	307
iNaturalist-Myrtaceae	iNaturalist_Myrtaceae_Observations	2017/01/01	2020/06/30	5,309	NA
iNaturalist-*Austropuccinia psidii*	iNaturalist_Myrtle_Rust_Observations	2018/03/25	2020/06/03	55	55
MRRI_AllSamples	MRRI_AllSamples	2017/03/30	2017/04/26	110	110
All sources		**2017/01/01**	**2020/06/30**	**340,249**	**5,964 (95 suspected)**

A ‘Confidential’ field was retained in the combined data set to ensure a record for entries which specifically requested geoprivacy. These are the ‘do not ID to landowner’ (MPI datasets) and ‘obscured’ (iNaturalist) attributes. We created a ‘data quality estimation’ field to capture any record of data quality from the original datasets. This is a field with mixed references; some refer to the confidence of the plant host species identification, some refer to the confirmation of myrtle rust infection (Table 3). In the future, many of these records could be quality checked against their respective photos. Further to data quality and completion checks, we created a field to count the number of empty (<null>) fields as an indicator of completeness (‘number of empty fields’).

Table 3. Definitions of the “data quality estimation” field created in the combined dataset. MRR= Myrtle Rust Reporter, DOC = Department of Conservation, PCL = Public conservation land.

Dataset	Name of original field	Meaning
DocPCLSurveillance Historic	‘Species Confidence’	Refers to confidence in the **plant species** identification (High, medium, low, no confidence)
iNaturalist_Myrtle_ Rust_Reporter	'num_identification_ agreements'	Number of times the record has been confirmed identified as the same species (in this case *Austopuccina psidii*)(values 0-7).
Restricted places (MPI)	‘Confidence’	Confidence in identification of Myrtle rust
iNaturalist Myrtle rust Observations	'num_identification_ agreements'	Number of times the record has been confirmed identified as the same species (in this case *A. psidii*)(values 0-7).
iNaturalist Myrtaceae Observations	'num_identification_ agreements'	Number of times the record has been confirmed identified as the same species (in this case Myrtaceae spp.)(values 0-7).

There were many records with inconsistent spelling and capitalisation, particularly in the species and genus fields. These were cleaned and completed consistently to provide consistency and accuracy. There were fields, such as % and height and counts, where data was cleaned to fit within realistic ranges (e.g. removal of negatives). Out of range records were reverted to missing data (<null>). The presence of these errors indicates the likelihood of other data entry errors in these fields.

2.1.1 Presentation of surveillance data

It is important that the presentation of these data do not compromise the privacy of the locations where MR was recorded and where landowners did not want to be identified. Therefore the entire combined dataset location points will need to be aggregated suitably to retain privacy, while simultaneously making important aspects of the data available. For the data presentation to be useful, attributes such as (but not limited to) plant host, presence of MR symptoms and the approximate location are necessary to be displayed clearly. The aggregation should be chosen carefully to not create opportunities for the data presentation to be misleading or misrepresentative (e.g. Modifiable areal unit problem/MAUP).

2.1.2 Methods for key data sets

MPI Hosts

These data were collected using Survey123 and managed by AsureQuality (Quentin Higgan). This was the first large scale formal data collection for the incursion. There were some attribute fields which were not logically named, such as ‘TreeID’ where this was actually the percentage of the tree surveyed, and ‘Survey Completed’, which was the main record of whether the plant had myrtle rust symptoms or not. Initially the ‘Survey Completed’ field only had the choices POS (positive for MR symptoms), Yes (no symptoms), LAB (sample sent to lab). Later additional categories REM (positive for MR, removed by MPI) and PSR (positive for MR, instructed landowner to remove) were added. Part way through the incursion, a revisits table (relational table) was introduced to capture return monitoring of the same sites and/or plants. The MPI biosecurity response captured all positive MR data from the start of the incursion (May 2017) to April 2018, even when the surveillance was carried out through one of the other providers (e.g. DOC) due to the reporting requirements. Therefore all positive records for this date period are captured within the MPI datasets. A list of the field names and brief explanation for the original MPIHosts dataset are shown in Table 4.

Table 4. Initial fields and definitions for the Ministry for Primary Industries (MPI) MPIHosts dataset. Those highlighted in green are those that were carried over to be included in the combined data set (*unknown definition).

MPIHosts	MPIHost_Explanation
'OBJECTID',	Unique identifier, used for relation to the revisit table (SurveyPlants)
'Date_Surveyed',	Date tree was looked at
'TreeID',	Percentage of tree surveyed
'Canopy_Impact_Score',	0-9 domain of how MR was impacting the plant canopy (same as DOC category- some rust, obvious rust etc.)
'Flowering_Activity_Score',	0-6 domain describing flowering plant phenology
'Tree_Branch_Score',	0-5 domain describing branch plant phenology, including infection
'Bud_and_Fruit_Score',	0-5 domain describing bud and fruit plant phenology
'Tree_Genus',	Plant genus
'GlobalID',	Unique ID
'CreationDate',	Date entered electronically
'Creator',	Person who created the data point electronically
'EditDate',	Date if data were edited
'Editor',	Person who edited
'SampleTaken',	Whether a plant sample was taken for laboratory identification
'Likelyhood',	Opinion of whether it was likely MR infection
'DateNegative',	Date negative results were available from laboratory testing
'ManagerComment',	Often species comments, pruning, location description, host
'Canopy_Revisit',	CanopyImpactScore (domain, in words) from a revisit of the same tree.
'Flower_Revisit',	Flowering activity score (domain, in words) from a revisit of same tree
'Tree_Branch_Revisit',	Tree branch score (domain, in words) from a revisit of same tree
'Bud_Fruit_Revisit',	Bud and fruit score (domain, in words) from a revisit of same tree
'Survey_Completed',	Myrtle rust present and tree disposal, or Survey completed (Domain: PSR, REM, POS, LAB, Yes)
'StatusCalc',	*
'RevisitDate',	Date tree revisited
'TreeHeight',	Estimation of tree height
'UNQ',	Geospatial referencing
'Shape',	Geospatial referencing
'REL_OBID',	Geospatial referencing
'SpatialEngine_ObID',	Geospatial referencing
'WorkInstruction'	Geospatial referencing

MPI Plant Survey (revisits table)

The revisits table was initiated to facilitate revisiting the same trees and/or locations also in the 'MPIHosts' dataset. The 'RelatedOBJ' was the foreign key to relate to the 'ObjectID' in the 'MPIHosts' data set. A list of the field names and brief explanation for the original MPIplant survey dataset are shown in Table 5.

Table 5**.** Table of initial fields and definitions for the Ministry for Primary Industries (MPI) "MPI Plant Survey" dataset. Those highlighted in green are those that were carried over to be included in the combined data set.

MPIPlantSurvey	MPIPlantSurvey_Explanation
'RelatedOBJ',	Foreign Key for relation to MPIHosts data
'DateSurveyed',	Date tree was looked at, this will be a revisit to a previously visited tree
'PercPlantI'nsp,	Percentage of tree surveyed
'CanopyImpactScore',	Presence and impact of MR. Same as DOC categories and MPIHost categories, but in words (domain).
'FloweringActivityScore',	Plant Phenology categories
'TreeBranchScore',	0-5 domain describing branch plant phenology, including infection
'BudAndFruitScore',	Plant Phenology categories
'SampleTaken',	Whether a plant sample was taken for lab identification
'Likelyhood',	Opinion of whether it was likely MR infection
'DateNegative',	Date negative results back from laboratory
'ManagerComment',	Manager comments
'TreeHeight',	Height estimation of tree
'ObjectID',	Geospatial identifier
'GlobalID',	Geospatial identifier
'created_user',	User who created data entry
'created_date',	Date date entry created
'last_edited_user',	User who last edited
'last_edited_date'	Date of last edit

MPI restricted places/biosecurity reporting hotline

The 'MPI restricted places' data was curated by the MPI/Biosecurity New Zealand MR long-term management (MRLTM) team. The dataset is possibly an amalgamation of reports during the response, used as a tracking spreadsheet to summarise reports by location. The quality of these data is not known; each report was recorded as a single row with species infected at a site listed in the same field and additional infections reported for the same location at new times were recorded in the comments field. Therefore infections were tracked per property, rather than per tree. For a variety of reasons, including those described above, we were unable to get the definitions of some of the fields. The photos confirmed in this data set were mostly confirmed by the trained phone technicians (not necessarily PHEL; location of these image files is also unknown). However, it is also the only record of reports collected via the MPI pest and disease hotline (0800 80 99 66). The hotline collected information on myrtle rust reports until 30 June, 2019 (inclusive, details supplied to MRLTM via email and recorded in the RestrictedPlaces dataset). All myrtle rust reports that came through the hotline post this date are no longer recorded, instead callers are advised to enter suspected MR infections to iNaturalist. A list of the field names and brief explanation for the original MPI restricted places dataset are shown in Table 6.

Table 6. Table of initial fields for the Ministry for Primary Industries (MPI) "MPIRestricted places" dataset and what they meant. Those highlighted in green are those that were carried over to be included in the combined data set (*unknown or uncertain definitions). RP= restricted places

RestrictedPlaces	RestrictedPlaces_Explanation
Classification	Property type/land area usage
Date_Reported	Date entry created
Address_Street	Physical address of reported plant(s), can be street address, description of place (park school etc), lat/long coordinates
City	Town or city of report location (cannot be blank)
Confidence	Confidence in identification of myrtle rust
Date_FollowUp	Date report followed up by MPI/AQ, can be blank
Date_Reinfection_Reported	Date of further reported infection(s) at the same address
Date_Restrictions_Off	Date when restrictions notice expired, can be blank
Date_RP_Issued	Date a restriction notice was issued*
Date_RP_revoked	*
FHQ_KK1_NP2	Field Headquarters*
Host_Species	Can be host to cultivar, species or genus level. Often multiple hosts listed in single cell (as per registers instructions)
Host_Species_2	Can be host to cultivar, species or genus level. Often multiple hosts listed in single cell (as per registers instructions)
ILTM_Designation_Status	Known infected area, outliers, pending status, Incidence long term management*
MR_Status	Whether MR confirmed or negative. Also includes if there is a photo for confirmation.
Name	Name of organisation associated with land where applicable, can be blank
Notes	Additional notes
OBJECTID	Geospatial identifier
Region	NZ region as based on regional councils of report location
Reinfection_Status	Whether record is/has a reinfection
Report_Origination	Where the report came from (hotline, DOC, MPI surveillance, others)*
Restricted_Places	Unique number assigned to each property reporting infections by the MPI myrtle rust; all species reported as potentially infected at a property were listed in the 'Host_Species' and 'Host_Species_2' fields
ReSurvey_HostSpecies1	*
ReSurvey_HostSpecies2	*
RP_Status	Management status, e.g. self management
RP_Status2	Management status, e.g. self management
Sample_Status	*
Tracing_Complete	Whether tracing was complete or ongoing
Treatment	Progress of any treatment

DOC historic

Department of Conservation (DOC) data were collected using Survey123 and managed by DOC MR team and Geographic information system (GIS) support. Additional fields were added to the original survey to capture more host specific data and additional information on disease severity (indicated by an * in Table 7). Surveillance sites were selected following the same logic as the wider response but with a Public Conservation Land (PCL) area focus. A list of the field names and brief explanation for the original DOC PCL historic dataset are shown in Table 7.

Table 7. Table of initial fields for the Department of Conservation (DOC) “DOC PCL surveillance historic” places dataset and what they meant. Those highlighted in green are those that were carried over to be included in the combined data set. * Additional fields added later to the original survey form. PCL = Public conservation land.

DocHistoric	DocHistoric_Explanation
OBJECTID	Unique identifier
globalid	Unique identifier
CreationDate	Date entered electronically
Creator	Person who created the data point via survey123
EditDate	Date if data were edited
Editor	Person who edited
Tree_Genus	List of tree genera (Myrtaceae genera present in NZ)
Tree_Height_m*	Tree height in metres
Canopy_Imp_Score	List of how much of the canopy is impacted by MR, whether it is tested in the laboratory or whether tree removed (domain 0-9).
Flowering_Fruit_Act	Plant phenology (domain 0-5). Can be null.
Tree_Branch_Score	Assessment of myrtle rust infection on leaves and stems. Domain, 0-4. Can be null.
Bud_Fruit_Score	Assessment of myrtle rust infection on flower buds and fruit. Domain, 0-4. Can be null.
Comment	Comments, various.
Plot_ID*	Can be place name, address, seed collection id code, or nothing (normally DOC specific, e.g. Tier1 plot ID codes or sentinel tree tag ID). Variable spelling and inconsistent use of capital letters, can be null
Tree_NVS_Code*	Six letter ID code specific for tree species and taxa recognised by DOC, can be null. Is sometimes in dataset as choice0 - choice29, or as the six letter codes defined in details
Species_Confidence	Confidence of observer on identifying the plant species/genus, can be null
Perc_Tree_Surveyed*	Approximate percentage of the tree visually surveyed by the observer, can be null
Specimen_Collected	Comment field for notes on other samples collected, e.g. for lab diagnostic testing
Other_Spec_Collected	One record has a comment in this field for seed collection (note on fruit ripeness and collection)
Seed_Collected	Was seed collected as part of the germplasm collection program - yes, no (allows null values)
POINT_X	Geospatial references
POINT_Y	Geospatial references
SHAPE	Geospatial references

DOC current

The more recent Department of Conservation (DOC) data were managed by DOC GIS support with observations recorded through Survey123. The majority of these observations are associated with the Myrtaceae seed collection project. The ad hoc surveillance is linked to scoping for seed collection as opportunistic observations rather than a targeted surveillance objective. This project ended in June 2019. A list of the field names and brief explanation for the original DOC current dataset are shown in Table 8. For the field indicating MR infection, surveillance collectors could only choose “suspected” MR infection options in the "myrtlerustpresent" field, infection was confirmed by the DOC MR team and PHEL staff when required.

Table 8. Table of initial fields for the Department of Conservation “DOC current surveillance” dataset and what they meant. Those highlighted in green are those that were carried over to be included in the combined data set. *unknown definition.

DocCurrent	DocCurrent_Explanation
OBJECTID	Unique identifier
Genus	List of tree genera (Myrtaceae genera present in NZ)
NVSCode	Six character code for identifying species
CreationDate	Date entered electronically, automatic field
MyrtleRustPresence	List of values similar to canopy impact score from other datasets. Field staff can mark as "no" or "suspected - …", then changed to confirmed by DOC staff after checking photographed symptoms (and or confirmed by PHEL)
Creator	Person who created the data point electronically
Editor	Person who edited
EditDate	Date if data were edited
CollID	Unique ID coded for each of the seed collections required
CollSite	Comment field for additional site names
Comments	Comment field
email	Automatic field from login details of creator
start	Start date of surveillance, scoping, or seed collection period
end_	End date of surveillance, scoping, or seed collection period
InfIntensity	Myrtle rust infection intensity. Domain, 0-2.
Latitude	Location
Longitude	Location
NativeExotic	Whether plant(s) surveyed were native or exotic
NoPlantsInfected	Number of plants infected
NoPlantsSurveyed	Number of plants surveyed
Organisation	Department of Conservation or Contractor
PartsPlantSurveyed	Parts of plants surveyed - can select multiple options
PercPopInf	Percentage of population infected from 10 to 100%, to the nearest 10
Phenology	Phenology of plant(s) surveyed, can select multiple answers
PlantHeight	Plant height or estimated average height of a population surveyed
PlantType	Growth form of the plant surveyed
PopSpecies	Size of the population, can select from set of predetermined ranges
ScopVar	Variety of *Leptospermum scoparium,* when known
SHAPE	Geospatial reference

DocCurrent	DocCurrent_Explanation
SiteID	Can be place name, address, seed collection id code, or nothing. Variable spelling and inconsistent use of capital letters
SiteNotes	Various other notes on site location and/or access to site
SurveillanceType	Surveillance for multiple or single plants. Written as "surveillance for:"
SurveyType	Description of what the survey is being completed for. Written as "what is this survey for?"
TypeSpecies	Comment field to enter if a herbarium samples were collected and checked/confirmed
username	Username of survey creator - automatic field generated from login details of the survey123 app
X	Geospatial reference
Y	Geospatial reference

Smaller data holders

PFR initiated a sentinel tree programme, planting five ramarama (*Lophomyrtus bullata*) and five pōhutukawa (*Metrosideros excelsa*) trees at 10 PFR sites (Kerikeri, Auckland, Pukekohe, Ruakura, Te Puke, Hawke's Bay, Palmerston North, Motueka, Marlborough and Lincoln). Sentinel trees were planted in August 2018 and monitored regularly until July 2019. Regular monitoring of the trees planted in Auckland and Marlborough is still continuing as part of the Ngā Rākau Taketake, Biological Heritage National Science Challenge. One of the purposes of the planted sentinel trees was to study regional differences in the development of MR on genetically uniform plants. It was also to gather geo-climatic data on rust and monitor host plant shoot growth. Existing trees around some of the PFR campuses were also recorded and monitored for MR presence in most cases weekly to monthly from July 2017 (until 2019).

Botanic gardens from Auckland, Wellington, Christchurch and Dunedin were contacted to request surveillance data. Surveillance data was received by email as an Excel spreadsheet with coordinates. Non-positive records were recorded and kept in-house, while positive records were entered into iNaturalist.

All the regional councils were approached by email to supply data. There was no reported coordinated data collection by the councils that was not already covered either by the initial MPI response or later recorded in iNaturalist. Singular records were emailed to R. Campbell to enter into the data set.

iNaturalist and the Myrtle Rust Reporter (MRR) app.

The MRR app (https://www.inaturalist.org/projects/myrtle-rust-reporter) was launched in November 2017 to assist and encourage public surveillance. Initially it did not include some key fields that researchers and community groups were interested in, therefore several fields were added or updated through the beta testing phase with the latest update, 16 April 2020, including host tree identity changes (specifically additional species listed and a host search and add feature). The MRR app is no longer supported developmentally (no current funding). The MRR is a project within the iNaturalist platform (https://www.inaturalist.org/), however MR records and Myrtaceae host plant records can be recorded directly into iNaturalist, bypassing the MRR app and the MRR additional information features can be added to any iNaturalist observation. Therefore three iNaturalist datasets are included in the combined dataset: Myrtaceae records, MRR records (curator output) and *A. psidii* records. MPI and DOC both promote iNaturalist to the public (including industry and DOC staff) to record MR surveillance.

2.2 Iconic trees layer of important myrtaceous plant hosts

A list of taonga/iconic trees belonging to the Myrtaceae family, in the public realm were compiled from literature and expert knowledge by Hone Ropata and Alby Marsh (PFR) as a component of the MBIE Endeavour funded Beyond Myrtle Rust programme, RA1.4: Kaitiakitanga – Māori-led solutions. These were assigned spatial coordinates based on the literature and expert knowledge. This was a desktop exercise and they are not ground truthed with accurate GPS location. Some “trees” are in fact stands of trees of significance, rather than a single tree, and therefore the area of importance is underestimated by assigning them a point location. There are also several records where only approximate location was possible during this approach and therefore a generic location in the area has been presented. This is not a complete list of myrtaceous taonga trees, there are expected to be many others of significance and importance to various hapū, iwi, ropu and rohe. Some of which are unlikely to share the significance to a public audience because of the privacy, respect and taonga. There are some maori names and taonga descriptions which were unable to be completed from knowledge available. These could be researched further with the appropriate maori partners. On discussion with Waitangi Wood (Wai Communications Ltd), it was more appropriate to label them as “iconic” trees rather than “taonga”, to highlight the importance of them to the general public, and not have misunderstandings with what labelling them “taonga” might infer.

2.3 Infection risk and latent period rasters created using the Myrtle Rust Process Model/climatic risk model

Specific methodology behind the creation of the Myrtle Rust Process Model is well documented in Beresford et al. (2018) and Beresford et al. (2020). Some factors for clarity and consideration are included in the “Introduction” in this document.

The updated latency period using the minimum latent period for *Lophomyrtus* sp. (Beresford et al. 2020) has been incorporated into all the model outputs, including back-dated to outputs for time periods prior to the 2017 New Zealand incursion. The output of the model is transferred to PFR as a set of .asc files. One for each week, for each of the epidemiological processes (Infection risk, latent period, sporulation risk). Infection risk is the most useful output for prioritising surveillance and management, in conjunction with latent period to understand the time lags between infection and visible symptoms. Sporulation risk is available in the same weekly .asc format, but is not requested for this contract.

The risk values are calculated from hourly weather forecast data and the MRPM sub-models and summarised for daily means (3pm NZ time). These are then summarised into weekly mean risk (Sunday to Saturday). Rasters are classified into five classes (described in Table 9) of defined intervals for interpretation of risk and latent period. Rasters are provided in the coordinate reference system (CRS) EPSG 2193/NZGD2000 Transverse Mercator, cell size 1000 (ncols 1051, nrows 1519).

For "infection risk", raster calculations to extract the data to the New Zealand coastline and reclassify the values into risk categories (Table 9) were carried out in ArcGIS Pro using ModelBuilder. For "latent period" rasters are supplied in .asc files (without extraction or reclassifying) due to time constraints and uncertainty of the requirements of the format to provide. A raster for the first week for each month between July 2017 and July 2020 are provided as a sample. It may be more meaningful to calculate mean risk values for monthly or seasonal periods due the weather pattern variations week to week. NIWA have recently published an update whereby the risk raster maps can be made available through ArcGIS Online https://niwa.co.nz/publications/water-and-atmosphere/water-atmosphere-24-july-2020/rust-coding.

Table 9. Infection risk and Latent period classifications, values and risk categories. Latent Period classifications updated from Beresford et al. (2018) with results from Beresford et al. (2020). Longer latency period is interpreted as lower risk (longer times between secondary inoculum availability).

Risk	Values	Risk categories	Reclassified as
Infection risk	< 0.2	Very Low	1
	0.2 - 0.4	Low	2
	0.4 – 0.6	Moderate	3
	0.6 – 0.8	High	4
	0.8-1.0	Very High	5
Latent period	< 10 days	Very High Risk	
	10-15 days	High Risk	
	15-30 days	Moderate Risk	
	30-50 days	Low Risk	
	> 50 days	Very Low Risk	

3 Results

The data from the data library is also partially presented in an interactive map managed by R. Campbell. This is currently under re-development to include the combined data set (was feeding off the multiple sources) (https://plantandfood.maps.arcgis.com/apps/webappviewer/index.html?id=db12ae762a0a4e3eb8c61b1f67120c3b)

3.1 Myrtle rust surveillance data compiled from multiple data sources

Comparison between the different data sources indicates that the MPI surveillance effort was orders of magnitude more than any other data sources during the incursion (Figure 1). Following the transition to long term management, all surveillance decreased to minimal levels and non-targeted areas (Figure 2, 3). Surveillance effort and detection of myrtle rust changed over time, seasonally and by surveillance effort (Figure 2, 3). There were peak numbers of MR detections around September 2017, April 2017, March 2018, January 2019 and January 2020 (Figure 2). Total surveillance over time was skewed to large numbers during the incursion management time frame (MPI data sets). The total surveillance had less distinct seasonal peaks than the MR detections, with increased surveillance records around September – October 2017 and mid 2018 (Figure 3).

There were higher densities of surveillance and detection in proximity to populated areas and also in areas predicted to be more at risk to MR infection from modelling and prior knowledge (Figure 4, 5). There could be potential bias by a snowballing effect whereby locations where MR was recently detected or publicised triggered an increase in surveillance in those areas. This is suggested by some of the temporal and spatial patterns of surveillance and MR detection over time (seen in the monthly maps produced throughout the incursion response, R. Campbell).

The most surveyed tree genera were Metrosideros, Leptospermum, Acca, Lophomyrtus, Callistemon, Kunzea, Syzygium, and Eucalyptus (Figure 6). Of these the most commonly detected infections were in Lophomyrtus (2869), Metrosideros (1604), Syzygium (801), Leptospermum (103), Callistemom (96), Acca (79), Kunzea (51) and Eucalyptus (35) (Figure 7).

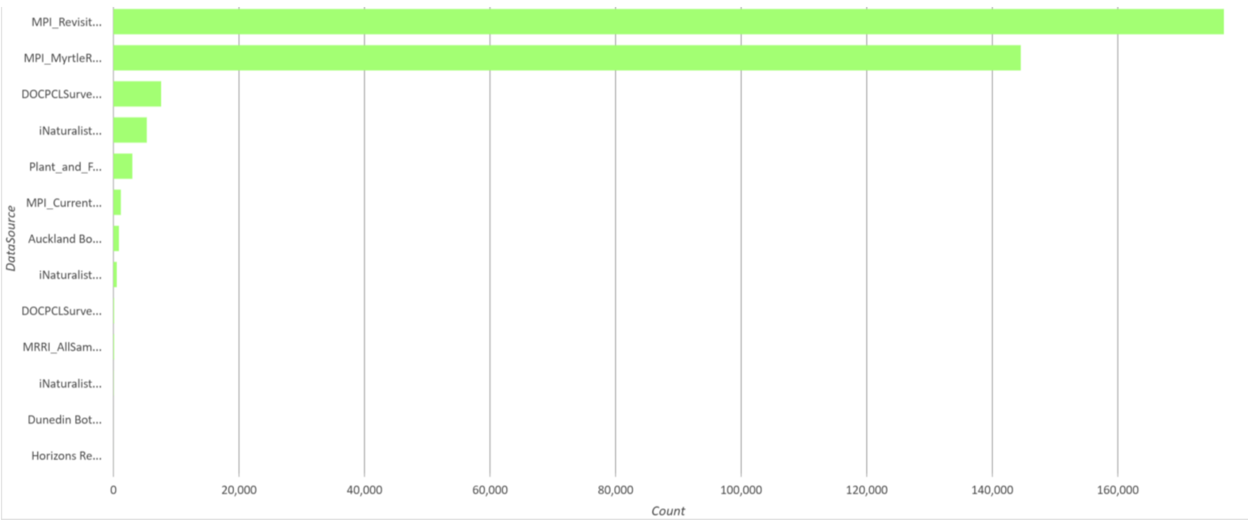

Figure 1. Distribution of surveillance records listed by data source (not all smaller data sources are shown). Data set names are provided in Table 2.

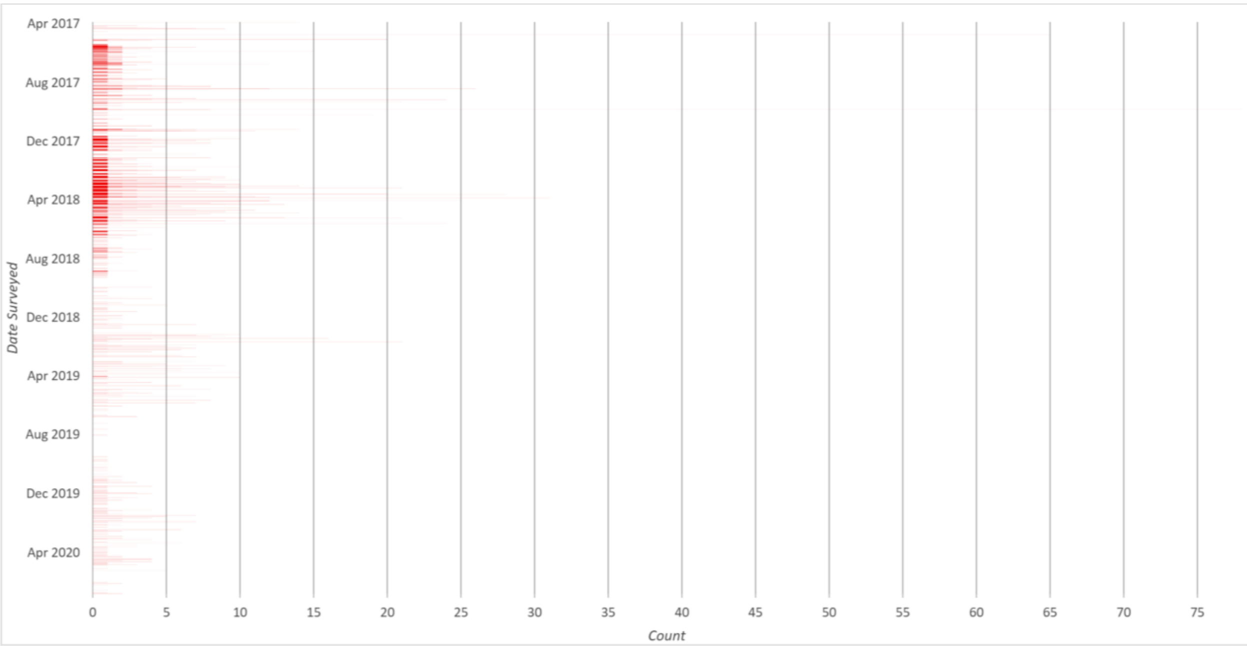

Figure 2. Number of reported “symptomatic” records of Myrtle rust presence over time for all data sets combined. Record peaks which are difficult to visualise occurred on 27/9/2017 with 78 records, 26/4/2017 with 65, 29/3 2018 with 31, 22/1/2019 with 21.

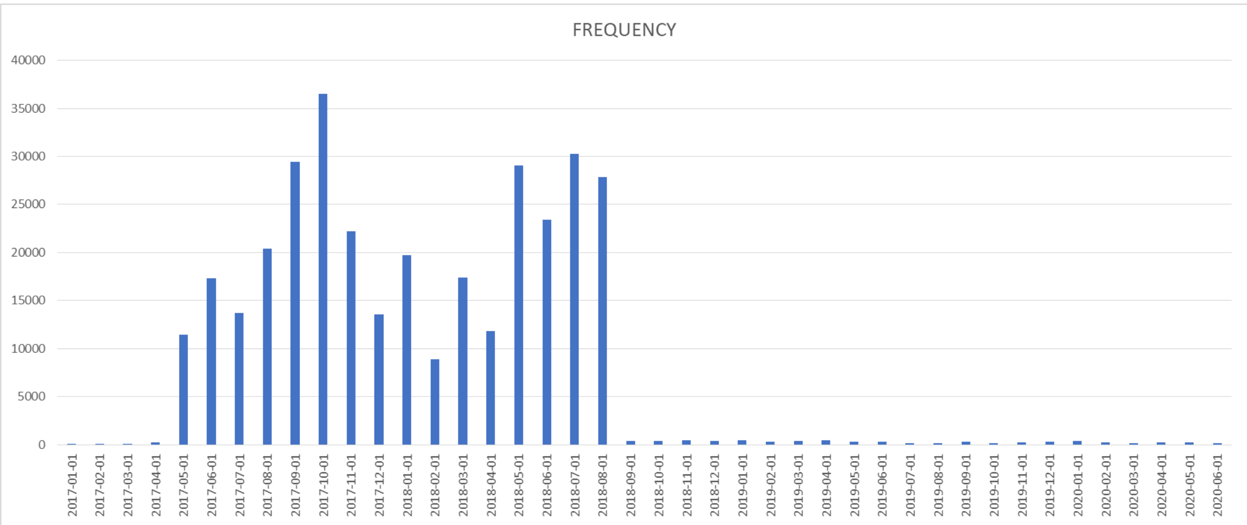

Figure 3. Distribution of surveillance effort over time. Data is grouped by month, with the first of the month shown as a time indicator. Later surveillance effort is overshadowed by the effort during the incursion response.

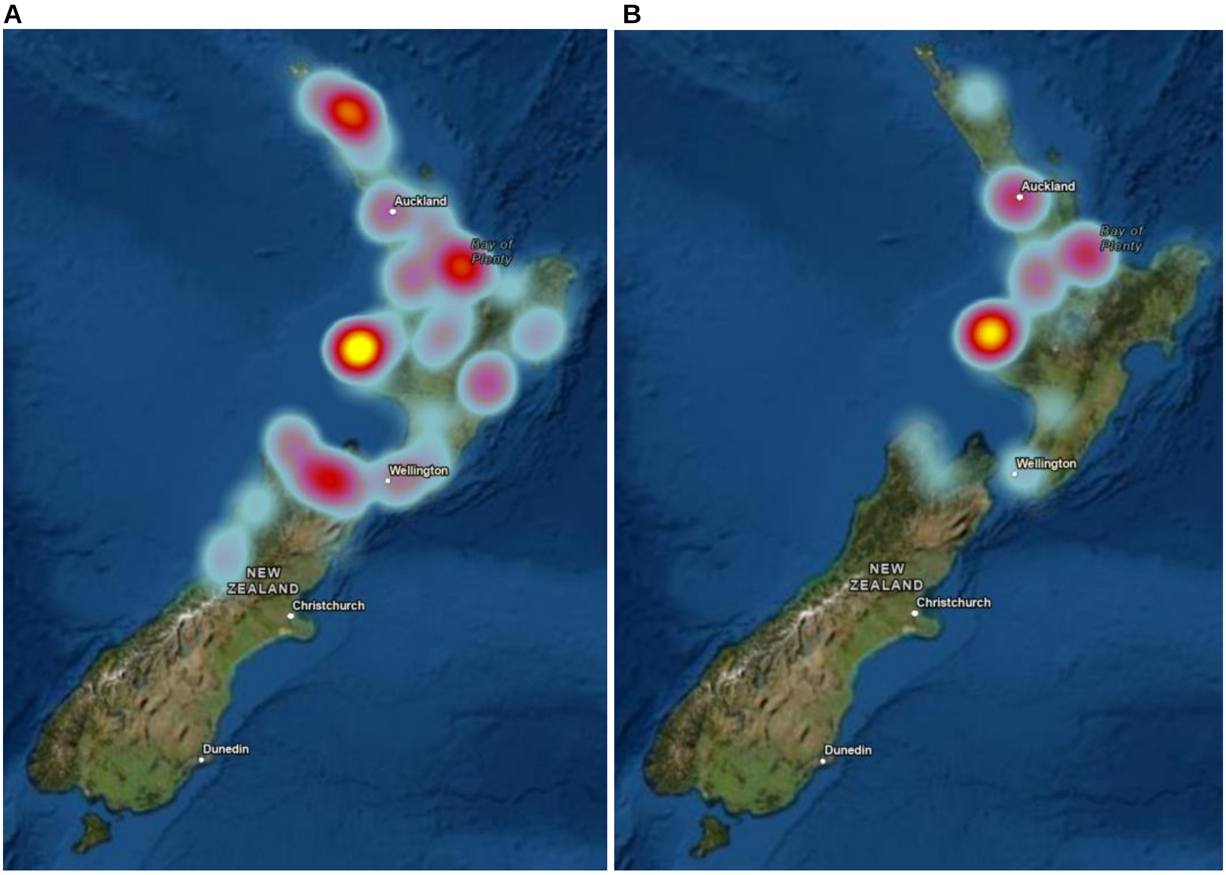

Figure 4. Density maps showing where surveillance (A) and symptomatic records (B) occurred nationally. Blue = sparse, yellow = dense. A) Total surveillance effort (positive and negative for Myrtle rust), B) Positive records only for myrtle rust.

Figure 5. Point records for national myrtle rust surveillance from January 2017 to June 2020. Pale points are locations of surveillance without myrtle rust detected (asymptomatic), red points show where symptomatic plants were detected.

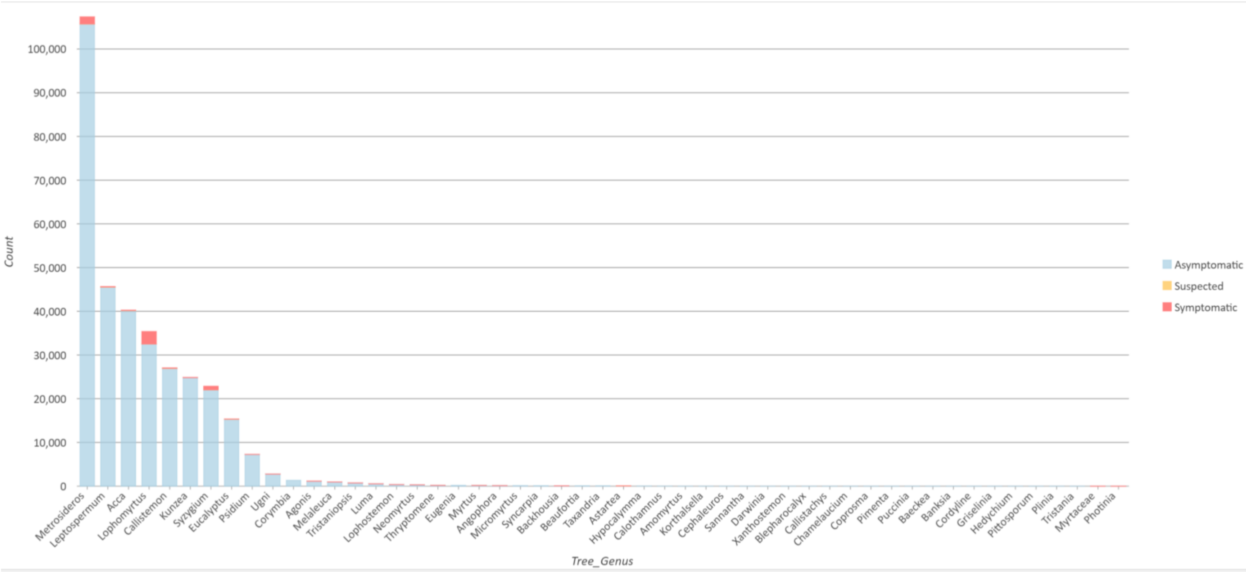

Figure 6. Surveillance effort for the various tree genera in relation to asymptomatic, symptomatic and suspected infection records, for the combined dataset.

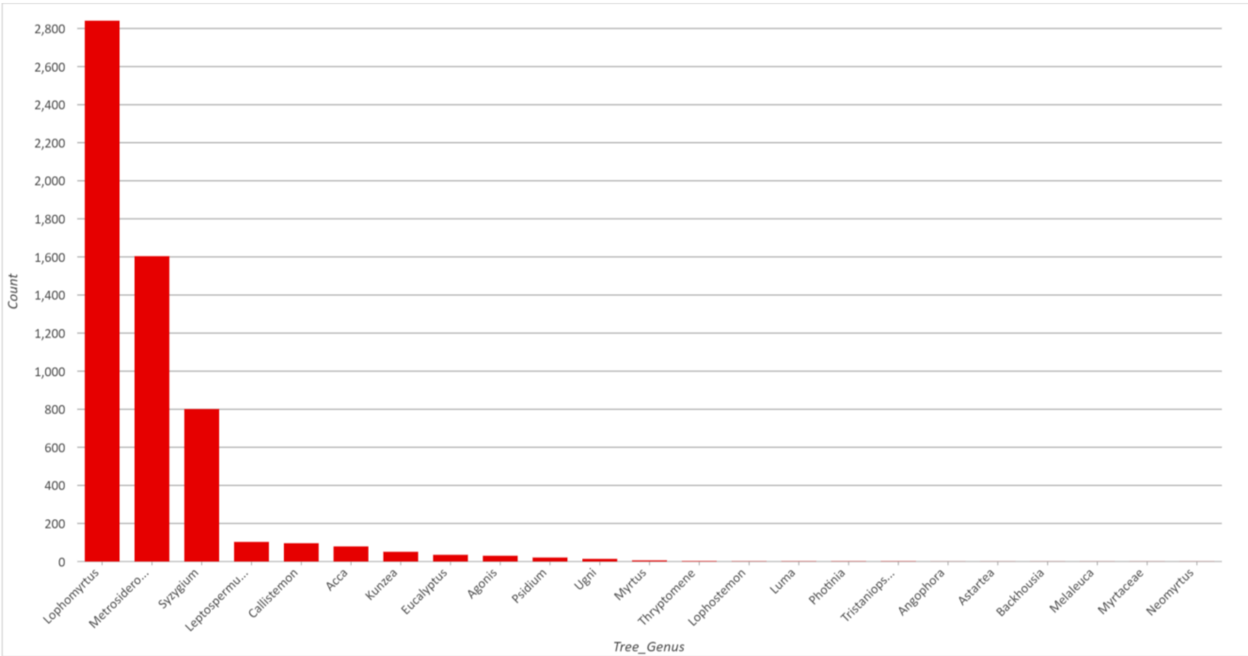

Figure 7. Plant genera ranked by number of symptomatic records only, for the combined dataset.

3.2 Iconic trees layer of important myrtaceous plant hosts

The list of iconic trees had 24 entries (Figure 8), the majority of which are significant pōhutukawa trees. All are in regions where MR has been detected as present at some time since the 2017 incursion response.

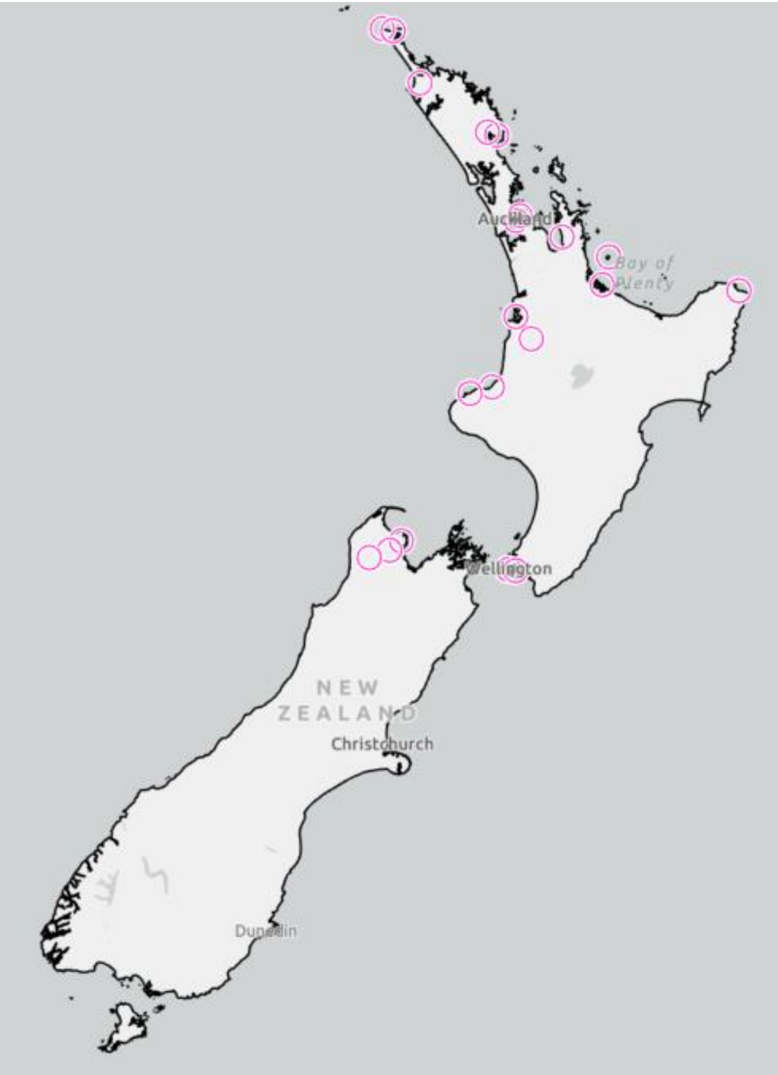

Figure 8. Locations of significant iconic Myrtaceae trees (in the public arena) in New Zealand.

3.3 Infection risk and latent period rasters created using the Myrtle Rust Process Model/climatic risk model.

Model output is available as .asc files summarising weekly mean infection risk or latent period. An .asc file has been produced for every week from 7 January 2017 and are currently being produced weekly https://niwa.co.nz/publications/water-and-atmosphere/water-atmosphere-24-july-2020/rust-coding. Provided here for the MR data library are examples of infection risk (e.g. Figure 9) and latent period (e.g. Figure 10) from January 2017 – July 2020.

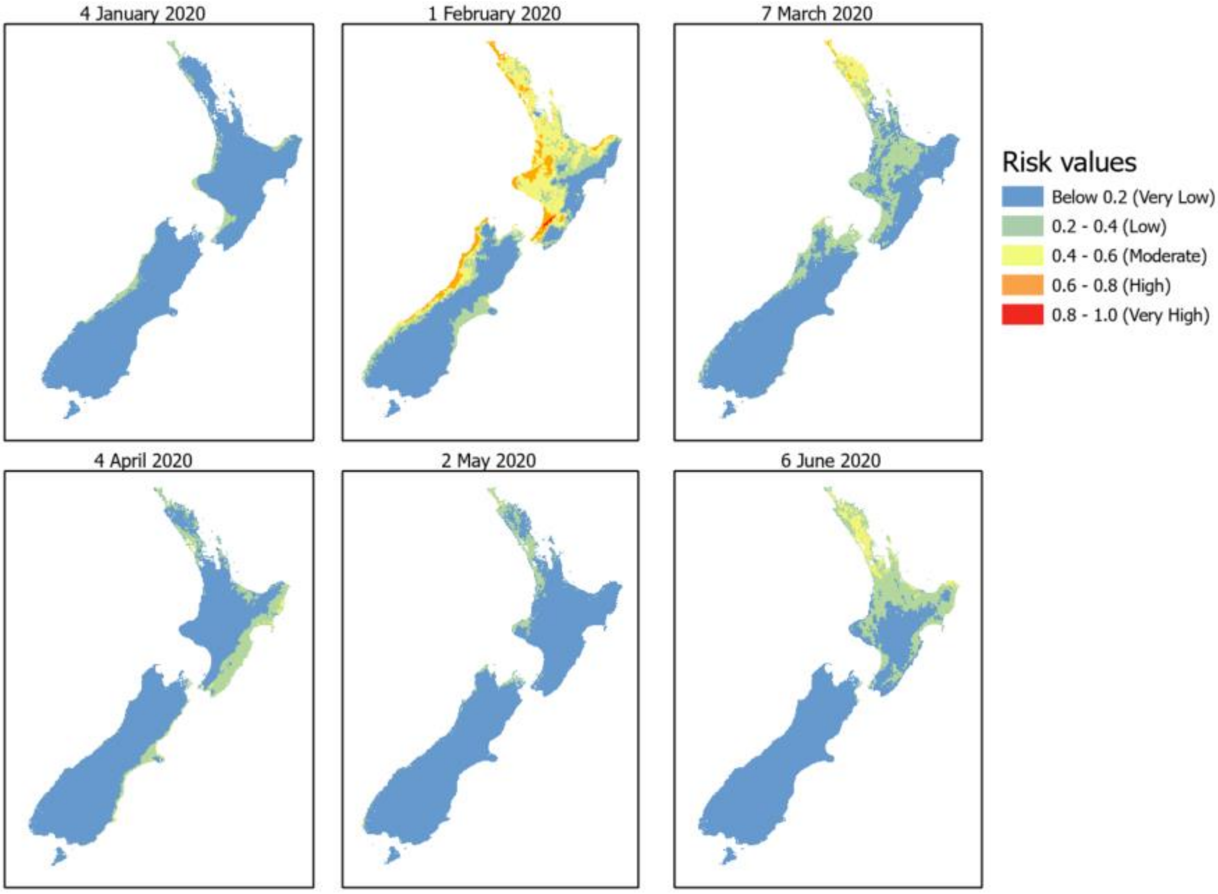

Figure 9. Examples of the infection risk maps for mean weekly myrtle rust infection risk based on the Myrtle Rust Process Model. Each map shows the mean risk values for the week prior to the date displayed.

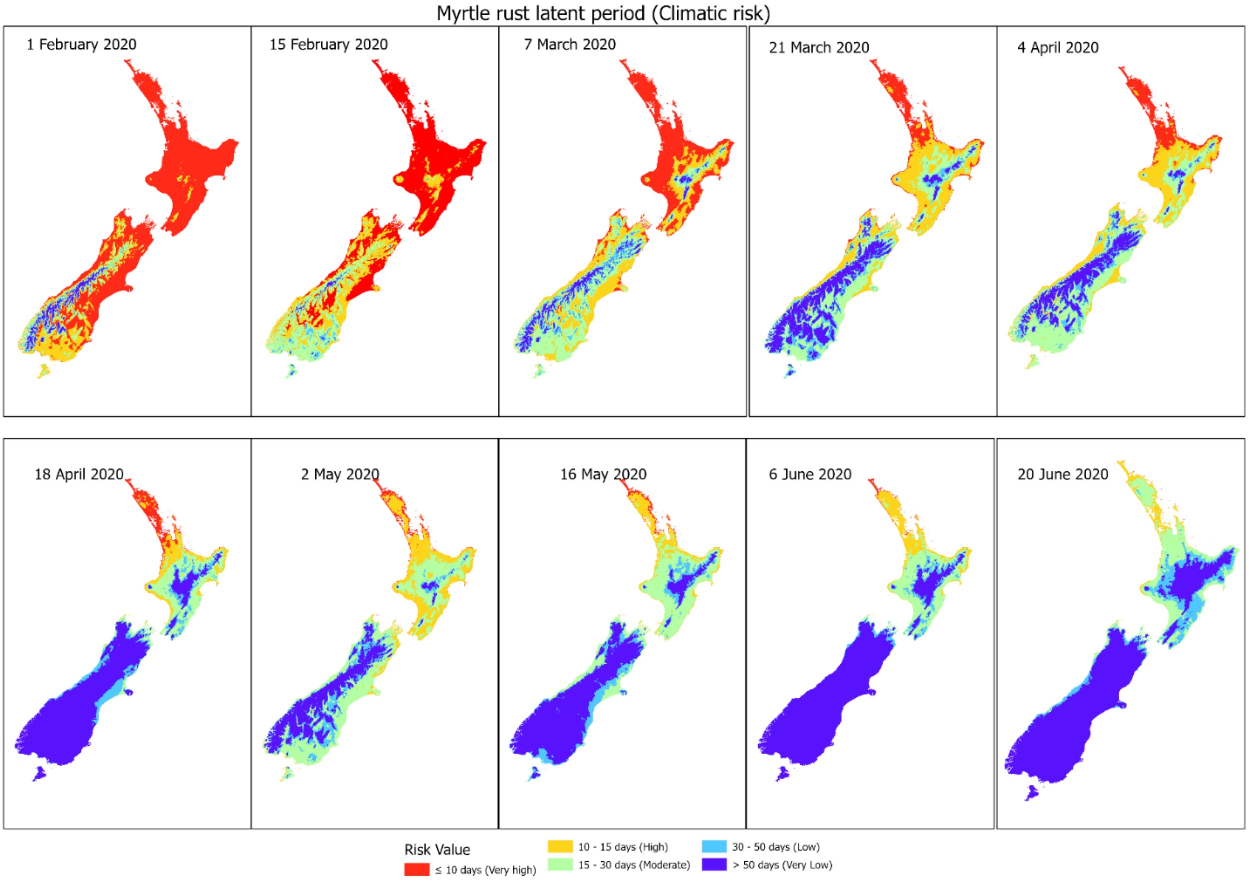

Figure 10. Examples of the Latent period predictions produced from the Myrtle Rust Process Model. Each map shows the mean latency period risk values for the 7 days prior to the date displayed.

4 Conclusions

The changes between incursion response and long term management resulted in an abrupt change in the data collection quantity and quality. Assumptions about the data are documented in the methods. These include where data were excluded or the quality of the data records where this information is available.

There is currently no existing data (other than within this dataset) to corroborate or validate the surveillance records. A national structured surveillance project would have to take place to gather extensive information of the current distribution of MR infection/symptoms. However, a number of records have photos associated with them which could be revisited to confirm or complete data records. Many of these photos are currently stored on ArcGIS Online (AGOL) from the incursion response, however linking them to records within the datasets was beyond the scope of this current project. The ‘Comments' fields within the dataset could also have more information which is not included/extracted into the correct data fields (this was also beyond the scope of this project). For example, host information, severity of infection or number of trees in the surveyed stand were sometimes only recorded in the comments field. Comments fields also sometimes contain specific location information which could be further quality checked. For example, S. Teasdale recognised that for some of the “MPI restricted places” data entries, the location description has been simplified to the start of access tracks (sometimes 8-10 km from the specific description in the location field).

Sampling errors will depend on the people who have contributed data. MPI data were confirmed in the laboratory in many cases or confirmed with photos by pathology and plant experts (Toome-Heller et al 2020). MR can also be present with no visible symptoms due to the latency and cryptic stages of the rust fungus, in which case, asymptomatic does not always mean that MR is not present. For other supporting information fields, we know there must be some data entry errors, as indicated by the presence of nonsensical values in some records (removed from combined data set).

We would also like to mention that plant species cannot be considered resistant based on absence of detection (Toome-Heller et al 2020). The number of plants infected from survey data can only give some tentative indication of susceptibility of these plants to the disease. Some geographical regions have more extensive and complete data (e.g. Taranaki) where these details may be more relevant. Susceptibility trials have been undertaken on a selection of native New Zealand Myrtaceae, and all species tested to date are considered susceptible, although variation in the levels of susceptibility and resistance within species varied (Smith et al. 2020). Furthermore, ongoing monitoring of Myrtaceae species in native forest showed that some plants were not infected until 2 years after the rust has been first detected at the site (Sutherland et al. 2020).

Other factors need to be quantified in the future to assess their effect on risk include host density, host growth and pathogen spore load (Beresford et al 2018). Ontogenetic resistance is closely linked to leaf expansion which is crucial to explaining the seasonality of MR development in the natural environment, therefore an inclusion of plant phenological stages would be helpful. Changes in plant vigour change the susceptibility of the host (young growth), therefore factors such as water availability or nutrients may also be important when considering the susceptibility to rust (Beresford et al. 2020).

5 Recommendations

There are several learnings from working with these incursion data that could help improve data collection in future incursion responses and also aspects which could better inform scientists and managers, helping them to respond and adapt to long term management more efficiently. These include:

- Metadata and data dictionaries of fields should be recorded more consistently from the beginning. High staff turnover in the involvement of MR management created knowledge gaps which were time consuming to resolve or required assumptions to be made about the quality of the data.
- More consistent methods for data collection could be employed from the initiation of the response, including between DOC and MPI. For example similar, but not the same, data fields were created in the DOC and MPI Survey123 data collection.
- Handover from ‘response’ to ‘long term management’ created a large data gap in space and time, of the collection of surveillance and monitoring data. This was probably most starkly driven by a sudden gap in funding.
- Climate change as a result of global warming is likely to increase the risk of MR in temperate climates through both a shortened *A. psidii* latent period and increased host active growth periods (Beresford et al. 2020).

6 Literature references

Beresford RM, Turner R, Tait A, Paul V, Macara G, Yu ZD, Lima L, Martin R, 2018. Predicting climatic risk of myrtle rust during its first year in New Zealand. NZ Plant Prot 71: 332 – 347.

Beresford RM, Shuey LS, Pegg GS, 2020. Symptom development and latent period of *Austropuccinia psidii* (myrtle rust) in relation to host species, temperature, and ontogenetic resistance. Plant Pathol 00: 1 – 11.

Toome-Heller M, Ho WWH, Ganley RJ, Elliott CEA, Quinn B, Pearson HG, Alexander BJR, 2020. Chasing myrtle rust in New Zealand: host range and distribution over the first year after invasion. Aust Plant Pathol 49: 327 – 328.

Smith GR, Ganley BJ, Chagné D, Nadarajan J, Pathirana RN, Ryan J, Marsh AT, Koot E, Carnegie AJ, Menzies T et al. 2020. Resistance of New Zealand provenance Leptospermum scoparium, Kunzea robusta, Kunzea linearis and Metrosideros excelsa to Austropuccinia psidii. Plant Dis. doi.org/10.1094/PDIS-11-19-2302-RE.

Sutherland R, Soewarto J, Beresford R, Ganley B, 2020. Monitoring Austropuccinia psidii (myrtle rust) on New Zealand Myrtaceae in native forest. NZ J Ecol 44(2): 3414.

Appendix 1. Myrtle-rust-combined-surveillance-dataset-methodology

1. Purpose

The purpose of this document is to provide an explanation of how the Myrtle Rust Surveillance Combined dataset was assembled, record what constituted a positive infection record, and provide names and locations of intermediate datasets.

The purpose of combining the dataset was to collate all New Zealand Myrtle Rust Surveillance data from the New Zealand incursion, for the period 1 January 2017 to 30 June 2020, into a single consolidated dataset, to facilitate analysis, improve data quality including detailed metadata improving the usability of the data in the future.

2. Context:

During the initial response to the Myrtle Rust incursion, several organisations began collecting information about where myrtle rust was found spatially. This was done using different systems, formats (schemas) and terminology, which are not easily compatible for analysis. Collating the data into single dataset could facilitate analysis. However, as there has been staff turn-over and as collation and interpretation of the data are retrospective, a record of decisions is recorded and explained here in case modifications are required in the future.

3. Copies of Original Source Datasets

Due to the rationalisation of the different schemas, some of the original fields (columns) were not included in the final combined dataset. However, copies of the original data are recorded if these excluded fields are required in future.

Copies of the original datasets are housed at:

i. the Department of Conservation (DOC)
ii. the New Zealand Institute for Plant and Food Research Limited

3.1 DOC Datasets

The original data is housed in NATIS1, in two feature classes:

i. NATIS1.NATISADM.OPERATIONAL_DOC_MyrtaceaePCLSurveillance
ii. NATIS1.NATISADM.OPERATIONAL_DOC_MyrtaceaePCLSurveillanceHistoric
iii. DOC_6_letter_codes

Location:
...MRDatasetMerge\a_Original_Data\DOC\DOC_MR_Data.gdb

3.2 MPI Host Plants Datasets

The original data is housed at AsureQuality. A "SQL Server Backup" copy of the MPI data from Quenten Higgan at AsureQuality with an intact relationship class (MYR_Host_Inspections) and coded domains (many).

Names:

i. HostPlant (feature class, aka MPI Host Plants)
ii. MYR_Host_Inspections (Relationship Class)
iii. MYR_SURVEYPLANTS (table, aka MPI Revisits)

Location:

...MRDatasetMerge\a_Original_Data\MPI\MYR_Archive_Mar20.gdb

3.3 MPI Restricted Places Dataset

Names:
MPI_Current_Restricted_Places_Dec19

Location:
MRDatasetMerge\a_Original_Data\MPI\MPI_Restricted_Places.gdb

3.4 iNaturalist / Myrtle Rust Reporter

Names:

i. myrtaceae_observations_99184_XYTableToPoint (data where taxon_family_name = "Myrtaceae")
ii. myrtle_rust_observations_99167_XYTableToPoint (data where species name = "*Austropuccinia psidii*")
iii. myrtle_rust_reporter_all_20200723_XYTableToPoint (data where project = "myrtle-rust-reporter")
iv. Taxon_Common_Names (table, plant names from NZTCS: https://nztcs.org.nz/)

Location:
...MRDatasetMerge\a_Original_Data\iNaturalist\iNaturalist.gdb

Notes:
Only records with quality_grade = 'research' were used, spanning 2017-01-01 to 2020-06-30.

As there was overlap between the three iNaturalist dataset downloads, data from the Myrtle Rust Reporter took preference, having additional information about the infection. Duplicate records in the other two data downloads where removed.

iNaturalistMyrtaceaeObservations gives an indication of the distribution of Myrtaceae species, which may be interesting particularly where Myrtle Rust has not yet been detected.

Where necessary, common names for host plants were obtained from NZTCS: https://nztcs.org.nz/).

3.5 Raoul Island Data

Name:
MRRI_AllSamples

Location:
...MRDatasetMerge\b_Intermediate_Data\Intermediate_Datasets.gdb

Notes:
Mostly positive infection records from Raoul Island.

3.6 Plant and Food Research

Name:
PFR_data_updates

Location:
...MRDatasetMerge\a_Original_Data\PFR\PFR_MR_Data.gdb

3.7 Auckland Botanical Gardens

Name:
AuckBotGard

Location:
...MRDatasetMerge\a_Original_Data\Auckland_Botanical_Gardens\AuckBotGard_MR_Data.gdb

3.8 Dunedin Botanical Gardens

Name:
DunedinBotGard

Location:
...MRDatasetMerge\a_Original_Data\Other\Other_Datasets.gdb

3.9 Horizons Regional Council

Name:
HorizonsRC

Loction:
...MRDatasetMerge\a_Original_Data\Other\Other_Datasets.gdb

4. Intermediate Datasets

Many of these datasets have coded domains, hence the requirement for intermediate datasets where the codes of the coded domains could be replaced with the domain description. Three fields were added to allow individual records to be traced back to the original datasets, these included: “Source” and “Source_OID”. As the presence of myrtle rust was recorded in different ways in the various datasets, a definitive field indicating the presence of myrtle rust was added called “Myrtle_Rust_Present”; with the values: "Symptomatic", "Suspected" and "Asymptomatic". Occasionally additional fields were included, e.g. a text field to record a GUID value for the MPI Hosts and MPI Revisits data.

Intermediate datasets are located here:
...MRDatasetMerge\b_Intermediate_Data\Intermediate_Datasets.gdb

4.1 DOC Datasets

Names:
OPERATIONAL_DOC_MyrtaceaePCLSurveillance
OPERATIONAL_DOC_MyrtaceaePCLSurveillanceHistoric

Myrtle Rust Symptomatic SQL:
DOC_NoCodedDomainsPCLSurveillance:
Myrtle_Rust_present_ IN ('ConfirmedYellowSpores' , 'ConfirmedLesions' , 'ConfirmedOlderLesionsPits')

DOC_NoCodedDomainsPCLSurveillanceHistoric:
Canopy_Impact_Score IN ('Confirmed positive by lab results' , 'Obvious leaf rust but no dead stems or branches' , 'Obvious leaf rust dead stems or branches' , 'Old and new leaf with some rust')

4.2 MPI Host Plants Datasets

Name:
MyrtleRustHosts_5thMarch2019NoDomains

<u>*Myrtle Rust Symptomatic SQL:*</u> *Canopy_Impact_Score IN (3, 4, 5) OR Survey_Completed IN ('POS' , 'PSR' , 'REM')*

MPI_Revisits

<u>*Myrtle Rust **Symptomatic** SQL:*</u> *CanopyImpactScore IN ('Obvious leaf rust but no dead stems or branches', 'Obvious leaf rust dead stems and branches', 'Obvious leaf rust dead stems or branches', 'Old and new leaf with some rust', 'Obvious leaf rust but no dead stems or branches', 'Obvious leaf rust but no dead stems or branches') And Likelyhood IN ('High', 'Low', 'Medium') And TreeBranchScore IN ('Extremely susceptible', 'Highly susceptible', 'Moderately susceptible', 'Relatively tolerant')*

<u>*Myrtle Rust **Suspected** SQL:*</u> *CanopyImpactScore IN ('Obvious leaf rust but no dead stems or branches', 'Obvious leaf rust dead stems and branches', 'Obvious leaf rust dead stems or branches', 'Old and new leaf with some rust', 'Obvious leaf rust but no dead stems or branches', 'Obvious leaf rust but no dead stems or branches') And Likelyhood IN ('High', 'Low', 'Medium') And TreeBranchScore IN ('Not infected or tolerant')*

<u>Notes:</u>
In order to transfer geometry (xy values) to the MPI_Revisits table to convert it into a feature class, the “Migrate Relationship Class_Data Management” tool was used to enable the transfer of xy values.
There were 914 records from the MPI Revisits dataset (20 positive, 1 suspected and 893 negative) that didn’t have a relationship to MPI Host Plants feature class, thus their actual location is unknown. These points were placed offshore at 1717066.68, 5536560.85 (xy) rather than allowing them to default to 0,0 in the “MR_Revisits” intermediate dataset, but these records were not added to the final MR_Combined_Surveillance dataset.

4.3 MPI Restricted Places Dataset

Names:
MPI_Current_Restricted_Places_Dec19
<u>*Myrtle Rust Infection SQL:*</u> *MR_Status IN ('Confirmed' , 'Confirmed (Photo)')*

4.4 iNaturalist / Myrtle Rust Reporter

Names:

iNaturalist_Myrtle_Rust_Reporter (data where project = “myrtle-rust-reporter”)
iNaturalist_Myrtle_Rust_Observations (data where species name = “Austropuccinia psidii” but not in Myrtle Rust Reporter)

iNaturalist_Myrtaceae_Observations (data where taxon_family_name = “Myrtaceae” but not in Myrtle Rust Reporter)

Notes:
Only records with quality_grade = ‘research’ were used.

iNaturalist_Myrtle_Rust_Reporter and iNaturalist_Myrtle_Rust_Observations
Myrtle Rust Infection SQL: scientific_name = 'Austropuccinia psidii'
Host data (tree genus and tree species) was usually extracted from the “Comments” field when recorded. The comments field was not altered.

4.5 Raoul Island Data

Name:
MRRI_AllSamples

Notes:
All records were symptomatic

4.6 Plant & Food Research

Name:
PFR_data_updates

Notes:
Myrtle Rust Infection SQL: MR_POS_NEG = 'POS'

4.7 Auckland Botanical Gardens

Name:
AuckBotGard

Notes:
Myrtle Rust Infection: symptomatic myrtle rust infections were recorded in iNaturalist, only asymptomatic records are recorded in AuckBotGard.

4.8 Dunedin Botanical Gardens

Name:
DunedinBotGard

Notes:
Myrtle Rust Infection: symptomatic myrtle rust infections were recorded in iNaturalist, only asymptomatic records are recorded in DunedinBotGard.

4.9 Horizons Regional Council

Name:
HorizonsRC

Notes:
All records are symptomatic

5. Final Dataset

Name:
MR_Combined_Surveillance

Location:
...MRDatasetMerge\c_Combined_Data\Combined_Surveillance.gdb

Notes:
The final cleaning of the data was undertaken. Records dated before 2017-01-01, which were mostly test entries were removed.

An AGOL web feature layer of the data published through ArcGIS Pro is available here:

http://deptconservation.maps.arcgis.com/home/item.html?id=77f0442ba4704edeaff2535752a345d5

AGOL Webapp map:

https://deptconservation.maps.arcgis.com/apps/webappviewer/index.html?id=5defa960ce24425ab59de932d5f7f55f

Appendix 2. Methodology-field-matching

Table A2: Field matching from the component datasets to the combined dataset. MPI = Ministry of Primary Industries, PFR = The New Zealand Institute for Plant and Food Research Limited, MRR= Myrtle Rust Reporter, MRRI = Myrtle Rust Raoul Island, DOC = Department of Conservation. The smaller datasets includes Regional councils, Botanic Gardens and PFR data.

Final Field Name	Description	DOCPCL Surveillance Historic	DOCPCL Surveillance	MPIHosts	MPIPlant Survey	iNaturalist_ Myrtle_Rust_ Observations	iNaturalist_ Myrtle_Rust_ Reporter	iNaturalist_ Myrtaceae_ Observations	Restricted Places	Raoul Island	Smaller datasets[1]
data_source	Who owns/collected the original data	DOCPCL SurveillanceHistoric	DOCPCLSurveillance	MPI_MyrtleRustHosts	MPI_Revisits	iNaturalist_Myrtle_ Rust_Observations	iNaturalist_Myrtle_ Rust_Reporter	iNaturalist_ Myrtaceae_ Observations	MPI_Current_ Restricted_Places _Dec19	MRRI_All Samples	Auckland Botanic Garden, Dunedin Botanic Garden, Horizons Regional Council, Plant_and_Food_Research
data_quality_ estimate	A gauge of the quality and trustworthiness of the data (various measures)	Species_ Confidence	NA	NA	NA	num_identification_ agreements	num_identification_ agreements	num_identification_ agreements	Confidence	NA	NA
planting_type	Captive/Cultivated, Nursery, Planted/Wild, etc.		NativeExotic			captive_cultivated	captive_cultivated	captive_ cultivated	Classification		DataSource
canopy_impact_ score	How impacted the canopy is by Myrtle rust infection	Canopy_Imp_Score		Canopy_Impact_Score	CanopyImpactScore		severityofinfection				
tree_genus	Genus of host plant	Tree_Genus	Genus	Tree_Genus	Tree_Genus	scientific_name	Fieldwhatplant speciesisinfected	scientific_ name			Tree_Genus
tree_species	Species of host plant when available	Tree_NVS_Code	NVSCode			scientific_name	Fieldwhatplant speciesisinfected	scientific_name	Host_Species		Tree_Species
date_surveyed	Date plant was surveyed or data were created (if two dates, the earlier was used)	CreationDate	CreationDate	Date_Surveyed, CreationDate	DateSurveyed, created_date	observed_on	observed_on	observed_on	Date_Reported	Created, Date_Found	Date_Surveyed
numberl_infected	Number of plants infected		NoPlantsInfected				Numberofplants infected			Num	
number_surveyed	Number of plants surveyed		NoPlantsSurveyed								
myrtle_rust_ symptoms	Symptom records, depending on data collected. E.g. yes, no, combination from canopyimpactscore, survey completed, scientific name (*Austropuccinia psidii*)	Canopy_Imp_Score	MyrtleRustPresence	Survey_Completed	ManagerComment	scientific_name	scientific_name		MR_Status	Status	MR_POS_NEG
myrtle_rust_ symptoms_2	Symptom records depending on data collected, some datasets recorded more symptom fields			Tree_Branch_Score	TreeBranchScore		presence_o				
creator	Creator of data	Creator	Creator	Creator	created_user	user_id	user_id	user_id			
creator_or_editor_2	Creator or editor additional to creator		username			user_login	user_login	user_login			

Final Field Name	Description	DOCPCL Surveillance Historic	DOCPCL Surveillance	MPIHosts	MPIPlant Survey	iNaturalist_ Myrtle_Rust_ Observations	iNaturalist_ Myrtle_Rust_ Reporter	iNaturalist_ Myrtaceae_ Observations	Restricted Places	Raoul Island	Smaller datasets[1]
editor	Editor additional to Creator or another editor	Editor	Editor	Editor	last_edited_user						
edit_date	Date data edited in the original dataset	EditDate	EditDate	EditDate	last_edited_date	updated_at	updated_at	updated_at			
photo	Links to available photos of plant and/or symptoms					image_url	image_url	image_url			
comments	Comments field, contains various information, sometimes information that should be in other fields	Comment	Comments	ManagerComment	ManagerComment	description	description	description	Notes	COMMENTS	Comments
revisits	Relational table parameters for revisiting sites/trees			RevisitDate	RelatedOBJ				Date_Reinfection_Reported, Date_FollowUp		
tree_height	Estimate height of tree in metres	Tree_Height_m	PlantHeight	TreeHeight	TreeHeight		heightoftrees				
percent_tree_ surveyed	Percent of the tree that was inspected	Perc_Tree_ Surveyed	PartsPlantSurveyed	TreeID	PercPlantInsp						
confidential	Level of confidentiality in addition to standard confidentiality. Particularly MPI data sets or within iNaturalist where users have specifically asked for geoprivacy.			Do not identify to landowner	Do not identify to landowner	geoprivacy	geoprivacy	geoprivacy	High confidentiality		
myrtle_rust_present	Summary of symptomatic, asymptomatic, suspected interpreted from single or multiple original fields	Canopy_Imp_Score	MyrtleRustPresence	Canopy_Impact_Score, Survey_Completed, Tree_Branch_Score	CanopyImpactScore, ManagerComment, TreeBranchScore	scientific_name	scientific_name	NA	MR_Status	Status	MR_POS_NEG

[1] As most of the smaller datasets where sourced in excel rather than a GIS, field matching wasn't required. Instead, excel spreadsheet column names were matched to those in the mr_combined_surveillance dataset to facilitate importing them.

Appendix 3. General data specifications for data suppliers to environmental reporting

This guidance sets out the basic specifications for how data should be formatted and supplied to MfE and Stats NZ's Environmental Reporting programme. These guidelines have been adopted to:

1. enable efficient and consistent workflows for Environmental Reporting's report production,
2. ensure compliance with the Declaration on Open and Transparent Government's open data policies.

Before final data are supplied, please fill out the following checklist:

File format

If data are tabular:

- ☐ Supplied in .CSV format.
- ☐ Alternative format has been agreed with relevant ER contract owner.
    - Format type: ____________________

If data are spatial:

- ☐ Supplied in .GPKG format (OGC GeoPackage).
- ☐ Vector spatial data supplied with a coordinate reference system of EPSG: 2193 (NZTM2000).
- ☐ Raster spatial data supplied in its native coordinate reference system (and this should be clearly identified in the metadata using an EPSG code (preferred), or if not available use proj.4) .
- ☐ Alternative format has been agreed with relevant ER contract owner.
    - Format type: Climate risk rasters as .asc files, EPSG 2193/NZGD2000 Transverse Mercator
- ☐ Alternative coordinate reference system has been agreed with relevant ER contract owner.
    - CRS: ____________________

Documentation

- ☐ A data dictionary supplied for every dataset. *The format of this should be a two column .CSV file, with the first column being the variable name and the second column being a short description of the variable. One row for each variable.*
- ☐ Metadata for spatial data supplied compliant with ISO 19115-3 and in standard XML.

File naming

- ☐ Files names contain only lowercase letters, numbers, and dashes.

Code

It is preferred that all analysis is undertaken using a reproducible, code-based workflow. And our preferred coding language is R.

Unless otherwise specified in your contract, it is expected that all relevant code be supplied at the time of data delivery. Relevant code includes but is not limited to; steps taken to clean, analyse, transform, model, visualise and export data.

- ☐ Code supplied is tidy, commented, and succinct.
- ☐ Where dependencies exist (e.g. packages or functions), these are supplied or publically available.

Data

- ☐ Data supplied with unambiguous site names or reference numbers *(e.g. it is not acceptable to record an observation as being from "Masterton" if there were multiple monitoring stations within Masterton).*
- ☐ All times recorded in NZST time – unless otherwise specified. *The date or time of observations should also be provided to the lowest unit appropriate to the data, e.g. if a read of ocean acidity was taken at 2020-05-26 13:45, this should not be shortened to the day only (i.e. 2020-05-26).*
- ☐ Column names should be lower case with words separated by underscores, and no special characters.

Licensing

- ☐ Data are appropriate for us to redistribute under Creative Commons CC BY 4.0 license.
- ☐ An alternative open license has been agreed with relevant ER contract owner.
  - License type: Surveillance/monitoring data licensed under CC only after aggregation and checking with original data suppliers as required.

Completed by

Name:	**Rebecca Campbell**
Organisation:	**Plant & Food Research**
ER contract owner:	
Date:	**31 August 2020**

A smart
green
future.
Together.