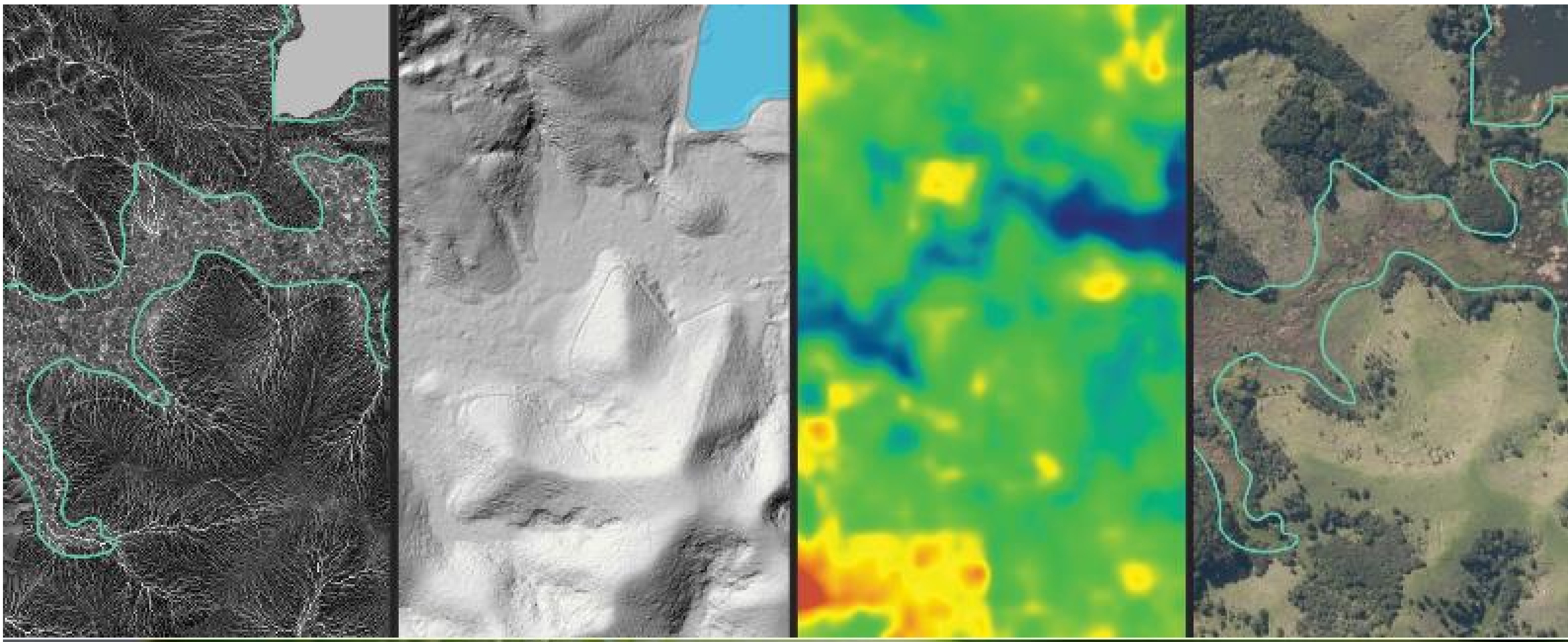
Engineers & Consultants

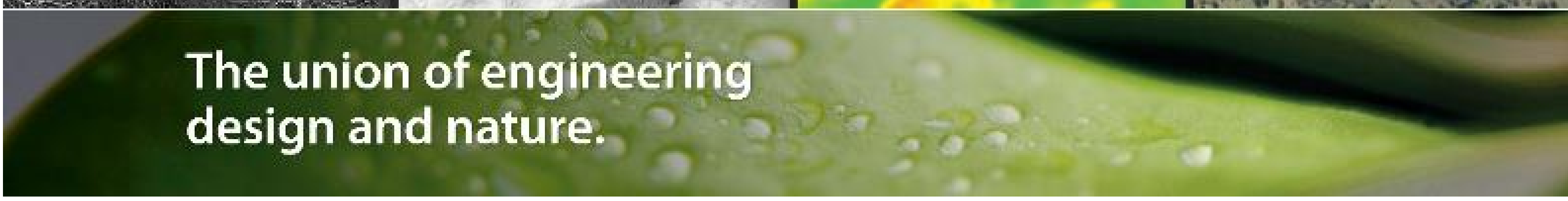
Proof of Concept for

Wetland Mapping Methods

Final

Prepared for Ministry for the Environment by Morphum Environmental Ltd and Lynker Analytics May 2021





09 377 9779 | info@morphum.com | www.morphum.com | Auckland | Waikato | Wellington | Nelson | Melbourne

Engineers & Consultants

Document Control

Client Name: Ministry for the Environment

Project Name: Wetland Mapping Methods

Project Number: P02662

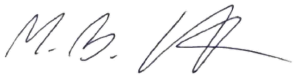
Document: Proof of Concept for Wetland Mapping

Revision History

Status	Date Issued	Authors	Reviewers	Releaser
Draft – for workshop	15/03/2021	Lynker Analytics: Matt Lythe Rātā Chapman Olsen David Knox Morphum Environmental: Mark Lowe Jacqui McCord	Morphum Environmental: Mark Lowe Lynker Analytics: Matt Lythe	Morphum Environmental: Stu Farrant
Final Draft	06/05/2021	Lynker Analytics: Matt Lythe Rātā Chapman Olsen David Knox Morphum Environmental: Mark Lowe Jacqui McCord	Morphum Environmental: Mark Lowe Lynker Analytics: Matt Lythe	Morphum Environmental: Stu Farrant
Final	28/05/2021	Lynker Analytics: Matt Lythe Rātā Chapman Olsen David Knox Morphum Environmental: Mark Lowe Jacqui McCord	Morphum Environmental: Mark Lowe Lynker Analytics: Matt Lythe	Morphum Environmental: Stu Farrant

Reviewed by:

Reviewer: M. Lowe **Signature:** 

Reviewer: M. Lythe **Signature:** 

Released by:

Reviewer: S. Farrant **Signature:** 

Recommended Citation: Lythe, M., Lowe, M., Farrant, S., Chapman Olsen, R., Knox D., McCord, J (2021). Proof of Concept for Wetland Mapping Methods. Prepared for Ministry for the Environment by Morphum Environmental and Lynker Analytics. Final, Version 1. Morphum Project Number: P02662.

Executive Summary

The National Policy Statement for Freshwater Management 2020 (NPS-FM) requires regional councils to identify and map natural inland wetlands 500 m^2 in extent or greater; or those naturally smaller in size where known to contain threatened species. However, discriminating wetlands is not a straightforward process given the ecological variability and transitional nature of these ecosystems.

This report summarises a proof of concept developed to delineate and classify wetlands in New Zealand using commonly available earth observational data and machine learning techniques. Conducted in Northland and Tasman, the pilot development applied a three-phase methodology comprising Masking, Classification and GIS post processing. All steps were designed to be run using open-access data, open-source algorithms, and commonly available software. Different data inputs were tested including the Sentinel-2A Multispectral Instrument with 13 bands, LiDAR and its derivatives, Aerial Photography (3 and 4 band) and radiometric data derivatives such as wetness gradient.

Five modelling scenarios were tested including semantic segmentation using convolutional neural networks (CNN) and classification decision trees using Random Forest. The best performing algorithm achieving an overall pixel accuracy of 0.83 was a CNN model using Sentinel-2A data, RGBI aerial photography and LiDAR. A Random Forest classification model using similar data inputs performed almost as well. The best performing CNN and Random Forest models were capable of discriminating wetlands at a pixel level from non-wetlands with an accuracy of 0.83. Spatial concordance of the modelled polygons with a small validation set was however relatively low. This incongruence reflects the difficulty of using a rules based GIS process to generate realistic polygons from the modelled pixels that approximate mapping by a trained ecologist.

Currently the model outputs are biased towards a high recall, low precision approach meaning they are attempting to predict all plausible wetland areas including likely false positives such as wet pasture, drainage channels and areas of vegetation resembling assemblages found in natural wetlands. However, confidence thresholds are included with the output polygons to enable a higher precision-based refinement of the dataset.

The findings suggest that an ensemble model of Random Forest combined with CNN could do better than any of the individual models tested thus far. This ensemble model would allow examination of input feature importance enabling more prediction emphasis to be placed on important features such as landscape form or canopy height rather than vegetation or water response where these might be less discriminatory. There is also the potential for a model with fewer output classes to perform better. Therefore, it is recommended that a 5 or 6-class aggregated model also be tested. Additionally, an approach that applies a separate model for pakahi and gumland may improve model outputs.

Augmenting the machine learning process with oblique photography for training and validation is likely to improve the mapping quality. Accurate GIS data describing impervious surface, soils, stormwater assets, ponds and vegetation in the post processing phase will also further improve the quality of the mapping approach.

Whilst it is possible to implement the recommended approach outlined in this proof of concept (as described in Appendix 4: Processing Pipeline) it is recommend that the refinements that have been identified to lift the accuracy and quality of the model outputs (as outlined in section 8: Recommendations) are tested and any potential improvements reflected in an updated processing pipeline along with associated algorithms and scripts.

In conjunction with the previous literature review undertaken as part of this project (Lythe et al., 2020), this proof of concept has confirmed that the proposed approach is potentially feasible to support the implementation of the NPS-FM mapping requirements. The proof of concept has identified a preferred model approach along with a series of recommendations to improve model accuracy and quality.

Contents

Figures

Tables

Glossary

Term	Definition
Convolutional Neural Networks (CNN)	A representative form of deep learning that is used for visual recognition. Convolutional Neural Networks (CNN)s utilise the spatial context of detected features to identify objects and classify scenes.
Decision Trees (DT)	A flowchart or way of structured thinking where each node represents a feature (attribute), each link (branch) represents a decision (rule), and each leaf represents an outcome. DT Algorithms include Random Forest, Boosted Regression and Classification and Regression Trees (CART).
Digital Elevation Model (DEM)	Graphical representation of elevation data to digitally represent terrain/land surface. LiDAR derivative.
Freshwater Ecosystems New Zealand (FENZ) geodatabase	National spatial dataset of rivers, lakes and wetlands. It has a minimum mapping unit of approximately 0.5 hectares and includes classification of only palustrine and inland saline hydrosystems.
Hydrosystem	Broad ecological category based on hydrological and landform setting, salinity, temperature.
Land Cover Database version 5 (LCDB v5)	National spatial dataset of land cover at 30 m resolution and minimum mapping unit of 1 ha.
LiDAR	Light Detection and Ranging (LiDAR) is an active remote sensing method that emits light in the form of a pulse laser and measures the way it is reflected from the earth's surface.
Near Infra-Red (NIR)	Spectroscopy method that is in the near infrared region of the electromagnetic spectrum (780 nm to 2500 nm)
Normalised Difference Vegetation Index (NDVI)	A measure of vegetation density.
Normalised Difference Water Index (NDWI)	An estimation of the leaf water content at canopy level
NPS-FM	National Policy Statement for Freshwater Management (2020)
Object based approaches	Partitioning an image into homogenous segments (called objects).
RADARSAT	Canadian satellite program run by the Canadian Space Agency (CSA)
Random Forest (RF)	Machine learning algorithm that employs an ensemble of decision trees
Remote Sensing (RS)	Remote Sensing is the field of observing the earth using sensors (such as cameras or LiDAR) from satellites or aircraft and the process of using this data to monitor or detect the physical characteristics of an area.
RGB / RGBI	An RGB image, sometimes referred to as a truecolour image, contains red, green, and blue colour components for each individual pixel. An RGBI image has an infrared band added. RGB/RGBI images in this report are aerial photography sourced.
Sentinel-1A and Sentinel-1B	Two satellites which share the same orbital plane with 20 m spatial resolution. Sentinel-1 is a Synthetic Aperture Radar (SAR) mission, providing continuous all-weather radar data.
Sentinel-2	Satellite imagery of 10m spatial resolution offering new imagery every 5 days.
Synthetic Aperture Radar (SAR) sensors	Sensors which acquire information under vegetation canopies and in cloudy conditions are useful with short-wavelength.

Term	Definition
The Landsat Program	This satellite program is a joint venture by National Aeronautics and Space Administration (NASA) and United States Geological Survey (UGGS) providing continuous imagery of Earth's land.
Topographic Position Index (TPI)	Algorithm used to determine topographic slope position. LiDAR derivative.
Topographic Wetness Index (TWI)	Metric for terrain determined variation in soil moisture. LiDAR derivative.

1. Introduction

Natural wetlands are productive habitats for a range of flora and fauna (Turpie et al., 2015) and provide a range of ecosystem services. However, they are also one of the most threatened ecosystems on earth (Mahdianpari et al., 2020). In New Zealand, approximately 10% of the historical extent remains (Ausseil et al., 2008; Ausseil et al., 2011; Myers et al., 2013; Robertson et al., 2018). Wetland loss and the resulting decline of the ecosystem services they provide have widespread negative impacts including vulnerability to climate change, biodiversity loss, water quality impacts, cultural effects, human health, amenity and recreational use.

The National Policy Statement for Freshwater Management 2020 (NPS-FM) requires regional councils to identify and map their natural inland wetlands to 500 m^2 in extent or greater, or those naturally smaller known to contain threatened species, and establish and maintain an inventory of wetlands that includes (at a minimum), the following information:

- identifier and location,
- area and Geographic Information System (GIS) polygon,
- classification of wetland type,
- any existing monitoring information, and,
- may include any other information, e.g. values assessment.

This NPS-FM requirement is the primary driver of this project. This technical report follows a comprehensive review of literature, data requirements and earth observation science techniques for wetland mapping at large scales (Lythe et. al., 2020) which outlined the technical approach to be used in this proof of concept (POC).

National wetland inventory development, and in turn wetland management, monitoring, and conservation, is one of the application areas that are expected to benefit from the increasing availability and capability of big data technologies including machine learning and cloud computing.

This technical report summarises a POC solution developed to identify, delineate, and classify wetlands at a national and/or regional council level using machine learning. It explains the design approach, algorithmic methods, data inputs, accuracy, limitations and cost implications of the approach.

The report layout is as follows:

1. Introduction
2. Objectives, study area, classification system
3. Methods and data processing
4. Model results and accuracy
5. Challenges and limitations
6. Deliverables
7. Key findings
8. Recommendations.

2. Objectives and Scope

In this section the scope and objectives of the POC are discussed along with the study area, data inputs and classification categories.

This technical report follows a comprehensive review of literature, data requirements and earth observation science techniques for wetland mapping at large scales (Lythe et. al., 2020) which outlined the technical approach to be used in this POC. The proposed method was presented and agreed upon with the Ministry for the Environment (MfE) and an appointed third-party reviewer. This literature review should be referred to by the reader for further context on the methods selected to be carried forward to the POC.

Limitations and risks were discussed during a workshop which covered the recommended approach and previous commissioning meeting; some of the limitations that remain are discussed within this technical report (Section 5).

2.1 Proof of Concept Project Objectives

The purpose of the POC was to design, build and test technical solutions to identify, delineate and classify natural inland wetlands down to a minimum mapping unit area of 500 m^2 in two geographically diverse regional areas of New Zealand.

The project intent was to:

- Develop a pragmatic and cost-effective method to map wetland extent to a minimum area of 500 m^2 or smaller that can be applied consistently at a national or regional scale and will be achievable for councils.
- Assess the utility and information value of a range of data inputs including medium resolution open access satellite imagery, radiometric survey data, LiDAR and high-resolution vertical aerial photography including RGB and RGBI bands.
- Assist in improving the understanding of the extent and location of New Zealand's remaining natural inland wetlands (including smaller degraded wetlands on private land) to aid and improve policy implementation, monitoring and compliance, and to also increase our capability for environmental reporting.
- Provide a tool to assist regional councils in carrying out obligations under the NPS-FM.

2.2 Scope

Specifically, the POC scope is to:

- Design and implement deep learning models to identify and delineate natural inland wetland (as defined by the NPS-FM) to an area of 500 m^2 or smaller using any of the available suite of data supplied by Northland Regional Council and Tasman District Council and/or any other data that is freely available.
- Develop methods to classify identified wetlands using the Johnson and Gerbeaux (2004) framework (hydrosystem and wetland class).
- Undertake a comparison of model results against existing and verified regional wetland mapping.
- Provide recommendations on optimum method and data inputs.
- Identify gaps or limitations including on levels of confidence in results and ability to apply the NPS-FM definitions remotely.

Note: The outputs from this POC include areas of public conservation land; however, as per the NPS-FM, regional councils need not map wetlands in these areas and may wish to exclude these areas to improve processing times. Also, the Coastal Marine Area (CMA) has been excluded from the POC analysis and map outputs (due to being excluded from the natural inland wetland definition in the NPS-FM). However, saltmarsh wetlands are still mapped in model outputs either where these wetlands naturally occur landward of the CMA due to backwatering of saline influence, or where the CMA is poorly mapped.

The following is not in scope of this project:

- The assessment of wetlands for biodiversity or functional values (current or potential).
- Extending the method to identify, delineate or classify former wetlands or other wetlands that do not meet the NPS-FM definition of natural inland wetlands including areas of improved pasture subject to temporary rain derived water pooling.
- Assessment of wetland form, structural class or vegetation assemblage.
- Creating new datasets for input into the model.

2.3 Study Areas

The POC was conducted in the Northland and Tasman regions. These areas have been identified as suitable for the following reasons:

- The diversity of wetlands,
- Availability of suitable data which is representative of a variety of wetlands on a national scale,
- Existing detailed mapping and classification,
- Access to expertise to undertake additional ground truthing.

These areas were also chosen as they are geographically different, with one area chosen in the North Island (Figure 1) and one in the South (Figure 2).

Additionally, a wetness gradient data set was available in Northland from a 2011 radiometric survey (Rissman et al., 2019). Using the chosen classification methodology this data could be tested for overall effectiveness also.

The study areas exclude the area seaward of the coastal marine area (mean high water spring, MHWS) and open water. For the Tasman region, the extent was limited to the areas covered by existing LiDAR data (which represents approximately 15% of the region) with only RGB (no Infra-red band) vertical aerial photography being available across the entire region. This is considered to be representative of a minimum viable data input situation that would be common across other regions in New Zealand.

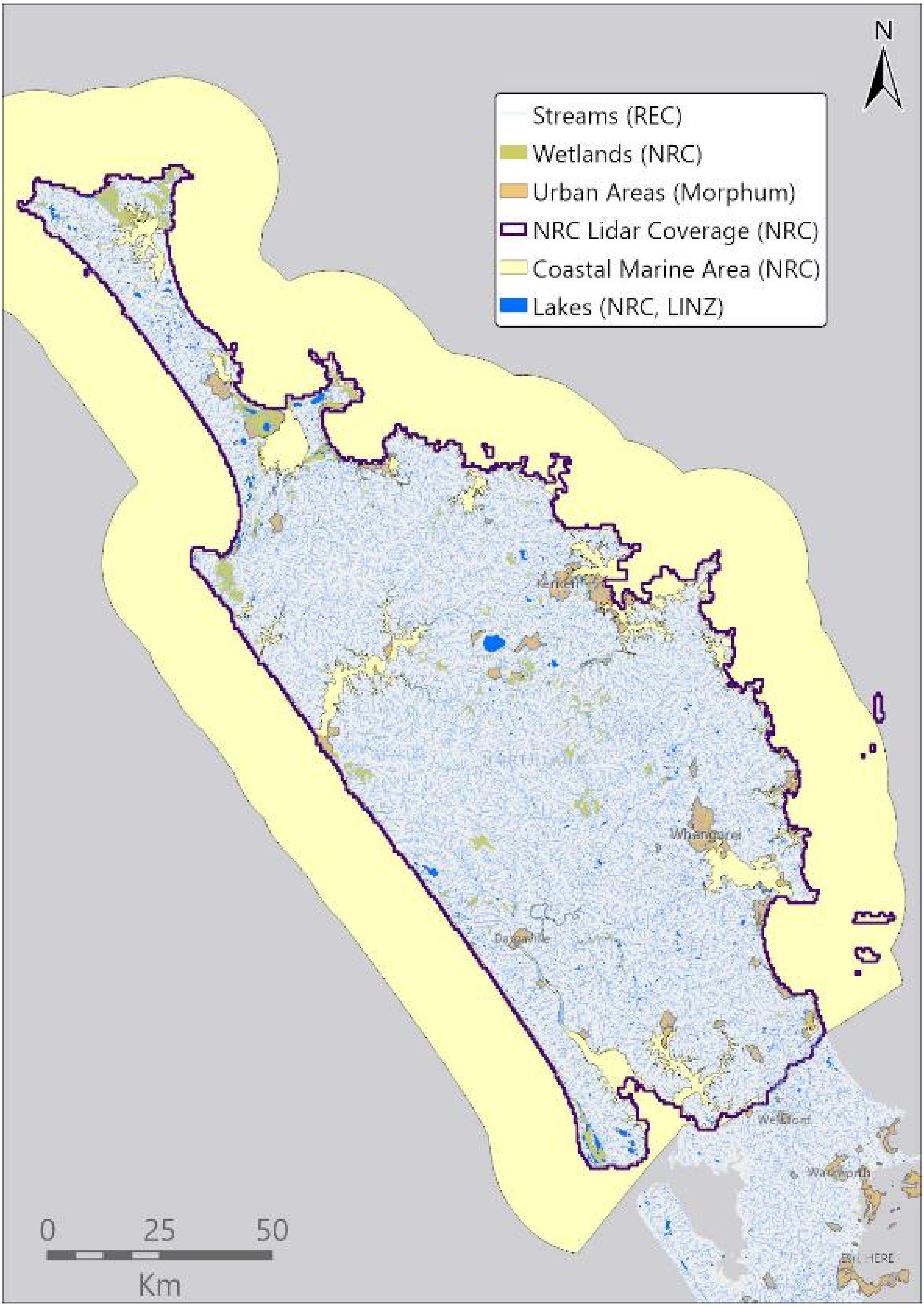


Figure 1: Northland Regional Council (NRC) study area showing the regional council boundary, urban areas, wetlands mapped by NRC using multiple sources of data from 1990 to 2020, and other information that has been used in the delineation of wetlands.

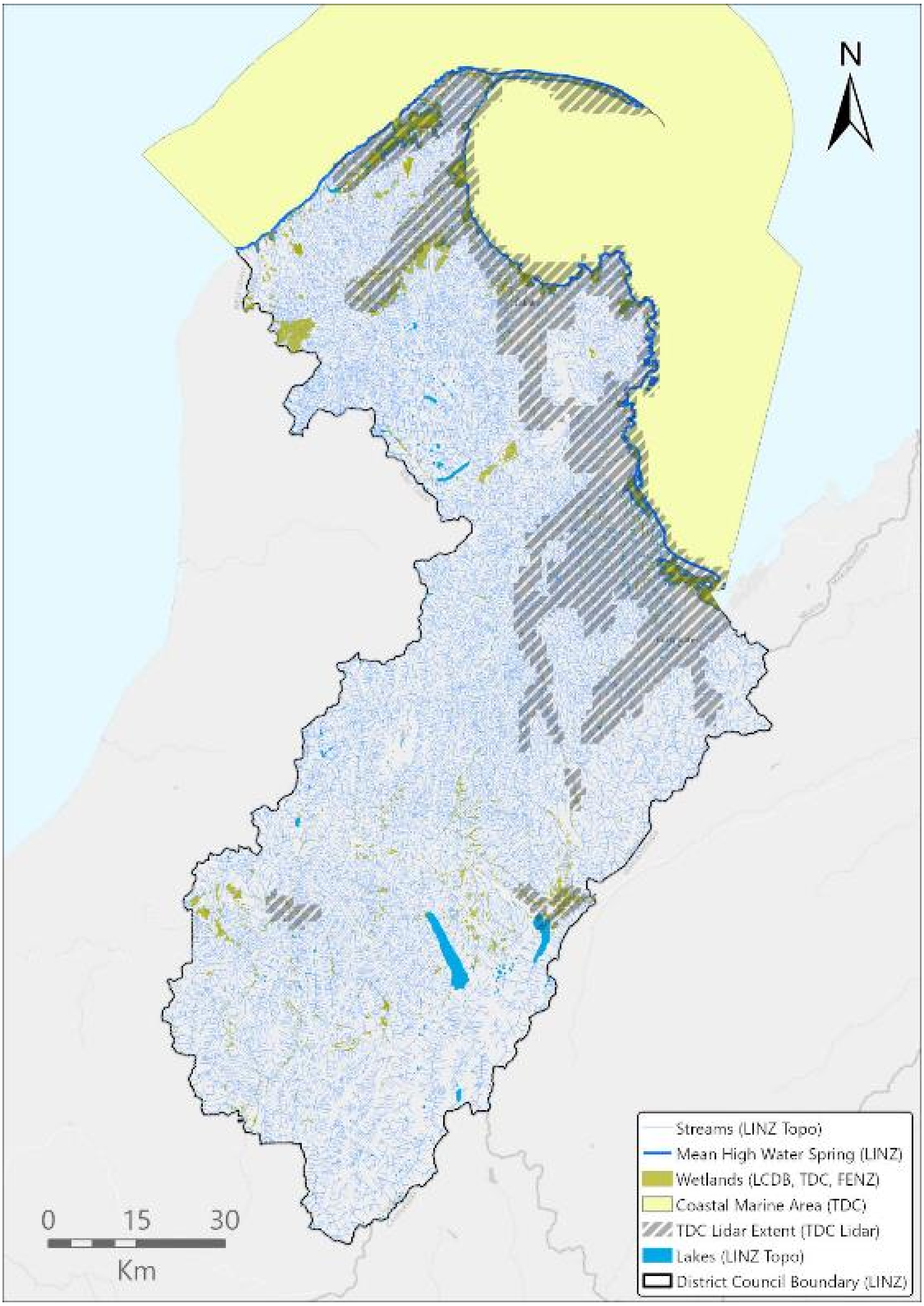


Figure 2: Tasman District Council (TDC) study area showing the district council boundary, urban areas, wetlands mapped using multiple sources of data from 1990 to 2020, LiDAR extent and other information that has been used in the delineation of wetlands. Note the three small southern LiDAR blocks were not processed in the POC.

2.4 Data Inputs

The total set of input data for each region is shown in Table 1. The data inputs considered to be commonly available in New Zealand include RGB or RGBI high resolution vertical aerial photography, LiDAR and its derived indices and data collected from the Copernicus programs by the European Space Agency (ESA) through the Sentinel missions. In Northland, a perennial wetness gradient layer from the 2011 radiometric survey was also able to be assessed. Regional councils have various levels of data availability for existing mapped natural inland wetlands and known artificial wetlands (Lythe et al., 2020); this data will also vary in completeness, accuracy and confidence.

Table 1: Input data for the study areas, Northland and Tasman.

Region	Data Description
Northland (Figure 1)	• RGBI Vertical Aerial Photography 40cm GSD (2019) • LiDAR 1m, (2019) and TWI, TPI, CH (derived from LiDAR) • Wetland geometries (NRC, 2020) • Sentinel-2 multispectral data (2015-2020) • Perennial wetness gradient layer from radiometric survey (NRC, 2011) • Coastal Marine Area (NRC) • Hydrology (NRC) • Stormwater assets (NRC)
Tasman (Figure 2)	• RGB Vertical Aerial Photography 30cm GSD (2018-19) • LiDAR 1m, (2015-19), TWI, TPI, CH (derived from LiDAR) • Wetland geometries (TDC, 2020) • Sentinel-2 multispectral data (2015-2020) • Coastal Marine Area (TDC) • Hydrology (TDC) • Stormwater assets (TDC)

2.5 Classification Scheme

Following a thorough national and international literature review covering the various means to classify wetland types (Amani et al., 2017; Brinson and Malvarez, 2002; Brooks et al., 2009; Grenier et al., 2007), the New Zealand classification method by Johnson and Gerbeaux (2004) was adopted for the POC. This is a semi-hierarchical classification system covering hydrosystem, wetland class, structural class and vegetation composition (Figure 3). The tiers of classification allow for wetlands to be recognised and described at different levels of detail, depending on what applications are intended.

Only the broader top tiers of the classification system (i.e. just hydrosystem and wetland class) are applied in the POC. The higher levels in the hierarchy are most applicable to broad-scale inventory, survey, or mapping, to sort wetlands into meaningful groupings for data storage, retrieval, and interpretation (Johnson and Gerbeaux, 2004). This method is also already adopted by some regional councils and it supports the possible aggregation of wetland classes, if required, to improve model output accuracy and confidence. Johnson and Gerbeaux (2004) note the four most important hydrosystems as estuarine, riverine, lacustrine, and palustrine. These hydrosystems align with those considered in this project (Lythe et al., 2020).

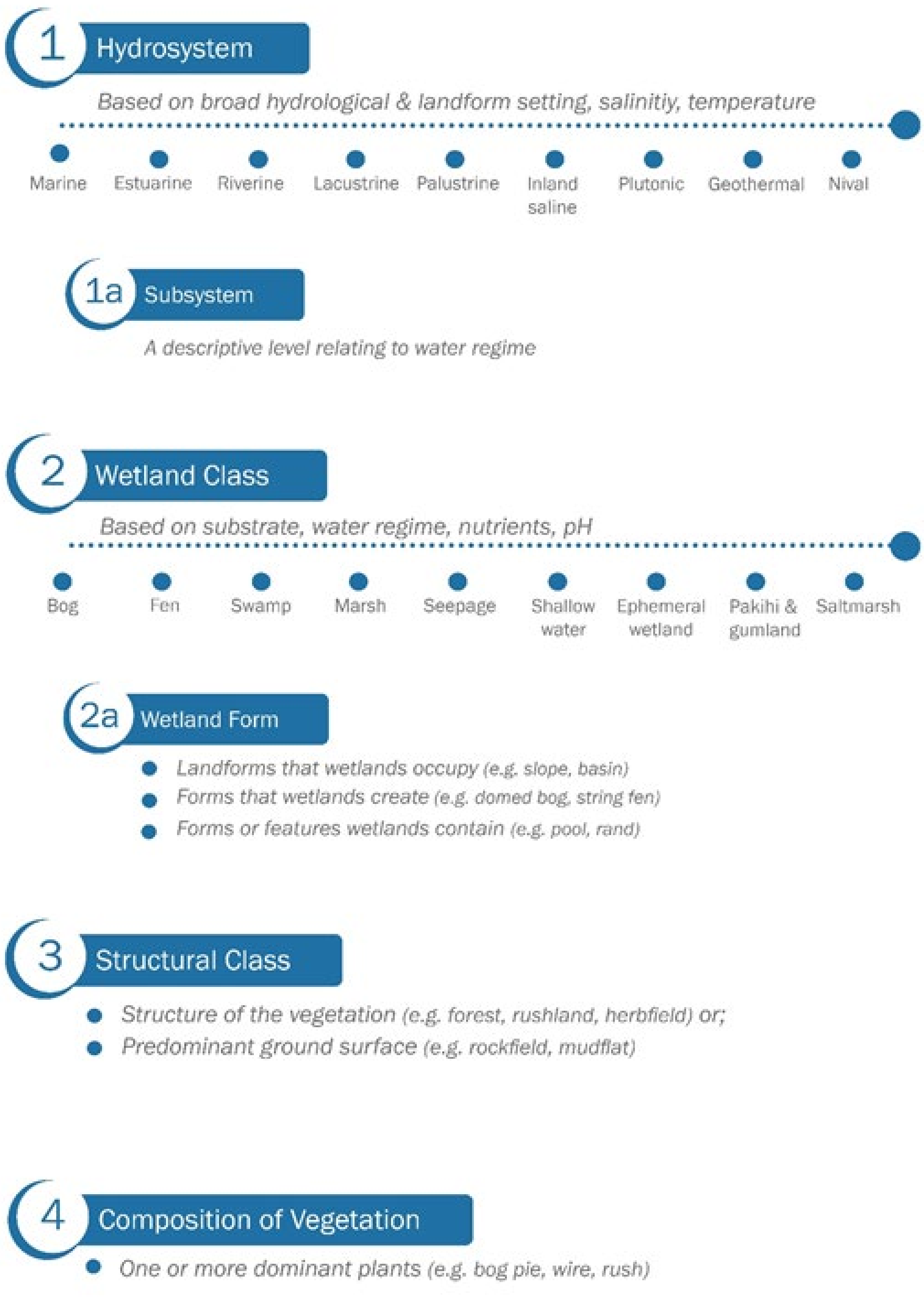


Figure 3: Semi-hierarchical classification system for New Zealand Wetlands (Adapted from Johnson and Gerbeaux, 2004).

3. Methodology

A three-phase approach was employed (Figure 4) consisting of:

1. Masking – medium resolution probability model of wetland probability,
2. Classification – high resolution classification of imagery into wetland class using semantic segmentation. Several data inputs were tested for effectiveness in multi-class wetland classification,
3. Post processing – including vectorisation, hydrosystem classification and validation of wetland class.

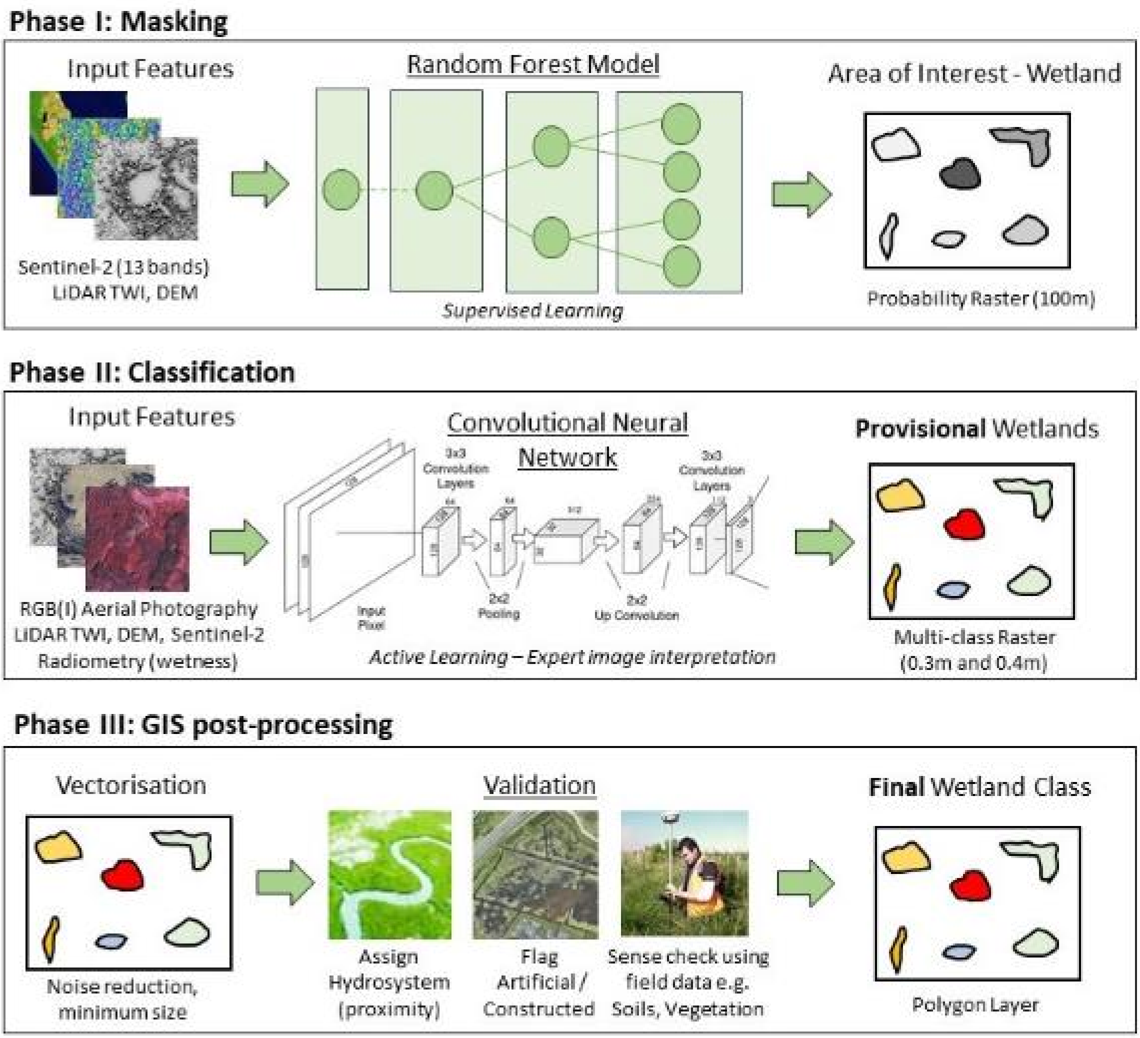


Figure 4: Solution Design

The results from each phase were used as input into the next phase with the resulting wetland polygons and attributes including hydrosystem assigned progressively. Each phase of this process is explained in this section.

3.1 Masking

The Masking phase is mainly a big data management procedure intended to reduce the target study area to a sub-region with a high probability of containing a wetland and therefore minimising the data set size and computational requirements.

A decision tree model was developed using a Random Forest algorithm. Random Forest is a supervised learning algorithm where the "forest" it builds, is an ensemble of decision trees. It is a bagging or aggregating technique with the trees in Random Forests run in parallel. There is no interaction between these trees while building the trees. It operates by constructing a multitude of decision trees at training time and outputting the mean prediction (regression) of the individual trees.

The Random Forest model was trained using a randomly selected subset of polygons from the FENZ and LCDB5 wetland datasets and considered the spectral properties of Sentinel-2 multi-band imagery and the LiDAR derivatives; Topographic Wetness Index (TWI) and Digital Elevation Model (DEM). The spatial resolution of processing output raster was 100 m^2 enabling the resulting data to work seamlessly with the subsequent classification model which is using 0.40 m input RGB(I) aerial photography. A low precision-high recall approach was taken to threshold the results. The outcome of this stage is a binary raster image where non wetland areas are masked for input into stage 2 (Figure 5).

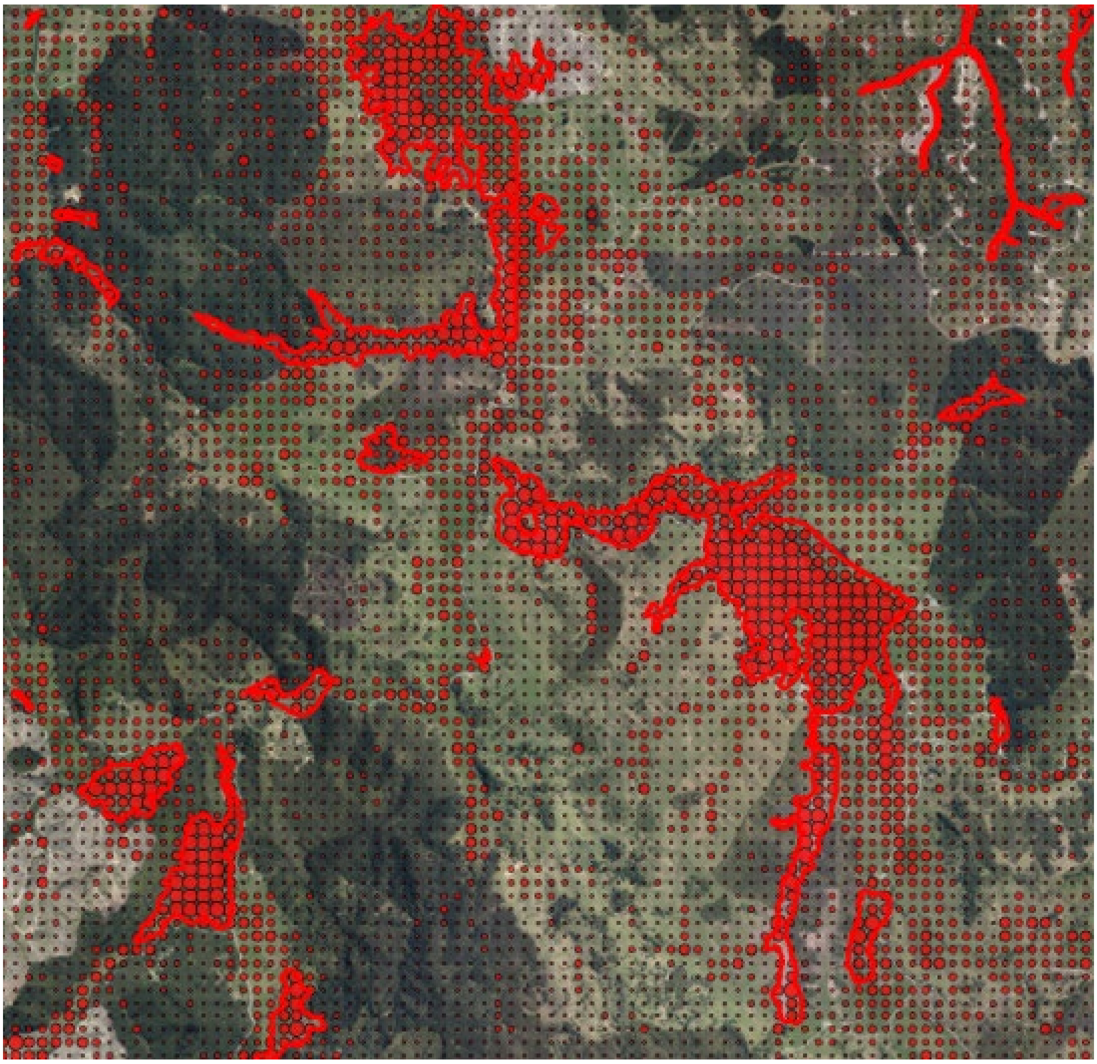


Figure 5: Random Forest Mask result in Northland area shown against NRC wetland polygons. Larger disc = high probability and smaller disc = low probability of this pixel containing a wetland.

The following table presents several additional statistical parameters including:

- True Positive (TP) = the model correctly predicts the positive class
- False Positive (FP) = the model incorrectly predicts the positive class
- True Negative (TN) = the model correctly picks the negative class
- False Negative (FN) = the model incorrectly predicts the negative class
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1 = 2 X ((Precision X Recall) / (Precision + Recall))

> Precision shows how precise the model is out of those predicted to be positive.
>
> Recall calculates how many of the actual positives the model capture through labelling it as True Positive.
>
> The F1 score conveys the balance between the precision and the recall. An F1 score reaches its best value at 1 and worst value at 0. A low F1 score is an indication of both poor precision and poor recall.

The impact of Precision and Recall is described visually in Figure 6.

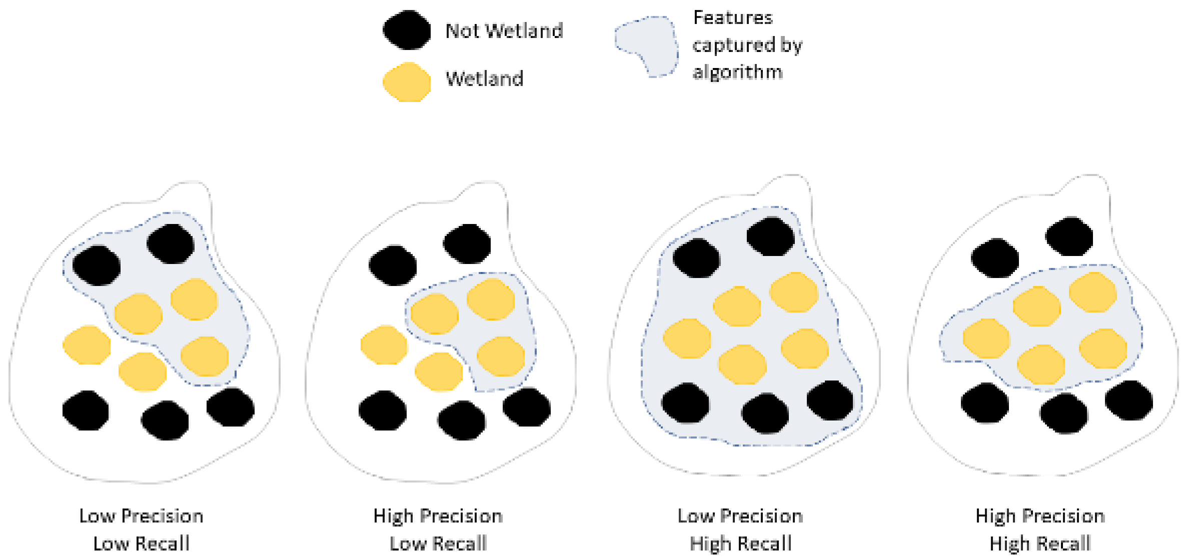


Figure 6: Illustration of Recall and Precision for wetland detection.

The Random Forest algorithm was validated against known wetland polygon data provided by NRC and TDC. In Northland, the accuracy and the overall weighted average accuracy of the mask was 83% while in Tasman the score was 92%. There is a higher-class imbalance in Tasman. Overall, these results are quite promising indicating these masks will be useful for filtering the CNN results.

One of the advantages of the Random Forest algorithm is that prediction is based on input features considered important for classification. This allows information gain to be understood at a feature level. Figure 7 shows the information gain for this model.

Table 2 Accuracy of Northland Random Forest Mask

	Precision	Recall	F1	Samples
Not wetland	0.83	0.88	0.86	5606
Wetland	0.82	0.77	0.80	4218
Accuracy			**0.83**	**9824**
Weighted Average	**0.83**	**0.83**	**0.83**	**9824**

Table 3 Accuracy of Tasman Random Forest Mask

	Precision	Recall	F1	Samples
Not wetland	0.87	0.73	0.79	285
Wetland	0.93	0.97	0.95	1005
Accuracy			**0.92**	**1290**
Weighted Average	**0.91**	**0.92**	**0.91**	**1290**

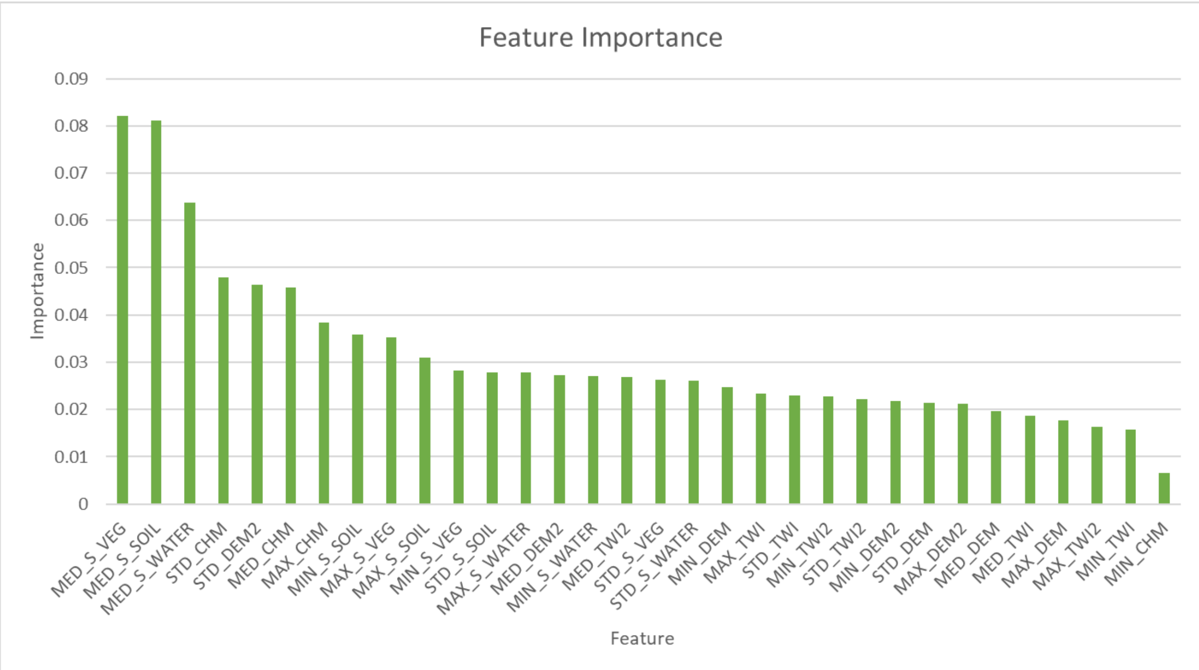


Figure 7: Information gain from Random Forest model.

It reveals that the three most important inputs to the model are composite indices derived from Sentinel-2 (S); vegetation index, soil index and water index. Canopy Height (CHM) and Digital Elevation Model (DEM) derived from the LiDAR survey were fourth and fifth. Med = Median, STD = Standard Deviation, DEM = 100 pixel radius, DEM2 = 40 pixel radius.

The outcome of this stage is a binary raster image where non wetland areas are masked (excluded) for input into stage 2 (Figure 5).

3.2 Classification

Wetland delineation and classification was undertaken primarily using semantic segmentation via a Convolutional Neural Network (CNN). CNNs are a class of neural networks in deep learning that are commonly applied to computer vision and image analysis. It resembles the neuron connectivity pattern in human brains. Specifically, a CNN is made up of one input layer, multiple hidden layers, and an output layer. The hidden layers structurally include convolutional layers, ReLU (activation function) layers, pooling layers, fully connected layers, and normalization layers. Compared to other classification algorithms, CNN requires much less pre-processing and can achieve better results as the number of training cycles increase.

Four separate CNN model scenarios were tested. A fifth Random Forest classification model was also assessed (Table 3).

Table 3: Classification model inputs

Model 1 - CNN	Model 2 - CNN	Model 3 - CNN	Model 4 - CNN	Model 5 - RF
RGBI Aerial Photography LiDAR Sentinel-2 Multispectral data	RGBI Aerial Photography LiDAR Sentinel-2 Multispectral data Radiometry (wetness)	RGB Aerial Photography	RGB Aerial Photography LiDAR Sentinel-2 Multispectral data	LiDAR Sentinel-2 Multispectral data

Active Learning was used to train the models – a methodology used to achieve high accuracy models using only the most essential training inputs (Figure 8). Model training was undertaken using approximately **94,000** points encompassing both regions and including all wetland classes. These were randomly sampled using a subset of the most confident datasets (based on information provided by Council personnel) available in Northland and Tasman containing **15,853** Wetland polygons.

The max-entropy sampling method is used to ensure only the most informative samples are reviewed and labelled by a human expert, leading to savings in human effort and processing time. The confidence in delineation and classification of the available existing mapped wetland areas was discussed with NRC and TDC. For TDC, a list of wetlands that were known to have been surveyed in the field was provided, and this was used to inform the subset of mapped wetlands with high confidence. For NRC, it was advised that wetlands from a variety of datasets have a high level of confidence, this included separate datasets for 'heathlands and 'saltmarsh', as well as, wetlands from an amalgamated 'known wetlands' dataset with confidence attribute scores of '2'; particularly the 'Top150' wetlands.

The CNN also needed to learn what is not a wetland and so **52,056** points were sampled outside of wetlands for the CNN to train on. The neural network models then propose labels on the areas that have the largest entropy (least confidently labelled or most uncertain).

In the training phase, these are iteratively reviewed and corrected by suitably experienced wetland experts and then added to the pool of labelled data for model retraining.

This continues until all high entropy areas are exhausted, which indicates that the model is now suitable for running against future, unlabelled data.

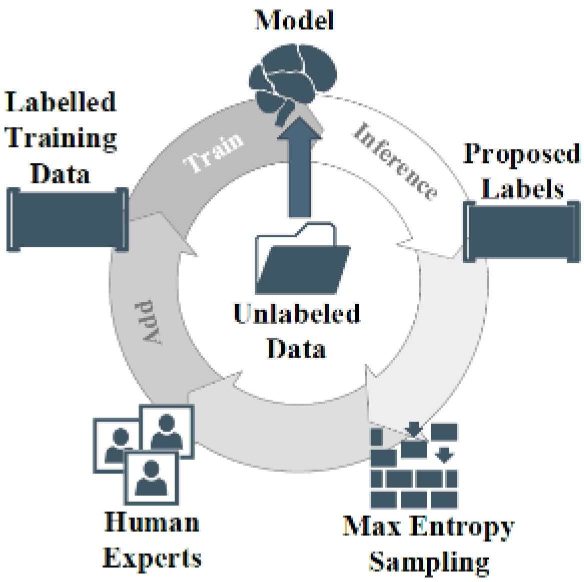


Figure 8: Active learning workflow

In total **12,963** new human input training annotations across all wetland classes were captured based on the optical orthophotography, slope surface and topographic wetness index (TWI). This work was conducted iteratively in between model inference cycles.

Table 4: Point annotation training data summary. NRC and TDC inputs were generated from existing Council wetland polygons. New were generated in the POC.

	Point Annotation			
Wetland Class	**NRC**	**TDC**	**NRC New**	**TDC New**
Bog	3,380	39	315	111
Ephemeral	48	3	150	57
Fen	541	45	210	15
Marsh	1,455	25	538	174
Pakihi & Gumland	9,662	62	493	50
Saltmarsh	16,489	93	621	262
Seepage	31	2	103	-
Shallow water	3,702	16	695	282
Swamp	6,553	60	1248	569
Total wetland	41,861	345	4,373	1,520
Not wetland	49,772	2,284	6914	156
Total Points	91,633	2,629	11,287	1,676

The remote sensing inputs used for training these models are shown below (Figure 9). These multi-image candidates were iteratively surfaced to the wetland ecology team using the max entropy sampling method. This sampling method also mitigates the effects of sampling bias in the training set or from the reviewers.

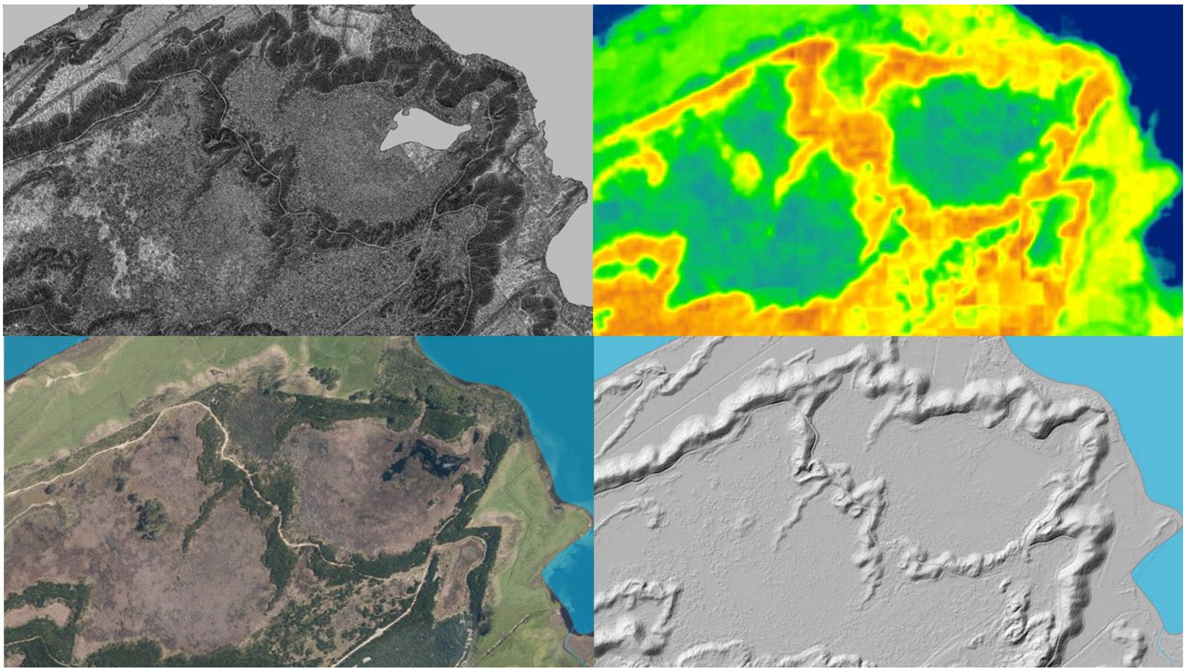


Figure 9: Imagery inputs used for training. Top left TWI, top-right NDWI, bottom left RGB, bottom right Slope.

Multiple semantic segmentation inference cycles were run using Python and Tensorflow/Keras. The result of this stage is a classified image with a spatial resolution of 2m depicting the predicted boundaries and class of wetland. A confidence score is applied to each wetland geometry.

3.3 GIS Post Processing

Following the machine learning phase, a GIS process is used to:

- Eliminate scatter and noise from the inference results,
- Vectorise and dissolve polygons,
- Flag excluded wetlands,
- Apply a decision process to assign hydrosystem classification, and,
- Validate wetland class.

The GIS rules, presented in Appendix 1, outline the rules that were trialled in this POC within the limitations of GIS data available or supplied as part of this project. Each regional and district council throughout NZ will hold GIS data useful for this phase with varying levels of resolution, completeness, and accuracy. Therefore, it is acknowledged that GIS post processing rules will need to be refined for each region based on available data. This applies particularly to the degree of mapped known constructed and artificial wetlands, however, also to the resolution and spatial accuracy of mapped watercourses. However, any regional refinement should always be an improvement from the national level rules. Furthermore, regional councils will be able to refine wetland inventories over time as GIS datasets are created or improved.

3.3.1 Vectorisation

The deep learning models produce multi-class raster images at a 0.3 or 0.4 m super-pixel resolution. These images spanned the entire masked image footprint. A GIS post processing method is used filter out noise (speckle), generalise, vectorise and clip to the target boundary.

A confidence threshold was used to determine a polygon boundary. For the NRC data this was set at a RF probability >35% (100% is higher confidence) and a CNN probability of >80%. This threshold was set to remove noise while retaining a high degree of recall. For the TDC this was set at a RF probability >40% and a CNN probability of >80%. RF and CNN probabilities are included in the model outputs as attributes to allow end users to further filter output polygons and/or prioritise further investigations. It should be noted that the RF and CNN probabilities included in the final model output are the mean probabilities attributed from the rasters.

It is recommended that further refinement of the probability thresholds applied to the raw undissolved model outputs is undertaken to optimise recall and precision. This can be assessed against Goodness of Fit analysis (Hargrove et al., 2006) results to assist in optimising the thresholds used. This process could involve iteratively filtering out raw machine learning output polygons based on each permutation of RF and CNN confidence scores in 1% increments. The resulting remaining polygons then dissolved and the Goodness of Fit (GOF) (Hargrove et al., 2006) tested against a set of held-out high confidence validation known wetland polygons to identify the combination of confidence score filters that returned the greatest GOF. It is likely that the optimal probability thresholds will differ between regions.

Once the raster is converted to a polygon feature and the confidence thresholding applied, an eliminate polygon process is used to eliminate small polygons. This was necessary to filter noise and some false positive results. Seepage, shallow water and ephemeral wetland outputs smaller than 100 m^2, and other wetland outputs smaller than 300 m^2 were removed. In the POC this process was performed on the raw model outputs prior to dissolving. This may have inadvertently removed, or fragmented, larger wetland areas comprised of multiple wetlands classes. It is recommended that future filtering on size to remove noise is undertaken on the dissolved feature dataset.

The polygons were then dissolved, merging them with the polygon from the surrounding features that it shares the longest boundary with. This merge is completed regardless of wetland classification with the final polygons including attributes reporting the wetland class percentage composition within each dissolved polygon. The dominate wetland class is also provided.

3.3.2 Excluded Wetlands

The NPS-FM definition of natural wetlands includes a series of exclusions. These exclusions include wetlands constructed by artificial means, geothermal wetlands, and areas of improved pasture[1]. Furthermore, there is no clear distinction within the RMA definitions between wetlands and lakes and as with between wetland classes, there is a natural gradient between wetland and lake systems. However, lakes can be defined as having a greatest dimension of 500 m (Irwin, 1975). Other general industry accepted thresholds include lakes having a water depth of at least 2 m, and an area of at least 0.5 ha.

There may be similar characteristics between lakes and some wetland classes, or components of wetland classes (i.e. similarities between open water areas associated with shallow water wetlands and swamps). There may also be similar characteristics between artificial and natural wetlands. Therefore, a GIS post

[1] that, at the commencement date, are dominated by (that is more than 50%) of exotic pasture species and is subject to temporary rain derived water pooling

processing approach was undertaken to flag potential lakes and artificial wetlands. The decision process is outlined in Appendix 1, Table 9.

The accuracy and validity of this flagging process is limited by the availability, completeness and accuracy of GIS input data. For example, no GIS data of known artificial wetlands including constructed stormwater ponds and wetlands was obtained when requested from any of the Northland district councils. Thus, the artificial and lake flagging process is considered an indication only and could be updated over time as regional councils refine wetland inventories to include constructed and artificial systems. There is also opportunity to refine this process where more accurate regional GIS data inputs currently exist or can be produced. Some further opportunities are also outlined in Table 9.

No GIS post processing rules were created for flagging geothermal wetlands as geothermal wetlands are non-existent or rare within the pilot study regions and it is expected that this wetland hydrosystem can be effectively excluded through the model training process.

Similarly, no GIS post processing rules were created for flagging output polygons that are located in areas of potential improved pasture (as defined by the NPS-FM). The decision to not include a decision process to exclude areas of improved pasture is a result of several factors including:

- The lack of industry agreement or guidance on the interpretation of the improved pasture definition at the time of undertaking this pilot project (i.e. agreed list of pasture species or guidance on what level of activity constitutes management of pasture). We understand that MfE is working though such implementation concerns and seeking to provide guidance to the industry.
- The level of evidence that may be required to demonstrate improved pasture may be beyond what can be assessed as part of a desktop assessment and/or model training exercise.
- The subjectivity in applying assessment scales when considering the dominance (more than 50%) of exotic pasture species and the limitation that this assessment is to be undertaken at the time of the commencement date of the NPS-FM.
- The fact that a certain level of exclusion of areas of managed pasture species dominated areas will occur inherently though the model training process.

We anticipate that this level of accuracy can be refined over time as regional councils manage and progressively improve their wetland inventories.

3.3.3 Hydrosystem

In practical terms, hydrosystems are of relevance for grouping wetlands over relatively large areas and on a regional basis (Johnson and Gerbeaux, 2004), with the four most important hydrosystems being estuarine, riverine, lacustrine, and palustrine. These hydrosystems align with those considered in this project (Lythe *et al.* 2020).

Taking result polygons from the vectorisation stage, we applied a sequential decision process to classify the wetland polygon into hydrosystem classes based on:

- Proximity to LCDB5 coastal vegetation types (estuarine)[2]
- Proximity to watercourses (riverine)
- Proximity to lakes and ponds (lacustrine)

[2] Note: Opportunities to improve the estuarine hydrosystem flagging rules using alternative data with a focus on elevation, salinity and or coastal proximity, rather than course LDCB vegetation data are being investigated.

Each target polygon is tagged with a hydrosystem class based on the criteria and decision sequence as outlined in Appendix 1, Table 10. Each of the dissolved model output polygons may be assigned more than one hydrosystem class.

There is considerable overlap between the units within each level of the classification system (including hydrosystem level) and boundaries of hydrosystems cannot be expected to be clearly definable on the ground (Johnson and Gerbeaux, 2004). Therefore, it is concluded to be unrealistic to employ a desktop exercise, including using simple GIS based rules, to apply wetland hydrosystem classification with a high degree of accuracy and confidence. This is further compounded by resolution and accuracy of GIS input data. Thus, the applied hydrosystem classes are considered an indication only and could be updated over time as regional councils refine wetland inventories. Some of the key constraints for each decision are indicated in Table 10. There is also opportunity to refine this process where more accurate regional GIS data inputs currently exist or can be produced. Some further opportunities are also outlined in Table 10.

3.3.4 Class Sense Check

In addition to validating the model wetland class predictions with known field validated wetland data (held out from the model training process), the project sought to utilise GIS input data to further sense check the model wetland class classifications.

Consideration was given to following a "fuzzy logic" weighting exercise similar to (or aligning with) that used by Ausseil *et al.* (2008). However, confidence in the accuracy of classification outputs from the fuzzy logic approach means that comparing wetland class outputs from the two approaches for the purpose of validation and sense checking would have little value. Based on field validation undertaken as part of the project, Ausseil *et al.* (2008) reported an overall agreement of 60%, however, this varied greatly between wetland classes. The report (Ausseil *et al.,* 2008) considered that overall wetland classification probably underestimated the marsh extent, overestimated seepages, and confused bog and fen. The accuracy of the results was likely a result of two factors; similarity and overlap between some wetlands classes in reality, as well as, the resolution of input GIS data (the fundamental soils layer (FSL) for example).

Therefore, rather than using GIS input data though a "fuzzy logic" weighting exercise or a decision rules process to re-assign or change the wetland class outputs from the machine learning process; it was decided to use a series of GIS decision rules to flag wetland classification outputs from the machine learning process that may be inconsistent with the class that may be expected based on GIS information. The decision process is outlined in Appendix 1, Table 11.

As with hydrosystems, there is considerable overlap between wetland classes (Johnson and Gerbeaux, 2004), and the accuracy and validity of this flagging process is limited by the availability, completeness and accuracy of GIS input data. Therefore, the polygons flagged though the sense checking rules decision process is anticipated to provide a tool to support regional council refine wetland inventories over time.

3.3.5 Stepwise procedure

The overall method is summarised in a stepwise manner here:

1. Assemble all input data. Note raster inputs vary by model.
2. Calculate indices – NDWI, TWI, DEM, NDVI, Soil Index.
3. Resample all inputs to a unified sample distance. This varies by model. Assemble raster data into multiband composite.

4. Gather local training annotations per class to calibrate model to region.
5. Train model using libraries in Appendix 3.
6. Run RF inference using model codebase supplied. Hold Mask raster for later use.
7. Run CNN inference using model codebase supplied choosing appropriate model for data available.
8. Remove noise from the CNN wetland-prediction raster using raster processing software specifically targeting outlying pixels and filling small gaps in wetland areas.
9. Convert the raster output to polygons in GIS software using the Raster to Polygon algorithm. All settings were left at default.
10. Apply confidence attributes to polygons using RF confidence and mean probability of "CNN not wetland". Use these parameters for filtering out false positives and optimising recall and precision, if required or desired.
11. Run GIS post processing tools. These are grouped into 6 ArcGIS Pro Models that run individual logical segments to dissolve the original data, classify the outputs, check against existing features and raster, such as slope. These tools do the bulk of the data analysis and preparation for the final output.
12. Apply polygon size filter to further reduce noise, if required or desired.

4. Model Results and Accuracy

The following tables and discussion examine the performance of the machine learning model predictions relative to hand labelled or ground checked validation data points within the area of interest.

The validation data was not used to train the machine learning models but was captured at the same time and using the same process as the training data. The per pixel statistics are calculated on the raw machine learning output and are before the GIS post processing which includes thresholding, vectorisation and simplification.

In addition, Goodness of Fit analysis (Hargrove et al., 2006) was undertaken on the model outputs, to compare the delineation against a subset of high confidence known wetland polygons. The comparison was made using model outputs following vectorisation, filtering on polygon size and probability values, and dissolving.

A random subset of 50 held out polygons (not used in the model training) from the NRC 'Top150' wetland dataset were used to compare spatial concordance against any model output polygon that intersected with the 50 held-out polygons. The Goodness of Fit analysis was undertaken on these datasets irrespective of class. The Goodness of Fit score was 31.5%, indicating a weak to moderate spatial concordance between the two datasets. The overall proportion of the model output subset intersected with the held-out subset was 36%. In contrast the overall proportion of intersect for the held-out subset was 87.5%. This is reflective of the high recall obtained by the probability thresholds set in the POC. It is recommended that further refinement of the probability thresholds applied to the raw undissolved model outputs is undertaken to optimise recall and precision using Goodness of Fit analysis.

The same analysis was undertaken for the TDC model using a subset of 57 held-out polygons from the mapped wetlands that were known to have been field validated. The Goodness of Fit score was 2.2%, indicating an extremely poor spatial concordance between the two datasets. The overall proportion of the model output subset intersected with the held-out subset was 10.8%. In contrast the overall proportion of intersect for the held-out subset was 20.7%. While the overall spatial concordance was poor, this is again reflective of the high recall obtained through the POC process in some locations. As above, it is recommended that further refinement of the probability thresholds applied is undertaken.

Overall, the spatial concordance of the modelled polygons with a small validation set was relatively low. While this is incongruent to the pixel level analysis reported below, it reflects the difficulty of using a rules-based GIS process (vectorisation) to generate realistic polygons from the modelled pixels that approximate mapping by a trained ecologist.

4.1 Model 1 – CNN including RGBI, Sentinel-2, TWI, DEM

Model 1 considers three open access data sources. These are arguably the most relevant and widely available open access imagery composition in the New Zealand setting. Most regional councils have a regular repeat visit orthophotography survey scheduled approximately every 3 to 5 years which increasingly includes the fourth infrared channel. Sentinel-2 data is open access and spans the period 2015 to the present while LiDAR coverage has been steadily increasing over New Zealand in recent years.

LiDAR returns can be interpolated to create high-resolution digital elevation models (DEM) and Topographic Wetness Index (TWI), from which wetland indicators based on flow convergence and near-surface soil moisture can be derived (Lang et al., 2013; Lang and McCarty, 2014; Millard and Richardson, 2013, 2015; O'Neil et al., 2018, 2019).

The statistics for this model are shown in Table 5. The support column refers to the number of samples of each class included in the validation process. Binary (wetland/not-wetland) statistics are followed by Multi-class statistics.

Table 5: Classification statistics for Model 1 - Northland

	Precision	Recall	F1 Score	Samples
Not Wetland	0.85	0.95	0.89	1171
Wetland	0.81	0.70	0.75	877
Accuracy			**0.82**	**2048**
Weighted Avg			**0.83**	**2048**
Bog	0.59	0.63	0.61	27
Fen	0.67	0.12	0.21	16
Marsh	0.50	0.02	0.04	44
Not Wetland	0.85	0.95	0.89	1171
Pakihi Gumland	0.86	0.76	0.81	222
Saltmarsh	0.88	0.88	0.88	336
Seepage	0.00	0.00	0.00	2
Shallow water	0.63	0.53	0.58	58
Swamp	0.80	0.59	0.68	172
Accuracy			**0.84**	**2048**
Weighted Avg	**0.83**	**0.84**	**0.83**	**2048**

In terms of binary classification, this model has an accuracy of 0.82 and a weighted average F1 score of 0.83 which suggests this method will be suitable for delineating wetlands at the scale required. The area shown in Figure 10 shows the model results with previously mapped wetlands (left) and modelled wetlands (right). The modelled data is reliably identifying known wetlands and identifying a wider area of wetland then previously mapped.

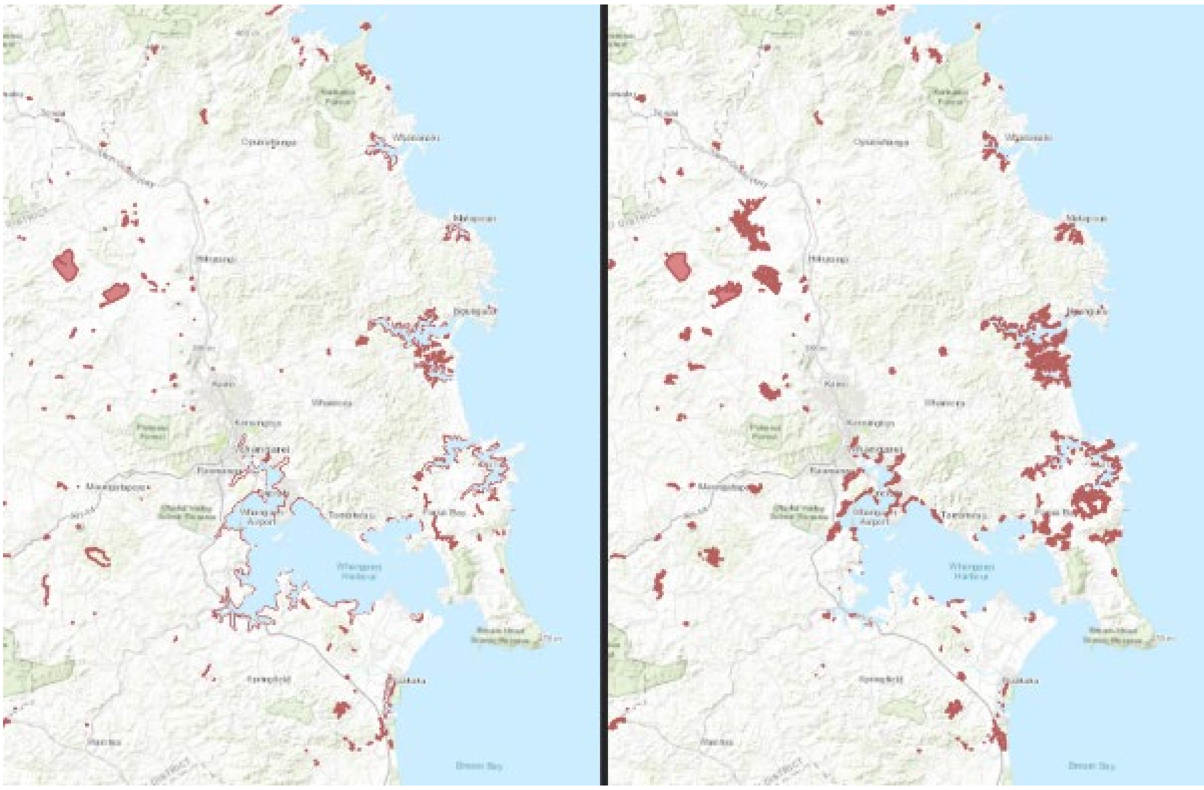


Figure 10: Model results showing extent of previously mapped wetlands (left) and modelled wetlands (right) in Northland.

This model reports an overall multi-class accuracy of 0.84 and a weighted average F1 score of 0.83. Both metrics are reasonable given the complexity of wetland ecosystems.

Looking at the individual classes, it is clear that some of the categories are more easily discriminated than others. Saltmarsh and Pakihi/Gumland have the highest F1 score at 0.88 and 0.81 respectively. Swamp, Bog and Shallow water are next best in the 0.68 through 0.58 range. Several of the classes such as Seepage and Marsh remain too sparse to return a statistically significant F1 score.

Overall, the variability in Precision and Recall is high indicating that there are probably too many classes under consideration by the model. The binary classification results suggest this model is useful for delineation but refinement of class categories may be required to lift individual class scores.

4.2 Model 2 – CNN including RGBI, Sentinel-2, TWI, DEM and Radiometry

Model 2 considers the same data sets as in Model 1 with the addition of the wetness gradient layer from a radiometric survey. This was only available in the Northland study area. It is noted that the radiometry data is sampled at a spatial resolution of 50 m, 125 times the resolution of the aerial photography. This data set was resampled to 0.4 m for modelling purposes.

The statistics for this model are shown in Table 6. The support column refers to the number of samples of each class included in the validation. Binary statistics are followed by Multi-class statistics.

Table 6: Classification statistics for Model 2 - Northland				
	Precision	**Recall**	**F1 Score**	**Samples**
Not Wetland	0.92	0.60	0.73	876
Wetland	0.76	0.96	0.85	11
Accuracy			**0.81**	**2048**
Weighted Avg			**0.80**	**2048**
Bog	0.50	0.05	0.10	19
Ephemeral	0.00	0.00	0.00	3
Fen	0.00	0.00	0.00	20
Marsh	0.44	0.13	0.20	31
Not Wetland	0.78	0.94	0.85	1181
Pakihi Gumland	0.74	0.68	0.71	197
Saltmarsh	0.89	0.79	0.84	348
Seepage	0.00	0.00	0.00	2
Shallow water	0.40	0.46	0.43	41
Swamp	0.74	0.27	0.39	206
Accuracy			**0.78**	**2048**
Weighted Avg	**0.77**	**0.78**	**0.75**	**2048**

In terms of binary (wetland/not wetland) classification, this model has an accuracy of 0.81 which suggests this method will be suitable for delineating wetlands at the scale required.

This model also reports an overall multi-class accuracy of 0.78 and a weighted average F1 score of 0.75. Both of these metrics are lower than the previous CNN model. The only class with a reasonable F1 score is Saltmarsh. Again, several of the classes such as Bog, Ephemeral, Seepage and Fen remain too sparse to return a statistically significant F1 score. It is possible that increasing the number of input layers introduced feature interactions that were harmful or that the capacity of the network to learn from the additional features was exhausted.

A subsequent project might examine the use of radiometric derived features in a Random Forest model where feature importance can be examined and compared to feature correlations.

4.3 Model 3 – CNN including RGB only.

Model 3 considers traditional visible band (RGB) vertical aerial photography only. Assessing this scenario helps build an understanding of the visible band data importance and the added impact of the LiDAR, Sentinel-2 and Infra-red data. The model statistics for this model are shown in Table 7. The support column refers to the number of samples of each class included in the validation. Binary statistics are followed by Multi-class statistics.

Table 7: Classification statistics for Model 3 - Northland

	Precision	Recall	F1 Score	Samples
Not Wetland	0.85	0.85	0.85	630
Wetland	0.68	0.50	0.57	394
Accuracy			**0.72**	**1024**
Weighted Avg			**0.75**	**1024**
Bog	0.00	0.00	0.00	10
Ephemeral	0.00	0.00	0.00	0
Fen	0.00	0.00	0.00	24
Marsh	0.00	0.00	0.00	5
Not Wetland	0.85	0.85	0.85	630
Pakihi Gumland	0.28	0.57	0.38	58
Saltmarsh	0.67	0.90	0.77	172
Seepage	0.00	0.00	0.00	1
Shallow water	0.27	0.06	0.10	64
Swamp	0.21	0.10	0.14	60
Accuracy			**0.72**	**1024**
Weighted Avg	**0.68**	**0.72**	**0.69**	**1024**

In terms of binary (wetland/not wetland) classification, this model has an accuracy of 0.72 and a weighted average F1 score of 0.75 which while lower than Model 1 suggests the RGB data is very useful for boundary definition.

This model reports an overall accuracy of 0.72 and a weighted average F1 score of 0.69. As expected, this is lower than Model 1. Saltmarsh is the only class with a reasonable F1 score. All of the other classes either have too few samples or are of low accuracy. This model does highlight however that RGB imagery is very suitable for masking as the "Not Wetland" class has a high F1 score of 0.85.

4.4 Model 4 – CNN including RGB, Sentinel-2, TWI, DEM

Model 4 is identical to Model 1 without the infrared channel available. This model when run in Tasman returns an accuracy and a weighted average F1 score of 0.79 which suggests this method will be suitable for delineating wetlands at the scale required (Table 8 and Figure 11). This result is slightly lower than Model 1 where the infrared channel was included. Precision rates are quite high for several classes however there is limited accurate validation data available meaning some class validation statistics are zero.

Table 8: Classification statistics for Model 4 - Tasman				
	Precision	**Recall**	**F1 Score**	**Samples**
Not Wetland	0.85	0.95	0.89	1171
Wetland	0.81	0.70	0.75	877
Accuracy			**0.82**	**2048**
Weighted Avg			**0.83**	**2048**
Bog	0.88	0.97	0.93	236
Fen	0.79	0.85	0.81	13
Marsh	0.22	0.05	0.08	40
Not Wetland	0.00	0.00	0.00	18
Pakihi Gumland	0.68	0.75	0.71	141
Saltmarsh	0.92	0.94	0.93	162
Seepage	0.92	0.72	0.81	261
Shallow water	0.00	0.00	0.00	1
Swamp	0.69	0.81	0.74	152
Accuracy			**0.79**	**1024**
Weighted Avg	**0.80**	**0.79**	**0.79**	**1024**

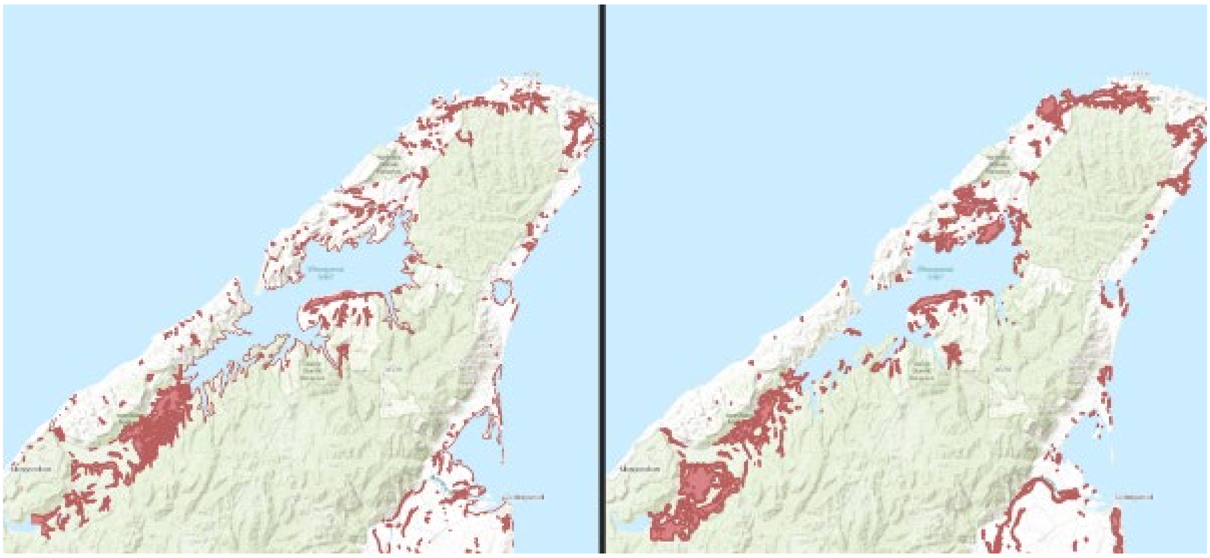


Figure 11: Model results showing extent of previously mapped wetlands (left) and modelled wetlands (right)

This model has a high recall amongst non-forested wetlands in the Tasman District. Overall, the Random Forest model generates higher confidence predictions meaning a slightly higher value was used to threshold the final polygon layer. The minimum polygon size threshold in the raster processing of 300 m^2 however resulted in a loss of smaller seepage and other wetlands.

4.5 Model 5 – Random Forest including Sentinel-2, TWI, DEM

Model 5 presents results from a Random Forest model including all of the inputs from Model 1 excluding the RGBI vertical aerial photography. It enables analysis of a Decision Tree algorithm, as well as the merits of LiDAR and Sentinel-2 data without aerial photography. The statistics for this model are shown in Table 9. The support column refers to the number of samples of each class included in the validation. Binary statistics are followed by Multi-class statistics.

Table 9: Classification statistics for Model 5

	Precision	Recall	F1 Score	Samples
Not Wetland	0.83	0.88	0.86	5606
Wetland	0.82	0.77	0.80	4218
Accuracy			**0.83**	**9824**
Weighted Avg			**0.83**	**9824**
Bog	0.91	0.82	0.86	121
Ephemeral	1.00	0.25	0.40	8
Fen	0.86	0.46	0.60	68
Marsh	0.78	0.26	0.39	179
Not Wetland	0.82	0.90	0.86	5544
Pakihi Gumland	0.82	0.73	0.77	1039
Saltmarsh	0.79	0.81	0.80	1754
Seepage	0.00	0.00	0.00	7
Shallow water	0.94	0.58	0.72	262
Swamp	0.85	0.63	0.72	842
Accuracy			**0.82**	**9824**
Weighted Avg	**0.82**	**0.82**	**0.81**	**9824**

In terms of binary (wetland/not wetland) classification, this model has an accuracy of 0.83 and a weighted average F1 score of 0.83, equivalent to Model 1.

This model reports an overall multi-class accuracy of 0.82 and a weighted average F1 score of 0.81. Both metrics are reasonable given the complexity of wetland ecosystems. The best performing classes are Bog and Saltmarsh, but Swamp and Shallow Water are also performing well. There are also some very high Precision and Recall rates which is promising in terms of this approach. Overall, this model is more consistent across the classes.

4.6 Comparison of Methods

A range of modelling scenarios have been considered across the two study areas. High, medium and low-resolution imagery and LiDAR data have been assessed in combination and in isolation. Semantic segmentation (CNN) and decision tree (Random Forest) modelling approaches have also been tested.

Based on the analysis, it is concluded that RGBI, LiDAR and Sentinel-2 input data are very influential in wetland classification. The CNN model including all of the open access data achieves the highest weighted average F1 score at 0.83. However, the Random Forest model is only marginally inferior with a weighted average F1 score of 0.81. Further, Model 5 (Random Forest) delivers better class specific accuracy and consistency than the CNN models.

The CNN model which includes radiometry data reports a lower overall accuracy than the CNN and Random Forest models. It is likely that increasing the number of input layers introduced feature interactions that were harmful or that the capacity of the network to learn from the additional features was exhausted.

All of the models are biased towards recall meaning there is generally over prediction. This means that areas of wet pasture, ditches and other areas bearing similarities with wetland vegetation are reported (Figure 12). The class predictions are less accurate than the binary prediction. This was expected at the outset and, while required under the NPS-FW provisions, is not as critical compared with the potential efficiencies of identifying and delineating wetlands at a desktop level; providing a platform for wetland inventory refinement overtime, including the attribution of accurate wetland classes.

Figure 12: Example of model including drainage channel in pastureland.

The dual modelling approach generates additional model confidence parameters that can be used by regional councils to filter and prioritise the output polygons. These include the Random Forest prediction confidence and the probability of "Not Wetland" from the CNN models. Applying these filters in isolation or in combination is a useful way to reduce false positives and prioritise important wetlands. It would also be advisable to consider using LiDAR derived Canopy Height and NDVI in conjunction with the polygons to further highlight wetlands with more biodiversity.

Overall, the findings from the POC suggest that wetlands can be delineated using AI applied to earth observation data however class discrimination is more difficult due to the variability and transitional nature of these ecosystems.

Forested wetlands, seepages and ephemeral wetlands are particularly challenging given their temporal variability and remote sensing characteristics. An ensemble model targeting consolidated categories where a Random Forest model is combined with a CNN model may do better than any of the individual models used in this POC. This approach would allow more emphasis to be placed on important features such as landscape form and canopy height rather than vegetation or water response. The CNN models are more difficult to tune towards the dominant inputs which has resulted in more false positive returns than is practically useful.

Optimising the confidence and probability scores will also improve the final polygon selection. The example below (Figure 13) shows the reduction in wetland polygon area on the right as increasingly aggressive thresholding rules are applied. The Random Forest confidence threshold is particularly useful in reducing false positives in many cases such as urban and farming areas.

Figure 13: Swamp in Northland with different polygon threshold rules applied. Left: No threshold. Middle: 16% RF and 79% CNN. Right: 40% RF and 79% CNN

Polygon examples from Model 1 are shown in Appendix 2. The project code, weights file, requirements e.g. python libraries are summarised in Appendix 3.

5. Challenges and Limitations

Several potential challenges were identified in the scoping of the POC, some of which were outlined in the Literature Review Document (Lythe *et al*. 2020) and discussed at workshops with the Ministry and stakeholders. Key outstanding limitations and challenges faced in the POC are discussed below.

Overlap of Hydrosystem Classification Classes

There is considerable natural overlap between the units within each level of the classification system (both for hydrosystem and wetland class) and boundaries of hydrosystems cannot be expected to be clearly definable on the ground (Johnson and Gerbeaux, 2004). Therefore, it is unrealistic to expect that a desktop exercise, including using simple GIS based rules, can apply wetland hydrosystem classification with a high degree of accuracy and confidence. This is further compounded by resolution and accuracy of GIS input data. Thus, the applied hydrosystem classes are considered an indication only and could be updated over time as regional council refine wetland inventories. There is also opportunity to refine this process where more accurate regional GIS data inputs currently exist or can be produced.

Overlap of Wetland Classification Classes

One functional use of wetland classification information is to enable the selection and prioritisation of wetlands for enhancement and protection in a manner that ensures a full range of wetland types and biodiversity values are protected. As with hydrosystems, there is also natural overlap of wetland classes. There is difficulty in distinguishing classes with confidence when observing and training remotely. In other studies, classification accuracies have generally been higher where fewer classes of wetland are used (Mahdianpari et. al., 2020). The models have produced accuracy results that vary by class. Forested wetlands, seepages and ephemeral wetlands report the lowest accuracy. The latter two type are mostly excluded based on size from the final data set. Amalgamation of class results may be required to more clearly discriminate classes using the CNN based approach.

Wetland and Lake Definition

There is no clear distinction within the RMA definitions between wetlands and lakes; and as with between wetland classes, there is a natural gradient between wetland and lake systems.

It is possible that the characteristics for some wetland types, such as shallow open water wetlands or open water components of swamps may resemble lake systems, and as such, the machine learning outputs may include lakes.

Our approach to this has been to acknowledge that lakes will also be included in the model outputs and then rely on a GIS post processing phase to flag potential lakes; with regional councils progressively improving their wetland inventories overtime.

Application of NPS-FM Definitions

The project scope was focussed on a POC method to identify, delineate and classify wetlands as per the NPS-FM definition. At the time of implementing this POC, there was some industry uncertainty regarding the interpretation and implementation of some aspects of these definitions, including the definitions of improved pasture and artificial wetlands, for example:

- Is there a clear distinction between artificial wetlands, and 'induced' wetlands (undefined in the RMA)?
- What constitutes a wetland that is constructed by artificial means? Does that extend to: pugged ephemeral channels; backwatering by culverts; surface discharges of water?
- Is there an agreed list of pasture species?
- What level of activity constitutes management of pasture?

We understand that MfE is working though such implementation concerns and seeking to provide guidance to the industry.

Therefore, there is a potential limitation on the accuracy of the model outputs to align with the NPS-FM definitions resulting from the uncertainty in the initial interpretation of the definitions, as well as the ability of both the model training and machine learning process to be able to distinguish between wetlands that are included and excluded by the NPS-FM definition (Figure 14).

The approach to this has been to:

- Rely on the model training process to contain a level of inherent exclusion of areas that are likely to be considered 'improved pasture',
- Train the model to identify wetlands acknowledging that certain artificial wetlands will also be included in the model outputs (where they have similar characteristics to natural wetlands) and then rely on a GIS post processing phase to flag potential artificial wetlands (dependant of accuracy and completeness of GIS data); with regional councils progressively improving their wetland inventories over time.

Figure 14: Example of model including wet pasture adjacent to a shallow water wetland.

Seasonal Variation

Wetland boundaries fluctuate seasonally and annually. The RGBI data used in the wetland classification POC is summer only, so the classification outputs may carry a seasonal bias. Counter-balancing this is the LiDAR based inputs which should be unbiased for seasonal variation. Multispectral data from the Sentinel satellites will also mitigate this bias.

Scale and Vectorisation

When assessing and determining wetland class in the field, the scale of the subject area, composition of the various matrices of soils and hydrology, as well as, the vegetative community responses, and the consideration of the wider landscape all form part of the consideration.

However, the machine learning algorithms report class at a pixel level base, meaning that the full considerations that would be taken into account in a manual and practical sense by a human are not fully applied. For example, the model may be confused where open water areas of swamps may have characteristics similar to shallow open water or ephemeral wetlands (when saturated); or the margins of a shallow water wetland may appear as a marsh or swamp.

Our approach to mitigate this limitation has been to report on the percentage of wetland class composition within each dissolved polygon. The dominant class is also reported. This reporting may assist regional councils in refining classifications at a later date.

Variable Spatial Resolution of the Machine Learning Model Inputs

Variable spatial resolution of the machine learning model inputs presents a challenge when mapping at sub-hectare scales. The spatial resolution of data inputs varies from 0.40 m (aerial photography) to 1 m (LiDAR) through to 10 m (Sentinel-2) and 50 m (radiometry). The coarser inputs have been resampled to the highest spatial resolution which preserves the information in the aerial photography. However, this also means that any variability across the 50 m radiometric pixel are not accurately represented.

Lack of Required GIS Input Data

The flagging of artificial wetland relies on GIS input data to be able to intersect the model outputs with known artificial wetlands. Very little data relating to known artificial wetlands including stormwater and/or wastewater wetlands or ponds and private constructed ponds and wetlands (for stock drinking, aesthetic reasons, or other) was sourced for the trial regions. With the exception of wetlands that were identified as artificial within one of the existing wetland datasets for the Tasman region, no data of known artificial wetlands was obtained. It is noted that increasingly well designed constructed wetlands will appear as natural systems from aerial photography alone with diverse vegetation and bathymetry making them hard to distinguish remotely. Risks with mapping and classifying constructed wetlands as 'natural' should be understood to avoid instances where required maintenance activities require consenting.

In order for the GIS post processing phase to accurately flag potential artificial wetlands with high confidence, Councils will need to develop and compile records of artificial wetlands and ponds. This exercise could include digitising stormwater and wastewater wetlands and ponds, compiling consented structures and compiling information from Farm Environment Plans. Such an exercise has other benefits and uses including for asset management, contaminant accounting and compliance.

Positional Accuracy, Completeness, and Quality of GIS Input Data

Assigning wetland hydrosystem and applying the GIS post processing sense check rules is limited by the positional accuracy, completeness and quality of GIS input data. For example, when making the assumption the wetland outputs that are in close proximity to larger continuous flowing stream and river channels are riverine wetlands, there is an assumption that the GIS geometry of the river channel is correct.

Some regional and district councils may have differing levels of accuracy, completeness and quality of GIS input data and there is an opportunity to improve this input data over time.

Confidence of Existing Wetland Datasets

The CNN model was initially trained using existing data of known wetlands. For Tasman, only data that was flagged as having been field validated was used. For Northland, data that was confirmed as having high confidence in delineation and class by NRC staff was used. Notwithstanding that, it is still possible that errors in delineation and classification exist in these datasets.

However, the availability of high confidence existing delineated and classified wetland data is a potential limitation in the initial model training; as well as, the validation of the model outputs (through held out data). The validity of the accuracy stats reported relies upon high confidence data to validate against.

Confidence in Remote Training

Classification of wetland class though human annotations was not always possible with a high degree of confidence using the rural specification vertical aerial photography (0.40 m). This included the assigning of class, the extent of the class footprint and the consideration of wetlands with matrices of wetland classes. Where a high degree of confidence was not possible, annotations were not placed; this included potentially small seepage areas or wetlands beneath terrestrial vegetation cover.

The low and medium resolution inputs also introduce noise into the final predictions as these spectral responses represent a much wider surface area than the high-resolution inputs.

6. Deliverables

The deliverables from this POC include the following:

A. An Esri File Geodatabase containing:
   i. Raster data file generated by Random Forest model used to mask each region. MASK_NRC, MASK_TDC. Spatial resolution 100m for NRC and 30m for TDC supplied in NZGD 2000 coordinate system.
   ii. A polygon feature class of raw wetland polygons for each region as generated through the machine learning phase. WETLANDS__RAW_NRC, WETLANDS_RAW_TDC.
   iii. A polygon feature class of wetland polygons following filtering, filtering, smoothing and attribute generation for each region as generated through the machine learning phase. WETLANDS_DISS_NRC, WETLANDS_DISS_TDC (Appendix 5).
   iv. ArcGIS Pro package which includes all ancillary layers which were used to create the final outputs and model builder tools.

B. An Esri File Geodatabase containing
   i. Geospatial data created in the POC, including wetland layer from dissolving FENZ and LCDB5 polygons; and NRC CMA boundary created through joining various model outputs supplied by NRC.

C. Python code including machine learning model and model parameters for:
   i. Random Forest Mask model.
   ii. CNN model variants

D. Final report (this document)

7. Key Findings

In this section the principal findings from the POC are summarised.

a. The Random Forest algorithm is an effective technique to identify an area with a high probability of containing a wetland. The Northland overall model accuracy as measured by the F1 score was **0.83** while the Tasman F1 score was **0.91**.

b. Using a CNN algorithm trained on open access and free data (Model 1) specifically; Sentinel-2 multispectral imagery, RGBI aerial photography and LiDAR, wetlands can be discriminated from non-wetlands with a weighted average F1 score of **0.83**.

c. Adding the Radiometry derived wetness gradient layer to the CNN (Model 2) didn't improve the classification result. This would indicate that either the sensor resolution is too coarse or that the wetness signal is already incorporated within the multispectral Sentinel-2 data.

d. A CNN model using RGB vertical aerial photography alone (Model 3) delivers an inferior result (weighted average F1-score of **0.75**) than when combined with the other open access data sets.

e. A CNN model trained on open access and free data (Model 4) specifically; Sentinel-2 multispectral imagery, RGB aerial photography and LiDAR, wetlands can be discriminated from non-wetlands with a weighted average F1 score of **0.79**.

f. A Random Forest multi-class model (Model 5), using the same inputs as Model 1, delivers a very similar result with a weighted average F1 score of **0.81** but more consistent and individually better class accuracies.

g. The multi-class CNN model has variable accuracy across the target wetland classes. Class decisions should be reviewed by local experts and these models used primarily for delineation.

h. The Random Forest model confidence is a useful tool to filter and reduce the polygon output. When used in combination with the CNN model confidence it enables the user to adjust the ratio of Recall and Precision.

i. Further modelling is recommended to move this work from POC to production use including testing of an ensemble model combining Random Forest with CNN as well as trialling a reduced class set.

j. The Random Forest and CNN methods can be consistently applied at a regional or national scale. It is recommended that local training annotations be added to localise and refine the base models for individual regional council use.

k. The machine learning libraries used in the modelling are open source and multiple widely available GIS software packages will be capable of applying the post processing workflow including QGIS, ArcGIS, MapInfo etc.

l. Attributing wetland class percentage composition for dissolved polygons containing multiple classes is a straightforward and efficient way of preserving the heterogeneity of complex wetland systems. It further allows for class aggregation to occur within a GIS process downstream.

m. The GIS post processing phase is a deterministic method whereby external GIS data are assessed to validate the results from the stochastic CNN model. Hydrosystem is determined in this way, as is whether the wetland may be artificial. Additional data such as impervious surface, oblique photography, soils data and vegetation can be used to sense check the CNN class result. In this POC, the GIS data is of variable quality and so this area represents a possible area of improvement and focus by regional councils. There is an opportunity to more fully test the GIS post processing rules in a Region with a richer GIS dataset.

n. Use of land use, impervious surface and other GIS data prior to vectorisation is also recommended. This will minimise false positives and ensure wetland polygons avoid intersecting the built environment.

8. Recommendations

In this section recommendations are made to move this work from POC to production use. This considers findings from the wide-ranging sensor and algorithmic scenario tests that have been developed in this POC. The chief recommendations are:

8.1 Key Recommendations

1. An ensemble model that integrates a Random Forest algorithm combined with CNN algorithm is the recommended modelling approach going forward. This will allow better fine-tuning of the image inputs to the terrains and vegetation assemblages.

2. Data inputs for wetland delineation should include as many of these data types as are available: Vertical rural aerial photography (RGBI), LiDAR, Sentinel-2 multispectral imagery.

3. Model accuracy is likely to be greater with fewer output classes. It is recommended that a 5 or 6-class aggregated model be used. Furthermore, it is recommended that a model for pakahi and gumland, if not all forested wetlands, is developed separately to the other classes to ascertain if this will also improve overall accuracy.

4. The Random Forest prediction confidence and CNN probability scores can be used to threshold resulting polygons for inclusion or exclusion. It is recommended that further refinement of the probability thresholds applied to the raw undissolved model outputs is undertaken to optimise recall and precision. This can be assessed against Goodness of Fit analysis results to assist in optimising the thresholds used. However, it is likely that the optimal probability thresholds will differ between regions.

5. GIS Post processing is an important phase post ML (as well as pre and post vectorisation) and relies on high quality inputs including impervious surface data, soils, hydrology, vegetation, stormwater assets. It is recommended to apply and assess the GIS Post processing phase in a region with rich and high confidence GIS data.

6. The incorporation of high-resolution oblique imagery for training and validation is recommended. Oblique images have proven very effective for human wetland identification in some regions as they offer views into the canopy.

7. It is probable that different models containing different weights and parameters will be required for different ecoregions to better approximate the vegetation, landscape and ecosystem characteristics of an area. The recommended modelling approach and inputs shouldn't change. Local training annotations however will need to be added to localise and tailor the POC models for individual regional council use.

8. Filtering on size to remove noise, if applied, should be undertaken on the dissolved feature dataset to avoid the potential for fragmentation and removal of multi class wetlands when applied on the raw machine learning outputs.

8.2 Overview of limitations

The best performing CNN and Random Forest models developed in the POC were capable of discriminating wetlands at a pixel level from non-wetlands with an accuracy of 0.83. Spatial concordance of the modelled polygons with a small validation set of wetland polygons from regional councils was however relatively low. This is due in large part to the model design which generates outputs that are biased towards high recall meaning they are predicting all plausible wetland areas including likely false positives such as wet pasture, drainage channels and areas of vegetation resembling assemblages found in natural wetlands. This over-prediction leads to many more polygons being generated than might be required under NPS-FW but it also ensures all probable wetlands are identified.

The modelling generates two confidence thresholds which can be used to enable selection, refinement, and reduction of the polygon dataset by councils. Due to variations in prediction confidence across wetland class however it is difficult to reliably use a single threshold uniformly. Overall, forested wetlands, seepages and ephemeral wetlands are least reliably predicted due to scarcity of examples and the challenges with observing beneath forest canopy. Furthermore, there are regional variations across New Zealand that will mean different models may be required to suit landscape, hydrological and vegetation characteristics. Several recommendations have been made to address these limitations which are summarised in Section 8.1.

9. References

Amani M., Salehi B., Mahdavi S., Granger J. E., Brisco B., and Hanson A., (2017) Wetland Classification Using Multi-Source and Multi-Temporal Optical Remote Sensing Data in Newfoundland and Labrador, Canada, Canadian Journal of Remote Sensing, 43:4, 360-373. doi: 10.1080/07038992.2017.1346468.

Ausseil A.G.E, Gerbeaux P., W. Chadderton L., Stephens T., Brown D., (2008). Wetland Ecosystems of National Importance for Biodiversity: Criteria, methods, and candidate list of nationally important inland wetlands.

Ausseil, A.G.E, W. Chadderton L., Gerbeaux P. ,Stephens T., Leathwick, J.R. (2011). Applying systematic conservation planning principles to palustrine and inland saline wetlands of New Zealand. Freshwater Biology, 54, 142 – 161.

Brinson, M., & Malvárez, A. (2002). Temperate freshwater wetlands: Types, status, and threats. Environmental Conservation, 29(2), 115-133. doi: 10.1017/S0376892902000085.

Grenier, M., Demers, A.M., Labrecque, S., Benoit, M., Fournier, R.A., Drolet, B., (2007). An object-based method to map wetland using RADARSAT-1 and Landsat ETM images: Test case on two sites in Quebec, Canada. Can. J. Remote Sens. 33. doi: 10.5589/m07-048.

Hargrove, W., Hoffman, F., Hessburg, P. (2006). Mapcurves: a quantitative method for comparing categorical maps. Journal of Geographic Systems, 8, 187-208.

Irwin, J. (1975(. Checklist of New Zealand lakes. New Zealand Oceanographic Institute Memoir 74. New Zealand Oceanographic Institute, Wellington. 161 p.

Johnson, P., Gerbeaux, P., (2004). Wetland Types in New Zealand. Department of Conservation, Wellington, New Zealand, Department of Conservation, Wellington, New Zealand.

Lythe M., Davis C., Lowe M., Farrant S., Chapman Olsen R., Stanley M., Knox D., (2020). Literature Review, Data Discovery and Recommended Approach for Proof of Concept for Wetland Mapping Methods. Prepared for Ministry for the Environment by Morphum Environmental and Lynker Analytics. Final, Version 1. Morphum Project Number: P02262.

Mahdianpari, M., Granger, J., Mohammadimanesh, F., Bahram Salehi, Brisco, B., Homayouni, S.,Gill, E., Huberty, B., Lang, M. Meta-Analysis of Wetland Classification Using Remote Sensing: A Systematic Review of a 40-Year .Trend in North America. Remote Sens. 2020, 12, 1882; doi:10.3390/rs12111882.

Myers S.C. Clarkson B.R. Reeves P.N., Clarkson B.D. (2013) Wetland management in New Zealand: Are current approaches and policies sustaining wetland ecosystems in agricultural landscapes? Ecological Engineering: 56 p107-120. ISSN 0925-8574.

Rissmann C, Pearson L, Lindsay J, and Couldrey M (2019). Mapping of Northland's Wetness Gradients utilising Radiometric and Satellite imagery – GIS Metadata. Land and Water Science Report 2019/38, Project Number: 19015.

Robertson, H.A., (2016). Wetland reserves in New Zealand : the status of protected areas between 1990 and 2013. New Zealand Journal of Ecology 40 (1): 1-11. https://www.jstor.org/stable/10.2307/26198722.

Turpie, K.R., Klemas, V. V., Byrd, K., Kelly, M., Jo, Y.H., (2015). Prospective HyspIRI global observations of tidal wetlands. Remote Sens. Environ. 167, 206–217. doi: 10.1016/j.rse.2015.05.008.

Appendix 1 GIS Rules for Post Processing

The GIS rules presented in the tables below outline the rules that were trialled in this POC, within the limitations of GIS data available or supplied as part of this project.

Each regional and district council throughout NZ will hold GIS data useful for this phase to with varying levels of resolution, completeness, and accuracy. Therefore, it is acknowledged that GIS post processing rules will need to be refined for each region based on available data. This applies particularly to the degree of mapped known constructed and artificial wetlands, but also to the resolution and spatial accuracy of mapped watercourses. However, any regional refinement should always be an improvement from the national level rules and regional councils will be able to refine wetland inventories over time as GIS datasets are created or improved.

Table 10: GIS rules for the flagging of potential lakes and artificial wetlands

Flagged as	Rule	Input Data	Reason	Constraints	Opportunity
Lake	Where wetlands intersect with a lake	FCDB5 LINZ Topo FENZ	Lakes are not wetlands but may contain wetlands at the margin	True lacustrine wetlands may be located around the margins of lakes and could be flagged depending of the location of wetland polygons outputs and input lake data	Development of a process to disaggregate lacustrine wetlands from the edges of lakes
Artificial	Where wetlands intersect with known reservoirs, weirs or dams or are connected solely by reticulated pipe networks	LINZ Topo	Indicate constructed wetland area	Lack of detailed data identifying artificial wetlands provided (stormwater and wastewater wetlands, consented dams, Farm Environment Plans etc)	Opportunity to refine this process where more accurate regional GIS data inputs currently exist or can be produced; alternative council can update wetland inventories over time.
	Where wetland intersects with known artificial wetland	Date supplied by TDC titled 'WetlandExtents_ Dissolved'	Wetlands known to be artificial		
Not a wetland	Where wetland intersects with an area judged not to be a wetland	Date supplied by TDC titled 'WetlandExtents_ Dissolved'	Area is not considered a wetland under Clarkson 2013 and Tasman Regional Management Plan		

Table 11 : GIS rules for assigning indicitive hydrosystem classification

Hydroclass	Rule	Input Data	Reason	Constraints	Opportunity
Riverine	If wetland is: Within 10m of a REC 4th order stream or above	REC	The dominant function in riverine wetlands is continually or intermittently flowing freshwater in open channels. These are typically associated with the larger stream and river systems. Palustrine wetlands may often intersect with geometries of smaller stream geometries.	The REC layer is not spatially accurate (A wetland within 10 m of the REC geometry may not be within 10 m of the river on the ground) but does allow filtering to the higher order streams.	Assign stream order to a more spatially accurate stream layer where the geometry exists in a connected network and/or use reach catchment size where available. Consider flood data if available
Lacustrine	If wetland is: Within 5 m of a lake or pond greater than 0.5 ha in area; Within 5 m of lake with greatest dimension greater than 500 m	LCDB5 LINZ Topo FENZ	These wetlands are associated with waters, bed and immediate margins of lakes. Lakes are arbitrarily defined in Johnson and Gerbeaux (2004) as having a major dimension of 500 m or more.	The lake data may be spatially inaccurate. This could result in false positives and negatives. The threshold value will need to be sense checked to determine error rate.	Explore options for also considering depth of the water body (i.e. wetlands associated with water bodies over 2 m in depth)
Estuarine	If wetland is: Within a polygon defined as mangrove, herbaceous saline vegetation or estuarine open water; Within or, within 10 m of, estuarine open water.	LCDB5 LINZ Topo FENZ	If the wetland has estuarine vegetation, then saline conditions can be assumed.	Relies on there being a degree of spatial accuracy in the LCDB5 classes relating to saline vegetation	Where regional councils have higher accuracy vegetation cover data this could be utilised Explore options for also considering elevation and distance from the CMA
Palustrine	All wetlands not meeting the riverine, lacustrine, or estuarine hydrosystem class rules.				

Table 12: GIS rules for applpying sense check

Flagged as	Rule	Input Data	Reason	Constraints	Opportunity
Potentially not a Bog	Where bog intersect with a stream	LCDB5 LINZ	Bogs are rainfall fed only		Opportunity to refine this process where more accurate regional GIS data inputs currently exist or can be produced (i.e. detailed OLFP or stream datasets especially where upstream catchment areas are known)
	Where bog intersects with a stream of 4th order or above	REC	Bogs are rainfall fed only	The REC layer is not spatially accurate	
	Where bog is on soil that is not peat	FSL	Bogs are only present on peat soils. Soils defined as peat have greater than 50% peat content.	FSL has a coarse scale and based on dominant soil type. S-map, which has better quality data, has limited coverage.	Better quality soil data will become available through S-Map
	Where bog is on ground with average slope below bog footprint of steeper than 4°.	Locally available LiDAR data	Bogs are located on almost level ground. Flat to gently undulating land is define as 0° to 3° by Landcare Research - Our Environment: Steepness of Slope data layer		
	Where bog intersects with a hydroclass that is riverine	Assigned hydroclass	Bogs are rainfall fed only whereas a riverine hydroclass has input from intermittently or flowing water	Relies on accuracy of data used to apply hydrosystem; this may flag an error in hydrosystem as well as wetland class.	
Potentially not a Fen	Where fen intersects with a stream of 4th order or above	REC	Fens are rainfall and groundwater fed only	The REC layer is not spatially accurate	
	Where fen intersect with a stream	LCDB5 LINZ	Fens are rainfall and groundwater fed only		Opportunity to refine this process where more accurate

Flagged as	Rule	Input Data	Reason	Constraints	Opportunity
					regional GIS data inputs currently exist or can be produced (i.e. detailed OLFP or stream datasets especially where upstream catchment areas are known)
	Where fen intersects with a hydroclass that is riverine	Assigned hydroclass	Fens are rainfall and groundwater fed only whereas a riverine hydroclass has input from intermittently or flowing water	Relies on accuracy of data used to apply hydrosystem; this may flag an error in hydrosystem as well as wetland class.	
Potentially not a Swamp	Where swamp is on ground with average slope below swamp footprint of steeper than 4°.	Locally available LiDAR data	Swamps are located mainly on valley floors, plains and deltas. Flat to gently undulating land is define as 0° to 3° by Landcare Research - Our Environment: Steepness of Slope data layer		
Potentially not a saltmarsh	Where saltmarsh intersects with a hydroclass the is not estuarine	Assigned hydroclass	Saltmarsh are in a saline environment so must have an estuarine hydroclass	Relies on accuracy of data used to apply hydrosystem; this may flag an error in hydrosystem as well as wetland class.	
Potentially not a seepage	Where seepage is on land with slope less than 21°	Locally available LiDAR data	Seepages are generally on moderate to steep hills. Moderately steep land is defined as being steeper than 21° by Landcare Research, Our Environment, Steepness of Slope data layer		
Potentially not an ephemeral wetland	Where ephemeral wetland is on peat soil.	FSL	Ephemeral wetlands are only located on mineral soils.	FSL has a coarse scale and based on dominant soil type. S-map, which has better quality data, has limited coverage.	Better quality soil data will become available through S-Map

Flagged as	Rule	Input Data	Reason	Constraints	Opportunity
Impervious Surfaces	Where wetland intersects with known impervious surfaces from: Building outline; Airport. Where wetland is within 1.5m of: Road centreline; Rail centrelines.	LINZ	Wetlands will not be situated on impervious surfaces and are unlikely to be near to manmade structures.	Data availability – regional councils will have various levels of accuracy and completeness of impervious data.	Rather than applying intersection with impervious service data as a flagging sense check stage, the process may be improved by applying an exclusion of this nature at the masking stage (depending on the level of confidence in the datasets used).

Appendix 2 Example Classification Results

Image: 1

Description: Model outputs identifying additional potential wetland

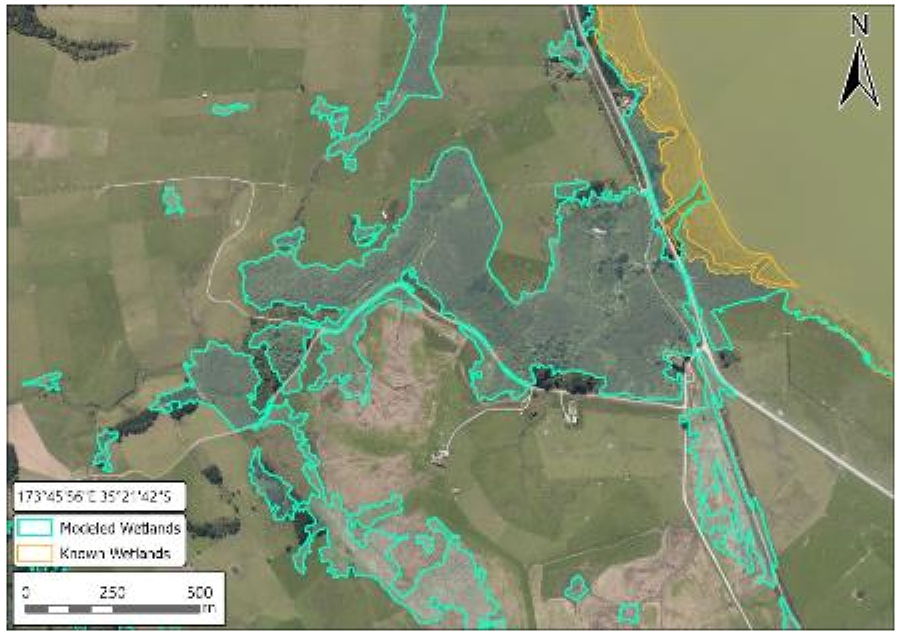


Image: 2

Description: Model outputs identifying additional potential wetland; possible false positive

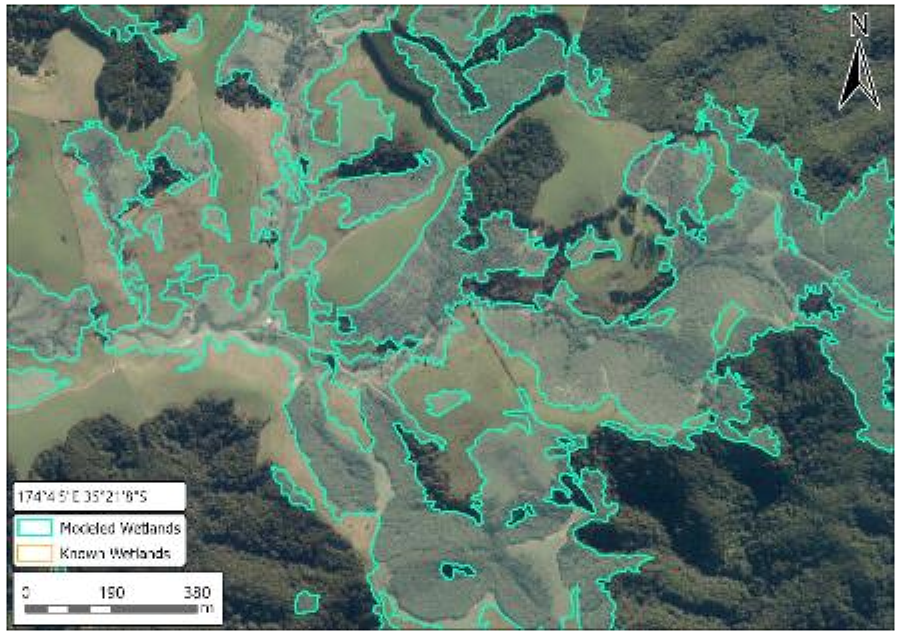


Image: 3

Description: Model outputs identifying additional potential wetland; possible false positive

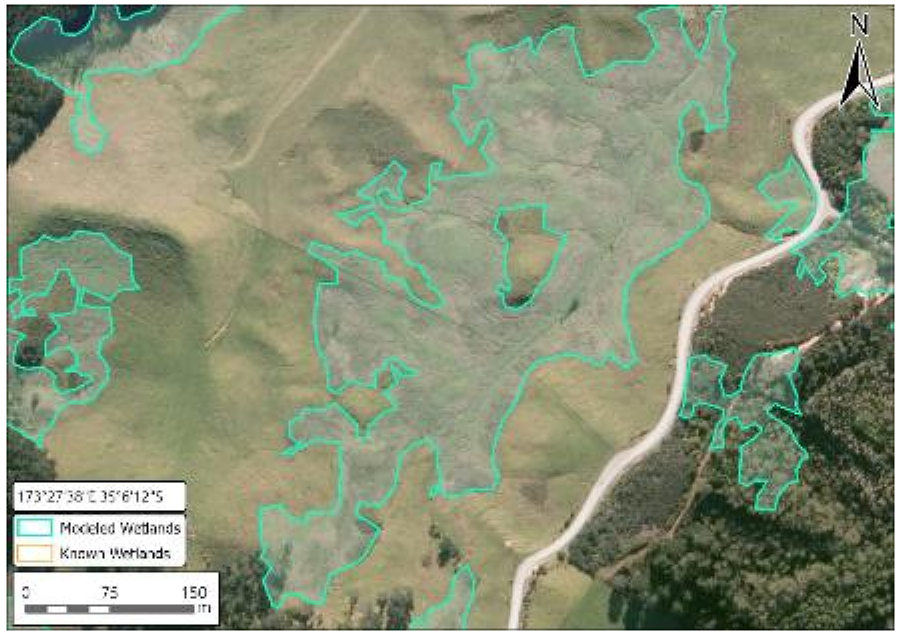


Image: 4

Description: High degree of 'insideness' of known wetland and high recall on model output including likely false positives

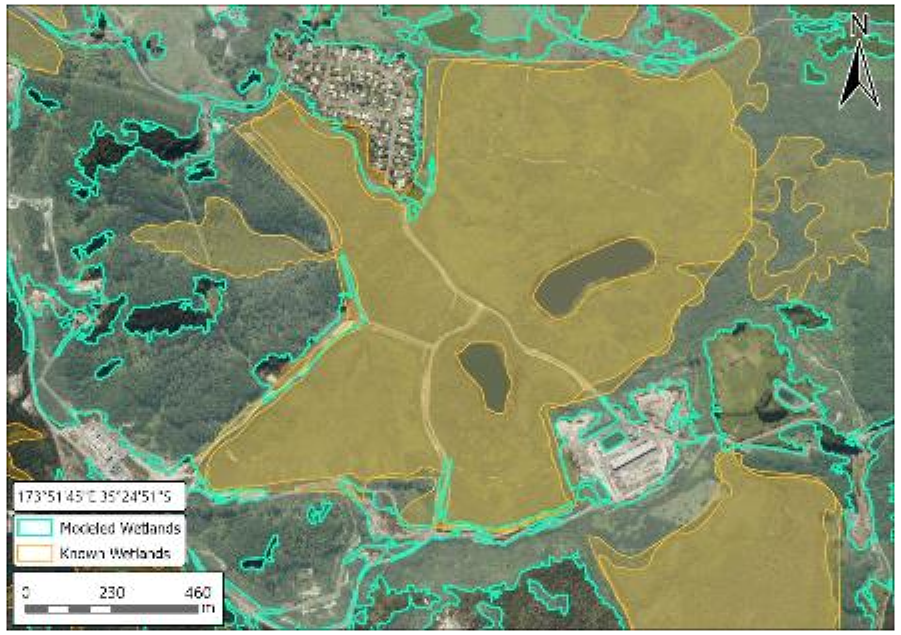


Image: 5

Description: High degree of 'insideness' of known wetland and high recall on model output

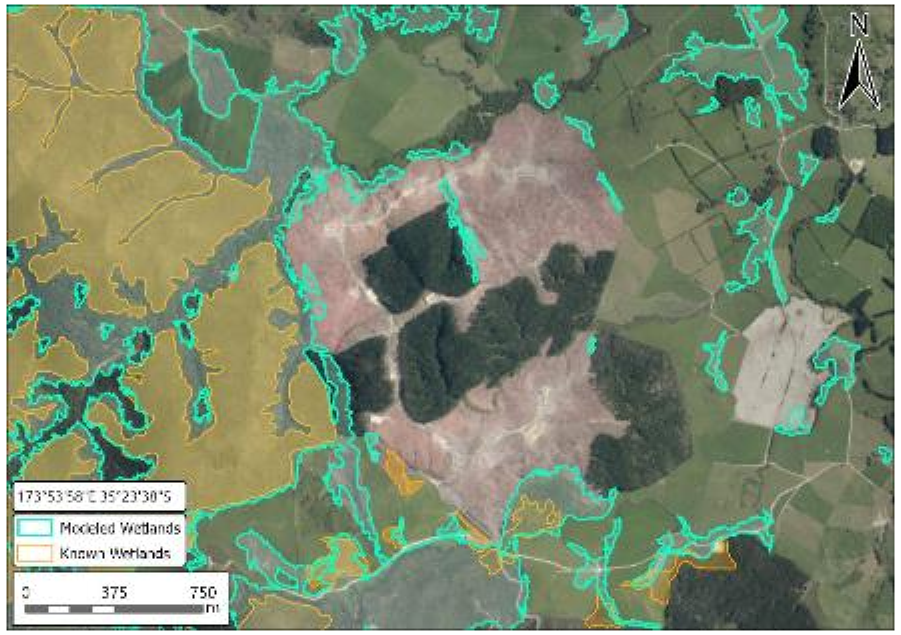


Image: 6

Description: High recall and likely false positive model outputs

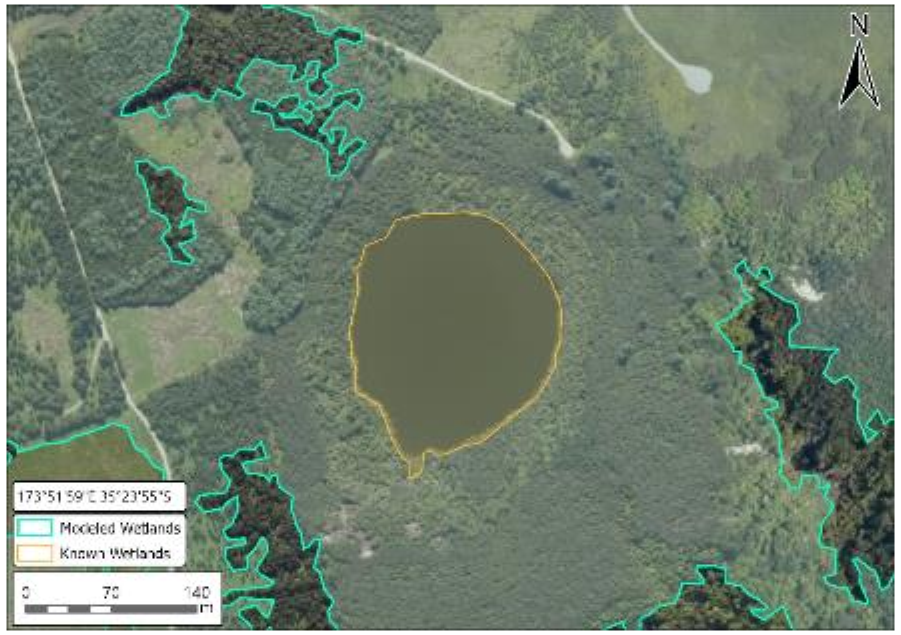


Image: 7

Description: Model outputs identifying additional potential wetland

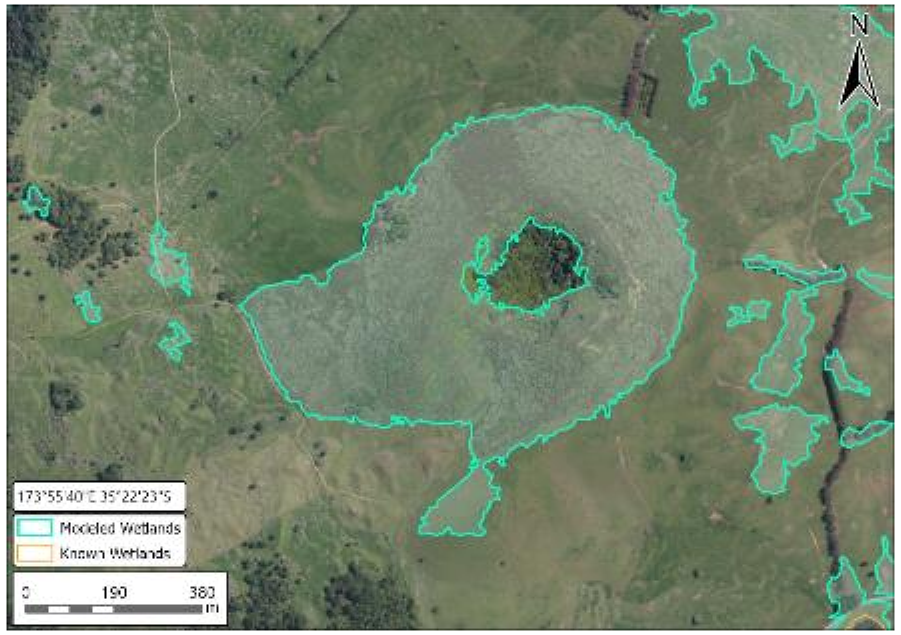


Image: 8

Description: High degree of 'insideness' of known wetland and high recall on model output

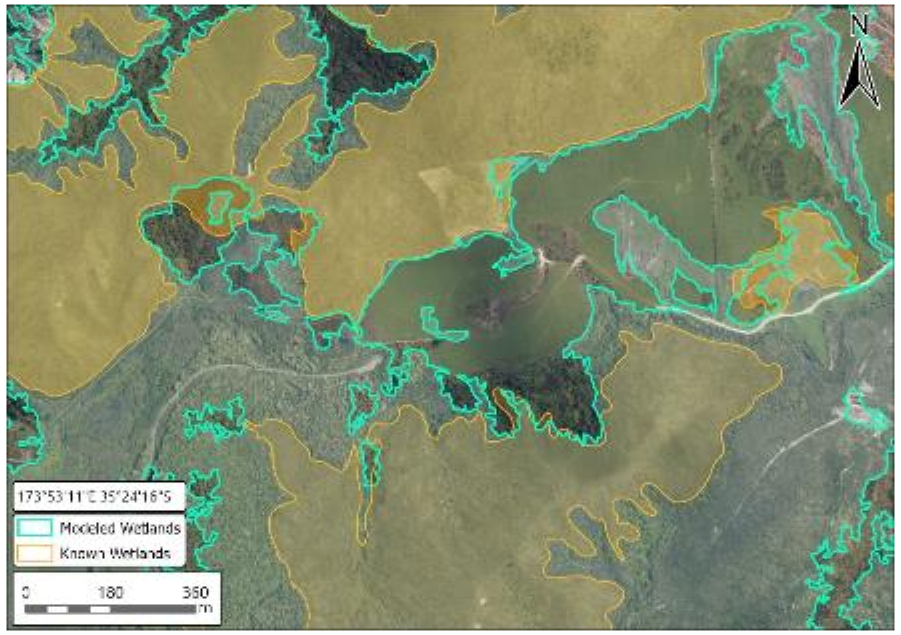


Image: 9

Description: High degree of 'insideness' of known wetland and high recall on model output (pakahi)

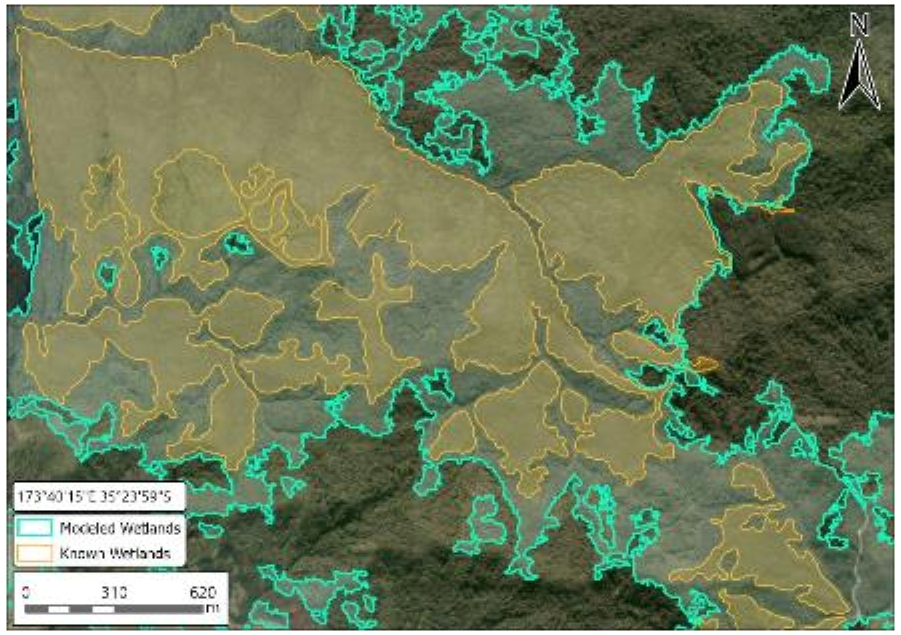


Image: 10

Description: Model outputs identifying additional potential wetland

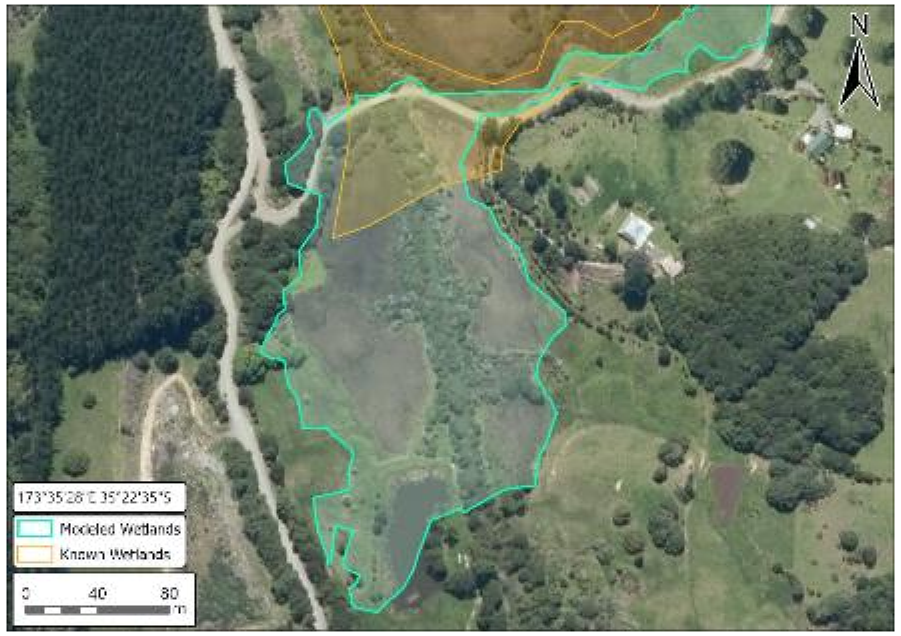


Image: 11

Description: High spatial concordance with mapped known wetland

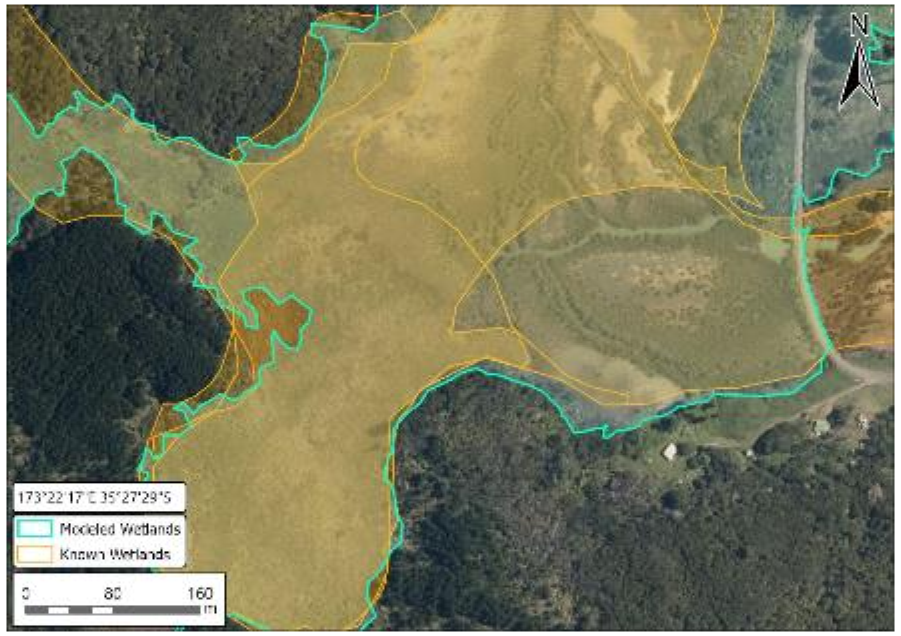


Image: 12

Description: Exclusion of CMA from model outputs

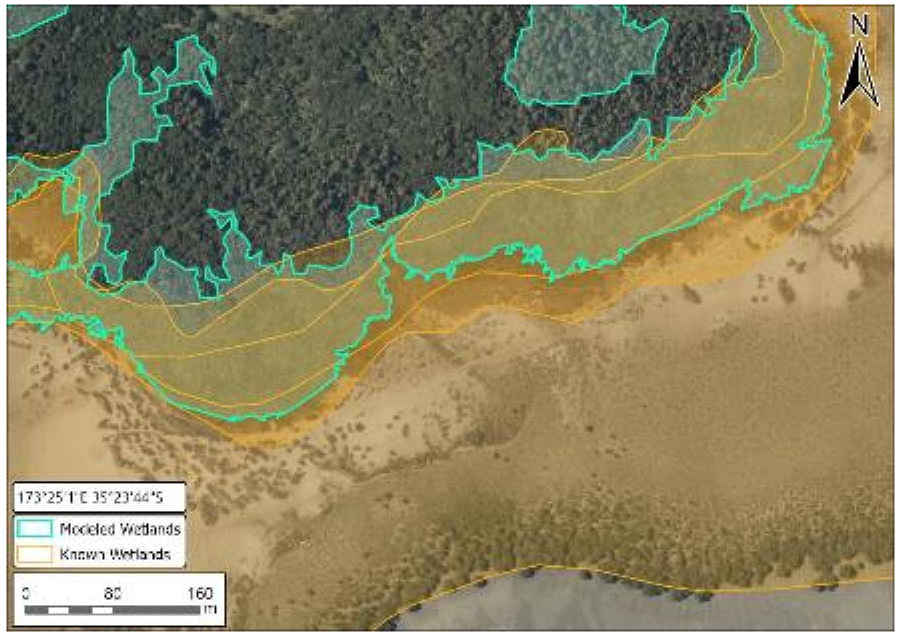
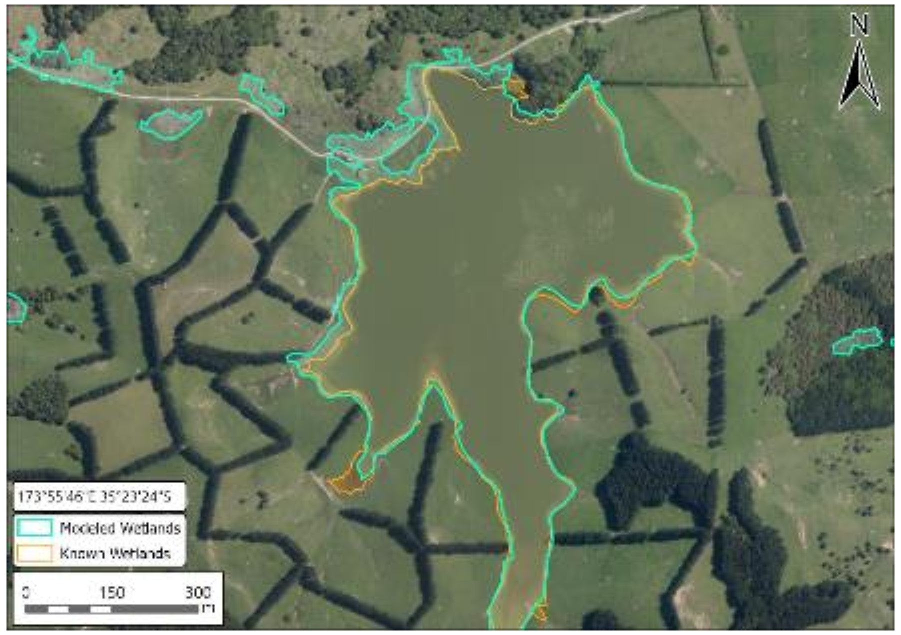
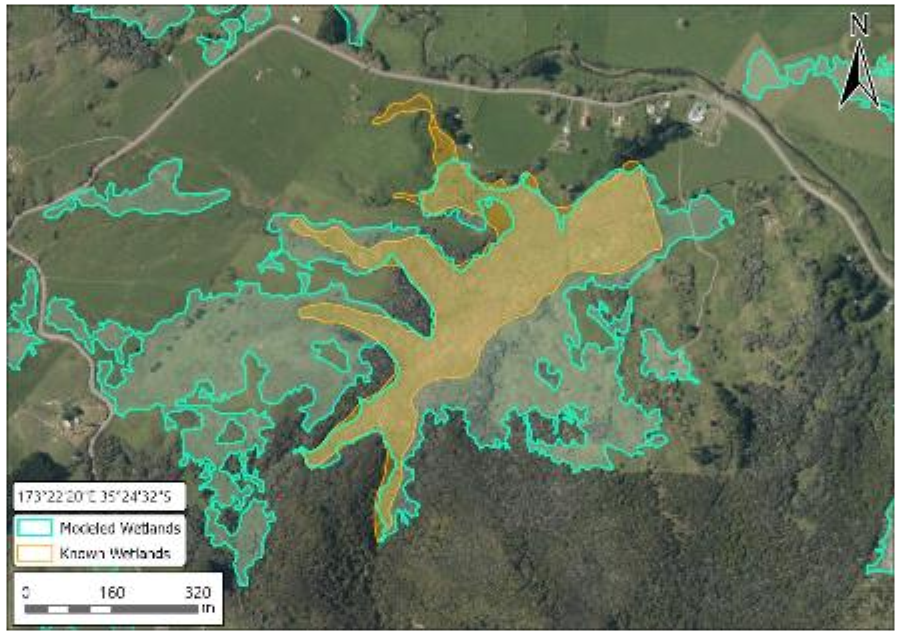


Image:	13	**Image:**	14	**Image:**	15
Description:	High spatial concordance with mapped known wetland and model outputs identifying additional potential wetlands	**Description:**	High spatial concordance with mapped known wetland	**Description:**	High degree of 'insideness' of known wetland and high recall on model output including likely false positive model results





Appendix 3 Machine Learning Algorithm

The project code will be copied to the Ministry's Github repository.

The repository consists of:

- Python code and a neural network model weights file.
- The python script will run the wetland inference over the imagery inputs to produce a wetland class raster.
- The models are trained on vertical aerial photography (RGB or RGBI), Sentinel-2 multispectral imagery, LiDAR and radiometric derived wetness.

The repository contains a requirements.txt file that lists the python libraries required to run the code and a README.md file that describes how to run the ML inference.

Appendix 4 Processing Pipeline

The following steps are required to perform the Machine Learning classification, polygon generation plus post processing procedure. The software requirements include: GIS processing software such as ArcGIS or QGIS, Rasterio, Python software libraries, Keras/Tensorflow and a GPU compute environment. The procedure can be run on a Windows or Linux operating system.

The processing steps are as follows:

1. Assemble all input data. Note raster inputs vary by model. Imagery should be in TIF format.
2. Calculate indices – NDWI, TWI, DEM, NDVI, Soil Index.
3. Resample all inputs to a unified spatial resolution. This varies by model and converges to the highest resolution. Assemble raster data into multiband composite.
4. Gather local training annotations per class to calibrate model to ecoregion.
5. Train model using libraries in Appendix 3.
6. Run RF inference using model codebase supplied.
7. Run CNN inference using model codebase supplied.
8. Remove noise from the CNN wetland-prediction raster using tools such as ArcGIS Desktop's Majority Filter, Boundary Clean, Expand, Shrink. Specifically targeting outlying pixels and filling small gaps in wetland areas.
9. Convert the raster output to polygons using software such as ArcGIS Desktop's Raster to Polygon tool.
10. Apply Random Forest and CNN confidence parameters to polygon layer and optimise thresholds for filtering out false positives and optimising recall and precision, if required or desired.
11. Apply confidence attributes to polygons using RF and CNN raster layers. RF confidence and mean confidence of "CNN not wetland". These confidences can be used to further filter the wetland polygon candidates.
12. Run GIS post processing tools. These are grouped into 6 ArcGIS Pro Models that run individual logical segments to dissolve the original data, classify the outputs, check against existing features and raster, such as slope. These tools do the bulk of the data analysis and preparation for the final output.
13. Apply polygon size filter to further reduce noise, if required or desired.

Appendix 5 Geodatabase Attribute List

Table 13: Summary of Wetland Polygon Attributes

Attribute	Description
OBJECTID	Unique object identifier for each dissolved wetland polygon
Shape	Feature type
Lake check	Identifies wetlands that could potentially be a lake by intersection with known lakes and ponds
Artificial check	Identifies wetlands that could potentially be an artificial (not natural) wetlands by intersecting with regional and district level mapped known artificial wetlands.
Known not wetland check	Identifies wetlands that could potentially not be a wetland based on known areas that have been designated as not wetland (for example, field assessed and mapped as not being wetland).
Bog check	Identifies a wetland assigned as a bog, that could potential not be a bog using the rules outlined in appendix 1
Fen check	Identifies a wetland assigned as a fen, that could potential not be a fen using the rules outlined in appendix 1
Swamp check	Identifies a wetland assigned as a swamp, that could potential not be a swamp using the rules outlined in appendix 1
Saltmarsh check	Identifies a wetland assigned as a saltmarsh, that could potential not be a saltmarsh using the rules outlined in appendix 1
Seepage check	Identifies a wetland assigned as a seepage, that could potential not be a seepage using the rules outlined in appendix 1
Ephemeral check	Identifies a wetland assigned as ephemeral, that could potential not be ephemeral using the rules outlined in appendix 1
Lacustrine	Assigns the wetland to the lacustrine hydrosystem based on proximity to a known lake
Riverine	Assigns the wetland to the riverine hydrosystem based on proximity to rivers (of stream order 4 or greater)
Estuarine	Assigns the wetland to the estuarine hydrosystem when the wetland is within an area of known saline vegetation types
Palustrine	All wetlands not assigned as a lacustrine, riverine or estuarine hydrosystem
Hydroclass	Name of assigned hydroclass
Impervious or similar	Identifies a wetland that could potential not be a wetland due to proximity with known impervious surfaces or buildings.
Bog %	Percentage of dissolved wetland polygon that is bog
Ephemeral %	Percentage of dissolved wetland polygon that is ephemeral
Fen %	Percentage of dissolved wetland polygon that is fen
Marsh (%)	Percentage of dissolved wetland polygon that is marsh
Pakihi and gumland %	Percentage of dissolved wetland polygon that is pakihi and gumland

Saltmarsh %	Percentage of dissolved wetland polygon that is saltmarsh
Shallow water %	Percentage of dissolved wetland polygon that is shallow water
Swamp %	Percentage of dissolved wetland polygon that is swamp
Seepage %	Percentage of dissolved wetland polygon that is seepage
Wetland class	The dominant wetland class based on the class percentages
Wetland Confidence (CNN)	Mean level of confidence in wetland delineation based of the final dissolved wetland attributed from the CNN raster using zonal statistics. Level of confidence ranges from 0 to 100 with a higher value attributed to higher confidence
Wetland Confidence (RF)	Mean level of confidence in wetland delineation of the final dissolved wetland extent attributed from the Random Forest model raster using zonal statistics. Level of confidence ranges from 0 to 100 with a higher value attributed to higher confidence
Shape length	Maximum length of dissolved wetland
Shape area	Dissolved wetland area