Manaaki Whenua Landcare Research

Northland Regional Council land-use classification: methodological report

Prepared for: Northland Regional Council

October 2023

Northland Regional Council land-use classification: methodological report

Contract Report: LC4345

Richard Law, James Ardo

Manaaki Whenua – Landcare Research

Reviewed by:

Laise Harris
Senior Researcher – Land Mapping Programme Leader

Robbie Price
Senior Technician – Informatics
Manaaki Whenua – Landcare Research

Approved for release by:

Sam Carrick
Portfolio Leader – Characterising Land Resources
Manaaki Whenua – Landcare Research

Contents

Summary

Project and client

Northland Regional Council (NRC) contracted Manaaki Whenua – Landcare Research (MWLR) to produce a land-use map, comprising two parts:

- a primary land-use layer to identify the major land-use types at the primary parcel scale, which will be made publicly available.
- a technical layer that includes additional existing underlying resource information, such as geology, soil, climate, and sub-parcel land-use.

The justification for this is that existing land-use information available to NRC is based on national land cover (LCDB) which has not been verified against region-scale information on land-use and land-use intensity. Therefore, a high-resolution geographic land-use layer with the capability to integrate key region-scale information is critical for improved freshwater accounting, economic impact assessment, policy development, and efficient implementation of land management mitigation measures.

Objectives

The objectives of the project were to:

- produce the primary land-use layer by applying the Australian Land Use and Management (ALUM) classification vocabulary to each primary parcel.[1]
- produce the technical layer by classifying an atomised discrete global grid system representation of the Northland region according to the same ALUM classification, including a set of additional attributes (generally of continuous raster data, such as rainfall).
- perform the above by applying a system of land-use classification that allows for the rapid reproduction and extension of land-use maps, and which can potentially apply future updates to input data.
- validate the primary land-use layer through a stratified random sampling technique ($n \approx 1,000$).[2]

---

[1] The ALUM classification is used in this case because New Zealand does not currently have an equivalent general-purpose land-use classification (including lists of farm commodities and management practices) that is ready for adoption.

[2] As will be explained later, we were unable to ground-truth such a large number of random positions, so our 'validation' exercise involved checking for visual consistency with recent aerial imagery. Alone this is insufficient as a rigorous validation for certain land-use classes that cannot be inferred from imagery alone. This is typically the primary distinction drawn between land use and land cover.

Methods

- Data included in the classification process include layers from many New Zealand central government departments, conservation groups, the Land Resource Information System (LRIS) portal, and data supplied directly from NRC.
- Each cell in the output data is intermediately classified as each class in the ALUM classification. The final classification is determined through a deterministic prioritisation process that combines a qualitative confidence score that captures trust in particular data sets and corroboration among multiple data sets; data currency; and geographic scale.
- Due to the volume and disparate provenance of source data used in the classification, we decided very early to use a discrete global grid system. Although considered an internal implementation detail, we consider this to be worth noting. This decision was made in order to:
    - ensure rapid prototyping through efficient combinatorial analysis and data partitioning, i.e. by being able to combine a large number of input layers and to apply expensive classification logic in an embarrassingly parallel fashion
    - avoid innumerable sliver polygons derived from overlaying a large number of data sets with arbitrary coordinate precision; such sliver polygons are the undesirable digital artefacts of combining vector geometries without considering uniform mapping scale or topology
    - allow easy integration of raster (numerical grids, such as continuous elevation data) and vector data (point, line, and polygon features with attributes, such as property boundaries)
    - allow data to be dynamically aggregated to any spatial unit (e.g. property parcel) or none (i.e. individual cells)
    - avoid human intervention (to allow automation for reproducibility).

Results

- A new, classified land-use data set was developed for NRC. This implemented the Australian Land Use and Management (ALUM) classification, Version 8 (ABARES 2016), which is licensed under the Creative Commons Attribution 3.0 Australia Licence (CC BY 3.0 AU). See Figure 1.
- This licence allowed us to use and adapt the ALUM classification to be compatible with tools built on the ALUM schema, while adapting a minimum set of extensions for specific issues of interest for land-use in Northland.
- The land-use classification combines multiple data sources into a single data set using a common spatial unit for classification (the property parcel), according to the ALUM schema, which includes the name of the class, commodities list, management practices, a confidence score, and provenance information (source data name, date, and scale). This is intended for public consumption.
- A technical layer of specific biophysical and climate attributes was also produced that uses the same data model but does not present the classification at any particular spatial unit, such as a parcel. It allows combined querying of both land-use and landform or climate attributes. Parcel-scale classifications are built from data classified

at the level of individual cells (or pixels); this layer does not aggregate up from cells, but instead entirely retains them. This is a higher-resolution land-use data set, intended for internal use at NRC. The average area of cells at the chosen resolution is approximately 44 m^2.[3] The average length of an edge of the hexagonal cells at this resolution is 4.1 m, and the total number of individual cells required to represent the region is 362 million.

Recommendations

- In the future, a New Zealand-specific classification system should be adopted in preference to ALUM. Although we consider ALUM useful, and it should be concordant with any future New Zealand land-use classification system (as it is with ANZSIC), there are several aspects of it that are only relevant to Australia, and several important omissions for the New Zealand context.
- To enhance updateability, data should be shared using standard spatial data APIs (such as WFS 2.0, OGC API – Features, or ArcGIS REST API) in preference to sharing static files. These do not have to be publicly accessible if that is inappropriate.

[3] H3 DGGS, cell resolution 13. This is not an equal-area DGGS, so cells may be as small as 26 m^2 and as large as 52 m^2 in the most extreme cases. See https://h3geo.org/docs/core-library/restable/

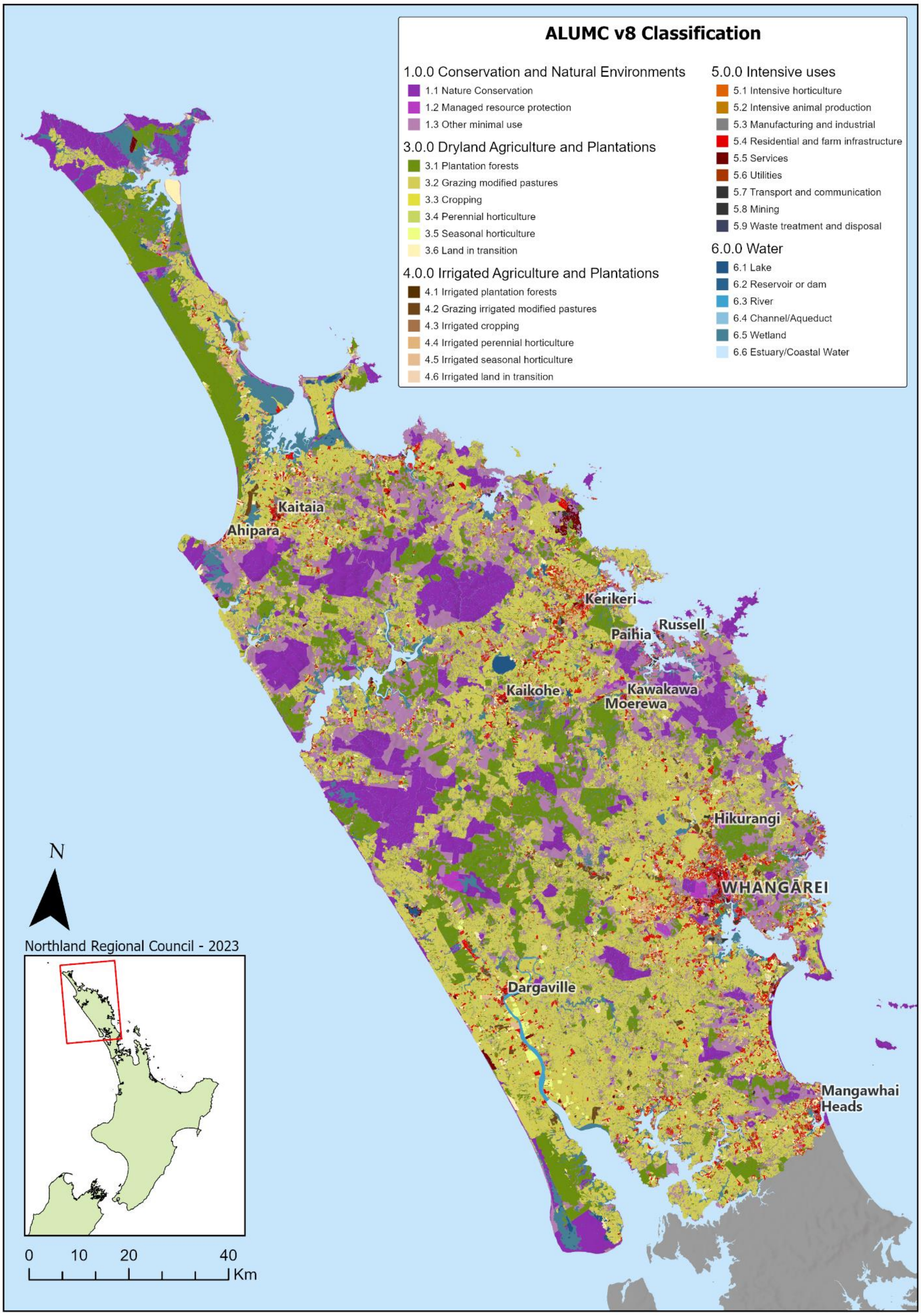


Figure 1. Land-use classification of Northland (secondary level). Note that Class 2 (Production from Relatively Natural Environments) is not considered (see section 4.4.3).

1 Introduction

Northland Regional Council (NRC) is interested in delineating the different land-uses in its region to support its regulatory role, particularly in relation to freshwater management. An understanding of current-day land-uses is also necessary for the regulation of intensive farming activities that have potential adverse effects on the environment.

Existing land-use information available to NRC is based on national land cover (LCDB) which has not been verified against region-scale information on land-use and land-use intensity. Therefore, a high-resolution geographic land-use layer with the capability to integrate key region-scale information is critical for NRC to perform its role.

2 Background

Northland Regional Council (NRC) contracted Manaaki Whenua – Landcare Research (MWLR) to produce a land-use map, comprising two parts:

- a primary land-use layer to identify the major land-use types at the primary parcel scale, which will be made publicly available
- a technical layer that includes additional existing underlying resource information, such as geology, soil, climate, and sub-parcel land-use.

An important requirement of this land-use map was that it was to be produced without the use of AsureQuality AgriBase data, as NRC does not have a license to use it. This influenced the project in two important ways. First, farm boundaries had to be obtained from NRC records or Land Information New Zealand (LINZ) property titles. Second, the range of possible commodities (livestock and crop types) was more limited than it otherwise might have been.

NRC also requested historical land-use maps to allow temporal comparisons of land-use. We considered this unfeasible for two reasons.

- Many important spatial data are made available through web portals that do not support requests for 'point-in-time' queries (i.e., the state of a particular spatial database at a particular point in the past), such as cadastral information.
- Regional-scale and quality-controlled LiDAR has only recently been flown and published for the first time.

NRC accepted this reasoning, and therefore a historical land-use map was considered out of scope for this contract. However, attention has been paid to maximising the computational efficiency of the process and minimising human intervention in the processing of input data, with the intention that, in the future, the land-use classification could be run with the same class definitions to enable a time-series record of land-use.

3 Objectives

The objectives of this project were to:

- produce the primary land-use layer by applying the Australian Land Use and Management (ALUM) classification vocabulary to each primary parcel
- produce the technical layer by classifying an atomised discrete global grid system (DGGS) representation of the Northland region according to the same ALUM classification, including a set of additional attributes (generally of continuous raster data, such as rainfall)
- perform the above by applying a system of land-use classification that allows for the rapid reproduction and extension of land-use maps, and which can apply 'upstream' updates to input data
- validate the primary land-use layer through a stratified random sampling technique ($n \approx 1{,}000$).

4 Methods

4.1 Adoption of the Australian Land Use and Management classification

The ALUM classification is used in Australia to provide a nationally consistent method of presenting land-use information. It is not intended for one specific use, unlike the Land Use and Carbon Analysis System (LUCAS) New Zealand Land Use Map, for instance. Instead, it is a general-purpose classification. The latest version (Version 8) adheres to the Australian Spatial Data Infrastructure (ASDI) standard for land-use data sets. ALUM continues to be used and revised using a collaborative approach involving all Australian states and territories (Lesslie et al. 2006).

We strongly encourage the reader to stop and become familiar with ALUM at this point (refer to ABARES 2016). We will assume in what follows that the reader is at least cursorily familiar with ALUM.

New Zealand does not currently have an equivalent general-purpose land-use classification ready for adoption. (Designing such a classification is a wider problem that is outside the scope of this project.) Similar recent regional land-use classifications in New Zealand have developed independent, customised and (unfortunately) incompatible land-use classification vocabularies and schemas (Herzig et al. 2020; Pearson & Couldrey 2016). For a thorough review of present and historical land-use classification systems developed and applied in New Zealand, see Rutledge et al. 2009.

Although we readily acknowledge that ALUM is not an ideal fit given New Zealand's idiosyncratic needs with regard to land-use information and our constitutional and legislative differences from Australia, ALUM is freely available and has proven to be robust across real-world application over several years and revisions. Further justification for its adoption is covered in the following points.

- ALUM includes rules for converting Australia and New Zealand Standard Industrial Classification (ANZSIC) 2006 codes to ALUM.
- Adopting ALUM obviates the need for us to spend time designing a data schema and classification vocabulary, and defining terms within the scope. Such tasks would have had a significant opportunity cost that would have affected the other quality aspects of the published Northland land-use data.
- Independently developing a customised land-use classification system for Northland would have further perpetuated land-use incompatibilities across regions.
- ALUM entails an extensive controlled vocabulary of 'commodities' and 'land management practices' intended to allow further distinction within land-use classes.
- The physical and economic geography of Tasmania (where ALUM has been used for some time) is not dissimilar to that of Northland; if it can be usefully applied there, it stands to reason that it might be usefully applied here.
- Tools for the visualisation of ALUM land-use data are freely available for download. Any software or visualisation rule sets designed for use with ALUM data will continue to function, even if the underlying information changes, while adhering to the same schema (e.g. for corrections, or across time).
- Adoption of a third-party, published but open standard allows for its repeated adoption, and therefore consistency across time, for land-use change analysis.

There are, however, disadvantages to adopting ALUM.

- The most obvious disadvantage is that there are several classes that are simply inapplicable to New Zealand.[4]
- The ALUM system does not reflect the principles of the Treaty of Waitangi or directly enable a Treaty-led, science-policy approach (Kukutai et al. 2021).
- Some design choices have been made in developing ALUM that, in our opinion, are inconsistent with its design philosophy.[5]
- There is a major top-level class distinction between irrigated and non-irrigated agriculture and plantations – despite the possibility of recording irrigation use and type as land management practices – which indicates an intense interest in matters of water allocation that are not as pressing in most regions of New Zealand.
- Any extensions to the classification system data schema (to adapt it for better application in New Zealand) might introduce an incompatibility that could mean losing several of the aforementioned benefits of its adoption. In section 4.1.1 several small accommodations are noted, but all of these accommodations are made within the ALUM schema by using comments rather than redefining classes or creating new classes. Thus, in this case we have avoided this potential disadvantage. At the same time, a number of ALUM classes are not applied in this classification, generally due to their inapplicability to New Zealand. These are simply omitted from consideration in the classification process.

---

[4] For example, 3.3.5 (Sugar), 3.3.6 (Cotton), and 3.3.7 (Alkaloid poppies).

[5] For example, there is a class, 3.4.8 (Citrus), which can equally be described at the same level in the hierarchy by 3.4.1 (Tree fruits), with the commodity recorded separately.

4.1.1 Specific extensions to ALUM

During the initial scoping phase, and again during a workshop held in April 2023, NRC described some specific requirements of the land-use data that didn't fit into ALUM. However, accommodations were designed that did not require adjustments to ALUM.

Sheep and beef cattle production systems

The first was to further classify land under classes 3.2 ('Grazing modified pastures') or 4.2 ('Grazing irrigated modified pastures') with sheep or beef livestock according to the Beef + Lamb New Zealand sheep & beef cattle production system (Table 1). This classification system makes a primary distinction between sheep & beef production systems depending on whether they are in the North or South Island. Thereafter, farms are generally distinguished based on slope, altitude, soil fertility, stocking rates, and body condition scoring. Finer distinctions are drawn by regional divisions. Table 1 includes some of the published summary information for North Island sheep & beef farm classes, specifically for the northern North Island (which includes Northland).

Determining whether and how pastoral lands should be classified into these classes is not well defined, especially in the complete absence of farm-level stocking rate (typically obtained from AgriBase with some unknown amount of error) and body condition scoring data. For the purposes of the present classification we have opted to use land-use capability classes[6] as an indicator, because these already embody slope and soil fertility, among other relevant geographical considerations.

Table 1. Sheep and beef cattle production systems *2021/22 (provisional) (northern North Island)*

Island	Class	Label	Est. farms	Total gross revenue	Total farm expenditure	SU*/ha
North	3	Hard hill country	920	$645,100	$459,100	7.8
	4	Hill country	3,055	$454,700	$330,000	9.8
	5	Finishing	1045	$592,200	$378,100	12.8
South	1	High country	200			
	2	Hill country	620			
	6	Finishing breeding	1,820			
	7	Intensive finishing	1040			
	8	Mixed finishing	465			

* SU = stock units.

Source: Beef + Lamb NZ Ltd 2023b.

[6] https://lris.scinfo.org.nz/layer/48076-nzlri-land-use-capability-2021/

Rather than define a quaternary level of classification that could include these classes, we opted instead to include the labels for these classes within the 'comment' field, and thus avoid any break with the ALUM schema, while allowing for users to select farms on this basis by performing an attribute filter on the 'comments' field.

There is insufficient information in the present case to record cattle and sheep commodities, due to the absence of AgriBase data, and consents equivalent to dairy effluent consents (which have the effect of locating dairy farms). Thus, the comments added to pastoral land which describe the nature of the land according to sheep & beef farms classes should not be taken as identifying the actual purpose of the farm.

Riparian vegetation

The second specific extension requested was the ability to identify riparian vegetation. Although the technical layer will include a canopy height model (CHM) which could be used to identify these areas, we decided to create some derived layers and apply simple logic to specifically label riparian vegetation on pastoral farms (i.e. classes 3.2 and 4.2). Again, we opted to do this using the 'comment' field rather than defining a specific class.

To identify riparian vegetation, we considered any land with a canopy height over 1.5 m (from a 5 m LiDAR-derived digital elevation model (DEM) (OpenTopography 2022). Then we further selected land that was also within 2 m (vertically) and 30 m (horizontally) of its drainage channel. To produce a smoothed and hydrologically-breached DEM,[7] a height above nearest drainage (HAND) map,[8] and a downslope distance to stream (DDTS) map,[9] we used WhiteboxTools, v2.3.0 (Lindsay 2016b).

4.2 Use of a discrete global grid system

A DGGS, in the simplest possible terms, is an imaginary mosaic that neatly divides the surface of the globe into discrete cells in a systematic way (Figure 2. The icosahedron-based hexagonal H3 DGGS showing the three coarsest resolutions (resolution 0 = green; 1 = blue; 2 = red)). The division can be made at different 'levels' or 'resolutions', as a nested grid of cells. It is discrete (each cell is distinct from every other cell, unique, and has a permanent identification) and forms a regular grid (there are no gaps between cells; cells have systematically identifiable neighbours, 'parent', and 'child' cells). A DGGS is global because each one is designed to be capable of covering the entire surface of a sphere or ellipsoid (Sahr et al. 2003).

---

[7] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#breachdepressionsleastcost

[8] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#elevationabovestream and Rennó et al. (2008)

[9] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#downslopedistancetostream

Another way to think of a DGGS is as a hybrid spatial data structure that is neither vector nor raster, but retains properties of both while adding certain advantages for spatial data storage, analysis, and modelling (Hojati et al. 2022; Purss et al. 2019).

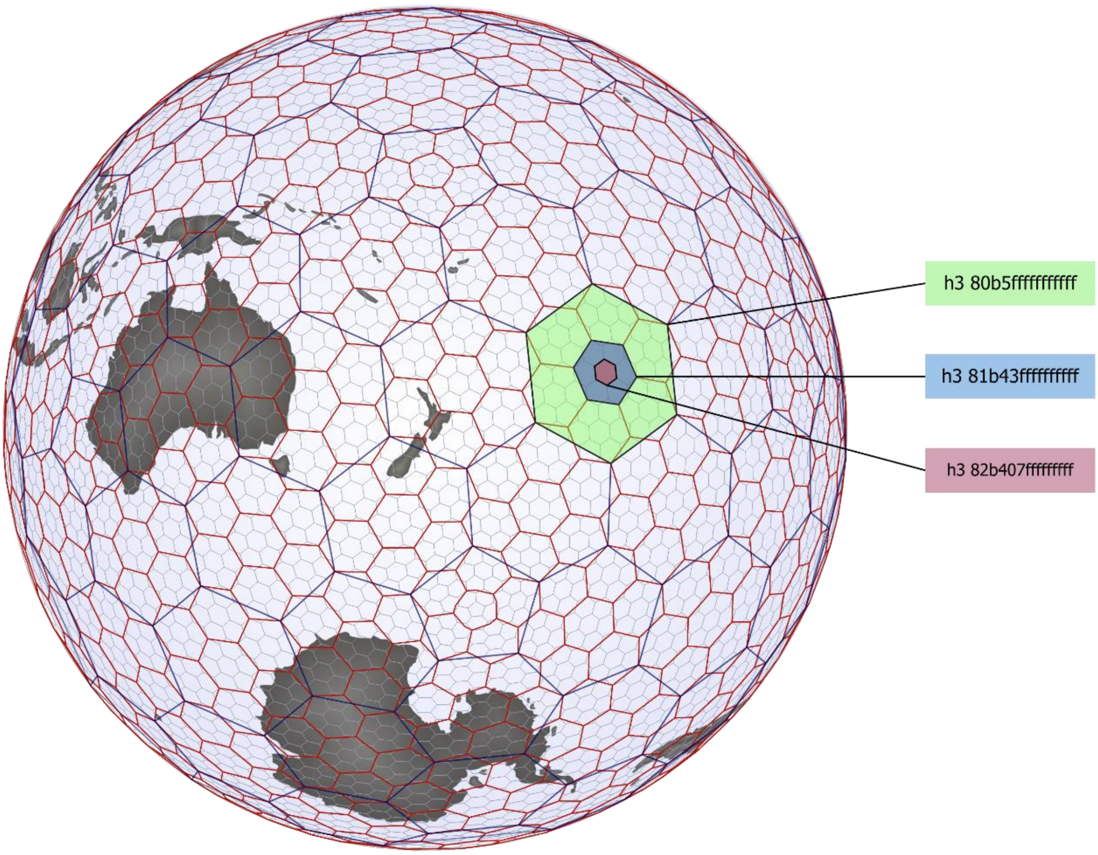


Figure 2. The icosahedron-based hexagonal H3 DGGS showing the three coarsest resolutions (resolution 0 = green; 1 = blue; 2 = red).
Note the non-congruent, rotational nesting of seven child cells within parent cells. Each cell has a unique 64-bit integer ID (H3 index), and the relationships between cells can be determined using these indices.

In prior land-use mapping projects we have noted the difficultly of working with large numbers of complex vector data sets (Herzig et al. 2020). The fundamental issue is that because vector data sets are produced independently and at different scales, even when they intend to model the edges of the same real-world physical objects (such as rivers or forests), all of them disagree about the precise location of the edges. A naïve intersection of vector data sets therefore produces a very large number of 'sliver polygons', which may be considered artefacts rather than real geographical information. The sheer number of these, which grows exponentially as more layers are combined, makes such an approach untenable. This is partly an issue of geographical scale and how this is modelled in the GIS (geographic information system).

Yet, at a fixed scale or tolerance, the edges of real-world objects can be considered equivalent. Therefore, an approach taken in Herzig et al. (2020) is to convert all input data for a land-use classification into raster data on a common grid. Land-use classifications can then be applied like raster algebra. By fixing the resolution of all inputs (in this case, to 5 m) the issue of sliver polygons is avoided. This is an efficient method; the trade-off is that it loses the advantages of retaining the vector data model. Among these advantages are: non-numeric attribute types and the ease of considering geographical objects (points, lines and areas, which may intersect and overlap).

A fundamental benefit of adopting a DGGS is that we can make the same transformation to a fixed-resolution grid while retaining the advantageous properties of vector geographical data. In other words, we can avoid intractable issues with sliver polygons, and keep data in a relational database. Also, because of the hierarchical nature of a DGGS, the classification workflow can be efficiently parallelised by working with partitions of data that are simply references to parent cells.

Logically, the classification is essentially the same as a raster algebraic combination of multiple layers, but by using Structured Query Language (SQL) we can more clearly express complex logical rules. A relational database management system (RDBMS) also allows us to define types that match the ALUM schema to perform type validation and apply overlay rules (particularly as ALUM gives priority to water classes).

More generally, a DGGS is an excellent fit for a land-use classification process for several important reasons.

- All input data become quantised (because they are modelled on a discrete, finite grid).
- Raster and vector data are modelled on a uniform grid, and so become interoperable.
- Classification logic can be conceptualised in the same way as raster algebra, but can adaptively use complex types (e.g. arrays, enumerated types or 'enums'), queries (e.g. text searches[10]), and custom functions within a RDBMS.
- Data are pre-prepared for partitioning on the basis of cell parents. By choosing a target resolution and using a fixed offset level as the basis for partitioning, classification queries can take a predictable amount of time (per partition) and can be computed in parallel without loading all data into memory.
- Data can later be dynamically (re-)grouped by feature identifiers associated with the original vector geometries (such as property boundaries), allowing the DGGS to remain an internal implementation detail, and for outputs to be expressed in a variety of 'spatial units' that do not need to be predetermined, such as parcels, farms, watersheds or a segmentation based on physical attributes.

[10] https://www.postgresql.org/docs/current/datatype-textsearch.html

General-purpose land-use classifications such as ALUM involve the combination and normalisation of wildly diverse sets of thematic input data. Each datum might be produced at a different scale and by a different organisation. To produce a land-use map for NRC, we have integrated over 30 different input layers, and adopting a DGGS approach allows us to efficiently scale as the number of data sets grows. The classification can therefore be maintained as a series of logical rules (written in SQL) rather than as a data set that is progressively and manually manipulated. This approach is therefore reproducible, as well as iterative and efficient to compute.

4.2.1 vector2dggs and raster2dggs: open-sourced tools

Despite clear benefits, the DGGS data model is not yet widely adopted within GIS. We developed tools for the efficient indexing of raster and vector data into a DGGS as part of this project. Our requirements for these tools were that they had a command-line interface, were easy to install, were multi-threaded and capable of using all CPU (central processing unit) cores, and would not use excessive amounts of read-only memory (RAM). They had to be capable of reading data from many common vector and raster formats, and write DGGS-indexed data out into a column-oriented data store,[11] which in turn may be copied into PostgreSQL.

The tools we created are called vector2dggs and raster2dggs (Ardo & Law 2023a, 2023b). We released these as open-source tools to leverage the experience of other developers in the community. We also wanted to offer these convenient methods back to the open-source community in reciprocity because we rely entirely on open-source software for other aspects of the classification workflow. Currently our tools support the H3 DGGS (Uber Engineering 2023), but they are intended to accommodate others in the future. See Figure 3 and Figure 4 for visual summaries of these tools.

Uber H3 is only one DGGS. It is based on a Dymaxion-oriented, Gnomonic-projected icosahedron that has 122 base cells, and it features an aperture seven subdivision (seven child cells for each cell) (Sahr et al 2003). Almost all H3 cells are hexagonal; and they are not strictly equal-area. The Uber H3 DGGS was chosen because it has many implementations and bindings in different languages and for different platforms, particularly in Python (Jahn 2023) and for PostgreSQL (Knudsen 2023).

4.3 Data sources

The production of a land-use map is an exercise in combinatorial analysis. Once a classification system has been chosen (ALUM), the task becomes one of assembling a collection of data sets that describe phenomena that can be mapped to these classes (or, if existing data are wholly inadequate and resources allow, developing a data collection methodology).

[11] Apache Parquet: https://parquet.apache.org/

A DGGS was adopted because it enables an adaptive and scalable approach to land-use classification. Over time this allowed for an iterative approach with tight feed-back loops to enable additional complexity, either due to a refinement to a rule or the inclusion of more data. At each change we could consider the influence of individual data sets and do further refinement, and ultimately rule them in or out according to our judgement of their contribution to the output.

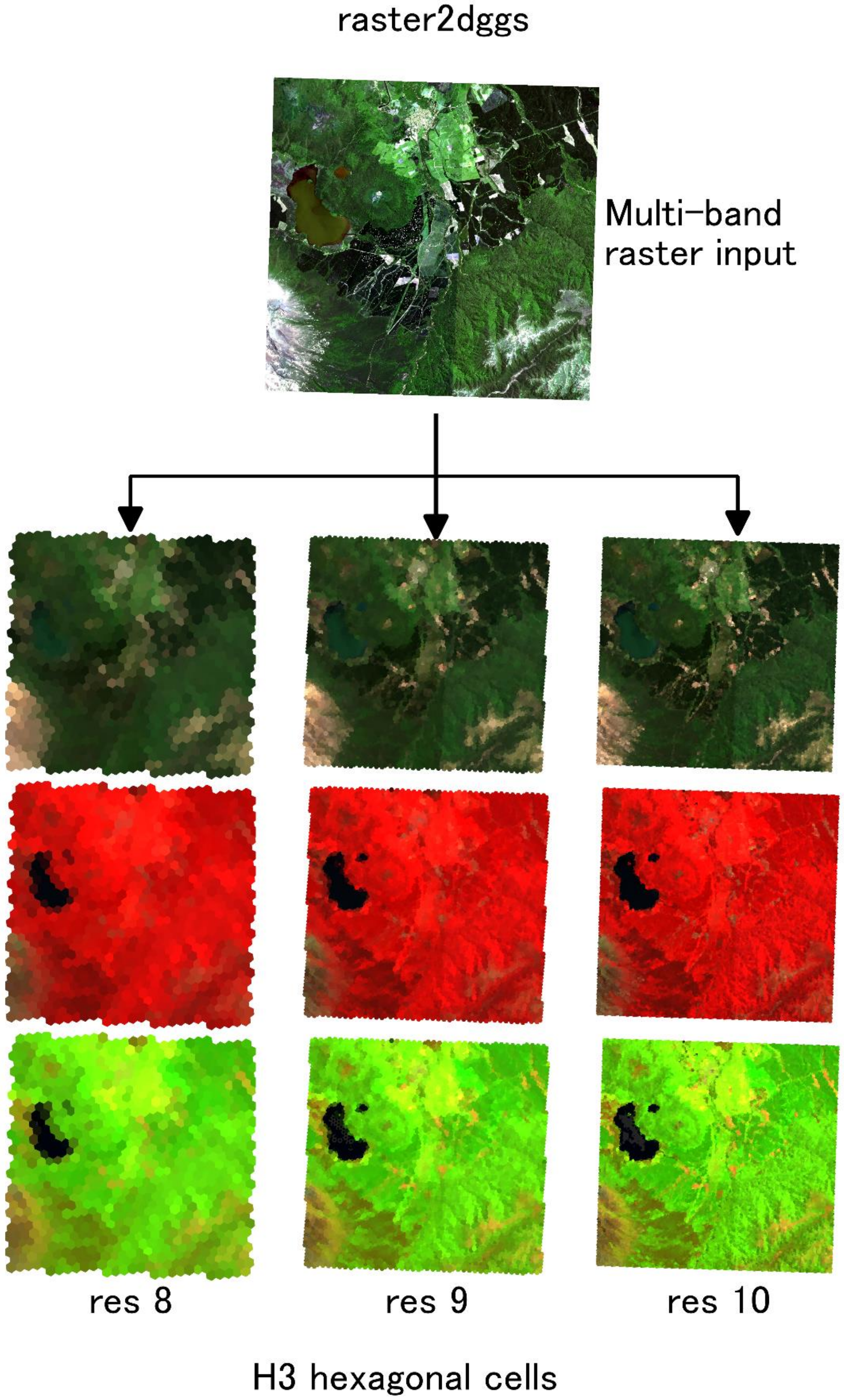
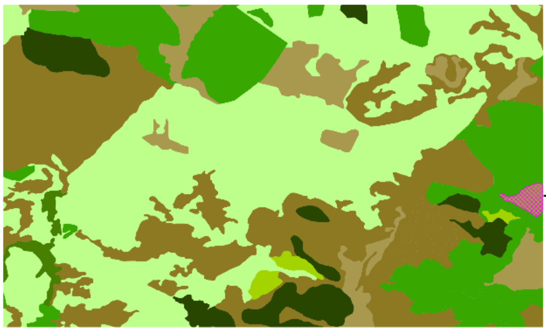


Figure 3. Illustration of the application of raster2dggs: converting a multi-spectral Sentinel-2 image to the H3 DGGS at three different resolutions.

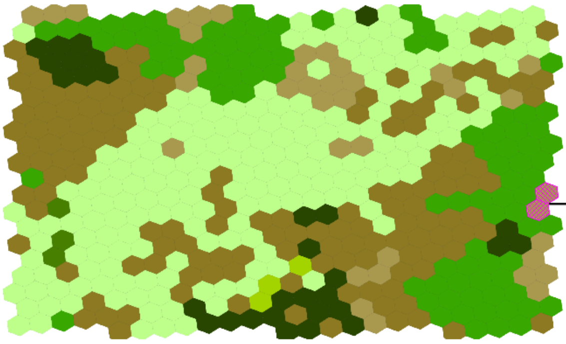
Vector input



fid	Name_2018
37	Manuka and/or Kanuka
38	Exotic Forest
39	Indigenous Forest
40	High Producing Exotic Grassland
41	Broadleaved Indigenous Hardwoods
42	Manuka and/or Kanuka
43	Manuka and/or Kanuka
44	Low Producing Grassland
45	Exotic Forest
46	Exotic Forest
47	Indigenous Forest
48	High Producing Exotic Grassland
49	Indigenous Forest

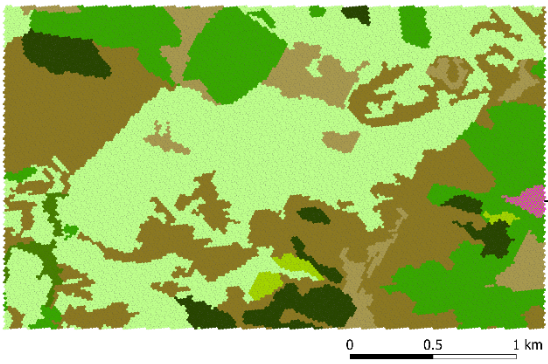
vector2dggs

H3 Resolution 10



fid	h3_10	Name_2018
39	8abb50235787fff	Indigenous Forest
39	8abb5023578ffff	Indigenous Forest
40	8abb502352d7fff	High Producing Exotic Grassland
40	8abb502352f7fff	High Producing Exotic Grassland
41	8abb50230317fff	Broadleaved Indigenous Hardwoods
41	8abb502303affff	Broadleaved Indigenous Hardwoods
42	8abb50230acffff	Manuka and/or Kanuka
42	8abb50230adffff	Manuka and/or Kanuka
43	8abb50230a47fff	Manuka and/or Kanuka
43	8abb50230a5ffff	Manuka and/or Kanuka
44	8abb50235677fff	Low Producing Grassland
44	8abb5023575ffff	Low Producing Grassland
45	8abb50230bb7fff	Exotic Forest

H3 Resolution 12



fid	h3_12	Name_2018
41	8cbb502300493ff	Broadleaved Indigenous Hardwoods
41	8cbb502300495ff	Broadleaved Indigenous Hardwoods
41	8cbb502300497ff	Broadleaved Indigenous Hardwoods
41	8cbb5023004b3ff	Broadleaved Indigenous Hardwoods
41	8cbb5023004bbff	Broadleaved Indigenous Hardwoods
41	8cbb502303021ff	Broadleaved Indigenous Hardwoods
41	8cbb502303023ff	Broadleaved Indigenous Hardwoods
41	8cbb502303025ff	Broadleaved Indigenous Hardwoods
41	8cbb502303027ff	Broadleaved Indigenous Hardwoods
41	8cbb502303031ff	Broadleaved Indigenous Hardwoods
41	8cbb502303033ff	Broadleaved Indigenous Hardwoods
41	8cbb502303035ff	Broadleaved Indigenous Hardwoods
41	8cbb502303037ff	Broadleaved Indigenous Hardwoods
41	8cbb50230303bff	Broadleaved Indigenous Hardwoods

Figure 4. Illustration of the application of vector2dggs: converting Land Cover Database coverage to the H3 DGGS at two different resolutions, and demonstrating how each cell retains attributes from the input.
Notes: Fidelity to the input increases with resolution. What is important to recognise is that sharp vector features merely hide the fact that boundaries in input data are often indeterminate or approximate. See Figure 4

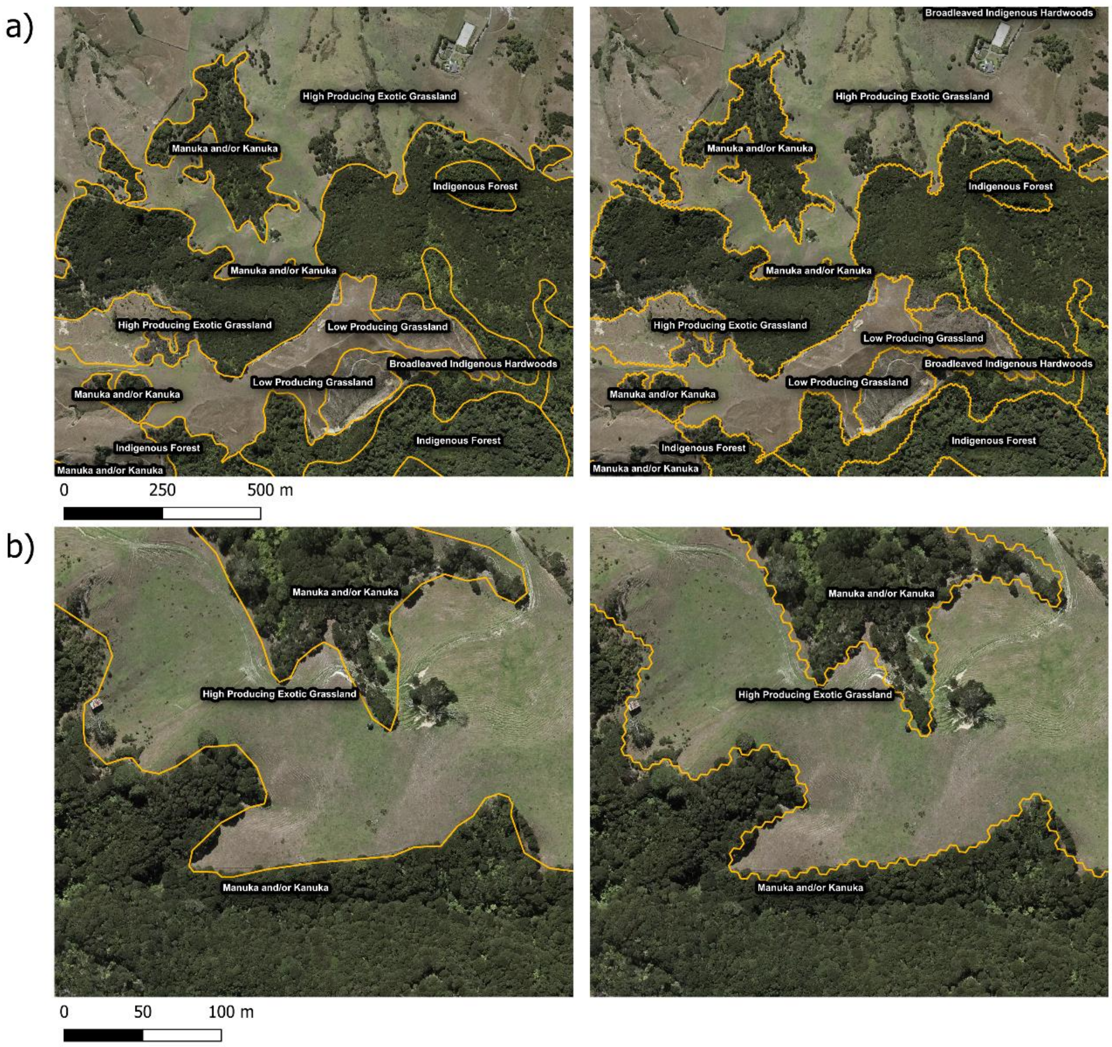


Figure 5. Demonstration of the fidelity of the H3 DGGS (resolution 13, right) to the Land Cover Database (produced at 1:50,000 map scale, left). Map a) is shown at 1:10,000 scale and b) at 1:2,500.

The data sets we have relied on are listed in Table 2. Aspects of these will be progressively discussed in sections 4.4.2–4.4.7 as they become relevant to specific ALUM classes. The data sets that are not publicly available will be discussed presently because they require special attention.

4.3.1 LiDAR-derived data sets

Three data sets were developed by MWLR from LiDAR collected by NRC during 2018–2020, and made available through OpenTopography (Northland Regional Council, Toitū Te Whenua Land Information New Zealand, 2022).[12] These are: a canopy height model (CHM), downslope distance to stream (DDTS), and height above nearest drainage (HAND).

[12] https://www.opentopography.org/

Table 2. Data sources used in the NRC land-use map

Source	Data set	Licence	Landing page
District councils[a]	Ratings database		
DOC	Ngā Whenua Rāhui	Public domain	nwr-open-data-nwr.hub.arcgis.com
ESA & MWLR	Winter 2022 land cover		
Land Information NZ (LINZ)	Central Record of State Land (CRoSL)		http://www.linz.govt.nz
LINZ	NZ Building Outlines	CC BY 4.0	data.linz.govt.nz
LINZ	NZ Facilities	CC BY 4.0	data.linz.govt.nz
LINZ	NZ Primary Parcels	CC BY 4.0	data.linz.govt.nz
LINZ	NZ Property Titles	CC BY 4.0	data.linz.govt.nz
LINZ	Protected Areas	CC BY 4.0	data.linz.govt.nz
LINZ	Topographic 1:50,000[b]	CC BY 4.0	data.linz.govt.nz
Ministry for the Environment (MfE)	Irrigated land area (2020)	CC BY 4.0	data.mfe.govt.nz
MfE	LUCAS NZ Land Use Map 1990, 2008, 2012, 2016, v011	CC BY 4.0	environment.govt.nz
MfE	Land Environments of NZ (LENZ) (level4)	CC BY 3.0	data.mfe.govt.nz
Ministry for Primary Industries (MPI)	Aquaculture Marine Farms	CC BY 3.0 NZ	maps.mpi.govt.nz
Manaaki Whenua – Landcare Research (MWLR)	Land Cover Database v5.0	CC BY 4.0	lris.scinfo.org.nz
MWLR	Revised extents of wetlands in NZ	CCBY-NC-SA4.0	datastore.landcareresearch.co.nz
MWLR	NZ Land Resource Inventory (NZLRI) (land-use capability)	Landcare Data Use	lris.scinfo.org.nz
NRC	Coastal zones		NRC ArcGIS REST API
NRC	Selected Land Use (SLU) Register		NRC ArcGIS REST API
NRC	Water take permits		
NRC	Dairy effluent consents		
NRC	Mangroves and saltmarshes		
NRC & MWLR	Canopy height model	CC BY 4.0	portal.opentopography.org
NRC & MWLR	Downslope distance to stream	CC BY 4.0	portal.opentopography.org
NRC & MWLR	Height above nearest drainage	CC BY 4.0	portal.opentopography.org
OpenStreetMap	OpenStreetMap[c]	ODbL	http://www.openstreetmap.org
QEII National Trust	QEII Covenants		qeiinationaltrust.org.nz
Stats NZ	Urban/Rural 2023	CC BY 4.0	datafinder.stats.govt.nz

Canopy height model (CHM)

The canopy height model is a LiDAR-derived digital surface model (DSM) with buildings subtracted. We used the LINZ buildings footprint data set to determine where these are, and added a small buffer.

Downslope distance to stream (DDTS)

The LiDAR-derived digital elevation model (DEM) is first down-sampled to 5 m, then processed to remove artefact topographic depressions and flat areas using a breach-filling method (Wang & Liu 2006). Five metres was chosen because the target DGGS resolution (H3 13) has an average edge length of 4 m, and also because some down-sampling was necessary to use WhiteboxTools algorithms, which load the input elevation models entirely into RAM (we used a machine that has 96 GB of RAM). Once suitably prepared, we first determined the location of stream channels with the `DInfFlowAccumulation`[13] tool, then ran the `DownslopeDistanceToStream`[14] tool to produce a raster data set, which we included as an input for the land-use classification.

Height above nearest drainage (HAND)

The HAND data set was produced using the same inputs as the DDTS data set and the `ElevationAboveStream`[15] tool.

4.3.2 2022 winter land cover

MWLR maps winter forage crops as part of contracts for the Ministry for the Environment (MfE), separately from this project. Although winter forage crops are not relevant for farm systems in Northland, as a by-product of accurately identifying winter forage elsewhere it is necessary to address potentially confusing land covers. Brassicas and kiwifruit can be readily confused in spatiotemporal, multispectral remote-sensing analyses because they are likely to exhibit a progression from green to brown at the same time (Dr Heather North, remote sensing scientist, pers. comm., 2022). Disentangling kiwifruit and avocados from winter forage is an intermediate step in the process of identifying winter forage crops. This intermediate output then becomes a valuable input to the present NRC land-use data, especially when combined with ancillary data to identify commodities more positively, since there are still many false positives when relying solely on satellite-derived information.

North et al. (2022) explain the methodology of mapping bare ground on rolling and hill country in New Zealand, including a wider variety of land-uses. The data set we use is an

---

[13] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#DInfFlowAccumulation
[14] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#downslopedistancetostream
[15] https://www.whiteboxgeo.com/manual/wbt_book/available_tools/hydrological_analysis.html#elevationabovestream

intermediate output of that work, and we consider only the identification of avocados, kiwifruit, and citrus (not further defined). North et al.'s (2022) project involved ground-truthing various types of orchards in many different regions of New Zealand. Although ground-truthing for this previous work was not conducted in Northland, data from this field work were used to build spectral libraries of different orchard crops, which were then used in an automated workflow exploiting the return period of Sentinel-2 images to analyse spatiotemporal signatures of different ground covers.

Due to the absence of ground-truthing in Northland in this related work, spectral signatures from the Bay of Plenty were applied to the imagery of Northland. Both avocados and kiwifruit are major horticultural crops grown in the Bay of Plenty, and ground-truthing data were readily available. A sample of current citrus orchards in Northland, identified manually from a combination of imagery and ratings data, was also used to develop a supplementary spectral signature for citrus orchards.

The intermediate data made available for use over Northland were derived from Sentinel-2 images. Spectral profiles were constructed based on Bay of Plenty ground-truthing data collected over approximately the same period. The data were available at 10 m^2 resolution, and for five dates in autumn and winter of 2022 (6 May, 11 May, 5 June, 7 June, and 11 August). Because these dates have different satellite swathes, not all parts of Northland are present on all dates; indeed, some places are only represented in one of these images. We calculate a mode across all 66 land-cover classes that are present in the intermediate data, ignoring nulls, in order to reduce the five images to one. Contiguous areas of fewer than 20 pixels were then removed to reduce the level of noise in the data. A sample of the output of this process is shown in Figure 6.

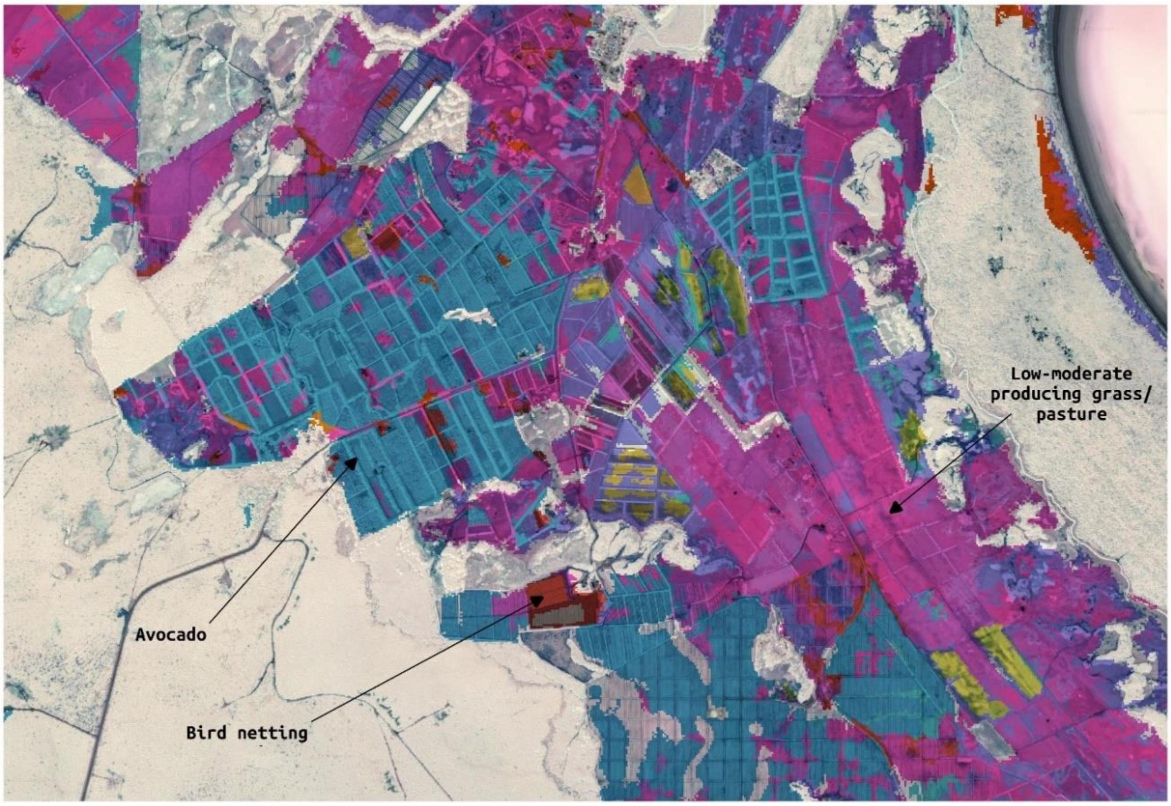


Figure 6. A sample of the composite winter-cover map used to assist with the identification of avocado, kiwifruit, and citrus land-uses.

4.3.3 OpenStreetMap

There were import classes for which we did not have adequate data to represent them. Using Overpass API[16] and Overpass QL, we created narrow queries against OpenStreetMap (OSM) to gather supplementary data to include in the final land-use classification.

OSM data are licensed under the Open Data Commons Open Database License (ODbL).[17] This licence requires that if 'Substantial' use of OSM is made, the 'Produced Work' also needs to be licensed under the ODbL. We consulted internally (with the Head of Data Management and Knowledge Services) whether the proposed uses were 'Substantial', and the judgement was that because the total extraction across all queries is fewer than 100 individual OSM features, the use is considered insubstantial, based on OSM guidelines. However, for a future (repeated) or national-scale classification, the use would reach the threshold for 'Substantial' and alternatives would need to be found to achieve a similar outcome in terms of class coverage if an ODbL licence is to be avoided.

A simple alternative for all these classes (except construction and industrial land-use, since they are readily enumerable) would be to manually identify and tag properties from the LINZ property title data set. This would negatively affect transparency and reproducibility of the land-use map, but if the present land-use map (or future produced works) *were* released under the ODbL licence by NRC, then no further consideration would be necessary.

Construction land-use

Within the NRC regional boundary extent we used Overpass QL to extract nodes, ways, and relations with a `landuse=construction` tag. This was considered necessary to capture land in transition, especially in the south of Whangārei, and near Marsden Point, where there are some ongoing residential developments that are not captured in any other data we have available.

Geothermal power plants

Within the NRC regional boundary extent we used Overpass QL to extract relations with `power=plant` and `plant:source=geothermal` tags, because of a new geothermal power plant that is not included in any other data we have available, and which we consider a significant landscape feature to include.

Industrial land-use

Within the NRC regional boundary extent we used Overpass QL to extract nodes, ways, and relations with a `landuse=industrial` tag. Although these ratings data include industrial land-use, we considered that, on their own, the ratings data have a low-

---

[16] https://wiki.openstreetmap.org/wiki/Overpass_API

[17] https://www.openstreetmap.org/copyright

confidence representation of industrial land-use. We used OSM data to corroborate and supplement the available information in order to build confidence in relevant classes.

Prisons

Within the NRC regional boundary extent we used Overpass QL to extract nodes, ways and relations with an `amenity=prison` tag. Prisons were not specifically included in any other available data, and we considered them a significant land-use, or at least one that is significantly different from surrounding land-uses, which merits special attention to avoid misclassification.

Oil refineries

Within the NRC regional boundary extent we used Overpass QL to extract relations with `landuse=industrial` and `industrial=refinery` tags. There is only one oil refinery in New Zealand, Marsden Point, which is currently being decommissioned. Oil refineries are a remarkably idiosyncratic land-use, the extent of which is ironically not simple to identify from other available data. The NRC Selected Land Use data set did identify the legal parcel on which the refinery stands, but with significant confusion with all other land that handled petrochemicals, including retail service stations.

4.3.4 Internal NRC data

NRC made several important internal data sets available to us for use in classification. However, not all of these data were ultimately used. It was especially challenging to use consent data that are represented as points (rather than associated with properties, as polygons). The two point-geometry data sets that were especially helpful were reservoirs and dairy effluent consents. These were generally straightforward to associate with the apparent intended property or feature boundary (although not perfectly so), since the point location of the feature is often coincident or very near to the relevant property. However, the absence of a reliable farm extent layer meant that, despite making positive associations between dairy effluent consents and apparent properties, we consider that the extent of the cattle dairy commodity is still under-represented in the output.

In contrast, water-take permits, represented as points where the water is taken rather than the area to which water is applied, meant that an otherwise valuable source of information was unable to be integrated into the classification system. If this source of data had relevant property boundaries linked to it, it would have been possible to use information (such as the name of the consented business) to infer more detailed land-use (e.g. irrigated horticulture) and even commodities (many horticultural businesses have the name of the relevant commodity in their title, such as 'King Avocado', even though this information is not otherwise formally captured).

We recommend (section 8) that, in the future, attention be paid to creating and maintaining a reliable property data set that accurately captures properties as combinations of specific identified LINZ parcels, and then associating all consents with particular property boundaries, so that consents can be represented as either points (locations relevant to the consent) or areas (properties related to the consent).

NRC data used were:

- coastline
- the Selected Land Use Register (often including polluting land-uses, and land contaminated by former land-uses)
- mangroves and saltmarshes
- dairy effluent consents
- LiDAR (data sets derived from).

4.4 Land-use classification

Having described the target land-use classification (ALUM, section 4.1), the technical method of making efficient combinations of input data (DGGS, section 4.2), and the specific input data (section 4.3), we now shift attention to the process of combining input data in a fashion that produces a land-use data set that adheres to the set of classes in ALUM. We begin with an identification of the analysis footprint (section 4.4.1), and then sections 4.4.2–4.4.7 consider each primary land-use class in turn.

Note that the nature of the exercise is that there is conflicting source information, and that an area of land may be captured under multiple diverse classes. The final class determination is made based on the sequential criteria of confidence, geographical scale, data recency, and, finally, a preferential hierarchy of classes (e.g. to prefer water in most circumstances).

The actual rules for defining a given location as any class are written in SQL. A brief attempt is hereby given in each case to capture our approach to mapping any given class. However, it is not always easy to succinctly express a complicated conditional logic statement as prose, especially when multiple input data sets are involved and there is a strict sequence of nested logical rules. Therefore, NRC is directed to supplementary materials where the actual SQL statements can be inspected. This is located alongside the delivered data. The written summaries here should be considered as an imperfect attempt to describe the nature of the actual applied logical rules. Some information may be omitted for the sake of brevity, or because it is considered to have a very minor influence on the classification outcome.

4.4.1 Analysis footprint

The chosen analysis footprint is identified from a specific data set, 'Regional Council 2023 (generalised)', sourced from the Statistics New Zealand Datafinder and licensed for reuse under CC BY 4.0.[18] The 'Northland Region' is extracted from this, which includes all of terrestrial Northland, the relevant territorial sea, and islands within this sea, but not Manawatāwhi / Three Kings Islands, which are identified by Stats NZ as being outside of any regional council.

[18] https://datafinder.stats.govt.nz/layer/111182-regional-council-2023-generalised/

4.4.2 Conservation and natural environments (1)

Because protected areas may overlie one another, giving additional levels of protection to land that is already protected under another instrument, priority is given to classes that are more protected (i.e. assignment to 1.1.1 takes precedence over 1.1.2, and so on). The 'comment' field may record one relevant name or other piece of information from the assigned class, but there may be other protected areas overlaying the same location. The LCDB v5 is used to also include the name of the assigned land-cover class, to enable queries for protected land of particular land-covers.

Class 1: Conservation and natural environments

This class includes land that has a relatively low level of human intervention. The land may be formally reserved by government for conservation purposes, or conserved through other legal or administrative arrangements. Areas may have multiple uses, but nature conservation is the prime use. Some land may be unused as a result of a deliberate decision of government or landowner, or due to circumstance.

Unlike the other classes in ALUM, only the tertiary classes of class 1 are used as the 'minimum level of attribution', so it is not possible to attribute classes 1.0.0, 1.1.0, 1.2.0, or 1.3.0 to any land.

Nature conservation (1.1)

For 'Nature conservation' (1.1), tertiary levels correspond to the International Union for Conservation of Nature (IUCN) 2008 categorisation. In general, we directly transformed features from LINZ Protected Areas to this class according to the relevant legislation and sections of that legislation. However, we also included QEII Covenants, Ngā Whenua Rāhui, and the Central Record of State Land (CRoSL).

Strict nature reserves (1.1.1)

This class corresponds to IUCN 2008 Category Ia. We classified as 1.1.1 all LINZ Protected Areas that come under the following pieces of legislation:

1. Reserves Act 1977:
   - s.20 Nature Reserve
   - s.21 Scientific Reserve
2. National Parks Act 1980, s.12 Specially Protected Area
3. Conservation Act 1987:
   - s.21 Ecological Area
   - s.22 Sanctuary Area
4. Marine Mammals Protection Act 1978, s.22 Marine Mammal Sanctuaries
5. Marine Reserves Act 1971, s.3 Marine Reserve

6 Wildlife Act 1953, s.9 Wildlife Sanctuary.

This class is also assigned when land is protected by the Reserves Act 1977 (s.22 Government Purpose Reserve, or s.23 Local Purpose Reserve) and the reserve has a listed 'purpose' of wildlife sanctuary, refuge, or habitat, or it meets the preceding criteria and its 'purpose' is not given but its title uses phrases such as 'bird sanctuary', 'wildlife sanctuary' or 'wildlife refuge'.

However, note that marine and other areas of water (class 6), following the ALUM guidelines, are ultimately classed as 6.•.1: an area of lake, river, marsh/wetland, or estuary/coastal water (6.1. 6.2, etc.) with a conservation influence (6.2.1, 6.2.1, etc.)

Wilderness areas (1.1.2)

This class corresponds to IUCN 2008 Category Ia. We classified as 1.1.2 all LINZ Protected Areas that come under the following pieces of legislation:

1 National Parks Act 1980, s.14 Wilderness Area

2 Conservation Act 1987, s.20 Wilderness Area.

None of these areas are located in Northland.

National Park (1.1.3)

This class corresponds to IUCN 2008 Category II. We classified as 1.1.3 all LINZ Protected Areas that come under the following pieces of legislation:

1 National Parks Act 1980, s.4 National Park

2 Conservation Act 1987:

- s.18 Specially Protected Area
- s.19 Conservation Park

3 Reserves Act 1977, s.13 National Reserve.

Natural feature protection (1.1.4)

This class corresponds to IUCN 2008 Category III. We classified as 1.1.4 all LINZ Protected Areas that come under the following pieces of legislation:

1 Reserves Act 1977:

- s.18 Historic Reserve
- s.19(1)(a) Scenic Reserve
- s.19(1)(b) Scenic Reserve.

We also included any land in the CRoSL with a name or statutory action matching the phrase 'historic reserve', or any land managed by Heritage New Zealand. However, we mapped these with lower confidence due to the use of text search.

Habitat/species management area (1.1.5)

This class corresponds to IUCN 2008 Category IV. We classified as 1.1.5 all LINZ Protected Areas that come under the following pieces of legislation:

1 Conservation Act 1987:

- s.25 Stewardship Area
- s.23B Wildlife Area

2 Wildlife Act 1953:

- s.14 Wildlife Refuge
- s.14A Wildlife Management Reserve

3 Reserves Act 1977:

- s.22 Government Purpose Reserve
- s.23 Local Purpose Reserve.

In the case of land under the Reserves Act 1977, we added the constraint that it must have a reserve purpose of one of the following:

- ecological management
- wildlife management
- environmental and landscape.

In the case where the reserve purpose is not given, we included a reserve if it had a name that included the phrase 'wildlife management'.

In this class we also included all reserves captured by QEII and Ngā Whenua Rāhui covenants, plus land identified by the CRoSL as managed by Fish and Game with a name or purpose matching the phrase 'wildlife management reserve'.

Finally, we specifically included the marine parts of the Whangārei Harbour Wildlife Refuge.

Protected landscape (1.1.6)

This class corresponds to IUCN 2008 Category V. We classified as 1.1.6 all LINZ Protected Areas that come under the following pieces of legislation:

1 Reserves Act 1977:

- s.17 Recreation Reserve
- s.23 Local Purpose Reserve[19, 20]

[19] But only if the reserve purpose is given as passive recreation, recreation, or walkway.

[20] Or, if the purpose is not given, and the other criteria are met, if the name of the reserve includes the phrase 'recreation'.

- s.22 Government Purpose Reserve.[16]

Other conserved area (1.1.7)

We classified as 1.1.7 all LINZ Protected Areas that come under the following pieces of legislation:

1 Conservation Act 1987:
   - s.7 Conservation Purposes
   - s.23A Amenity Area
2 National Parks Act 1980, s.15 Amenities Area
3 Reserves Act 1977:
   - s.16(11) Recreation Reserve (Racecourse)
   - s.22 Government Purpose Reserve[21]
   - s.23 Local Purpose Reserve.[18]

Managed resource protection (1.2)

Biodiversity (1.2.1)

We classified as 1.2.1 all LINZ Protected Areas that come under the following pieces of legislation:

- Fisheries Act 1983, cl.68 Freshwater Fisheries Regulations – Faunistic Reserves
- Wild Animal Control Act 1977, s.27 Recreational Hunting Area
- Reserves Act 1977, s.22 Government Purpose Reserve, if the purpose is for fisheries management.

Surface water supply (1.2.2)

We classified as 1.2.2 all LINZ Protected Areas that had water supply or reservoir as the stated reserve purpose. We also included land recorded in the CRoSL that had a name or statutory action that matched a search for 'water conservation' or 'water supply'. This was included at a lower confidence. The CRoSL includes territorial and regional government water conservation and water supply reserves.

Groundwater (1.2.3)

We did not consider there were sufficient data from any source to map to this class. Although there are places within Northland that are facing pressures on aquifers, we are

---

[21] But only if the reserve purpose is given and is considered amenable to this class (e.g. an amenity, green belt, landing site, general government reserve).

not aware of any land that is specifically and primarily used to manage an aquifer and that does not have another primary use (such as commercial forestry).

Landscape (1.2.4)

We classified as 1.2.4 all LINZ Protected Areas that come under the following pieces of legislation:

- Waitangi Endowment Act 1932, s.2 Local Purpose Reserve
- Reserves Act 1977, s.12 Local Purpose Reserve, if the purpose is given as 'landscape'
- Reserves Act 1977, s.22 Government Purpose Reserve, or s.23 Local Purpose Reserve, if the purpose is in the nature of coastal protection, estuarine buffer, riverbank protection, soil conservation, etc; or if the purpose is not specified but the name of the reserve matches the phrases 'soil conservation' or 'river control'.

Traditional indigenous uses (1.2.5)

We did not consider there were sufficient public data to adequately map this class in any extensive way. Also, its definition reflects Australian indigenous cultures and not the Māori world view. As noted above, ALUM does not reflect Treaty principles or enable a Treaty-led science-policy approach to land-use classification in New Zealand.

However, we did include LINZ Protected Areas under Reserves Act 1977 (s.22 Government Purpose Reserve or s.23 Local Purpose Reserve) if the stated reserve purpose was a Māori burial ground or urupā. We considered these could also be mapped as class 5.5.2 ('Public services'), which includes 'cemeteries and crematoria' in its definition, but that the specifically indigenous nature of urupā could have been lost if we had done so. A more appropriate New Zealand-specific classification system could allow for the inclusion of both burial site types.

We also included marae reserves (from LINZ Protected Areas) and land identified from the CRoSL meeting a similar definition. These may also fit under 5.5.2 ('Public services') or 5.5.3 ('Recreation and culture'), but again we decided to include them in 1.2.5 in an attempt to capture their specifically indigenous character. A future land-use classification developed specifically for use in New Zealand may better capture land that we have assigned to this class under ALUM. It may also include information about land *tenure*, which the present map does not.

Other minimal use (1.3)

This secondary class (which is not a minimum level of attribution) includes areas of land that do not have a primary use, but rather have ancillary uses.

Defence land – natural areas (1.3.1)

We classified as 1.3.1 all LINZ Protected Areas that had a reserve purpose of 'defence area' or 'reserves base'. We also included areas identified in the CRoSL with a name or statutory actions related to defence (except defence housing). Finally, we included land identified from the NRC Selected Land Use Register as being training areas set aside primarily or exclusively for the detonation of explosive ammunition.

Stock route (1.3.2)

We did not consider there were sufficient data from any source to map to this class.

Residual native cover (1.3.3)

This class attempts to capture land that is under native cover and is unused or has an indeterminate use. However, we made the decision to include areas of vegetation under this class when they coincide with specific land-uses (classes 3.2.0, 3.2.4, 3.6.0, 4.2.0, 4.2.3, 4.6.0, or 5.4.3), and when the canopy is sufficiently tall (canopy height above 1.5 m). We do not know if such vegetation is native, but we do consider it residual when co-located with these classes. The justification for doing this is that there is no other adequate class in which to include such areas of residual vegetation within farms (or lifestyle properties), and because NRC specifically mentioned that residual vegetative cover is important to identify.

In existing ALUM applications the target spatial unit is larger than most small areas of residual vegetation and would not be mapped to this class. We considered that since we were using an unsupervised semi-automated approach, and that we had access to a canopy height model produced from recent LiDAR, we would include any DGGS cell that met this criterion under this class. However, we are unable to ascertain whether livestock are excluded from any such areas.

Other land that may be mapped to this class includes areas identified by the Land Cover database (LCDB) as sand or gravel, landslide, gravel or rock, depleted grassland, flaxland, fernland, gorse or broom, mānuka or kānuka, broad-leaved indigenous hardwoods, (exotic) deciduous hardwoods, or indigenous forest. Also, residual reserves identified from the CRoSL are captured here. However, in both cases we use a lower confidence in the assigned class if the canopy height model indicates that the canopy is less than or equal to 1.5 m.

Rehabilitation (1.3.4)

In this class we included land areas from LINZ Protected Areas noted as quarry reserves (i.e. former quarries, many of which are quite old), or if the relevant legislation is the Reserves Act 1977 (s.22 Government Purpose Reserve or s.23 Local Purpose Reserve) and the reserve purpose is afforestation, beautification, or plantation.

Land areas from the CRoSL are included in cases where the recorded statutory actions relate to the phrases 'quarry', 'stream improvement' or 'harbour improvement'. We assign a lower confidence in such cases due to the use of mostly unsupervised text search.

4.4.3 Production from relatively natural environments (2)

Class 2: Production from relatively natural environments

This class includes land that is subject to relatively low levels of intervention. The land can not be used more intensively because of its limited capability. The structure of the native vegetation generally remains intact despite deliberate use, although the floristics of the vegetation may have changed markedly. Where the native vegetation structure is, for example, open woodland or grassland, the land may be grazed.

Where native grasses have been deliberately and extensively replaced with improved species, the use should be treated under class 3, 'Production from dryland agriculture and plantations'.

We did not consider there were sufficient data from any source to map to this class, at any level from the primary to the tertiary. We note that this category, more than any other in ALUM, is more relevant in the Australian context than in Northland.

4.4.4 Production from dryland agriculture and plantations (3)

Class 3: Production from dryland agriculture and plantations

This class includes land that is used principally for primary production, based on dryland farming systems. Native vegetation has largely been replaced by introduced species through clearing, the sowing of new species, the application of fertilisers or the dominance of volunteer species. The range of activities in this category includes plantation forests, pasture production for stock, cropping and fodder production, and a wide range of horticultural production. If there is evidence of irrigation infrastructure, land should be mapped under class 4, 'Production from irrigated agriculture and plantations', even if irrigation water has not been applied in the current growing season.

Fallow or ploughed land should be assigned to the most likely land use based on the dominant activity conducted in comparable nearby areas. Fallow or ploughed land should be allocated to the relevant pasture, cropping or horticulture class (rather than using land in transition). Record the fallow or ploughed status in the management field.

Plantation forestry (3.1)

This class includes pre-1990 planted forest and post-1989 forest from LUCAS LUM, but only where it is also identified as being composed of unspecified exotic species. Where LUCAS LUM identifies the species (*Pinus radiata* or Douglas fir), land is classified as 3.1.2.

Pre-1990 planted forest with an unknown species is included in this class too, since, as a secondary class, it is intended to accommodate ambiguous cases that cannot be determined at the tertiary level. However, where these cases coincide with parcels identified as having a forestry land-use in the NRC ratings database, and the LiDAR canopy height model indicates the presence of any vegetation, we assign a confidence of 3 rather than 4, due to the added corroboration.

We include parcels from the ratings database if they have a forest land-use, but not if they are indicated as being a protected forest of any kind. We improve confidence in cases where the canopy height model offers corroborating evidence of vegetation.

We integrate (with low confidence) land areas identified by the CRoSL if the statutory actions match the phrases 'state forest' or 'crown forest', but not if they are also noted as being for conservation. We improve confidence in cases where the canopy height model offers corroborating evidence of vegetation.

Hardwood plantation forestry (3.1.1)

The two species of exotic forestry identified in LUCAS LUM are *Pinus radiata* and Douglas fir, both of which are softwood plantation species. We are aware of commercial blue gum (*Eucalyptus globulus*) plantations in Southland and Bay of Plenty, but not Northland. Therefore, we did not include this class.

Softwood plantation forestry (3.1.2)

We base the identification of this class on the pre-1990 planted forests, and post-1989 forests in LUCAS LUM, when the species type is identified (either *Pinus radiata* or Douglas fir). However, we also require the NRC ratings database to identify an area of land as being used for exotic forestry before giving a positive identification.

Other forest plantation (3.1.3)

We did not consider there were sufficient data from any source to map to this class.

Environmental forest plantation (3.1.4)

This class is defined as:

> *area managed for environmental and indirect production uses (e.g. prevention of land degradation, windbreaks, shade and shelter). This also includes trees planted for carbon credits.*

We do not have data on whether any forest is planted for the purposes of obtaining carbon credits. However, we did use this class extensively to represent sub-areas on farms that are forested. This captures windbreaks, riparian vegetation, regenerating native bush, and possibly silvopastural practices and agroforestry.

This class is treated as a *special case* and is only applied on top of land that is penultimately determined to be one of the following classes:

- Grazing modified pastures (3.2, or 4.2, irrigated)
- Pasture legume or grass mixtures (3.2.4, or 4.2.3, irrigated)
- Land in transition (3.6, or 4.6, irrigated)
- No defined use (3.6.4, or 4.6.4, irrigated)
- Urban residential without agriculture (5.4.3).

The use of this special case is to allow us to define much forested land broadly within this class (as an intermediate identification), but then only apply it narrowly in cases where we consider it likely to be used for environmental purposes, or as an adjunct to pastoralism (e.g. as a windbreak).

This class may also include areas of gorse and/or Scotch broom, mixed exotic shrubland, and deciduous (exotic) hardwoods (all as identified in the LCDB), noting that the 3.1 category does not indicate whether the trees or shrubs are native or exotic.

Within the map this class tends to have a high level of spatial detail because it also includes areas of canopy greater than 1.5 m, as identified in the LiDAR-derived canopy height model, leading to the inclusion of small patches of shrubs that are far beyond the mapping scale of existing national data sets (LCDB and LUCAS LUM).

Grazing modified pastures (3.2)

The primary identification of this class comes from the ratings database land-uses:

- 11 (Rural industry: Dairy)
- 12 (Rural industry: Stock finishing)
- 14 (Rural industry: Store livestock)
- 16 (Rural industry: Specialist livestock).

We include the canopy height model (CHM), height above nearest drainage (HAND) and downslope distance to stream (DDTS) at this point to add an additional and specific comment, 'Riparian vegetation', when the canopy height is above 1 m, the cell is vertically proximate to a water channel (2 m or less), and the cell is horizontally proximate to a water channel (30 m or less). (See Figure 7 for a visual explanation of this.) Note that the identification is, in practice, only made on land classed as 3.2 or 4.2, because the objective is to determine riparian vegetation only when it is an adjunct to pastoral land-uses.

We also include commodity information ('cattle dairy') when this is known from available sources, which include dairy effluent consents. 'Specialist livestock' (as identified from the ratings database) is not specific enough to include anything in the commodities field, but is included as a comment, where relevant, as is 'Store livestock' and 'Stock finishing'. Most grazing land that is left without a commodity is probably sheep & beef farming (i.e. it would be some combination of the ALUM commodities cattle, cattle meat, cattle stud, sheep, sheep meat, sheep stud, or sheep wool), but this information was not available.

Other commodities for specialist livestock would have been included if it were possible to do so; these could still be edited into the map by council staff (we recommend using the controlled ALUM vocabulary, and separating commodities with commas, to retain interoperability). Note that there is an exclusion area within Northland where deer may not be farmed, but this is not recorded in this land-use map.[22]

Irrigated land area (from MfE, dated 2020) is used to overlay this land and class it as 4.2 rather than 3.2. The irrigation type is recorded as a management practice.

A special comment is added to land that is classed as 3.2.0 to attempt to further disaggregate productive pasture into the classes used by Beef + Lamb New Zealand.[23] The three North Island classes, and their conditions for classification, are:

- North Island finishing – when LENZ level 4 slope is less than 3, and NZLRI CCAV[24] is 15 livestock units or above
- North Island hard hill country – when LENZ level 4 slope is greater than 7 and NZLRI CCAV is less than 10 livestock units
- North Island hill country – all other cases.

This sub-classification is not explicitly applied as a quaternary level to the hierarchical ALUM classification, although it could be loosely treated as such, noting, however, that this classification is made for all land classed as 3.2 (or 4.2) and is not limited to sheep & beef production systems. (Even known dairy farm areas have this comment.) This is because that information is not available, hence the absence of 'sheep meat' and 'cattle meat' as commodities. It is also useful to retain this on dairy farms because they may not remain as dairy farms, whereas this information is based on static physical attributes and does not change.

[22] https://www.fedsnews.co.nz/northland-deer-farmers-urged-to-register-with-doc/

[23] https://beeflambnz.com/data-tools/farm-classes

[24] Stock carrying capacity, expressed in sheep-per-hectare (breeding ewe-equivalents), estimated average for all farmers

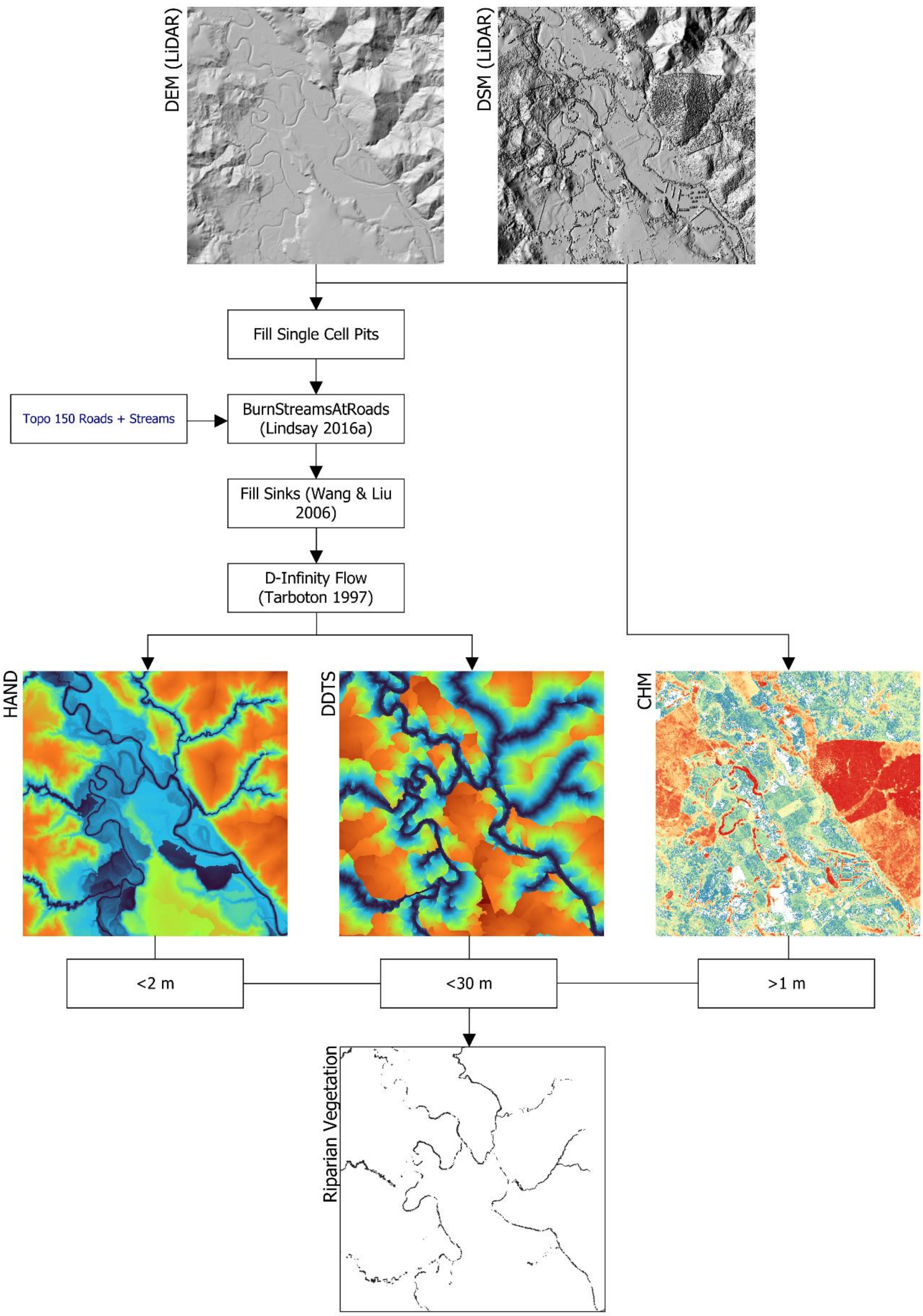


Figure 7. A visual summary of the identification of riparian vegetation.
Notes: DEM = digital elevation model; DSM = digital surface model; CHM = canopy height model; HAND = height above nearest drainage; DDTS = downslope distance to stream.
Sources: Lindsay 2016a; Tarboton 1997; Wang & Liu 2006.

The classes are included under the comments field of the data, and an analyst would need to perform text matching to filter for these classes. Note that as a comment it may be either the only remark or one of multiple remarks (e.g. with 'Riparian vegetation'), so text matching (rather than whole-text equality) is the appropriate way to filter for these land areas. Comments are separated with a comma and a single white-space character.

Native/exotic pasture mosaic (3.2.1)

We did not consider there were sufficient data from any source to map to this class.

Woody fodder plants (3.2.2)

We did not consider this class relevant in the New Zealand context.

Pasture legumes (3.2.3)

We did not consider there were sufficient data from any source to map to this class.

Pasture legume/grass mixtures (3.2.4)

We did not consider there were sufficient data from any source to map to this class.

Sown grasses (3.2.5)

We did not consider there were sufficient data from any source to map to this class.

Cropping (3.3)

Our primary definition of this class is the ratings database, class 13 ('Arable farming'), but also LCDB, class 30 ('Short-rotation cropland'). Irrigation[25] is included and determines whether an area of land is identified as 3.3 or 4.3. The irrigation type is recorded as a management practice.

We did not consider the tertiary classes under class 3.3 relevant in the Northland context: 'Cereals' (3.3.1), 'Beverage and spice crops' (3.3.2), 'Hay and silage' (3.3.3), and 'Oilseeds' (3.3.3), 'Cotton' (3.3.6), 'Alkaloid poppies' (3.3.7), Pulses (3.3.8).

Sugar (3.3.5)

Although we have been made aware (Robbie Price, Landcare Research senior GIS technician, pers. comm., 2023) that there is (or was) a trial sugar cane field in Northland near Whangārei, this class was not mapped due to insufficient data.

[25] https://landcareresearch.sharepoint.com/sites/sp00208/dataandanalyis/data.mfe.govt.nz

Perennial horticulture (3.4)

Our primary definition of this class is LUCAS LUM, class 77 ('Perennial cropland'), which includes all orchards and vineyards. We also integrated ratings data, land-use class 15 ('Market gardens and orchards'). Irrigation[26] is included and determines whether an area of land is identified as 3.4 or 4.4.

Tree fruits (3.4.1)

We only had remote-sensing data available for identifying avocado, citrus, and kiwifruit, and because citrus is captured under class 3.4.8 and kiwifruit is a vine fruit (class 3.4.4), all land classed as tree fruit with high confidence is avocado orchard. This is specifically captured in the commodity field. Confidence is considered high when ratings data, LUCAS LUM, and winter land cover all concur.

A lower confidence classification under 3.4.1 is made when ratings data indicate that land is horticultural and suitable for stone or pip fruit but there is no recorded commodity.

Olives (3.4.2)

Using publicly available data from the regional olive growers' association,[27] olives were included, with the 'olive' commodity also noted.

Tree nuts (3.4.3)

We did not consider there were sufficient data from any source to map to this class in an unsupervised way, although we were made aware of one specific macadamia orchard, so this was specifically included, with the 'macadamia' commodity noted.

Vine fruits (3.4.4)

All land classed as vine fruit with high confidence is kiwifruit orchard. This is specifically captured in the commodity field. Confidence is considered high when ratings data, LUCAS LUM, and winter land cover concur. This information comes from MWLR 2022 winter land cover mapping using Sentinel-2 imagery and ground-truthing from Bay of Plenty.

The same source offers a classification of bird netting as a distinct cover. The presence of bird netting is always taken to indicate the presence of kiwifruit orchards, and this is captured in the management practices (semi-protected rather than open growing) and in the commodity.

[26] https://landcareresearch.sharepoint.com/sites/sp00208/dataandanalyis/data.mfe.govt.nz
[27] https://oliveti.co.nz/about-oliveti/

A lower confidence classification under 3.4.4 is made when ratings data indicate that land is horticultural and suitable for kiwifruit, but without agreement from winter land-cover mapping (as this layer does have known false negatives).

Shrub berries and fruits (3.4.5)

Ratings data are used to identify land-use with a potential use as horticulture (berries). Specifically, in the absence of other information for identifying berries, we use the ratings data 'Property category code' (Office of the Valuer-General 2010, Appendix F). This code gives a 'property category that broadly describes the nature of the property'. This is assigned by valuers on the basis of the 'highest and best use, or the use for which the property would be sold given the economic conditions prevailing at the effective date of valuation'.

Although this may differ from the actual property use, in practice when narrow categories such as this are assigned it appears to be because the existing property use (i.e. for commercially growing berries) also appears to naturally be the 'use for which the property would be sold'. We presume this might be because of existing capital investments, location, and the nature of the soil.

When the berry plantations identified from ratings data coincide with irrigated land, or LUCAS LUM (class 77, perennial cropland), we assign this a higher confidence than if the only evidence is ratings data.

Perennial flowers and bulbs (3.4.6)

This is identified identically to class 3.4.5, except that we consider commercial flower growing as identified by ratings data.

There is potential for confusion between this class and class 3.5.2 (seasonal flowers and bulbs). The available data do not allow us to distinguish between perennial and seasonal uses within this class. We have chosen to leave 3.5.2 unassigned.

Perennial vegetables and herbs (3.4.7)

We did not consider there were sufficient data from any source to map to this class.

Citrus (3.4.8)

A high-confidence classification under 3.4.8 is made when ratings data, LUCAS LUM, and winter land cover all concur about the presence of citrus. A lower confidence classification is made when ratings data indicate that land is horticultural and suitable for citrus, but without agreement from winter land-cover mapping (as this layer does have known false negatives).

Grapes (3.4.9)

Land identified from the ratings database with a potential best use (category code) of vineyard is assigned to this class. However, if the land is zoned multi-use within rural industry, it requires corroboration from LCDB (class 33, 'Orchard, vineyard or other perennial crop') in order to be assigned to this class. There are a limited number of vineyards in Northland, which meant it was feasible to inspect the application of these criteria, and we were satisfied that in all cases vineyards were captured (comparing against recent aerial images and Street View images). However, this process gives no scope for finding vineyards not identified in the ratings database. The 2022 winter land-cover map did not detect any vineyards in Northland (false negatives) and was not considered when making this classification.

Seasonal horticulture (3.5)

The definition of this class includes data from the ratings database (land-use 13, 'arable farming') and LCDB (class 30, 'Short-rotation cropland'). We assign a higher confidence when these two sources of information aligned, and a lower confidence otherwise.

Seasonal fruits (3.5.1)

We did not consider there were sufficient data from any source to map to this class.

Seasonal flowers and bulbs (3.5.2)

This class is left unassigned: see the definition of class 3.4.6 ('Perennial flowers and bulbs').

Seasonal vegetables and herbs (3.5.3)

We received a list of 22 addresses from Warwick Simpson, Vegetables NZ director and a Northland kūmara grower. His full list of 40 addresses was filtered (by him) to exclude those with addresses not considered usable for geocoding. Of the 22 addresses provided, we made confident matches on 20 using the LINZ address database. This was performed manually given the small number of locations. All of the 20 addresses were on land that had otherwise been classified as 3.5.0, and we therefore further classified the land as 3.5.3 (or 4.5.3 with evidence of irrigation) and added 'kumara' to the commodity field.

Land in transition (3.6)

This class is intended to capture areas where the land-use is 'unknown and cannot reasonably be inferred from the surrounding land-use'. The documentation for ALUM Version 8 recommends using this class 'sparingly'. However, there are data with which to positively identify vacant or transitioning land, which we chose to map to this class. Specifically, we used the vacant land-use classes in the ratings database to indicate land in transition, supplemented by data from OSM on current construction sites. A comment is included, where possible, to indicate whether the vacant land is residential, commercial, industrial, etc.

Degraded land (3.6.1)

We did not consider there were sufficient data from any source to map to this class.

Abandoned land (3.6.2)

We did not consider there were sufficient data from any source to map to this class.

Land under rehabilitation (3.6.3)

We did not consider there were sufficient data from any source to map to this class.

No defined use (3.6.4)

Land is assigned to this class when it cannot be assigned to any other class, and when the canopy height is 0. This is because the class definition states that this represents land that is cleared of 'intact native vegetation'.

Where land is unable to be assigned to any class, and it *does* have a canopy (of any height), it is assigned to the special 'unknown' class, 0.

Abandoned perennial horticulture (3.6.5)

We did not consider there were sufficient data from any source to map to this class.

4.4.5 Production from irrigated agriculture and plantations (4)

Class 4: Production from irrigated agriculture and plantations

This class includes agricultural land uses where water is applied to promote additional growth over normally dry periods, depending on the season, water availability and commodity prices.

This includes land uses that receive only one or two irrigations per year, through to those uses that rely on irrigation for much of the growing season. Baxter and Russell (1994) created this primary class because of the degree of intervention involved in irrigation, and its potential impacts on hydrology and geohydrology.

If there is no evidence of irrigation (infrastructure or active irrigation) then the area should be mapped to the appropriate dryland class.

The distinction between class 3 and class 4 land is more marked in Australia than in Northland. This distinction is not as critical in New Zealand, where land may not be routinely irrigated because it does not typically have a soil moisture deficit. We based our decision on whether land is irrigated entirely on data published by MfE. For many purposes, ALUM classes 3 and 4 should be considered collectively in the New Zealand context. Almost all classes in 3 have a corresponding class in 4.

Irrigation methods are included in the 'manage_desc' field (management practices), according to the concordance presented in Table 3.

Table 3. Concordance between Ministry for the Environment (MfE) irrigation types and ALUM management practices

ALUM management practice	MfE irrigation type
irrigation surface	Border dyke/Borderdyke Wild flooding
irrigation spray	Gun K-line/Long lateral Lateral Linear boom Pivot Rotorainer Side roll Solid-set
irrigation drip	Drip/micro

Irrigated plantation forests (4.1)

Class 4.1 is not included in the land-use map. We did not consider that plantation forests are irrigated in New Zealand, and did not include the MfE irrigation data in the application of class 3.1, or in the corresponding tertiary classes. This is the exception to the rule for class 4, where MfE irrigation data are the determining factor for whether an area of land is class 3 or class 4.

Grazing irrigated modified pastures (4.2)

See class 3.2.

Irrigated woody fodder plants (4.2.1)

See class 3.2.2.

Irrigated pasture legumes (4.2.2)

See class 3.2.3.

Irrigated legume/grass mixtures (4.2.3)

See class 3.2.4.

Irrigated sown grasses (4.2.4)

See class 3.2.5.

Irrigated cropping (4.3)

See class 3.3.

Irrigated perennial horticulture (4.4)

See class 3.4.

The irrigation type is recorded as a management practice. The very co-presence of drip/micro irrigation is considered a factor in improving the identification of land as one of the tertiary classes under 4.4.

Tree fruits (3.4.1)

See class 3.4.1.

Olives (3.4.2)

See class 3.4.2.

Tree nuts (3.4.3)

See class 3.4.3.

Vine fruits (3.4.4)

See class 3.4.4.

Irrigated shrub berries and fruits (4.4.5)

See class 3.4.5.

Irrigated perennial flowers and bulbs (4.4.6)

See class 3.4.6.

Irrigated perennial vegetables and herbs (4.4.7)

See class 3.4.7.

Irrigated citrus (4.4.8)

See class 3.4.8.

Irrigated grapes (4.4.9)

See class 3.4.9.

Irrigated seasonal horticulture (4.5)

See class 3.5.

Irrigated turf farming (4.5.4)

We did not consider there were sufficient data from any source to map to this class (which does not have a class 3 equivalent under ALUM).

Irrigated land in transition (4.6)

See class 3.6.

4.4.6 Intensive uses (5)

Class 5: Intensive uses

This class includes land uses that involve high levels of interference with natural processes, generally in association with closer settlement.

The level of intervention may be high enough to completely remodel the natural landscape—the vegetation, surface and groundwater systems, and the land surface. If crop and or animal type is known, record this in the commodity field.

Intensive horticulture (5.1)

There is less available information about the most intensive forms of horticulture compared to extensive horticulture. Intensive horticulture includes plant production systems with special purpose improvements, such as glasshouses.

For this secondary class we attempt a broad classification by identifying ratings data with a land-use code with the suffix `HG`, which indicates horticultural land with glasshouses. However, this is not a reliable indicator on its own. Therefore, we also include a consideration of the NRC Selected Land Use Register, filtering for land recorded as having bulk storage or use of persistent pesticides. However, this is again at best an indirect indicator, and the classification system requires that either LCDB or LUCAS LUM add a corroborating classification into a broad class for orchards, vineyards, or perennial crops (LCDB class 33; LUCAS LUM class 77).

Production nurseries (5.1.1)

We did not consider there were sufficient data from any source to map to this class.

Shadehouses (5.1.2)

Shadehouses are captured at the ratings unit scale based solely on ratings data, where all the following conditions must be met:

- land-use type 15 (prefix 1, rural industry; suffix 5, market gardens and orchards)
- land-use category type prefix `HG` (horticulture: greenhouse)
- improvements description must contain the description `SHADE HSE` (shadehouse).

Glasshouses (5.1.3)

See class 5.1.2, but searching ratings data improvements descriptions for `GREEN HSE` rather than `SHADE HSE`.

Glasshouses — hydroponic (5.1.4)

We did not consider there were sufficient data from any source to map to this class.

Abandoned intensive horticulture (5.1.5)

We did not consider there were sufficient data from any source to map to this class.

Intensive animal production (5.2)

This secondary class is not intended to capture land-used for grazing. Most such features are smaller than the minimum mapping unit or scale of the available input data or are otherwise not identified or identifiable.

We did, however, include land identified in the NRC Selected Land Use data set as having a livestock dip or spray-race operations.

Dairy sheds and yards (5.2.1)

NRC provided a point geometry layer of dairy effluent consents to assist in the identification of dairy farms. (This forms a strong basis for assigning 'cattle dairy' as a commodity on relevant class 3 land.) The location of this point was not merely an address or an arbitrary point on the relevant farm, but rather the location of the milking yard. We did not consider it feasible to then delineate the boundary of the yard itself (as distinct from nearby spaces such as pasture or vehicle access), but instead used an existing data set of buildings identified from LiDAR, and joined all buildings within 50 m of this point. Visual inspection indicated this was a safe threshold because dairy yards tend to be isolated features, but it probably left as unclassified some buildings that are functionally part of the same complex. Buildings captured by this are also assigned a commodity of 'cattle dairy', as per ALUM recommendation.

Feedlots (5.2.2)

We did not consider there were sufficient data from any source to map to this class.

Poultry farms (5.2.3)

We did not consider there were sufficient data from any source to map to this class.

Piggeries (5.2.4)

We did not consider there were sufficient data from any source to map to this class.

Aquaculture (5.2.5)

To map to this class, we directly relied on a layer representing current marine farms, produced by MfE. This layer includes an indicator of species that was transformed into an ALUM commodity (i.e. 'crustaceans', 'finfish', or 'molluscs').

Horse studs (5.2.6)

We did not consider there were sufficient data from any source to map to this class.

Saleyards/stockyards (5.2.7)

We did not consider there were sufficient data from any source to map to this class.

Abandoned intensive animal production (5.2.8)

We did not consider there were sufficient data from any source to map to this class.

Manufacturing and industrial (5.3)

Land is assigned to this secondary class based on combining the following layers:

- NRC Selected Land Use Register, with the following Hazardous Activities and Industries List (HAIL) categories:
    - a A6: Fertiliser manufacture or bulk storage
    - b A12: Pesticide manufacture (including animal poisons, insecticides, fungicides or herbicides) including the commercial manufacturing, blending, mixing or formulating of pesticides
    - c A15: Printing including commercial printing using metal type, inks, dyes, or solvents (excluding photocopy shops)
    - d A16: Skin or wool processing including a tannery or fellmongery, or any other commercial facility for hide curing, drying, scouring or finishing or storing wool or leather products

   e A17: Storage tanks or drums for fuel, chemicals or liquid waste
   f B2: Electrical transformers including the manufacturing, repairing or disposing of electrical transformers or other heavy electrical equipment
   g C1: Explosive or ordinance production, maintenance, dismantling, disposal, bulk storage or re-packaging

- ratings data (any ratings unit with a land-use type in the range 70–79 (all industrial land, including vacant).
- LCDB (class 1, 'Built-up area (settlement)')
- OSM industrial areas.

Land assigned industrial in the ratings database is included first; a higher confidence is assigned when the land is not vacant and is also recognised by LCDB as 'built-up area'.

Land from the NRC Selected Land Use Register is considered next. The highest confidence is assigned in only one specific case: where there is no more than one assigned HAIL category (and it must be A17), and it is noted as being 'current and verified'.

Finally, OSM industrial land-uses are included. This late inclusion is intentional, such that the classification isn't over-reliant on OSM data, but they are still present because OSM offers greater spatial precision than the other sources of generic industrial land.

General purpose factory (5.3.1)

The definition of a general purpose factory is based solely on data from the NRC Selected Land Use Register, classes:

- A1: Agrichemicals including spray contractors' commercial premises
- A2: Chemical manufacture, formulation or bulk storage
- A4: Corrosives including formulation or bulk storage
- A9: Paint manufacture or formulation (excluding retail paint stores)
- A14: Pharmaceutical manufacture including the commercial manufacture, blending, mixing or formulation of pharmaceuticals, including animal remedies or the manufacturing of illicit drugs with the potential for environmental discharges
- B1: Batteries including the commercial assembling, disassembling, manufacturing or recycling of batteries (but excluding retail battery stores)
- B3: Electronics including the commercial manufacturing, reconditioning or recycling of computers, televisions and other electronic devices
- D1: Abrasive blasting including cleaning and disposal
- D2: Foundry operations including the commercial production of metal products by injecting or pouring molten metal into moulds
- D3: Metal treatment or coating including polishing, anodising, galvanising, pickling, electroplating, or heat treatment or finishing using cyanide compounds
- D4: Metalliferous ore processing including the chemical or physical extraction of metals, including smelting, refining, fusing or refining metals

- D5: Engineering workshops with metal fabrication
- E2: Asphalt or bitumen manufacture or bulk storage
- E3: Cement or lime manufacture
- E4: Commercial concrete manufacture or cement storage
- F2: Brake lining manufacturers, repairers or recyclers
- F3: Engine reconditioning workshops
- F4: Motor vehicle workshops
- F5: Port activities including dry docks or marine vessel maintenance facilities
- G2: Drum or tank reconditioning or recycling.

Confidence is highest (value of 2) when there is only one HAIL category. Confidence increases to a value of 3 when there are two applicable classes, and is assigned 4 when there are three or more HAIL classes. The count of HAIL classes is based on the full set of categories, not just the subset above.

Finally, there is a condition on the size of the property. When a property meeting the preceding criteria is considered 'large', then it is assigned to a different class (5.3.3, 'Major industrial complex'). If it is not 'large', it is assigned to this class, 5.3.1. 'Large' in this context is defined as property over 1.4 ha in size.

Food processing factory (5.3.2)

We did not consider there were sufficient data from any source to map to this class.

Major industrial complex (5.3.3)

See class 5.3.1.

Bulk grain storage (5.3.4)

We did not consider there were sufficient data from any source to map to this class.

Abattoirs (5.3.5)

We did not consider there were sufficient data from any source to map to this class.

Oil refinery (5.3.6)

The Marsden Point oil refinery, although it very recently closed, is identified by this class. Its boundary is taken directly from OSM, as other sources of industrial land did not include enough attribute information to distinguish the oil refinery from other industrial spaces.

Sawmill (5.3.7)

This class is mapped directly from the NRC Selected Land Use Register, HAIL class A18 ('Wood treatment or preservation including the commercial use of anti-sapstain chemicals during milling, or bulk storage of treated timber outside'). This is assigned the highest confidence where this is the only applicable HAIL class, and it is both current and verified.

Abandoned manufacturing and industrial (5.3.8)

We did not consider there to be sufficient data from any source to map to this class.

Residential and farm infrastructure (5.4)

This secondary class is not at the minimum level of attribution within ALUM, so land under this class may only be classified at the tertiary level (classes 5.4.1–5.4.5).

Urban residential (5.4.1)

The definition of this class begins with an exclusion: properties considered part of a small urban area according to the 2023 Statistics NZ definition, if they are nevertheless greater than 0.7 ha in size, are not classified as 5.4.1.

Following this, remaining land may be classified as 5.4.1 if it is assigned a residential or lifestyle class in the ratings database (classes 20–29, or 90–99, under the LINZS30300 standard for valuation data). A lower confidence is assigned if the same source notes that the land is vacant, or its category code indicates bare land or parking. However, the property must also be within a 'major', 'large', 'medium', or 'small' urban area (but not a 'rural settlement') according to Statistics NZ. It must also be on a property in which the ratings database includes an improvements description that indicates the presence of something like a dwelling, flat, apartment, bach, boarding house, incomplete dwelling, motel, or hotel.

Land may also be assigned 5.4.1 if it is urban and is managed by Housing New Zealand according to the CRoSL. This is an important inclusion, because Housing New Zealand land is not included in the ratings database.

Rural residential with agriculture (5.4.2)

Although it is possible to identify rural residential (often referred to as 'lifestyle') properties using existing data, it is more difficult to discern whether agriculture is or is not occurring on such a property, especially given that this class is intended to capture agriculture at 'sub-commercial or hobby' scales (i.e. agriculture is not the primary form of income for the household).

In the absence of this information, and the inability to use class 5.4 (below, the minimum level of attribution), we opted to assign all rural residential properties to class 5.4.3, although it is clear on inspection that some properties captured by this do have small-scale agriculture, but there are no available data to systematically determine this.

Rural residential without agriculture (5.4.3)

This is fundamentally a very similar classification to class 5.4.1, but properties are only considered if they are indicated as being within a 'small' urban area, 'rural settlement', and 'rural other' (as defined by Statistics NZ). Note that there is some potential overlap with 5.4.1 in small urban areas, but under class 5.4.3 a low confidence is assigned to land areas that otherwise meet the conditions but are smaller than 0.2 ha in size. Then when properties are greater than 2 ha, a lower confidence is again assigned because we consider there is increased potential for commercial-scale agriculture as property size increases (although there is no other available information to support this suspicion).

Remote communities (5.4.4)

This class was not considered relevant in the New Zealand context because its definition is explicitly based on Australian standards.

Farm buildings/infrastructure (5.4.5)

Buildings identified from LINZ building outlines derived from LiDAR are classified under this class if they are located on a ratings unit associated with any form of rural industry or a large (above 2 ha) lifestyle property. However, a lower confidence is assigned in the latter case.

Services (5.5)

The definition of services at the secondary level is based on data from the NRC Selected Land Use Register, covering classes:

- A3: Commercial analytical laboratory sites
- A5: Dry-cleaning plants including dry-cleaning premises or the bulk storage of dry-cleaning solvents
- A11: Pest control including the premises of commercial pest control operators or any authorities that carry out pest control where bulk storage or preparation of pesticide occurs, including preparation of poisoned baits or filling or washing of tanks for pesticide application
- F7: Service stations including retail or commercial refuelling facilities.

Other sources of data could be considered in this specific class (e.g. the NZ Facilities layer), but they are not because they can readily be assigned a tertiary class under 5.5.

Commercial services (5.5.1)

This layer is defined with direct reference to the ratings database, land-use types 80–84. This includes all commercial land according to the ratings database, including 'Multi-use within commercial' but not 'Car parking' (85) or 'Vacant commercial' (89).

Public services (5.5.2)

This class captures a wide variety of service land, from schools to crematoria. The range of input data to support it is correspondingly broad.

The NZ Facilities layer (from LINZ, which compiles data from authoritative sources) captures schools and hospitals.

The ratings database is filtered for community services: educational, medical and allied, personal and property protection, halls, cemeteries and crematoria.

From the LINZ record of protected areas, land is included if any of the following conditions are met.

- It is protected under the Conservation Act 1987, s.60 Administration Purpose.
- Generally, when the reserve purpose is administrative buildings, education, health, police, post offices, workshops, etc.
- It is protected under the Reserves Act 1977, s.22 (Government Purpose Reserve) or s.23 (Local Purpose Reserve), and the reserve purpose is something that seems appropriate to this class (e.g. library, fire station, community centre, crematorium, ambulance station). If the reserve purpose is not specified, the land area may still be captured if the name of the protected area matches a text search for phrases such as 'community facility', 'school', 'civil defence', 'public utility', or 'cemetery reserve'.

A text search was also applied to the CRoSL, for terms matching phrases such as 'kinder', 'kohanga reo', 'playcentre', 'fire', 'plunket', etc. The text search in this case, and in all cases throughout, is performed with PostgreSQL text search types, known as 'tsvector' (text organised for efficient searching) and 'tsquery' (queries into tsvectors).[28] These don't match exact strings, but rather use the concept of lexemes to afford flexibility in querying text. For example, both 'play centres' and 'playcentre' would match the tsquery 'play centre'.

Recreation and culture (5.5.3)

This class includes religious, and recreational land-uses obtained from the ratings database. The recreational uses may be both passive or active indoor, or active outdoor. Passive outdoor recreation is instead captured in class 1.1.7.

LCDB class 2 ('Urban parkland/open space') is included in this category, as are golf courses, and sports fields, as identified in the LINZ topographic series 1:50,000 scale. Gun clubs (identified from the NRC Selected Land Use Register) are included in this category.

[28] https://www.postgresql.org/docs/current/datatype-textsearch.html

Finally, the CRoSL is searched for land areas where either the given statutory action or common name is akin to 'recreation reserve', 'esplanade reserve', 'war', 'gallery', 'camp', 'museum', 'scout', 'guide', 'showground', etc.

Defence facilities – urban (5.5.4)

There was not sufficiently accurate information to isolate all urban land managed or owned by the New Zealand Defence Force, but we did include protected areas with a reserve purpose of 'air training', as identified from the LINZ protected areas layer.

Research facilities (5.5.5)

We were unable to identify a systematic method of discovering all land areas devoted to research facilities. Instead, we specifically included the location of the NIWA Northland Aquaculture Centre[29] by reference to its property title.

Utilities (5.6)

At this broad secondary level, this class includes land identified by the ratings database as any kind of utility (except water supply: as class 1.2.2 'Surface water supply' includes these areas).

We also capture the Ngāwhā Geothermal Power Station through OSM.

A broad text search through the CRoSL for public utilities complements these other sources.

Fuel powered electricity generation (5.6.1)

We did not consider this class to be relevant for Northland.

Hydro electricity generation (5.6.2)

We did not consider this class to be relevant for Northland.

Wind electricity generation (5.6.3)

We did not consider this class to be relevant for Northland, yet. We are aware of two proposed wind farms (Kaiwaikawe and Rototuna), but neither are built.

Solar electricity generation (5.6.4)

We did not consider this class to be relevant for Northland, yet. We are aware of multiple proposed solar farms (e.g. Pukenui, Lodestone – Kaitaia, Dargaville, Ruakākā Energy Park,

---

[29] https://niwa.co.nz/aquaculture

and other sites), but none are complete. These will, however, be captured as 'Land in transition' (class 3.6.0 or 4.6.0). Notably, Lodestone is designing its facilities to allow multiple uses by having adequate space between rows of panels to allow for continued arable or pastoral use of the land. ALUM does not explicitly recognise multiple uses.

Electricity substations and transmission (5.6.5)

Pylons used for electricity transmission are included within this class, as 10 m buffers around the point location of the pylon. Although minor, this was included because pylon wires themselves were present as a digital artifact in the canopy height model layer, and this was corrected by ignoring the canopy value where it coincided with electricity transmission with pylons. Thus, the classification was already accounting for the presence of pylons, and it was trivial to explicitly include them here.

The NRC Selected land Use Register was included where land had a HAIL ID of B4, indicating substations and switchyards.

The CRoSL was searched for electrical substations, which were included on that basis.

Finally, land areas from the ratings database were included where the given land-use type was '62: Electric utility services'.

Gas treatment, storage and transmission (5.6.6)

Land areas from the ratings database were included where the given land-use type was '63: Gas utility services'.

Water extraction and transmission (5.6.7)

This class represents the 'extraction, purification, treatment or supply of fresh water for public, domestic and commercial use'.

The LINZ protected areas layer gives some indication of land used for water extraction: land that is managed under the Reserves Act 1977, s.22 (Government Purpose Reserve), or s.23 (Local Purpose Reserve), where the reserve purpose is drainage, flood mitigation, or pump station. Alternatively, a positive match against the following text queries can also qualify land for inclusion: 'stormwater retention', 'stormwater reserve', 'drainage reserve', 'drain reserve'.

Land areas identified by the ratings database as having a land-use type of 65 ('Utility services: sanitary') are incorporated.

Land from the CRoSL is included when it matches text searches for terms like 'water', 'waterworks', etc. if they do not also match terms such as 'sewerage'.

Transport and communication (5.7)

This broad secondary class includes car-parking areas identified from the LINZ protected areas layer (Reserves Act 1977, s.22 Government Purpose Reserve, or s.23 Local Purpose Reserve, reserve purpose of 'car park' or 'parking'; or the name matches a text search for 'carpark'). It augments this with ratings data, where the recorded land-use is 32 ('Transport: car parking'), 85 ('Commercial – car parking'), or 98 ('Residential – car parking'). Ratings data that meet these criteria are treated with less confidence if the improvements description does not support this by including a reference to parking.

The NRC Selected Land Use Register further supplements this because there is a HAIL category (F8) that captures 'transport depots or yards including areas used for refuelling or the bulk storage of hazardous substances.' Confidence is improved if this is the only HAIL category applied to a land area, and when it is current and verified.

Finally, we include a text search across the CRoSL for land with a name or statutory action matching a query for 'car-parking'.

Airports/aerodromes (5.7.1)

Airports and aerodromes are captured through a similar combination of data from the LINZ protected areas layer, CRoSL, and NRC Selected Land Use Register, as follows:

- protected areas, when the reserve purpose is given as 'Aerodrome', or the name of the reserve matches the phrase 'aerodrome reserve'
- NRC Selected Land Use Register, when the HAIL category is F1, 'Airports including fuel storage, workshops, washdown areas, or fire practice areas'; confidence is improved if this is the only HAIL category applied to a land area, and when it is current and verified
- the CRoSL, when the name or statutory action matches a query for 'aerodrome' or 'airport'.

Roads (5.7.2)

Roads are captured directly from the LINZ 1:50,000 scale topographic map road data set. These road centrelines are buffered by a constant and a function of the number of lanes on a road, n, such that the buffer width is $\mathrm{w} = 1.5(n+1)$. At the target resolution, H3 13, the hexagonal cells have average edge lengths of 4.09 m, so some roads will appear discontinuous at the chosen map scale, but the largest roads will not.

Road parcels are also included, at a low confidence, but only when these parcels intersect roads present in the LINZ topographic data. This is to avoid including unformed legal roads ('paper roads'). Although under common law the public has the same right of access to unformed roads as formed roads, in practice the land for such roads was set aside in the period of provincial government (1854–1876) and thus, due to their age, or difficult terrain, they are not apparent on the ground, and it is likely few people exercise their rights to use them. As a map of land-use rather than tenure (or access rights), these are intentionally not captured.

Railways (5.7.3)

Railway centre lines are also obtained from the LINZ topographic data 1:50,000 series. Where a track is noted as single track, we buffer it by 5 m so that it has an area, and 10 m otherwise.

The NRC Selected Land Use Register is also included by filtering for land with HAIL category F6, 'Railway yards including goods-handling yards, workshops, refuelling facilities or maintenance areas'.

Ports and water transport (5.7.4)

Ports and water transport areas are captured by an NRC data set of coastal areas. Those areas noted as 'mooring', 'port facilities', and the special case of Marsden Cove Marina, are all captured under this class, which is intended to capture places where ships dock to transfer cargo or people to land. However, identified areas may be superseded later in the classification by ALUM class 6.6.3 ('Estuary or coastal waters – intensive use'), as ALUM prioritises water when there are conflicting possibilities.

Navigation and communication (5.7.5)

We did not consider there were sufficient data from any source to map to this class.

Mining (5.8)

It was possible to search the NRC water-take permits for current water permits (diversions) where the name of the entity given the permit included the word 'quarry'. As water take permits are point locations, and not the location of the quarry, we join these preferentially to an existing mine site from LCDB (if it is within 500 m), and if that does not result in a match, we expand the search to quarries from LINZ quarries (1:50,000 scale topographic map data), retaining the 500 m spatial search condition. In the unusual case that there are two matches to one water permit, only the closest would be retained. These quarries are considered with lower confidence if they are noted as being disused.

The mines from LINZ and LCDB are included in the classification even if they do not have a corresponding water permit, but at a lower confidence.

The NRC Selected Land Use Register is also considered, where land is assigned this class if the HAIL category ID or IDs include E5,[30] E6[31] or E7.[32]

---

[30] Coal or coke yards.

[31] Hydrocarbon exploration or production, including well sites or flare pits.

[32] Mining industries (excluding gravel extraction), including exposure of faces or release of groundwater containing hazardous contaminants, or the storage of hazardous wastes, including waste dumps or dam tailings.

Finally, the protected areas layer is considered for reserves under the Reserves Act 1977, s.23 Local Purpose Reserve, when the reserve purpose is ‘gravel’, ‘local purpose (metal) reserve’, ‘metal’, or ‘road metal’. This is a relatively low-confidence classification: many of these reserves appear to be historical and are likely to be classified as a different class where there are data to support a different classification with higher confidence.

Mines (5.8.1)

We did not distinguish mines and quarries from the broader secondary class, ‘5.8 Mining’. Although in practice most land areas identified as 5.8 in Northland are considered much more likely to be quarries than mines, we did not consider there were clear, available data to support this universally.

Quarries (5.8.2)

See above.

Tailings (5.8.3)

We did not consider there were sufficient data from any source to map to this class.

Extractive industry not in use (5.8.4)

Protected land areas protected under the Land Act 1948, and Reserves Act 1977, with a reserve purpose (in both cases) of ‘gravel’, were considered to indicate historical gravel quarries. This is a medium-low confidence classification (confidence value 3).

Waste treatment and disposal (5.9)

Protected land areas protected under the Reserves Act 1977, s.23 Local Purpose Reserve, with reserve purposes of ‘land fill’, ‘refuse disposal’ or ‘waste recovery centre’, were considered to indicate land areas used for disposal facilities.

Layered over this is land identified from the NRC Selected Land Use Register, with the following HAIL category IDs:

- G3: Landfill sites
- G4: Scrap yards including automotive dismantling, wrecking or scrap metal yards
- G5: Waste disposal to land (excluding where biosolids have been used as soil conditioners)
- G6: Waste recycling or waste or wastewater treatment.

Effluent pond (5.9.1)

This class was identified directly from LINZ 1:50,000 topographic data, a layer of ponds where the use is recorded as being ‘oxidation’, which indicates waste stabilisation ponds.

Landfill (5.9.2)

Although considered an imperfect source for this purpose, we searched the CRoSL for land areas under local government control where the statutory actions note that a piece of land has been acquired for landfill purposes. Because the CRoSL indicates an entire parcel, we consider that this class probably has low confidence, in terms of both the appropriateness of the class overall and spatial precision.

Solid garbage (5.9.3)

We did not consider there were sufficient data from any source to map to this class.

Incinerators (5.9.4)

We did not consider there were sufficient data from any source to map to this class.

Sewage/sewerage (5.9.5)

LINZ topographic data include a layer of ponds, many of which have a recorded use. When this use is 'sewage treatment' or 'settling', it is included in this class.

We augmented this identification with data from the CRoSL, where we singled out a record for a sewage treatment site near Kawakawa, which was noticed during validation but not captured in any relevant data set. This is the parcel with the title NA35B/1243. The classification system we have developed for this land-use classification, although it does not involve manual identification in the sense of one-off, non-reproducible, point-and-click operations, does allow us to define very narrow criteria for inclusion for a given class, as we have done in this case.

Finally, we included land from the LINZ protected areas layer if it is managed under the Reserves Act 1977, s.23 Local Purpose Reserve, and if the reserve purpose is 'sewage treatment', 'wastewater treatment', or 'water supply and sewage utilities'.

However, as much of the geographic space identified as 5.9.5 is actually water (e.g. settling ponds), it is more often captured as class 6.1.3 (Lake – intensive use), because of the preferential identification of water features under the ALUM classification system. The system developed to support this classification overlays water features (class 6) only after all a classification under classes 1-6 has first occurred. This is done to correctly identify which subclass of water feature to apply.

4.4.7 Water (6)

Class 6: Water

Water features are regarded as essential to the ALUM Classification because of their importance for natural resources management, agricultural production and as points of reference in the landscape. At the secondary level, the classification identifies water features, both natural and artificial. Tertiary classes relate water features to intensity of use and conservation status.

Because water is a land cover rather than a land use, water classes may not be mutually exclusive with other land use classes at particular levels in the classification. Generally, water classes should take precedence so that, for example, a lake in a conservation reserve will be classed as 6.1 'Lake' or 6.1.1 'Lake–conservation', rather than as 1.1 'Nature conservation'. Water features where a conservation tertiary class applies may be attributed using the comments field.

It may be surprising that a land-use map includes water as a primary class. Indeed, it breaks the established pattern of ALUM, in that the five preceding primary classes are ordered by increasing levels of human intervention or potential impact on the natural landscape, to be followed by water rather than something with even more intense modification compared to a natural landscape.

Despite the recognition that water is a type of cover rather than a use, water is considered essential to ALUM. Indeed, the advice is that where water is not mutually exclusive with another land-use class, 'generally, water classes should take precedence so that, for example, a lake in a conservation reserve will be classed as 6.1, "Lake – conservation", rather than as 1.1, "Nature conservation"' (ABARES 2016, p. 23).

The approach taken here is that water classes always take precedence over other classes where we have data to locate the water features. Therefore, we first classify land under all other possible classes, and then overlay water features.

In most secondary classes organised under 6 (6.1, 6.3, 6.5, and 6.6), the tertiary class assignment is a ternary distinction between conservation, production, and intensive use; that is, relating to classes 1, 2, and 5. (However, recall that we didn't classify anything under class 2, 'Production from relatively natural environments', see section 4.4.3.) Therefore, when this overlay occurs, if a water feature intersects land that has been assigned under classes 1 or 5, it will become classified under class 6, with the appropriate ternary class for conservation or production. When a water feature either overlaps with no other classification, or class 3 or 4 land, it is not assigned a ternary class but rather remains at the secondary level in the hierarchy. For example, non-protected, natural (non-reservoir) lakes on pastoral farms (classes 3 or 4) are typically assigned 6.1 ('Lake').

Lake (6.1)

Lakes are captured primarily from LCDB (class 20, 'Lake or pond') and secondarily from LINZ topographic data ('ponds'). In the latter case, ponds are only included if the recorded use is not 'oxidation', 'sewage treatment' or 'settling'.

Lake – conservation (6.1.1)

Where an identified lake intersects land that is otherwise classified as class 1 ('Conservation'), it is assigned this tertiary class.

Lake – production (6.1.2)

No areas are assigned to this class due to the decision not to classify any land under the primary class 2.

Lake – intensive use (6.1.3)

Where an identified lake intersects land that is otherwise classified as class 5 ('Intensive uses'), it is assigned this tertiary class.

Lake – saline (6.1.4)

This class is mapped directly from LINZ topographic data, 'lagoons', which are saline by definition.

Reservoir/dam (6.2)

Reservoirs are identified as a subset of lakes defined by the LINZ topographic series. Using NRC water permits, selecting for current permits with the 'dam' activity sub-type, we selected the lake (if any) that is within 500 m of the permit location.

We did not consider the tertiary reservoir classes ('Reservoir', 6.2.1; 'Water storage – intensive use/farm dams', 6.2.2; 'Evaporation basin', 6.2.3). It may nevertheless be possible to search the authorisation name for water take permits to infer whether a given reservoir is used for farms.

Other lakes that do not have a corresponding water-take permit may still be classified as reservoir/dam when the recorded LINZ lake use is 'reservoir'.

River (6.3)

Rivers are captured primarily from LINZ topographic data, both lines and polygons. Line features are first buffered by 5 m, and therefore may appear discontinuous in the land-use map at the target resolution.

This is supplemented by LCDB class 21 ('River'), which only captures wide lowland rivers, and which we include at a low level of confidence only because we consider the LINZ topographic data delineation to be superior.

River – conservation (6.3.1)

Where an identified river intersects land that is otherwise classified as class 1 ('Conservation'), it is assigned this tertiary class.

River – production (6.3.2)

No areas are assigned to this class due to the decision not to classify any land under the primary class 2.

River – intensive use (6.3.3)

Where an identified river intersects land that is otherwise classified as class 5 ('Intensive use'), it is assigned this tertiary class.

Channel/aqueduct (6.4)

This class is mapped directly from LINZ topographic data: 'Drains'.

Supply channel/aqueduct (6.4.1)

We did not consider that there was sufficient data from any source to map to this class.

Drainage channel/aqueduct (6.4.2)

We did not consider there were sufficient data from any source to map to this class.

Stormwater (6.4.3)

We did not consider there were sufficient data from any source to map to this class.

Marsh/wetland (6.5)

Currently, the best available spatial data set of wetlands with national coverage is the revised extent of wetlands in New Zealand (Dymond et al. 2021).[33] This layer is a combination of the Wetlands of National Importance (WONI) and LCDB, and an attempt to reconcile the two while recognising that LCDB is considered to have better boundary delineation while WONI includes more wetlands. Note that neither includes wetlands smaller than 0.5 ha, so this land-use is not able to represent them either.

---

[33] https://datastore.landcareresearch.co.nz/dataset/revised-extent-of-wetlands-in-new-zealand

We supplement the above data with wetlands provided by NRC, and with LINZ topographic data ('Swamps'). Used together, the boundary delineation of wetlands is perhaps biased upwards (i.e. larger than reality), but rests on the authority of the most reliable data available.

Marsh/wetland – conservation (6.5.1)

Where an identified wetland intersects land that is otherwise classified as class 5 ('Conservation') it is assigned this tertiary class.

Marsh/wetland – production (6.5.2)

No areas are assigned to this class due to the decision not to classify any land under the primary class 2.

Marsh/wetland – intensive use (6.5.3)

Where an identified wetland intersects land that is otherwise classified as class 5 ('Intensive use') it is assigned this tertiary class.

Marsh/wetland – saline (6.5.4)

Saline marshes are identified directly from an NRC layer provided to us. This layer represented areas of mangroves and salt marshes in great detail and therefore high confidence. The vegetation type is recorded as a comment.

This is nevertheless augmented by LINZ topographic data (mangroves). Where such areas also coincide with LCDB class 70 ('Mangroves'), the confidence is increased to the highest level (1). LCDB mangrove areas alone are also included, at a confidence level of 2.

Finally, reference is made to LUCAS LUM, class 79, 'Wetland (open water').

Estuary/coastal waters (6.6)

The coastline itself is an important consideration in constructing a land-use map. In this case we have defined it by reference to the coastline as defined by an NRC data set, with high confidence. (This is the same layer that assists in defining class 5.7.4, 'Ports and water transport'.) Inland areas of open saline water are included by reference to LCDB class 22 ('Estuarine open water').

Estuary/coastal waters – conservation (6.6.1)

Where an identified estuary or coastal water intersects land that is otherwise classified as class 1 ('Conservation') it is assigned this tertiary class.

Estuary/coastal waters – production (6.6.2)

No areas are assigned to this class due to the decision not to classify any land under the primary class 2.

Estuary/coastal waters – intensive use (6.6.3)

Where an identified estuary or coastal water intersects land that is otherwise classified as class 5 ('Intensive use') it is assigned this tertiary class.

4.5 Commodities

In ALUM, agricultural commodities are assigned to ALUM classes where it is possible to do so. The agreed list of commodities is a controlled vocabulary with several hundred entries (ABARES 2016, p. 31). The list of agreed terms was compiled from lists such as the Australian Bureau of Statistics, United Nations trade codes, Australian and New Zealand Standard Industrial Classification (ANZSIC), and others. The commodity description is intended to provide additional information beyond what is recorded in the ALUM classification of primary land-use. It is possible, in some cases, to specify commodities with greater or lesser levels of detail; e.g. 'cattle' vs 'cattle dairy' – the latter used when cattle are known to be used for milk production, and the former implying a level of uncertainty about what the cattle are for (the other options being 'cattle meat' and 'cattle stud').

These data are recorded in the 'commod_desc' field. If two or more commodities are recorded, they are separated with a comma followed by a white-space character. It is valid according to the ALUM schema to use the value `UNKNOWN`, but we have opted always to record a null value to indicate that a possible commodity is unknown due to lack of available data, not due to some other form of uncertainty.

Only a small number of commodities have been included, on the basis of available information, and according to priorities set by NRC:

- `avocado`
- `cattle dairy`
- `finfish`
- `grapes`
- `kiwifruit`
- `kumara`
- `macadamias`
- `molluscs`
- `olives`

However, note that under ALUM citrus is also captured as a distinct class, and not as a commodity. If it were possible to disambiguate, specific types of citruses could have been recorded as commodities (using the terms 'lemons', 'oranges', etc.), Although we could capture areas of citrus, there are insufficient data to further classify into specific citrus commodities.

Vineyards are distinct in that they are captured as a distinct class, but there is also an allowance to record 'grapes' as a commodity. Like citrus, although it is possible to locate viticulture we did not opt to further distinguish 'grapes dried', 'grapes table', or 'grapes wine', although most identified vineyards in Northland may be taken as existing for wine production.

Sheep & beef farms represent most of the pastoral land that is not 'cattle dairy'. However, in the absence of AgriBase (which has its own limitations), there is no available database that records farms and their livestock systems. We do record, as a comment, which Beef + Lamb NZ farming system an area of pastoral farm probably corresponds to, but this is based entirely on the physical characteristics of the land and must not be taken to definitively record that sheep or beef are grazed.

More information about the process for identifying these commodities can be read in the relevant sections of this report, according to the relevant ALUM class.

4.6 Management practices

Land management practices may also be recorded within ALUM, again using an agreed list of terms (ABARES 2016, p. 49). It is recorded in the fiel 'manage_desc'. Presently there is very little information with which one could capture management practices in Northland automatically, across the entire region, without field visits or manual inspection of aerial imagery. (The latter remains a possibility for augmenting the delivered data, which is one of the recommendations made in section 8.) As such, the only management practices that can be recorded are:

- horticulture:
  - fully protected
  - open growing
  - semi-protected
- irrigation:
  - irrigation drip
  - irrigation spray
  - irrigation surface

When there are multiple management practices recorded, these are separated by a comma and one white-space character (e.g. `open growing, irrigation drip`).

With respect to the captured horticultural land management practices, we assume `open growing` as a default. Where there is evidence of bird-netting from Sentinel-2 imagery mosaics, we record `semi-protected`. `fully protected` is a possibility when ratings data indicate glasshouses or greenhouses, while shadehouses are attributed as semi-protected. However, much of this information (except bird-netting) is derived from information at a property level, so it is likely that this information, even if correct, is upwardly biased in terms of spatial coverage. That is, if a significant fraction of a parcel contains glasshouses, horticultural land anywhere on the property could be labelled fully protected even if there is a mixed regime.

More information on irrigation system types is available in section 4.4.5, particularly Table 3, p. 34.

4.7 Reproducibility

A primary objective of this project has been to develop an automated process of land-use classification. An automated process is a reproducible process. However, the workflow developed to create the land-use is proprietary: although the workflow itself depends entirely on various open-source projects, it is not itself open source. (However, as mentioned earlier, we have made two key components open source: raster2dggs and vector2dggs.)

The basic premise of the workflow can, nevertheless, be laid out. We have developed a 'land-use information system' with three primary components:

1. a relational database management system (PostgreSQL, with PostGIS and H3 DGGS extensions)
2. a scalable, reproducible workflow, written in Snakemake (Mölder et al. 2021)
3. a system of overlaying configuration files that encapsulate four concepts:
   a. data that can be used as inputs, from different sources and in different formats
   b. the rules of a particular system of land-use classification
   c. output files
   d. profiles, which express the environment where a workflow is run (such as available computing resources), and the region of interest for which to produce a spatial land-use data set.

The first two components are entirely separate from the third. There is a configuration interface defined, to which configuration files in the third component must adhere. Source control is used to be able to return to a point in time to reproduce a land-use map using the same environment and dependent software that were used at the time the data for a particular classification were produced. Basic provenance information is implicitly recorded using source control (git), such as who wrote particular classification rules, and when.

When a workflow is run a PostgreSQL database is created, into which data are imported, then mutated following the rules defined by the configuration (such as the computation of derived columns, or filters). Snakemake rules are run in a specific sequence to download, transform, and combine desired data according to explicit rules. This combines some relatively popular and simple command-line tools with custom software, which can be shown abstractly in a directed acyclic graph (Figure 8). The workflow software is responsible for controlling the order of operations according to the directed acyclic graph, and it also allows most parts of the workflow to be executed in parallel, scaling with central processing unit (CPU) cores. Each rule describes a particular environment of software it needs to run, perhaps within a containerised, virtual operating system (using

Docker[34]), or described with Conda.[35] In both cases, exact versions of software are recorded, from which we can reliably rebuild the environment in the future even as these independent software projects change.

Ultimately, output data are written to disk in standard, stand-alone GIS formats (vector formats such as GeoPackage or File Geodatabase, or raster formats such as GeoTIFF and KEA). There is control over the 'spatial unit' for aggregated output, which in the present case was set as the parcel, produced in addition to the (default) cell-level output, which has no aggregation.

At the conclusion of the workflow execution the database can be stopped, and its data removed: the relational database management system is considered an ephemeral store of intermediate data, although it, too, could be archived and restored. Data that were downloaded and indexed to DGGS *are* retained on disk as stand-alone files, as are all output files. Alongside log files, these are compressed into a single archive, which can be stored for potential retrieval at a later point if the need arises to inspect the input data, or to reproduce the output without relying on external data services. This later point is important because, even though GIS web data services are convenient, they are online resources so they are not guaranteed to be persistent.

[34]https://www.docker.com/

[35]https://conda.io/

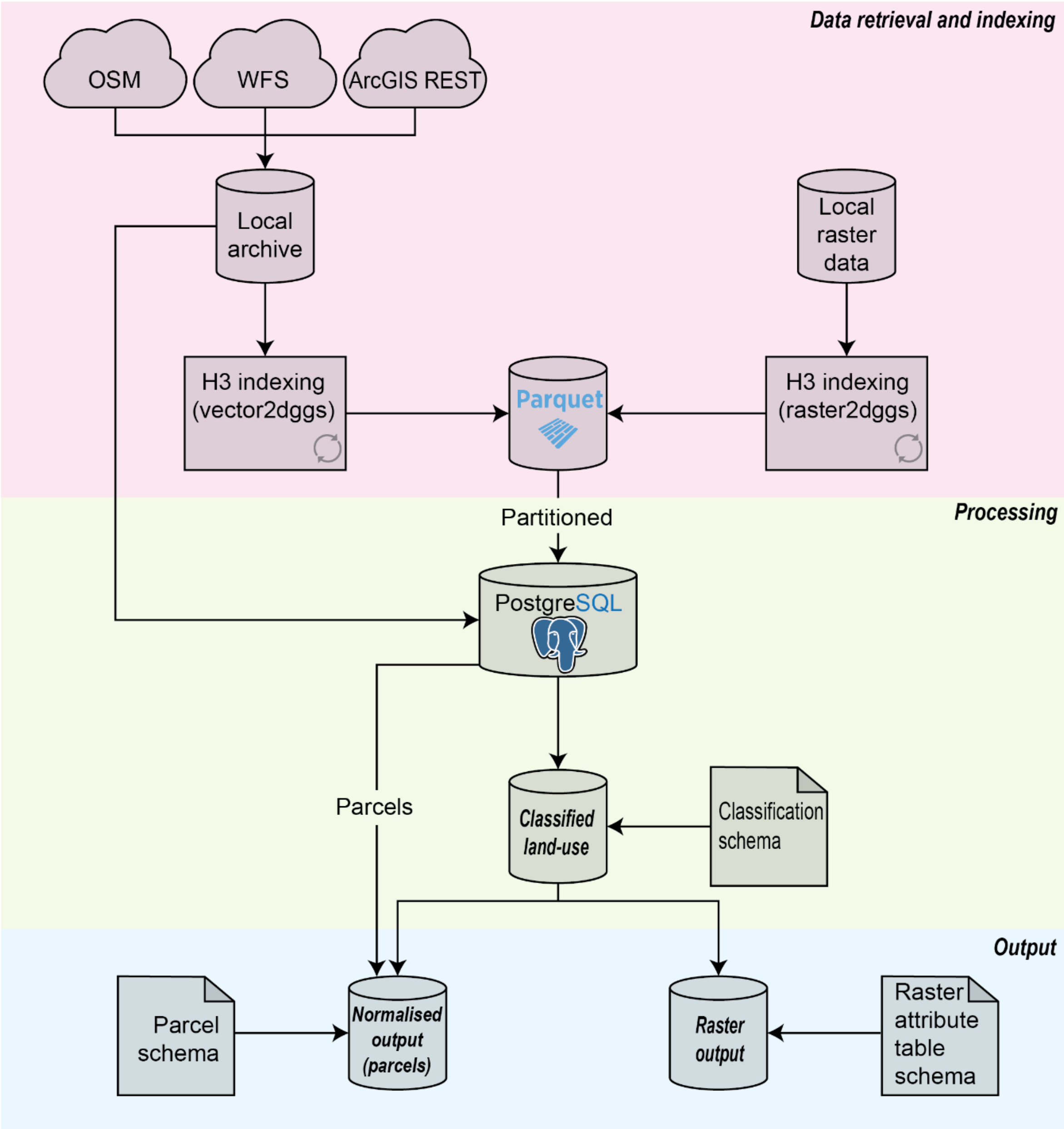


Figure 8. Diagrammatic directed acyclic graph representing the classification workflow from beginning to end, including downloading data via web services, DGGS indexation, importing data to PostgreSQL, partitioning, classification, and writing output files.
Notes: OSM = OpenStreetMap; WFS= web feature service; ArcGIS REST= ArcGIS representational state transfer API.

The workflow is largely designed to be executed in chunks of fixed size (spatial partitions) composed of groups of DGGS cells according to parent–child cell relationships. The workflow can then scale by handling one partition per CPU core while RAM consumption is managed within the available limit. The land under each spatial partition is calculated by one core, before the data from all partitions are re-assembled at the end.

Throughout the development of this land-use data set, the workflow was run hundreds of times (albeit mostly on small areas, and at lower resolutions), but the final workflow was still able to be conducted on a developer laptop with 20 CPU cores but only 16 GB of RAM, and was able to complete in approximately 24 hours, which includes all download time (which is not performed in parallel, to address the potential for network throttling and dropped packets) and DGGS indexing time (which is itself expensive).

Although it is possible, theoretically and practically, to reproduce this data set, the software with corresponding configuration files with which it has been done only exists in-house as proprietary software.

Another form of reproducibility is a less strict kind. Rather than being able to produce an exact replica of the final land-use data set, it would involve producing an equivalent data set in the future. The difficulty with this is that dependent web services may cease to exist, others may change their schema, and others may cease to be maintained, meaning the data becomes stale. These issues could be managed, but they are where attention would need to be focused if a future version were to be created.

5 Technical layer

In addition to the primary output data sets, which apply the ALUM schema both to parcels and at an individual cell level, we have prepared a technical layer intended for internal use by an analyst. The technical layer is not a classification: it is a collection of data sets that may help NRC produce queries that can identify land areas that meet certain conditions and relate these to data in the cell- or parcel-level ALUM output.

The data included in the technical layer are as follows:

1 NIWA (climate variables, average over 1991–2020):[36]

- a growing degree days (base 5˚C) (`GDDbase5_Annual`)
- b growing degree days (base 10˚C) (`GDDbase5_Annual`)
- c heating degree days (`HDDbase18_Annual`)
- d potential evapotranspiration (`TotalPenmanPET_Annual`)
- e sunshine hours (`TotalSunshine_Annual`)
- f rainfall (`TotalRainfall_Annual`)
- g rain days (`WetDays1mm_Annual`)
- h relative humidity (`RelHumidity0900_Annual`)
- i days of soil moisture deficit (`SMDDays_Annual`)
- j solar radiation (`GlobalIrradiance_Annual`)
- k average daily temperature (`MeanTemp_Annual`)
- l average daily maximum temperature (`MaxTemp_Annual`)

[36] https://niwa.co.nz/climate/research-projects/national-and-regional-climate-maps

m average daily minimum temperature (MinTemp_Annual)

n earth temperature (EarthTemp10cm_Annual)

o wind speed (WindSpeed10m_Annual)

2 NRC:

a mean annual rainfall (NRC_meanAnnual_Rainfall)

b freshwater management unit (FMU)

c digital river network

i HydroID

ii HydroCat

3 MWLR:

a canopy height (CHM)

b slope (Slope)

c elevation (DEM)

d NZLRI

i land-use capability (luc)

ii lithology baserock (baserock)

iii NZLRI Edition 2 lithology (rock2)

e New Zealand Fundamental Soil Layer

i soil series (SERIES)

ii soil type (SOILTYPE)

iii NZ soil classification (NZSC)

4 MfE:

a LENZ Level 4 category (LENZ_LV4).[37]

In addition to these attribute fields, for each record there is a geometry field, which is one H3 DGGS cell, at resolution 13, represented as a hexagon in the NZGD2000 / Mt Eden 2000 projection, EPSG:2105. There is also the H3 cell ID itself (a 64-bit integer represented in hexadecimal). Figure 9 visualises three of these technical layers.

[37] Which is itself a summary of 15 climate, landform, and soil variables specifically chosen for their relevance to biological distributions, segmented, and labelled according to a hierarchical classification system with 500 categories: https://data.mfe.govt.nz/layer/52358-land-environments-new-zealand-lenz-level-4-polygons-2009/

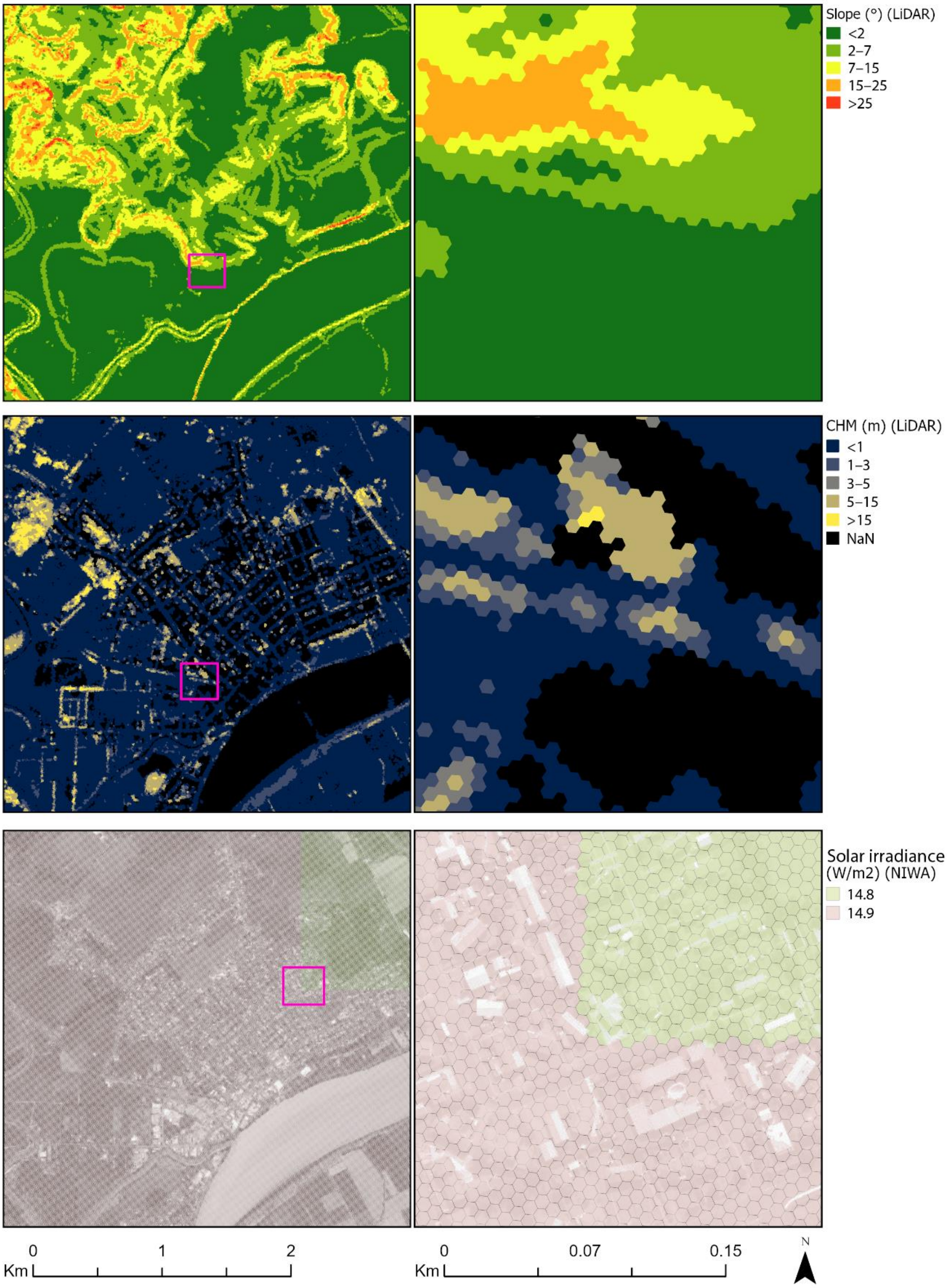


Figure 9. Sample of three layers included in the technical layer. The inset maps on the right-hand side show the map at a small scale, where individual H3 DGGS cells are visible. These discrete cells allow the association of multiple layers.
Note: CHM = canopy height model.

5.1 Using the technical layer

To associate these data to the cell-level data, there is another ID attribute present in both the technical layer and the cell-level output: `unit_id`. This Unit ID represents a unique combination of all attributes in the ALUM data, including parcel IDs, such that it can operate as a unique identifier for joining the ALUM classification to any of the attributes in the technical layer. It is calculated as a truncated MD5 hash of all the attributes, and then converted into a big integer. Using the Unit ID as a join attribute between the technical and ordinary data, it is possible to perform queries to locate specific areas of interest; for example, pastoral land (classes 3 or 4) with a certain amount of annual rainfall, in a specific freshwater management unit.

The data are provided as an Esri File Geodatabase, and in a more specialised format: Apache Parquet.[38] The File Geodatabase is considerably larger than the Parquet data, although they both contain the same information, and the difference in size is explained by the methods of compression inherent in each format. The volume of data is large and may be somewhat difficult to work with: there are approximately 300 million DGGS cells at H3 resolution 13 over Northland, and all of these are included. The reason Parquet is much smaller on disk than the File Geodatabase is that Parquet files are stored in column-orientation, rather than row-orientation; and this has benefits for compressing data, particularly for spatially auto-correlated data when it is organised by spatial partitions. However, both data sets are exactly equivalent and either may be used. We recommend using the Parquet representation, if possible, or else uploading the data into a relational database management system and creating relevant indices on the various attributes to avoid loading the entire data set into memory or requiring a full sequential scan when querying the data.

What follows is a series of worked examples of using the Parquet-format technical data, both on its own and in association with the ALUM data. This relies on an open-source Python library called Dask.[39] Dask allows users to use idioms associated with Pandas DataFrames, but to scale them across available CPU cores.

The first example demonstrates opening the technical layer for reading (Figure 10). This is done 'lazily', such that nothing is put into memory until the compute method is called on the Dask DataFrame. Before this is done, however, we use the query method to first isolate DGGS cells that are part of the Awanui freshwater management unit (FMU), with a slope less than 10° and a canopy height of 2 m or less. The result is a set of approximately 6 million rows, reduced from the several hundred million in the full set (which could not fit into memory). This is computed in parallel and took under 2 minutes to compute with a fairly dated 8-core CPU.

[38] https://parquet.apache.org/

[39] https://www.dask.org/

```
>>> import dask.dataframe as dd
>>> df = dd.read_parquet('/media/lawr/T71/nrc/technical/Deliverables/Technical_all_variables.parquet'
>>> df.columns

Index(['EarthTemp10cm_Annual.', 'GDDbase5_Annual', 'GDDbase10_Annual',
       'GlobalIrradiance_Annual', 'HDDbase18_Annual', 'MaxTemp_Annual',
       'NRC_meanAnnual_Rainfall', 'MeanTemp_Annual', 'MinTemp_Annual',
       'RelHumidity0900_Annual', 'SMDdays_Annual', 'TotalPenmanPET_Annual',
       'TotalRainfall_Annual', 'TotalSunhine_Annual', 'WetDays1mm_Annual',
       'WindSpeed10m_Annual', 'FMU', 'CHM', 'Slope', 'DEM', 'LENZ_LV4',
       'HydroID', 'HydroCat'],
      dtype='object')

>>> awanui_flat = df.query(f'(FMU == @catchment) & (Slope < @slope) & (CHM < @chm)', local_dict={
        "catchment": "Awanui", "slope": 10, "chm": 2}
    ).compute()

                 EarthTemp10cm_Annual.  GDDbase5_Annual  GDDbase10_Annual  ...  LENZ_LV4   HydroID
h3_13                                                                      ...
8dbb58d8049a4ff                   15.7      4175.799805       2294.199951  ...     G1.1a  233907.0  A
8dbb58d8069003f                   15.7      4176.700195       2294.500000  ...     G1.1a   53488.0
8dbb58d8069007f                   15.7      4176.700195       2294.500000  ...     G1.1a   53488.0
8dbb58d806900bf                   15.7      4176.700195       2294.500000  ...     G1.1a  233742.0
8dbb58d806900ff                   15.7      4176.700195       2294.500000  ...     G1.1a  233742.0
...                                ...              ...               ...  ...       ...       ...
8dbb5ebadad54ff                   14.3      3645.699951       1935.800049  ...     D1.1a  227554.0
8dbb5ebadad897f                   14.3      3645.699951       1935.800049  ...     D1.1a  227612.0
8dbb5ebadadc2ff                   14.3      3645.699951       1935.800049  ...     D1.1a  227554.0
8dbb5ebadadd93f                   14.3      3645.699951       1935.800049  ...     D1.1a  227424.0
8dbb5ebadadd97f                   14.3      3645.699951       1935.800049  ...     D1.1a  227438.0

[6159195 rows x 23 columns]
```

Figure 10. Example demonstrating how to read the technical layer (Apache parquet format), ultimately selecting for land in the Awanui catchment, below 10˚ in slope, and with a canopy height of less than 2 m.

The second example (Figure 11) takes the DataFrame computed in the first step and counts how many DGGS cells there are according to groups of LENZ Level 4 categories, indicating that the most common LENZ Level 4 category in this set is A2.1a, which is described as follows:[40]

> *This environment is restricted in extent, occurring only on sites with little slope and at low elevation in the far north, often in close proximity to Environment A1. As with the previous environment, A2's climate is typified by very warm temperatures, very high solar radiation and moderate annual water deficits. Sand and peat are the dominant soils, with some estuarine alluvium. Soils are very poorly-drained and of very low natural fertility.*

This description implicitly supports our selection criteria, as the Awanui catchment is in the Far North, and we have not selected land with slope exceeding 10˚.

[40] https://www.landcareresearch.co.nz/assets/Tools-And-Resources/Maps/LENZ/LENZ_Technical_Guide.pdf

```
>>> awanui_flat.groupby(['LENZ_LV4'])['LENZ_LV4'].count()

LENZ_LV4
A1.1a        362
A1.1b     170026
A1.1c     504389
A2.1a    1919982
A4.1a     142250
A5.1a    1671484
A5.1c      58583
A5.2a      67571
A6.1a      60645
A6.1b     587356
A6.1c      94776
A6.1d         39
A7.1a       1162
A7.2a       7267
D1.1a      44046
D1.1b       1751
D1.2b      37321
G1.1a     170509
G3.1a     245211
G3.1b     192136
G3.2a       3634
Name: LENZ_LV4, dtype: int64
```

Figure 11. Taking the previous result and counting the number of cells in each LENZ Level 4 category.

The third example (Figure 12) demonstrates how quick visualisations of the data can be made, using supporting libraries (in this case, matplotlib). The example takes the existing area and makes a histogram of NRC's mean annual rainfall data, in 15 bins (Figure 13). This shows that around 1.2 million of the resolution 13 H3 DGGS cells (approximately 52.6 km^2) in our area of interest have mean annual rainfall of around 1,200 mm.

```
>>> import matplotlib.pyplot as plt
>>> fig, ax = plt.subplots()
>>> fig.tight_layout()
>>> awanui_flat.hist(column='NRC_meanAnnual_Rainfall', bins=15, ax=ax)
>>> fig.savefig('rainfall.png')
```

Figure 12. Sample Python code to produce a histogram of a variable in a dataframe using matplotlib.

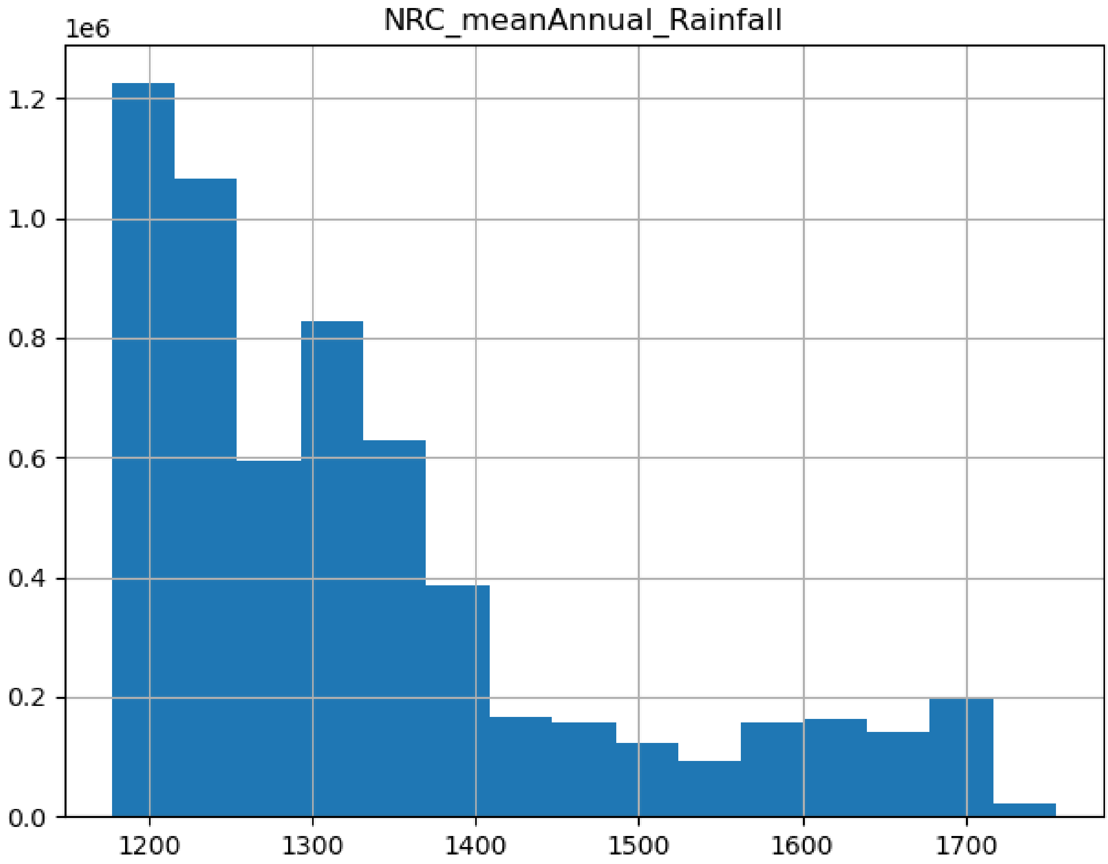


Figure 13. Example histogram plot of mean annual rainfall, only including those parts of the region that meet the preceding criteria.

Finally, the last example demonstrates a 'join' between the technical layers and the ALUM classified data (Figure 14 and Figure 15). This allows the production of a contingency table between the ALUM land-use classes and the LENZ Level 4 classes (i.e. generalised physical attributes). To do this, a new library called Polars[41] is introduced. This is akin to Pandas and Dask but considerably faster especially when joining across data stored in the Apache Parquet format. It is necessary to perform two joins: the first on the H3 DGGS cell ID and performed between the technical layer and a disaggregate layer of H3 DGGS cells; the second on the `unit_id` attribute and the disaggregate H3 DGGS cells, and the ALUM classified data, which comprise DGGS cells that have been geospatially dissolved. The `unit_id` attribute is a truncated MD5 hash of all ALUM attributes, such that the ALUM data are ready to be re-associated with individual DGGS cells. The result is also illustrated in a map (Figure 16).

[41] https://www.pola.rs/

```python
>>> import polars as pl
>>> import matplotlib.pyplot as plt
>>> from matplotlib.colors import LogNorm
>>> import pandas as pd
>>> import seaborn

# Load technical layer with Polars, and filter
>>> df = pl.read_parquet('/media/lawr/T71/nrc/technical/Deliverables/Technical_all_variables.parquet')
>>> awanui_flat = tech.filter((pl.col("CHM") < 2) & (pl.col("FMU") == "Awanui") & (pl.col("Slope") < 10))
>>> awanui_flat.collect()

# Joining ALUM (aggregated cells) to disaggregate H3 DGGS cells
>>> alum = pl.scan_parquet('luis-202606-nrcluc.parquet').drop(columns=['fid','parcel_id'])
>>> alum = alum.unique(subset=['unit_id'])
>>> cells = pl.scan_parquet('luis-202606-nrcluc-cells.parquet')
>>> cells = cells.with_columns(pl.col('unit_id').cast(pl.Int64, strict=False))
>>> alum_cells = cells.join(alum, on='unit_id',how='left')

# Joining technical layer to ALUM data
>>> awanui_flat.join(alum_cells, on='h3_13')

# Draw a contingency table
>>> crosstab = pd.crosstab(awanui.lu_description, awanui.LENZ_LV4, normalize=True)
>>> seaborn.heatmap(
        crosstab.transpose()*100, annot=True, annot_kws={'fontsize': 'x-small'},
        fmt='.1f', cmap='YlGnBu', vmin=0, vmax=60, xticklabels=True,
        square=False, norm=LogNorm)
>>> plt.show()
```

Figure 14. Python code for producing a cross tabulation of ALUM classes and LENZ Level 4 classes, after applying a filter on the technical layer, and joining the technical layer to the land-use layer.

Figure 15. Contingency table between ALUM classes and LENZ Level 4 classes in the filtered area. For example, this shows that 16.5% of the area is both 'Grazing modified pastures' (ALUM) and A2.1a (LENZ Level 4). Note that the colour scale is log-transformed.

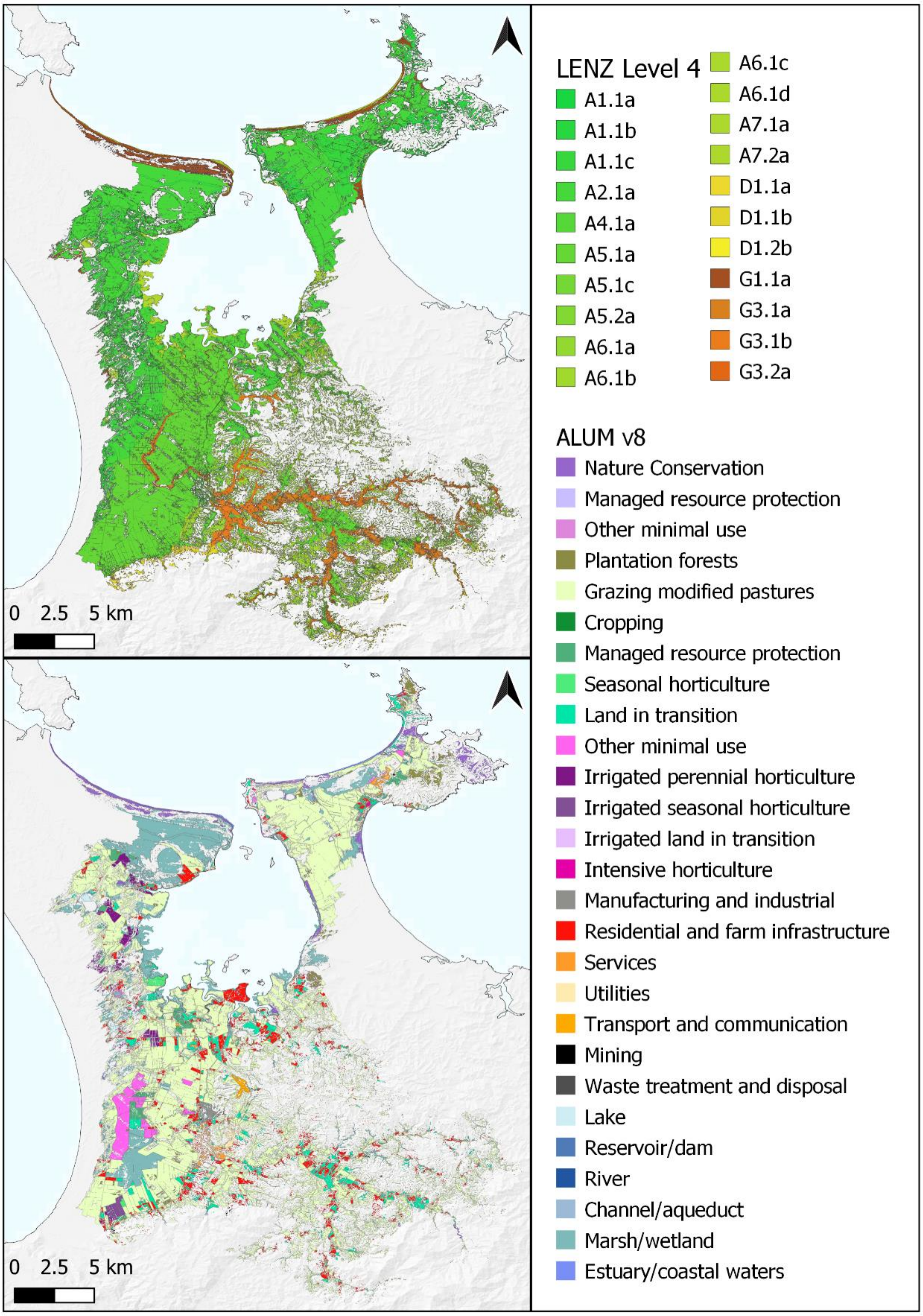


Figure 16. Map of the filtered area showing LENZ and ALUM classifications.

6 Usage

6.1 Overview

6.1.1 Projection

The data were delivered in the NZGD2000 / Mt Eden 2000 projection (EPSG:2105), although most input data are natively available in NZGD2000 / New Zealand Transverse Mercator 2000 (EPSG:2193) and are (must be) transformed to WGS 84 (EPSG:4326) before they are indexed in the H3 DGGS. The final transformation from WGS 84 to Mt Eden 2000 has an accuracy of 1 m. Mt Eden 2000 is the official NZGD2000 meridional circuit projection that cadastral surveyors are required to use under the Cadastral Survey Rules 2021 in Northland, Auckland, and Waikato.

6.1.2 Selecting data

The most fundamental operation to be performed on these data will probably be a query to select subsets of the data, exploiting the hierarchical nature of the classification system to meet specific operational needs of council staff. Examples include:

- selecting all land of a specific tertiary class (e.g. softwood plantation forestry)
- selecting all land of a specific secondary class, including all tertiary classes under that secondary class (e.g. intensive horticulture)
- selecting all land of a specific secondary class, including all tertiary classes under that secondary class, with specific management practices or commodities (e.g. intensive horticulture, with irrigation).

Other considerations (e.g. confidence, geographical subsetting) may also enter the query.

Data may be selected on their own to geographically identify land meeting certain criteria. They may also be selected to be aggregated to answer numerical questions, such as:

- How much land is of a specific class? For example, what is the total area of land used for softwood plantation forestry)?
- What is the average number of rain days on pastoral land in a specific catchment? For example, what is the average number of rain days on dairy farms in the Awanui catchment? In this case, the 'Awanui catchment' would need to be independently determined as a set of catchment IDs to use as a filter on the technical layer's `HydroID` attribute, or by using one of the predetermined freshwater management units (FMUs).

The technical layer is designed to allow consideration of ancillary data to support more specific and advanced queries, particularly those focused on environmental applications; for example, to use NIWA long-term climate data to select pastoral land (land-use classification) with dairy cattle (commodity), where the land is irrigated (management practice), and where the number of days of soil moisture deficit is above some relevant threshold (technical layer). To achieve this, the technical layer is joined to the land-use data using the H3 DGGS cell ID as a key, on a 1:1 basis. We recommend applying a filter to the technical layer first, before performing an inner join, to reduce the number of potential joins to be considered by the computer.

6.1.3 Adding additional data

The set of data included in the technical layer was decided at a workshop held with NRC in March 2023. It may be necessary to add additional data to this set. One option is to use the land-use classification data as one would any standard GIS layer; i.e. spatial overlays. However, the benefit of using a DGGS is that it affords the opportunity to use the H3 DGGS cell ID to efficiently join disparate data in an implicit spatial join. We recommend using our command-line tools (raster2dggs and vector2dggs, described earlier) to convert geospatial data into H3, and then use data-frame manipulation tools such as Pandas (in Python) or Arrow (in R) to perform the association.

6.2 Sample maps

This section presents a sample of maps demonstrating the ALUM-classified land-use data for parts of Northland. In each of the following figures the legends are different, because they only include items for features that are in the map.

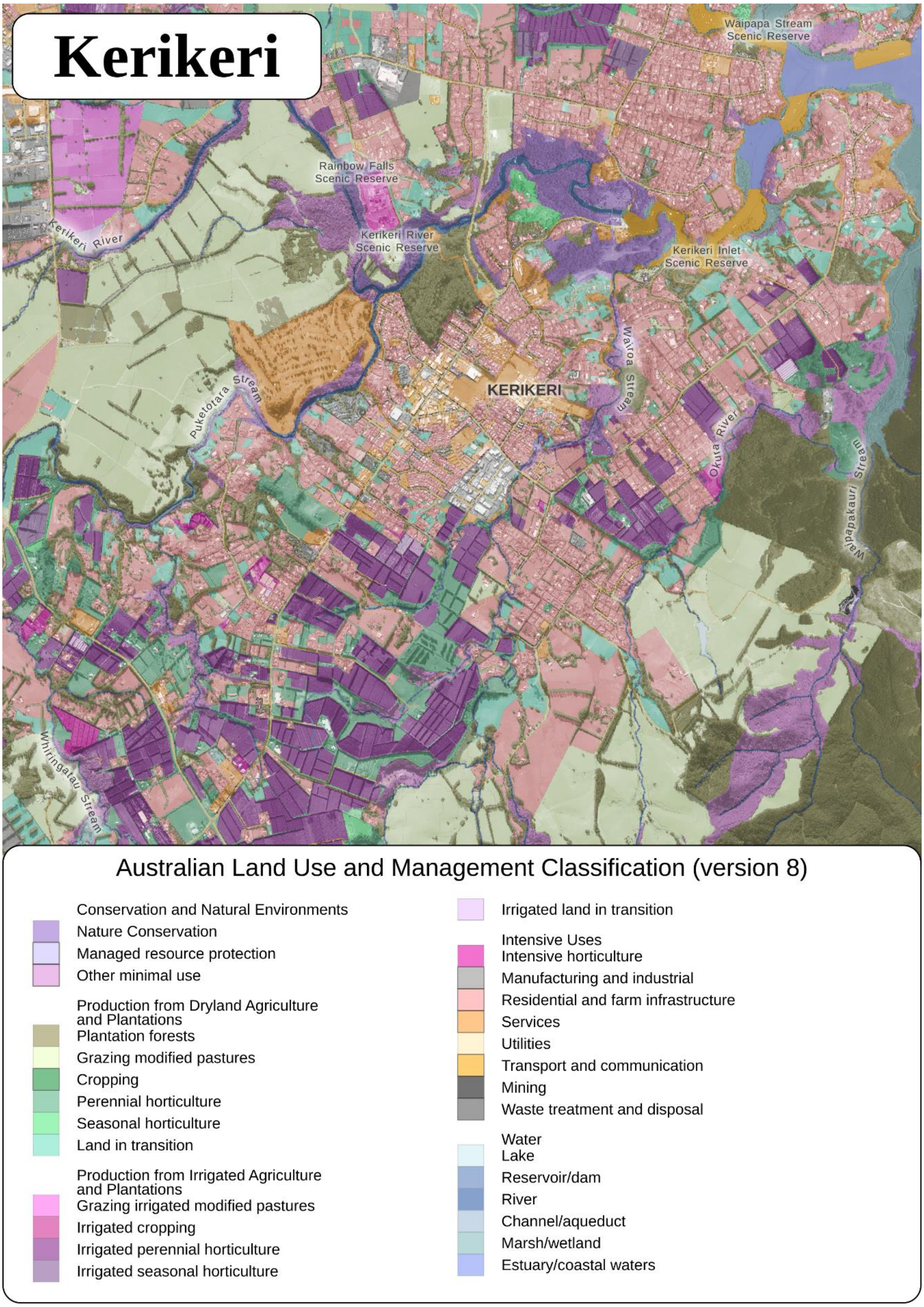


Figure 17. Land-use map of Kerikeri (secondary level).

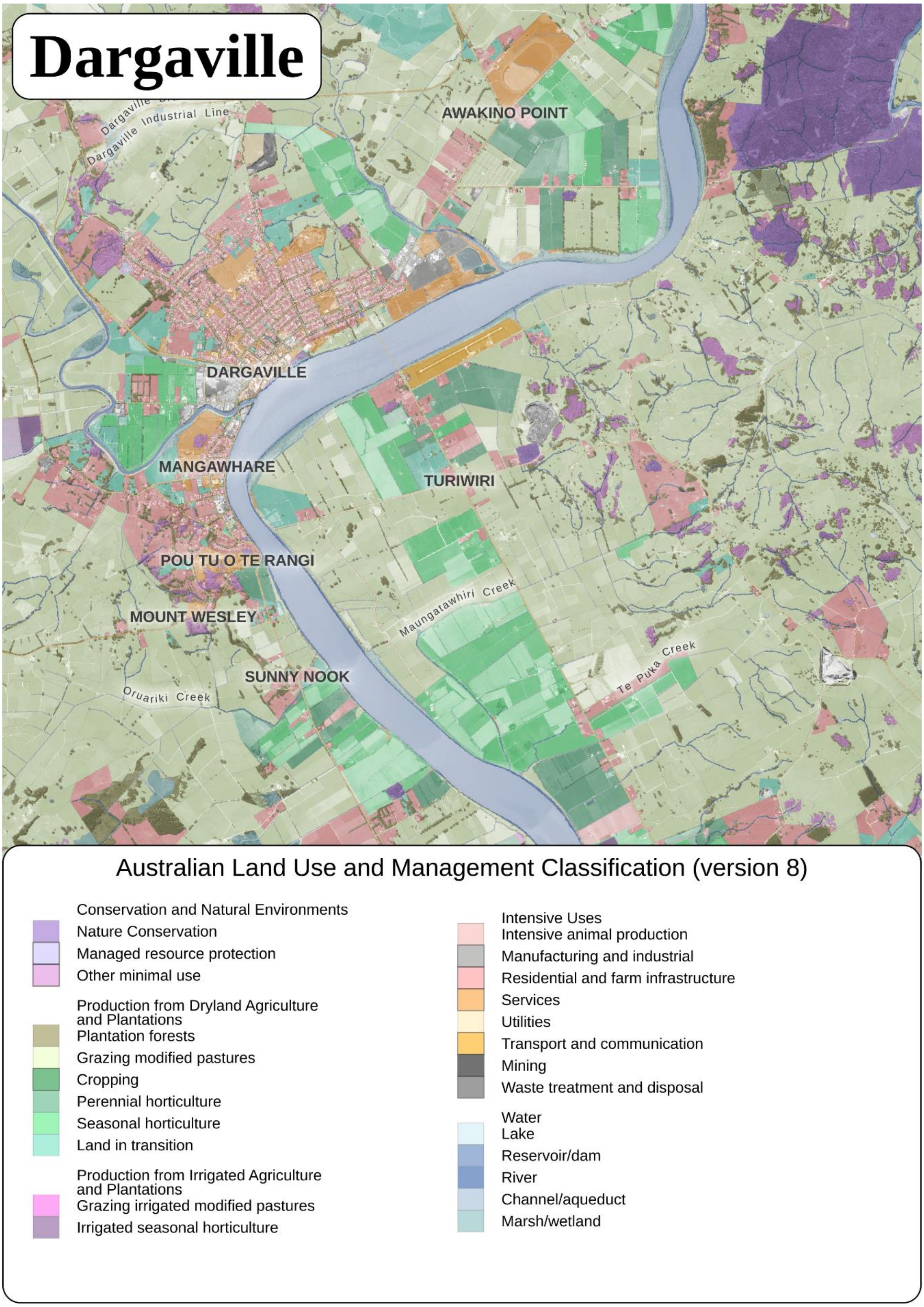


Figure 18. Land-use map of Dargaville (secondary level).

Figure 19. Maps showing different selections of commodities (a–d), management practices (e, f), and custom content embedded in comments (g, h). These are not necessarily mutually exclusive.

7 Assumptions and limitations

This land-use classification is not a map of land tenure. Specifically, we did not intend to capture whether land is owned by the state or by Māori.

Although the general trustworthiness of input data is accounted for through the mechanism of noting a qualitative confidence score and geographical scale, both spatial and categorial errors in input data can be influential in the output. Input data are basically assumed to be correct, although not every feature carries the same confidence weight or can be filtered out. In the future, a hypothetical land-use change analysis may not be able to determine between a real land-use change and one that has come about from corrections to input data collections, especially as many of these collections are maintained by third parties. In the case of spatial error, the same difficulty of change attribution may apply simply if the positional accuracy of input data improves, or if the nominal scale increases.

7.1 Class-wise confidence

ALUM requires that a qualitative confidence score is included for each classified land area. This is a value from 1–4, where 1 is 'very confident', 2 'confident', 3 'reasonably confident', and 4 'not very confident'. It is intended to be attached to the primary land-use classification, and not to commodities or management practices.

The automated process used to produce this classified map relies on these qualitative confidence scores internally: because each cell is potentially classified as every possible class, the final determination is made by performing a database sort operation, with priority given to higher confidence, lower scale, and more recent source data. Because confidence is qualitative and subjective, there are no objective criteria on which a given confidence assignment stands.

The general principle we sought to apply was an expert assessment of the reliability of data (categorially and spatially), sometimes mediated by attribute queries (e.g. to use higher confidence for validated data within a data set), coupled with a focus on corroboration. Corroboration between two sources of data that describe the same or complementary information is considered to increase the confidence of a given classification; for example, we were more confident in mixed-reliability ratings data that declared a land-use of forestry if the canopy height model independently indicated the presence of a canopy above a given height. This is performed per cell, such that a ratings unit with a declared land-use of forestry might have sub-areas with different confidence in the assigned forestry class.

Figure 20 through to Figure 27 visually summarise the assigned confidence values for each class. These are determined for the dissolved cell-level output, not parcels. In brief, there are a few key takeaways that should guide the use of some specific classes.

- Some classes, such as 'Plantation forestry' (3.1.0) and 'Grazing modified pasture' (3.2.0) have a spread of confidence values. Where they are not captured with a high degree of confidence, this indicates a lack of corroborating evidence across multiple datasets, or conflicting evidence.

- Rural residential without agriculture (5.4.3) is known with mid or low confidence and appears to be overstated. Many areas with this classification appear to have significant agricultural activities.
- Rivers are more difficult to confidently identify than other forms of water (class 6). Over time, different delineations of river channels and riverbanks have occurred; as these are dynamic landscape features, there is less confidence when there is less mutual identification, and vice versa, leading to a spread of confidence values.
- Conservation land is confidently identified, especially for land with the most stringent protected, but less so for conservation land that it subject to various activities, for example 'Surface water supply' (1.2.2).

Even when confidence is low, the assigned land-use is still considered the most appropriate classification. Moreover, high confidence is not a guarantee of correctness.

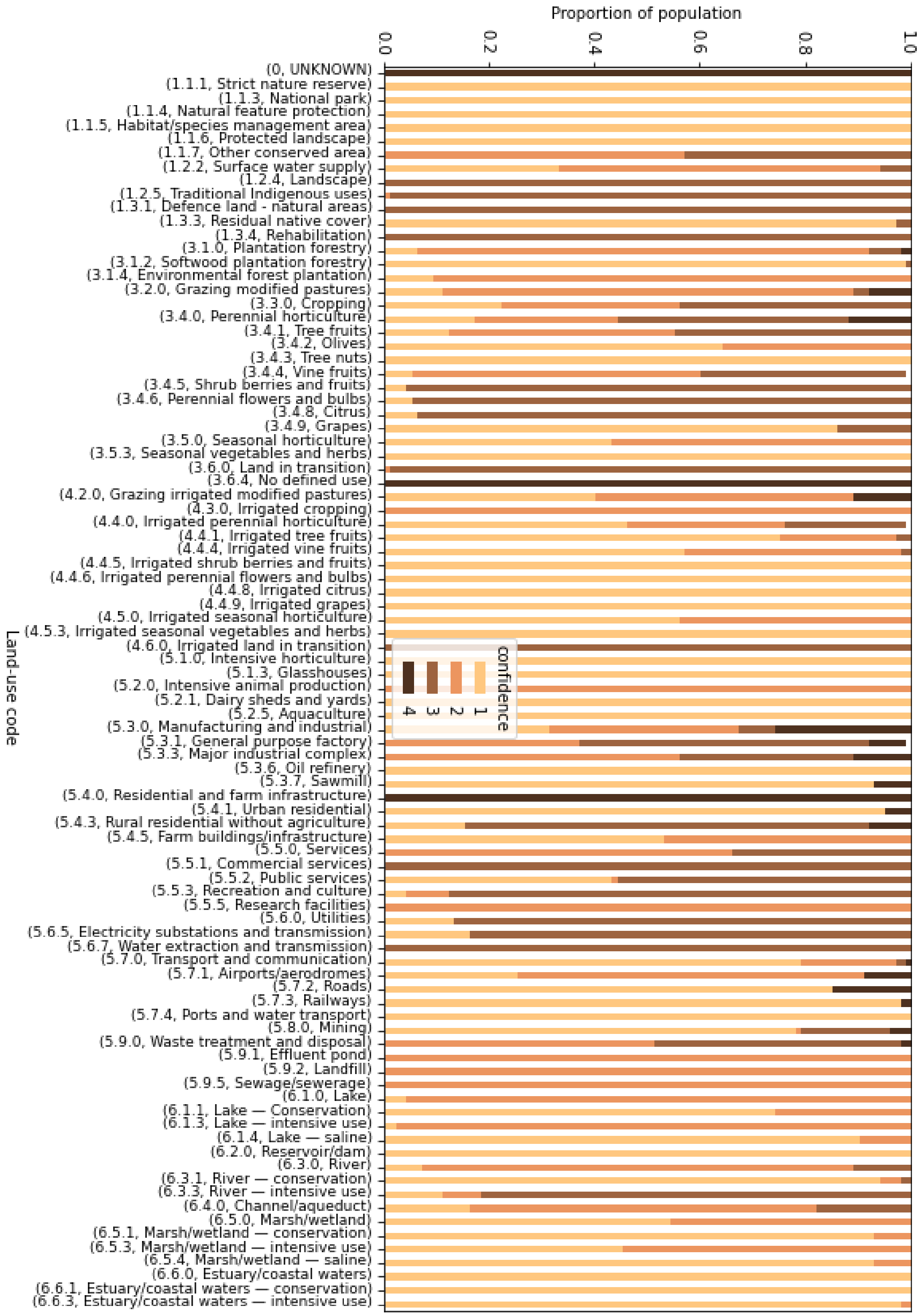


Figure 20. Class-wise confidence values, as a proportion of the land area assigned to the class.

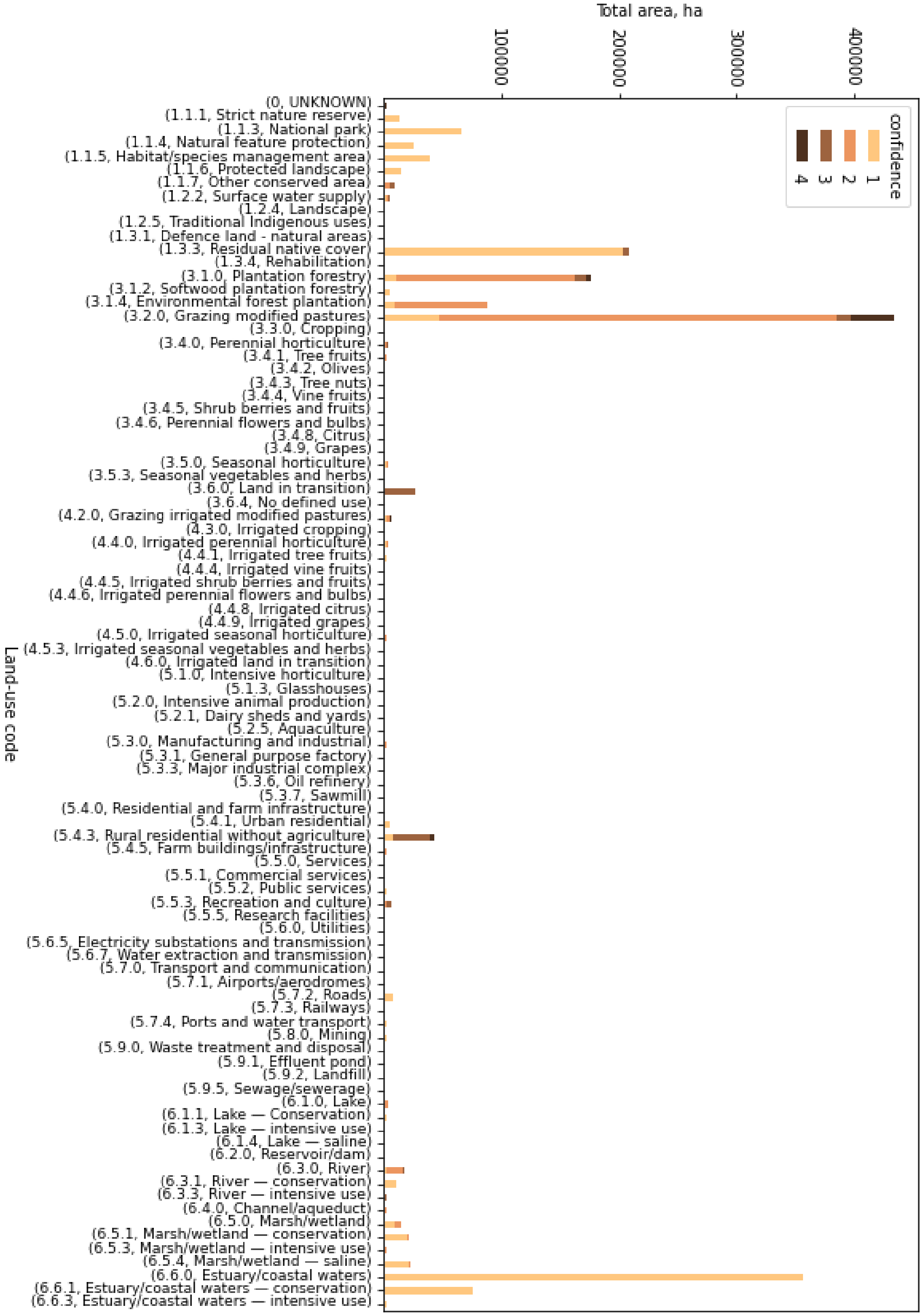


Figure 21. Class-wise confidence values, as a total of the land area assigned to the class.

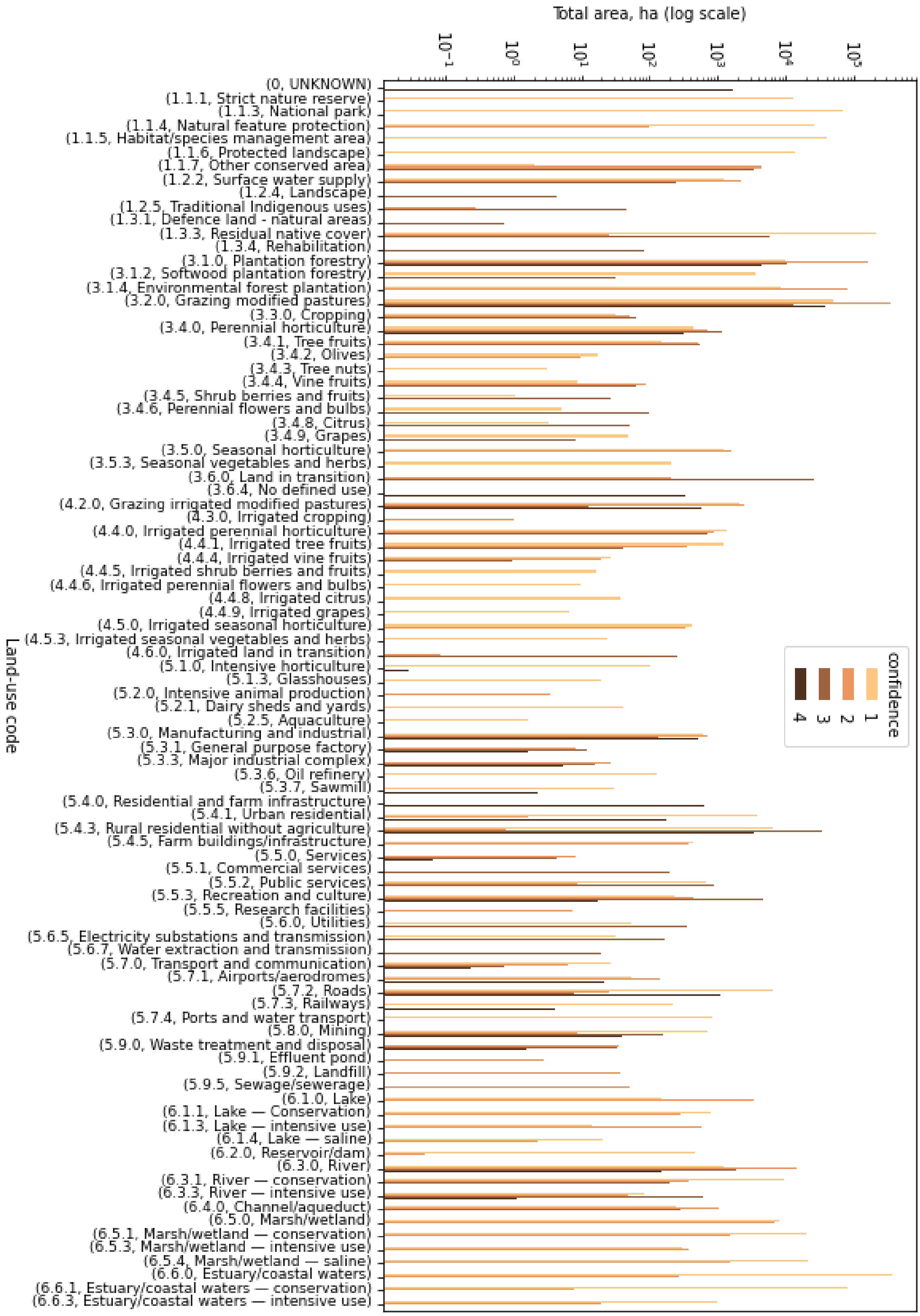


Figure 22. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale.

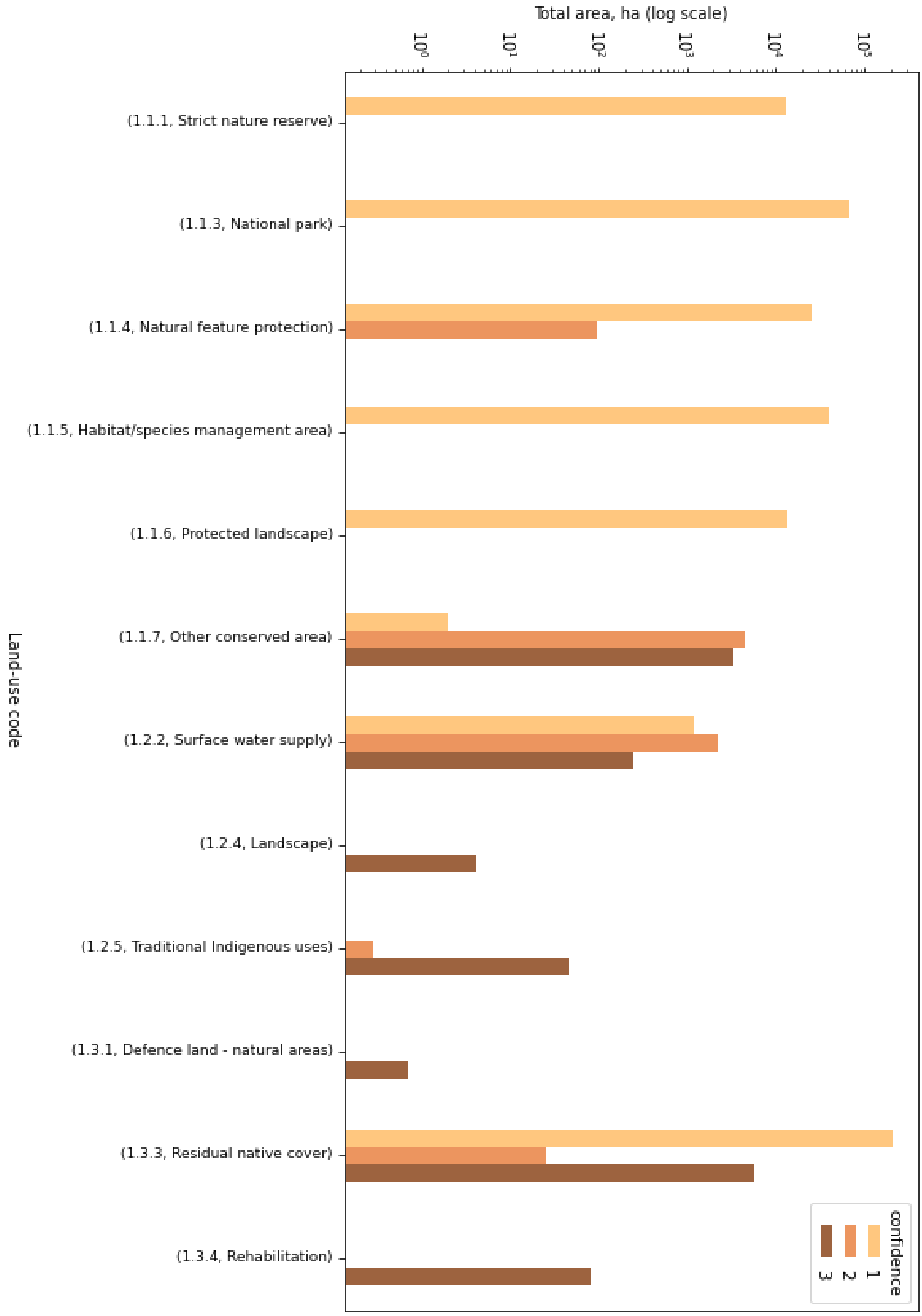


Figure 23. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale, for Class 1 ('Conservation and natural environments').

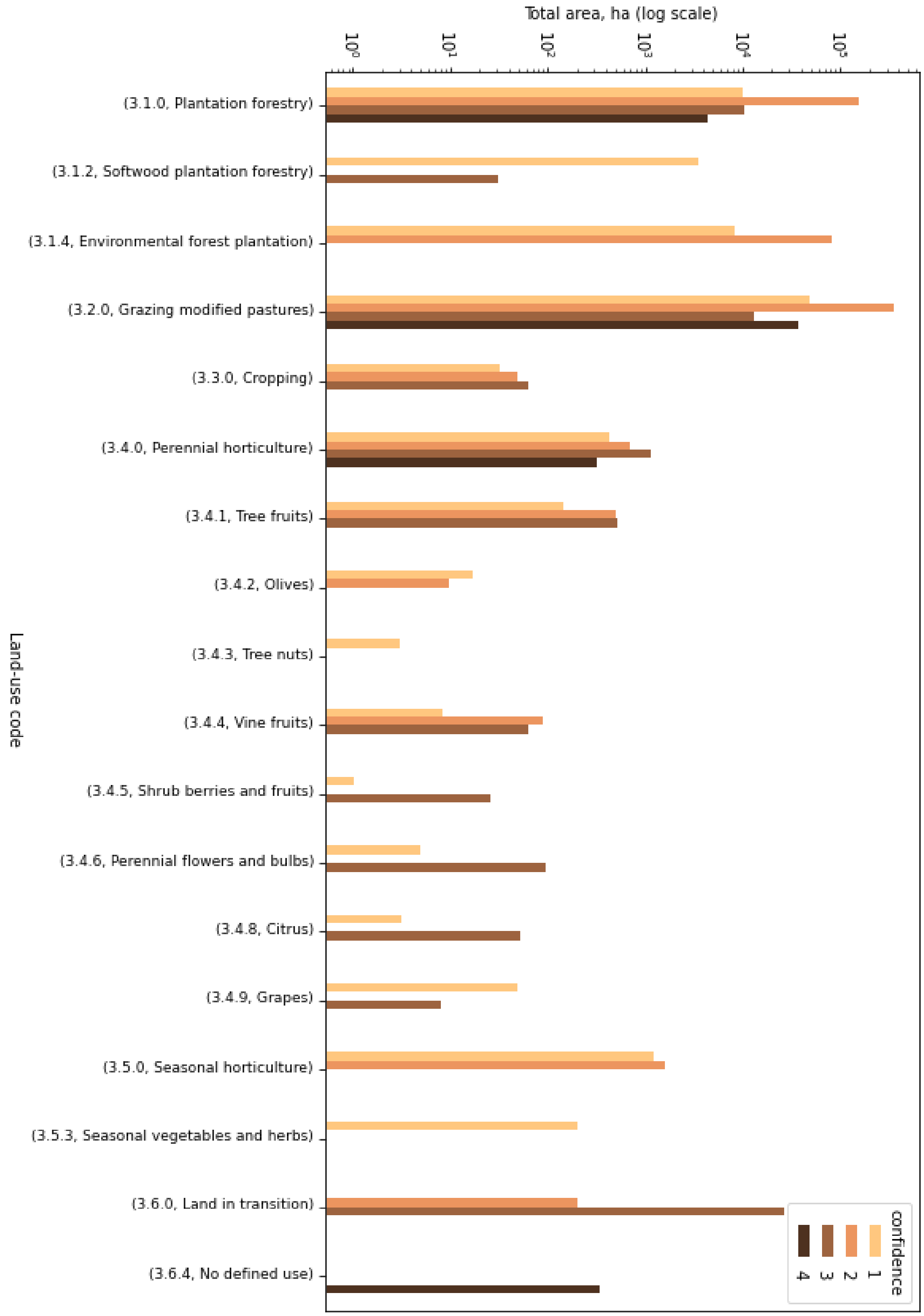


Figure 24. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale, for Class 3 ('Production from dryland agriculture and plantations').

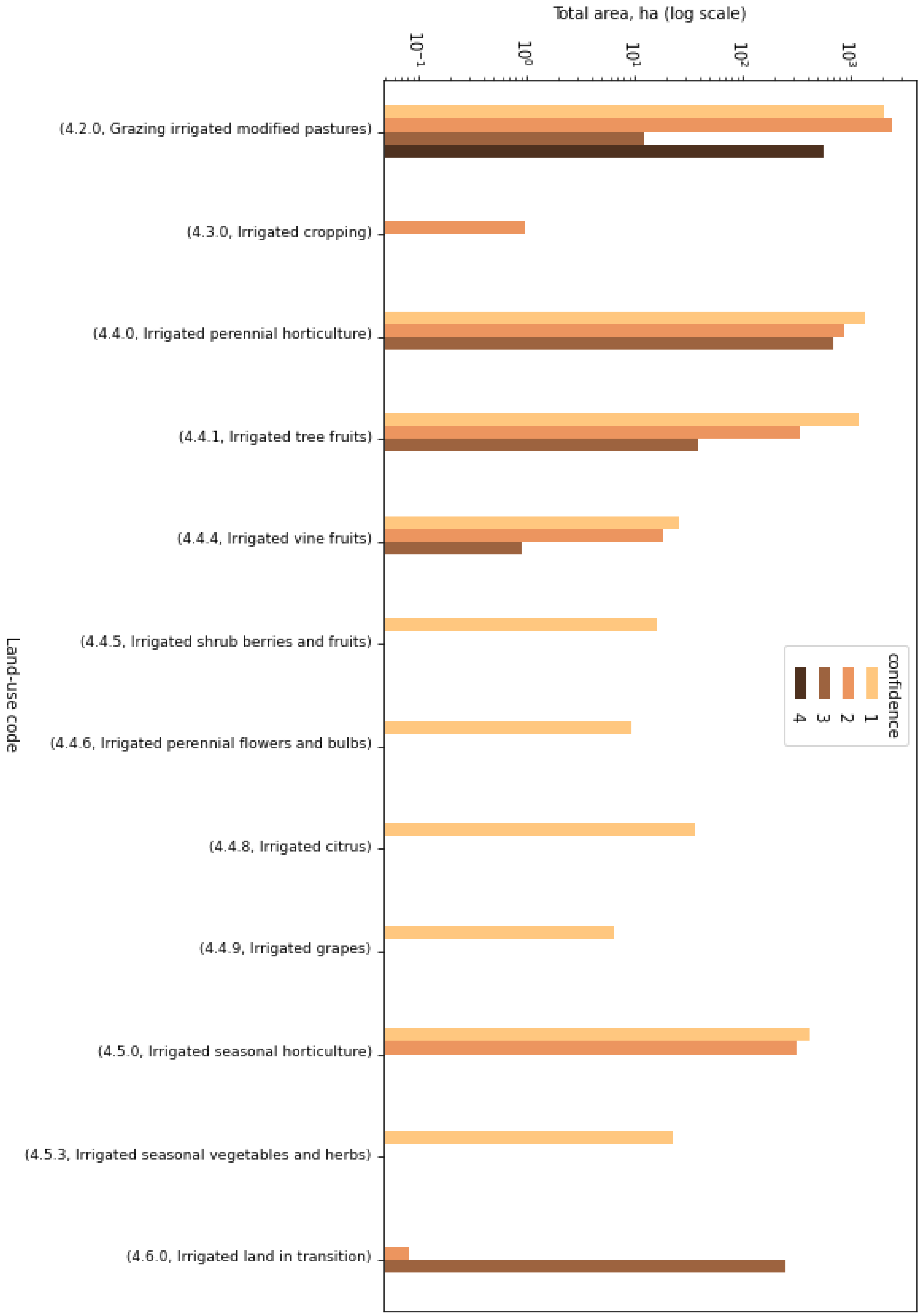


Figure 25. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale, for Class 4 ('Production from irrigated agriculture and plantations').

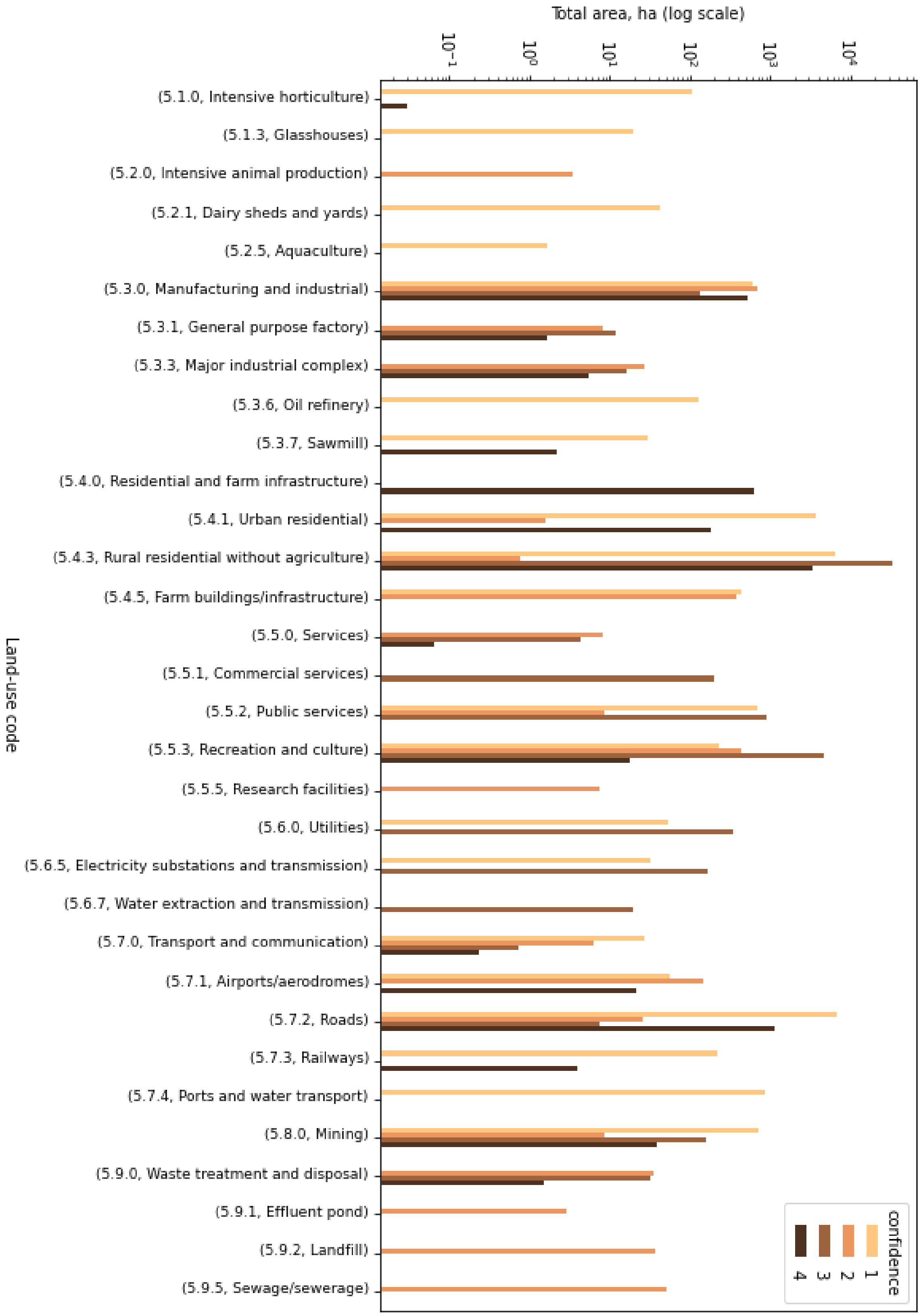


Figure 26. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale, for Class 5 ('Intensive uses').

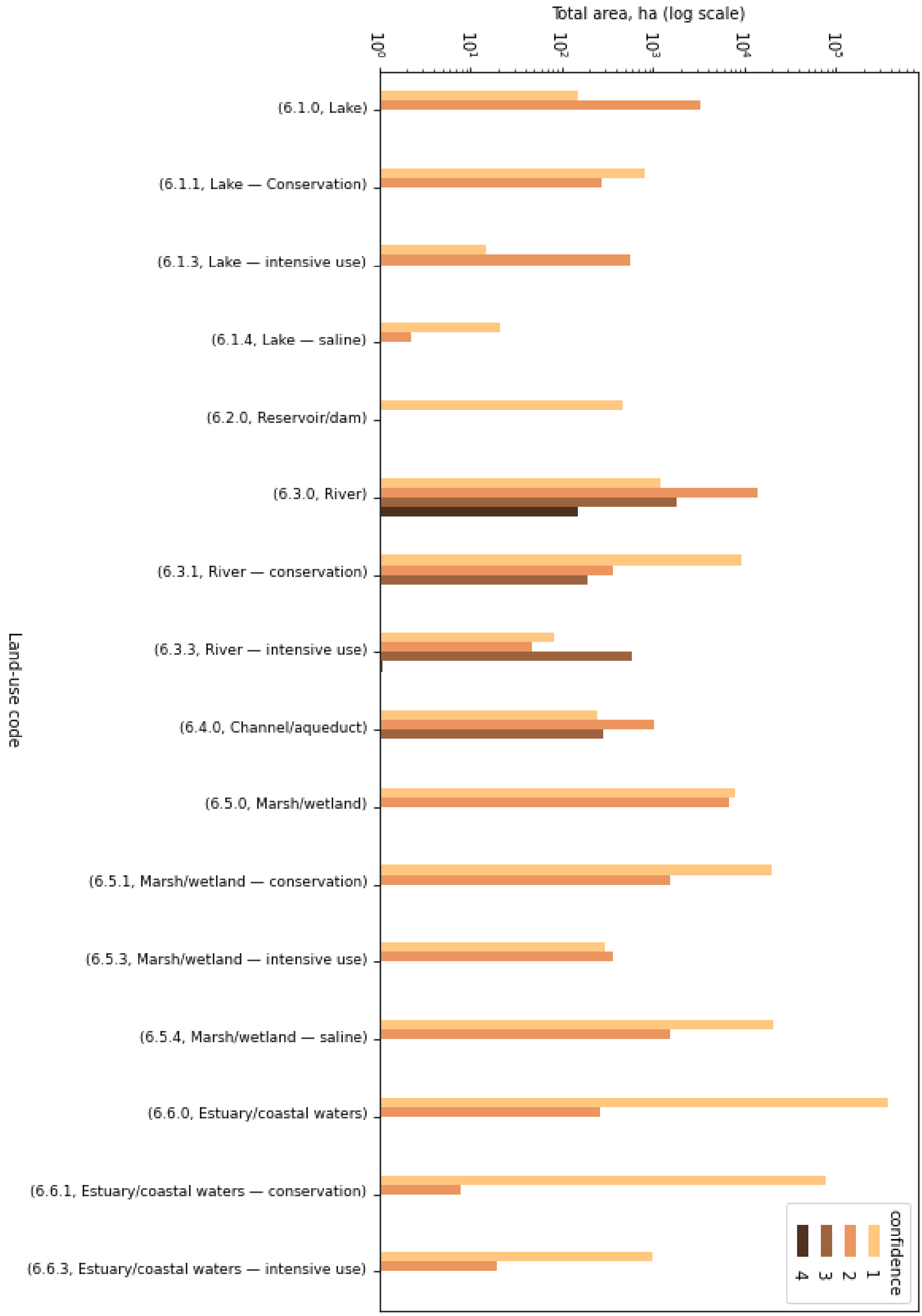


Figure 27. Class-wise confidence values, as a total of the land area assigned to the class, on a log scale, for Class 6 ('Water').

7.2 Validation

True validation of these data requires ground-truthing. This was not possible for this project. Instead, validation was performed against 40 cm aerial imagery taken between 2014 and 2016.[42] This is the most recent available imagery for Northland, but it is already 7–9 years out of date. Depending on the degree of change of a particular land-use class, some errors are likely to be due to the validation source information rather than the assigned land-use class, so it is important to take care when interpreting the results of this validation.

The validation was conducted as a stratified random sample of parcels, ensuring that all classes are represented in proportion to their presence, and that the total size of the sample is approximately 1,000. Two analysts took an equal share of the sample and used an in-house geospatial validation tool called 'mapaccuracy', which presented each analyst with the structured ALUM information (class, confidence, comments, etc.), and allowed the choice of one of three validation classes: true, false, and maybe. This ternary classification was chosen over a binary true/false classification because land-use classes can be ambiguous when presented with only aerial imagery as validation evidence.

The analysts understood that 'true' meant that the assigned class is consistent with the aerial imagery, 'false' that it is not, and 'maybe' that it is unclear either way. For example, a forested area classified as 1.1.1 ('Strict nature reserve') would be *consistent* with the assigned class; a built-up area would not, and some area with an even mix of conflicted classes could be assigned 'maybe'. However, it is not possible to truly *know* (without local knowledge) that a specific area of forest is indeed protected using aerial imagery alone.

The sample is drawn from parcels, and parcels are assigned land-use classes according to a majority (or plurality) rule. This presents difficulties, particularly for land-use classes representing natural areas that are typically sub-parcel in extent (such as lakes), as these are unlikely to have been assigned to represent entire parcels (Figure 28). A hypothetical parcel with 30% pasture cover, 30% forest cover, and 40% lake cover would be assigned to the lake class. Whether the lake class is the correct class to apply to the parcel is moot; parcels are arbitrary boundaries, and many instances of incorrect classification at the parcel scale ultimately come about from this arbitrariness and lack of consideration for the possibility of multiple uses and covers.

[42] Northland 0.4m Rural Aerial Photos (2014-2016) | LINZ Data Service

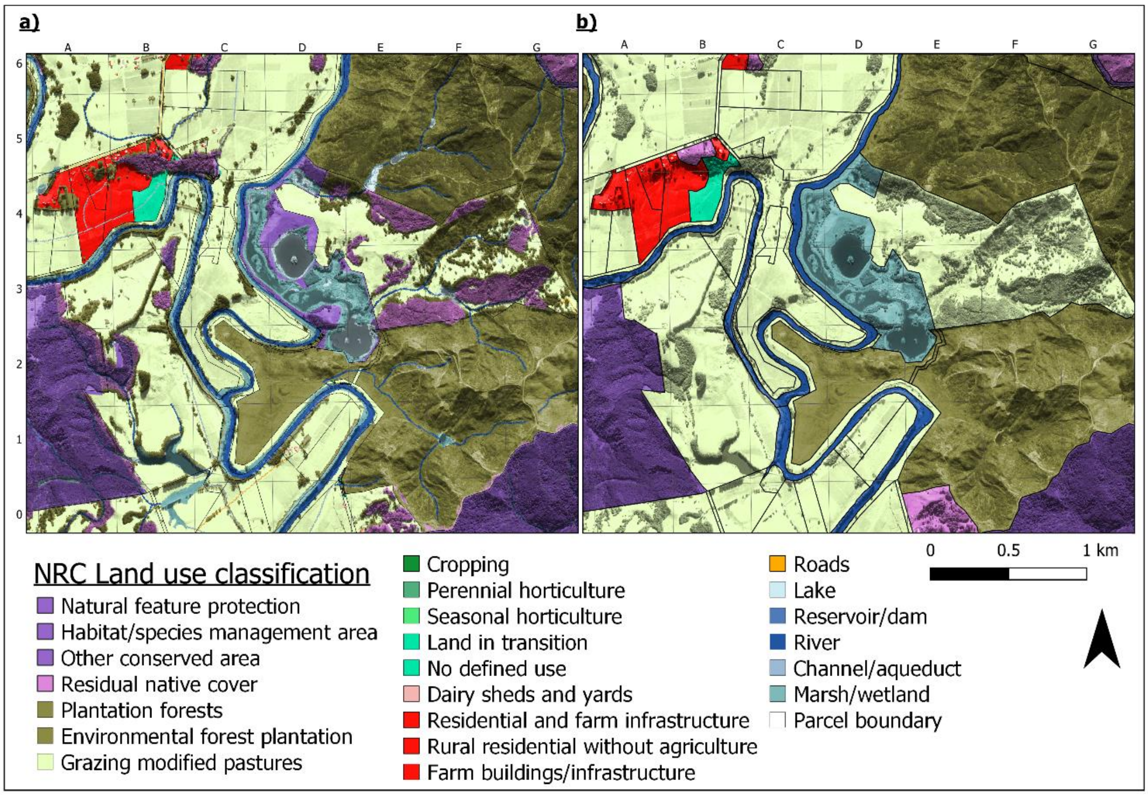


Figure 28. Demonstration of the limitations associated with assessing the parcel-scale classification (b) which is defined as the majority (or plurality) land-use class at the cell-level (a). Note cases such as the lifestyle property ("Rural residential without agriculture") in B5 which is dominated by an area of "Residual native cover" which comes to represent the entire parcel. This is not clearly an incorrect classification, but rather reflects the difficulty of mapping to parcels, which often have arbitrary boundaries.

Using parcels as the sample population for validation also means that exceptionally large parcels have an equal likelihood of being chosen as any other parcel, biasing the validation sample selection towards smaller parcels, which are often more difficult to assess for correctness than larger parcels. One alternative would be to use individual DGGS cells as a validation unit; this was ruled out for the sake of consistency with an earlier validation of preliminary output data, which followed the same protocols, and which had been performed on parcels. For the same reason a second alternative was rejected: to perform a stratified random sample with area weighting such that larger parcels are more likely to be included in the sample.

The validation process does not address what the ideal classification of a parcel is. Rather, the assessment is only concerned with determining whether the assigned class is plausible and consistent with what is apparent in the visible image.

Figure 29 through to Figure 36 show the results of the validation. Following are some key insights about the worst-performing classifications.

- One classes had a complete failure rate: 1.3.1 ('Defence land – natural areas'). This is a class with very low representation in the sample because it is both unusual, and unlikely to meet the parcel-majority criteria. For defence land, we relied on the CRoSL, which, at the time of classification, among other examples of error, declared that some parcels making up Whangarei Airport were managed by the New Zealand Defence Force. This case is probably erroneous, and plausibly explained by the airport being used as a WW II airbase in 1942.

- Other classes, such as 'Landfill' (5.9.2), 'Research facilities' (5.5.5), 'Tree nuts' (3.4.3), and 'Shrub berries and fruits' (3.4.5) also have small populations in the full data set, and only a few parcels met the parcel-majority criterion, whereby most of the parcel is identified as a vineyard or berry cropland. Some of these errors are false negatives, explained by reliance on old aerial imagery used for validation (new orchards have opened, or expanded, since the imagery was taken). The same applies to 'Citrus' (3.4.8), 'Glasshouses' (5.1.3), and others.
- 'Residential and farm infrastructure' (5.4.0) as a class is primarily applied at the scale of buildings; an entire parcel identified as this class suggests an error or corner case. Thus, we would expect much of a random sample of parcels of this class to be erroneous. On the contrary, if a sample were taken of sub-parcel features taken from the non-parcel output, we would expect a very successful validation.
- 'Marsh/wetland' (6.5.0 and including 6.5.1) are the worst performing classes under 'Water' (6.0.0). Largely this is again due to the effect of arbitrary parcel boundaries, which encompass parts of both a legitimate marsh, but also other landscape features including roads and farms. Parcel boundaries often appear to represent the boundaries of natural features (and perhaps they once did), but dynamic landscape features, among which marshes are a perfect example, do not clearly correspond these human-created boundaries and this leads to classification confusion at the parcel level.

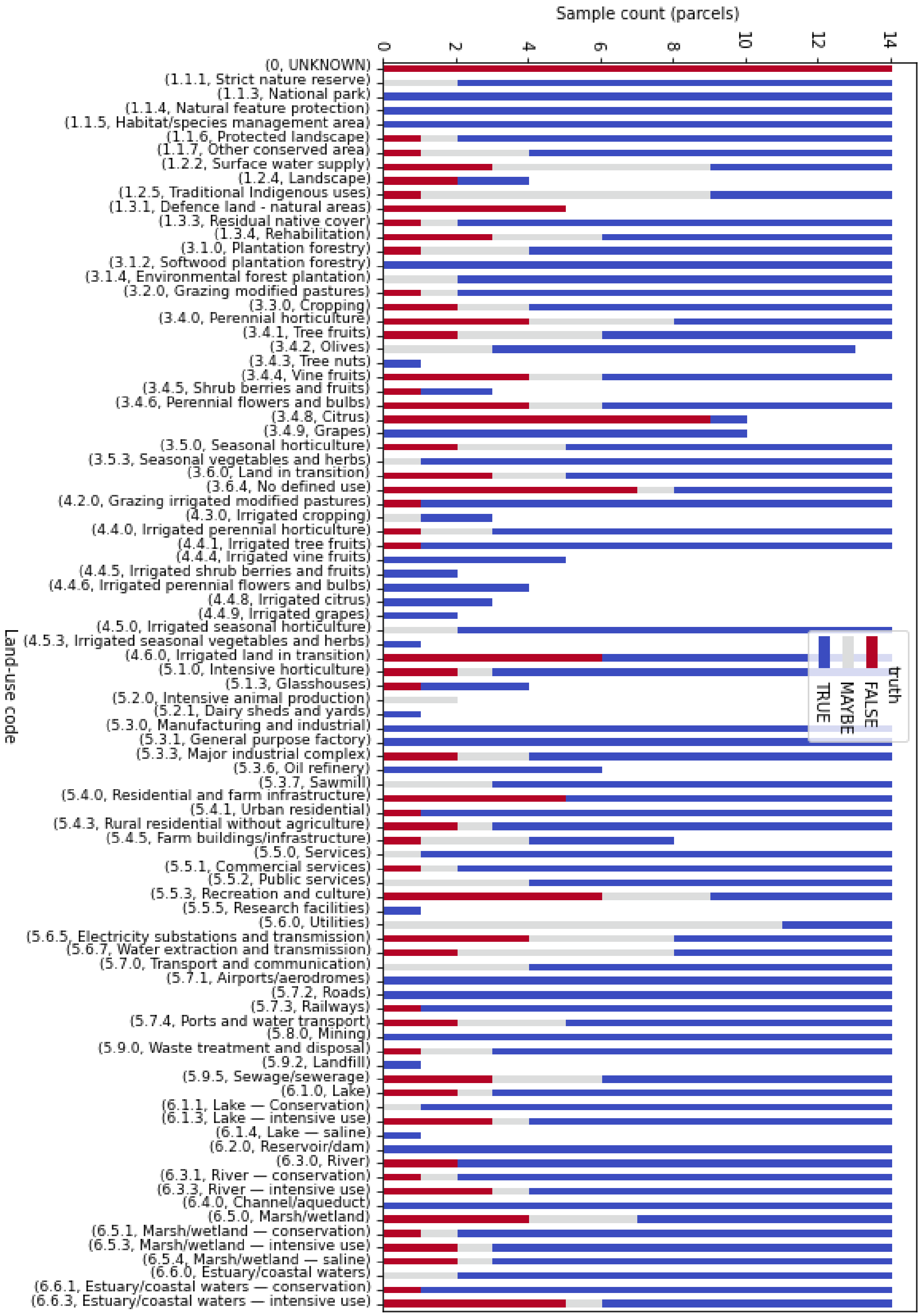


Figure 29. Validation results, by class, as a proportion.

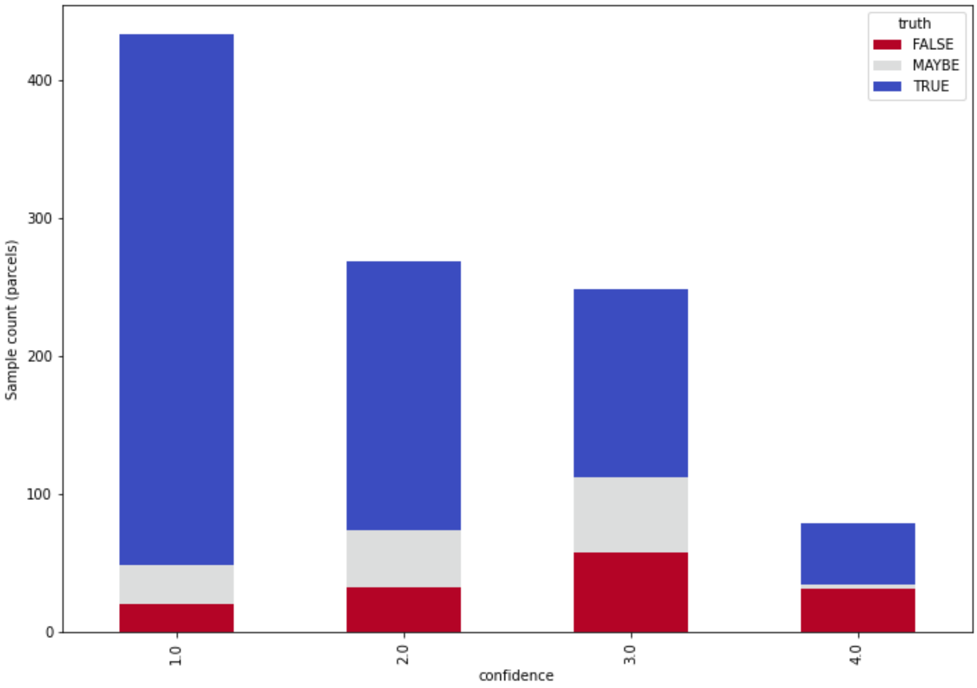


Figure 30. Validation results, by confidence category.

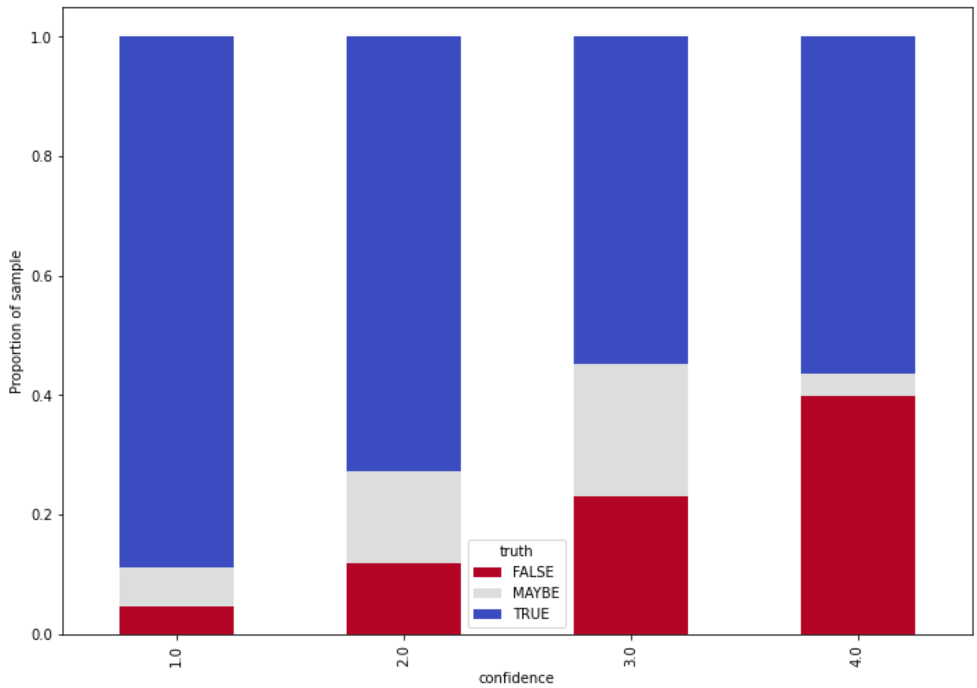


Figure 31. Validation results, by confidence category, proportional.

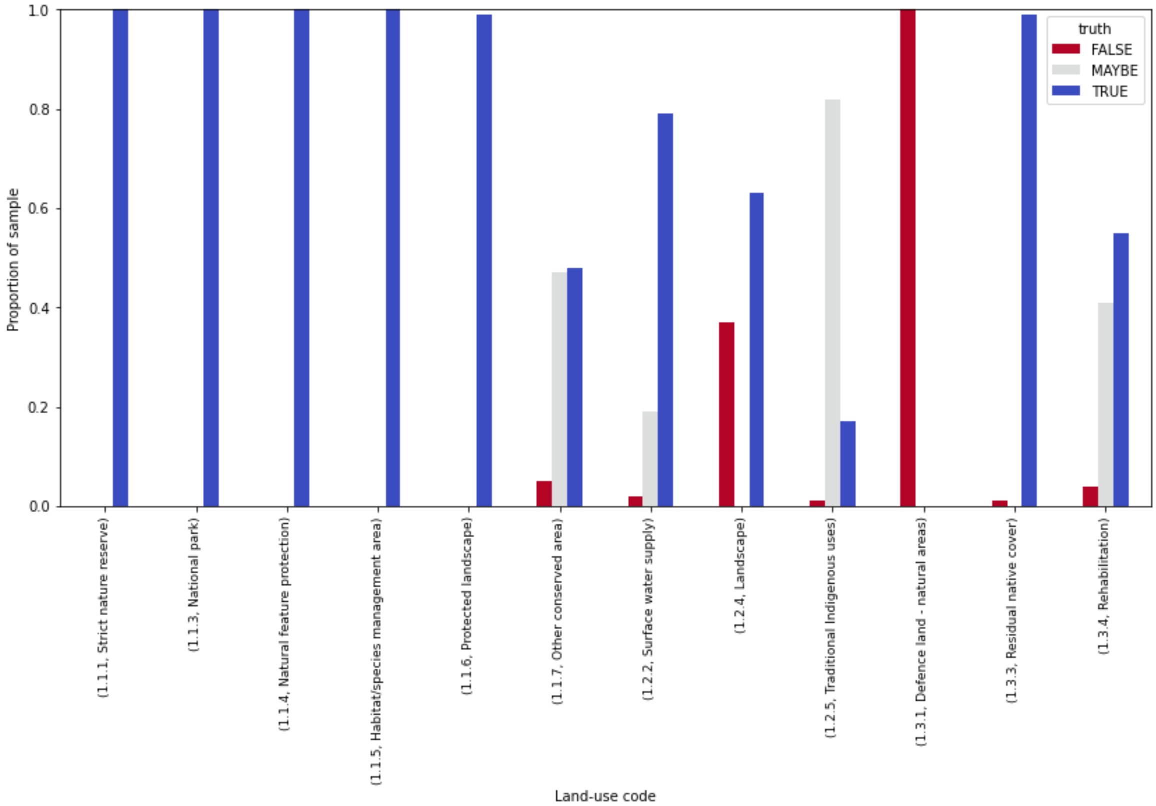


Figure 32. Validation results (proportional), class 1 ('Conservation and natural environments').

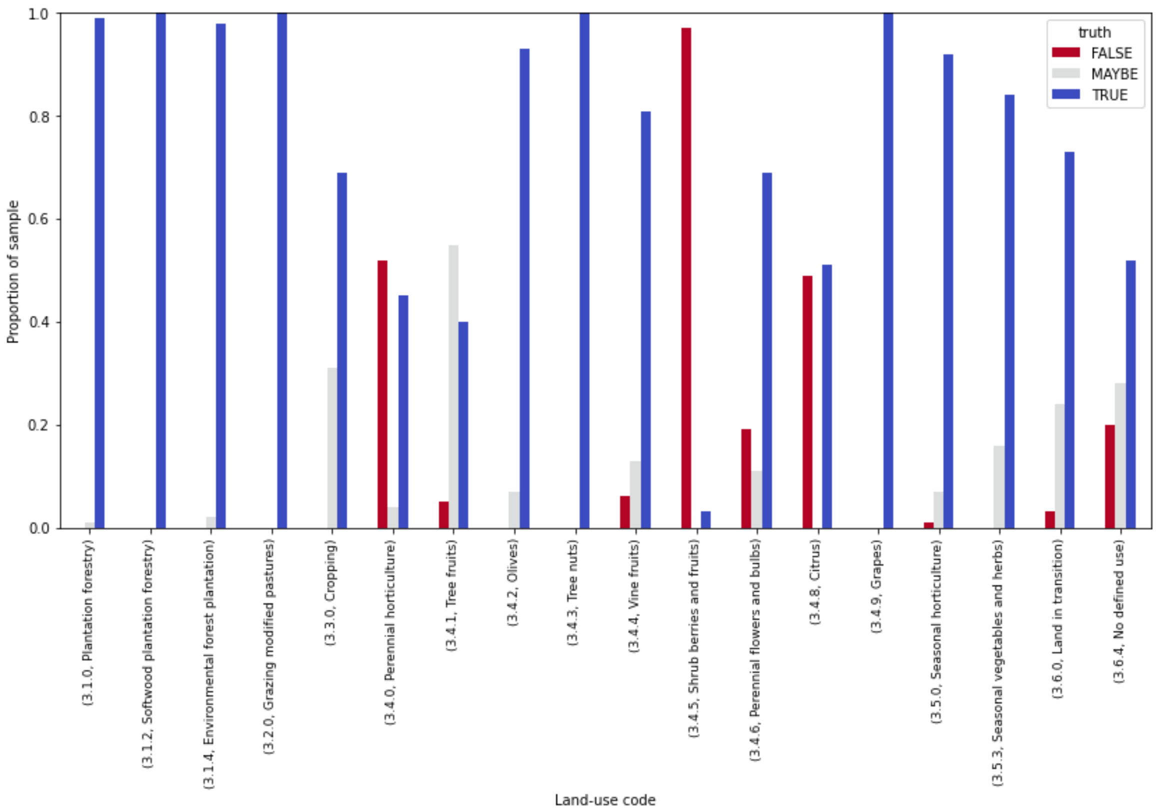


Figure 33. Validation results (proportional), class 3 ('Production from dryland agriculture and plantations').

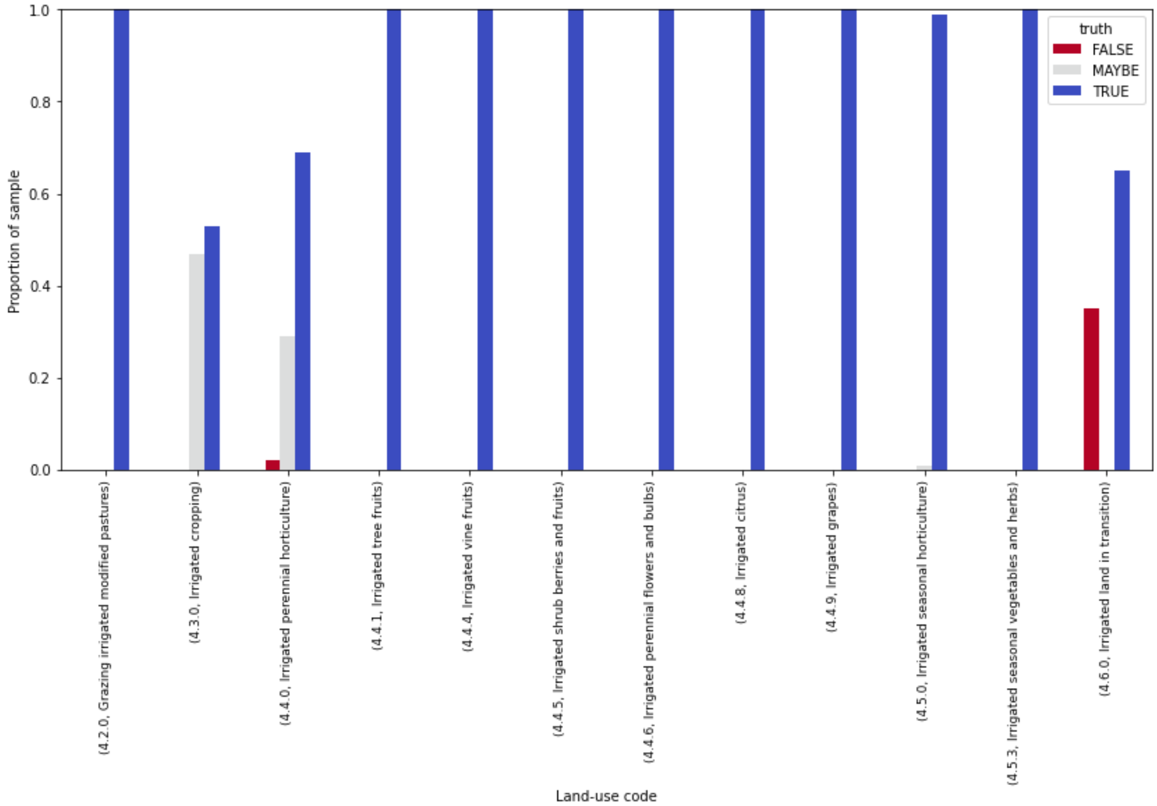


Figure 34. Validation results (proportional), class 4 ('Production from irrigated agriculture and plantations').

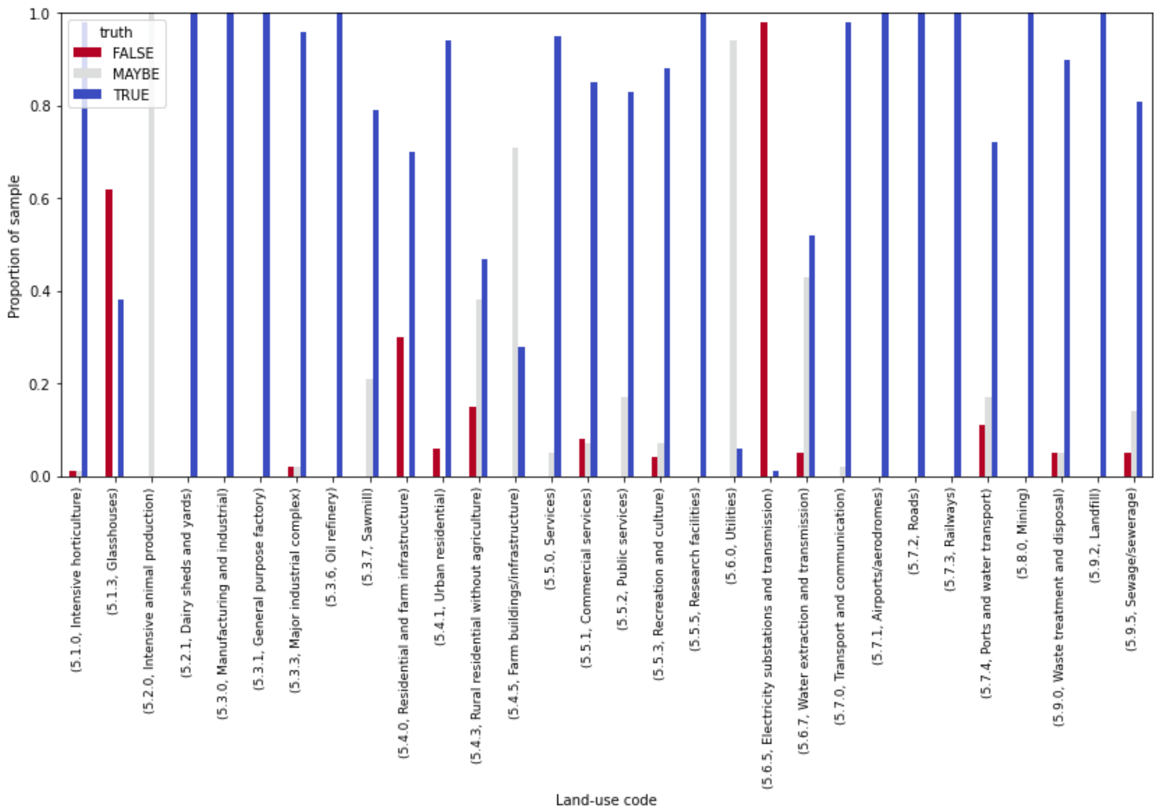


Figure 35. Validation results (proportional), class 5 ('Intensive uses').

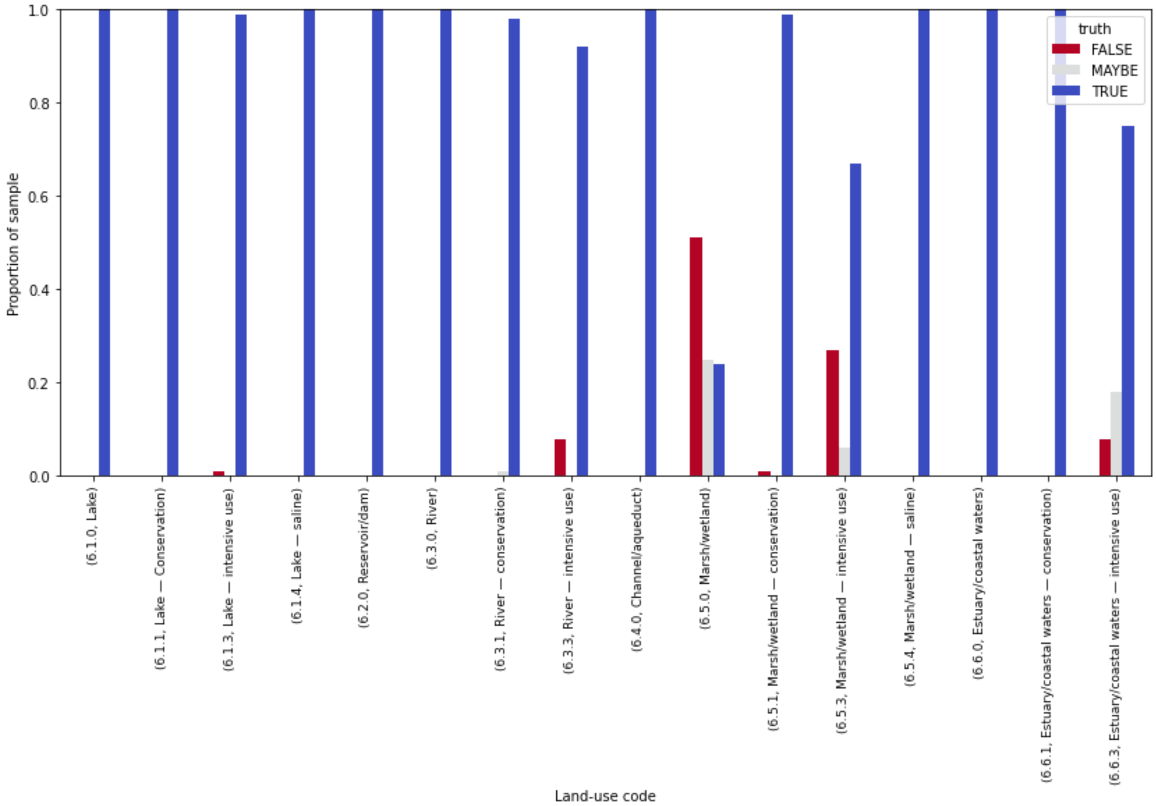


Figure 36. Validation results (proportional), class 6 ('Water').

8 Recommendations

8.1 Curation of a property boundary data set

As noted in section 4.3.4, we recommend that council resources be put towards creating and maintaining a reliable property boundary data set that accurately captures rateable properties as combinations of specific identified LINZ parcels. This is so that consents can be explicitly linked to these property boundaries, rather than only to a point feature such as a reservoir.

8.2 Use of web services for sharing data

We recommend that data sharing prioritise standard services for geospatial data on the web over stand-alone files. Rather than making internal data sets available as files that can be shared (e.g. via Microsoft SharePoint or as email attachments), web services (such as ArcGIS Server with a REST API endpoint, or the Koordinates platform exposing a Web Feature Service endpoint) should be used to share data with external collaborators. Access can be limited with an access-control list to users with specific API keys. These keys, which only give permissions to see a curated list of layers, can then be shared for the duration of a project to access the data. This is intended to ensure data currency and thereby to enable future updates of the land-use map. It also allows corrections, deletions, and additions to the data made during the lifetime of the project to be inherited. This can be particularly useful for data that are changed frequently, such as consents. It also provides an ongoing single source of truth for authoritative data.

8.3 Operationalised manual correction, augmentation, and updates

We suggest that manual augmentation (and correction) be performed by council staff, if resources permit. This is in addition to those updates that could occur in the future as region-wide, point-in-time updates to the land-use data. For example, one could capture the milking system (management practice) from aerial imagery on a case-by-case basis. The objective for the present data set was all-of-region coverage in an automated and repeatable fashion. This meant that manual inspection was out of scope for this contract, but this should be performed to augment the data set and make it more useful for council purposes. It could be done in places that are priorities for management (e.g. priority catchments), with an analyst working through a small subset of the full data set for manual augmentation.

Alternatively, the data could be embedded into systems, such that updates or corrections are performed by council employees who are working in the field, on an *ad hoc* basis. In the future, the augmented land-use data could be used as an input for an updated land-use map in a process that respects existing confidence levels (high, due to manual inspection) and recency of update.

One example of this approach is Bay of Plenty Regional Council, who also make parcel-scale land-use information available via a public viewer.[43]

8.4 Use the confidence values to prioritise data collection

Although the confidence values are qualitative and subjective, they are potentially very valuable for the purposes of prioritising amendments. This can be performed for specific classes of interest, or generally across an area of interest.

Overall, class 3.1 ('Plantation forests') is not determined with high confidence. Major sources of these data are LUCAS LUM, which is produced at only a 1:50,000 scale, and ratings data, which are inherently low-confidence because they provide classifications for entire ratings units when only part of a property may be in forest. Plantation forestry as a target for remote sensing is also challenging due to its inherent dynamism, with saplings difficult to discern from regenerating native bush, and the impossibility of determining – on aerial or satellite evidence alone – whether a harvested forest will be returned to forestry or represents an actual change in land-use.

An alternative frame of reference would be to select sub-regions of interest, such as priority catchments, and focus attention on manually identifying the correct or most appropriate land-use for an area of land, beginning with land that has the lowest confidence classification. In this way, a modest expenditure on manual labour can efficiently deliver high-value improvements to the data.

---

[43] https://gis.boprc.govt.nz/BayMaps/?appid=defe8819f0274d008c15b733563356f1

9 Acknowledgements

We sincerely thank Northland Regional Council for their hospitality during a workshop held at the midpoint of the development of the land-use layer.

We thank Warwick Simpson for his assistance locating kūmara farms in Northland.

Christine Harper and Laise Harris have been instrumental in offering advice and personal connections to ensure the success of the project.

Robert Gibb deserves special mention for his pioneering work on DGGSs while at MWLR, which has been, and will continue to be, very influential within GIS (Gibb 2016).

The work of two Dutch geomedia and design internship students from the Aeres University of Applied Sciences (AUAS), Anouk Witteman and Carlijn Zwartjens, in developing a cartographic style for ALUM data that is more appropriate to New Zealand, and applying this in various works of cartography, is greatly appreciated. ArcGIS style files included with the land-use data were made by these two interns.

10 Glossary

ALUM: Australian Land Use and Management Classification, Version 8.

Australian and New Zealand Standard Industrial Classification (ANZSIC): a system of industrial classification developed jointly by the Australian Bureau of Statistics and Statistics New Zealand. The 2006 revision is intended for alignment with the International Standard Industrial Classification of All Economic Activities (Revision 4) (ABS 2013).

Beef + Lamb New Zealand Ltd: the farmer-owned, industry organisation representing New Zealand's sheep & beef farmers. '[It] invest[s] farmer levies in programmes that grow the sheep and beef industry and provide sustainable returns now and for future generations.' (Beef + Lamb NZ Ltd. 2023a).

CHM: canopy height model, a measurement of trees, buildings and other objects that are above the ground topography. It can be calculated by subtracting a digital terrain model from a digital surface model, but in this project we also excluded building footprints such that our canopy height model explicitly aims to capture the height of the tree canopy above ground topography.

CPU: central processing unit.

CRoSL: Central Record of State Land, which identifies all Crown land in New Zealand and provides contextual information on the land, such as parcel/title information, ownership information, and legislative information. License: Attribution 4.0 International (CC BY 4.0). See https://www.linz.govt.nz/products-services/data/types-linz-data/crown-land - data/central-record-state-land

DEM: digital elevation model.

DGGS: discrete global grid system, a hierarchical, recursive, and space-partitioning mosaic used as a geospatial data structure that transcends the raster and vector models for spatial information. (See Sahr et al. 2003.)

Downslope distance to stream (DDTS): the distance from each grid cell in a raster to its nearest channel cell, measured along the downslope path. This is properly calculated on a digital elevation model that has had hydrological corrections applied (such as breaching depressions).

geocoding: a process whereby a text description of a place (e.g. a postal address) is fed into a function that returns corresponding geographical coordinates.

GIS: geographic information system, software for the storage, manipulation, analysis, transformation, and visualisation of digital geographical information.

HAIL: 'The Hazardous Activities and Industries List, a compilation of activities and industries that are considered likely to cause land contamination resulting from hazardous substance use, storage or disposal. The HAIL is intended to identify most situations in New

Zealand where hazardous substances could cause, and in many cases have caused, land contamination. The HAIL groups similar industries together, which typically use or store hazardous substances that could cause contamination if these substances escaped from safe storage, were disposed of on the site, or were lost to the environment through their use' (Ministry for the Environment 2021).

Height above nearest drainage (HAND) (or elevation above stream): a terrain index that measures the relative height of each pixel in a raster to its nearest stream cell, measured along the downslope flow path. This is properly calculated on a digital elevation model that has had hydrological corrections applied (such as breaching depressions). (See Rennó et al. 2008.)

IUCN: International Union for Conservation of Nature.

Land-use: 'the purpose to which the land cover is committed. Some land uses, such as agriculture, have a characteristic land cover pattern. These usually appear in land cover classifications. Other land uses, such as nature conservation, are not readily discriminated by a characteristic land cover pattern. For example, where the land cover is woodland, the land use may be timber production, grazing or nature conservation' (ABARES 2021).

LCDB: Land Cover Database version 5.0, New Zealand. (See https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/)

LiDAR: an acronym for 'light detection and ranging', a technology for remote sensing in three dimensions by using laser scanning, applied in this case from airborne sensors and used to make high-resolution terrestrial elevation models, canopy-height models, and derived data products (as in hydrology).

LRIS: Land Resource Information System portal (https://lris.scinfo.org.nz.).

LUCAS LUM: Land Use and Carbon Analysis System Land Use Map, a land-use map intended primarily to map four key types of 'woody' land-use, intended for meeting the requirements of international reporting of carbon stocks. (See Newsome et al. 2018.)

MfE: Ministry for the Environment.

MWLR: Manaaki Whenua – Landcare Research.

NIWA: National Institute for Water and Atmospheric Research.

NRC: Northland Regional Council.

NRC Selected Land Use Register: the local government register of significant land-use that uses the Hazardous Activities and Industries List (HAIL). Parcels are identified as relating to one or more of the activities on the HAIL, which can give an indication of the appropriate land-use classification under ALUM.

ODbL: Open Data Commons Open Database License: a share-alike licence for data, used by OpenStreetMap. (See https://www.openstreetmap.org/copyright and https://opendatacommons.org/licenses/odbl/)

OSM: OpenStreetMap, a free, collaborative, and open global geographical database, which is maintained by a community of volunteers.

Overpass QL: Overpass Query Language, a procedural, imperative programming language written with a C-style syntax, which is used for querying the OpenStreetMap database and producing output files in several different formats. (See https://github.com/drolbr/Overpass -API)

RAM: random-access memory.

Relational database management system (RDBMS): a database management system based on a relational model of data management, whereby data are organised as tuples with IDs that allow relationships to be expressed. In the present case, a DGGS cell ID is a ready-made ID for expressing a direct spatial relationship between different data sets.

SQL: Structured Query Language, a domain-specific language for manipulating data held in a relational database management system. In this case we store data in PostgreSQL and our land-use classification is expressed entirely in SQL, with supplementary configuration files used to establish which inputs must be obtained (automatically) so that the SQL can then be executed. Because SQL is declarative, we can express what output we want based on our inputs, but it is the relational database management system that works out how best to execute the query. In other words, a human uses SQL to express what to obtain from a database, but not how to retrieve it.

technical layer: the highest-resolution view of the land-use data that has been produced. Whereas the primary land-use layer is presented at a parcel scale (each parcel has one assigned class), the technical layer records individual DGGS cells and additional attribute information (such as elevation from a digital elevation model).

WONI: Wetlands of National Importance.

11 References

ABARES 2016. The Australian Land Use and Management classification (Version 8). Canberra, ACT, Australian Bureau of Agricultural and Resource Economics and Sciences.

ABARES 2021. ACLUMP: definitions. https://www.agriculture.gov.au/abares/aclump/definitions

ABS 2013. Australian and New Zealand Standard Industrial Classification (ANZSIC). https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-industrial-classification-anzsic

Ardo J, Law R 2023a. Raster2dggs (v0.2.5) [Computer software]. https://github.com/manaakiwhenua/raster2dggs

Ardo J, Law R 2023b. Vector2dggs (v0.5.2) [Computer software]. https://github.com/manaakiwhenua/vector2dggs

Baxter J, Russell L 1994. Land use mapping requirements for natural resource management in the Murray-Darling Basin. Project M305: Task, 6.

Beef + Lamb NZ Ltd 2023a. Company profile. https://beeflambnz.com/about-blnz/company-profile

Beef + Lamb NZ Ltd 2023b. Data & tools. https://beeflambnz.com/data-tools/sheep-beef-farm-survey

Dymond J, Sabetizade M, Newsome P, Harmsworth G, Ausseil A-G 2021. Revised extent of wetlands in New Zealand. New Zealand Journal of Ecology. https://doi.org/10.20417/nzjecol.45.32

Gibb RG 2016. The rHEALPix Discrete Global Grid System. IOP Conference Series: Earth and Environmental Science 34: 012012. https://doi.org/10.1088/1755-1315/34/1/012012

Herzig A, Manderson A, Jolly B, Barnes M, Baish L 2020. Opportunities and constraints on intensive land-use expansion in the Horizons Region – Land-Use Mapping (Tech. Rep). Horizons Regional Council.

Hojati M, Robertson C, Roberts S, Chaudhuri C 2022. GIScience research challenges for realizing discrete global grid systems as a Digital Earth. Big Earth Data 6(3): 358–379. https://doi.org/10.1080/20964471.2021.2012912

Jahn D 2023. H3pandas (0.2.4) [Computer software]. https://github.com/DahnJ/H3-Pandas

Knudsen Z 2023. H3-pg [Computer software]. https://github.com/zachasme/h3-pg

Kukutai T, McIntosh T, Boulton A, Durie M, Foster M, Hutchings J, Mark-Shadbolt M, Moewaka Barnes H, Moko-Mead T, Paine S-J, Pitama S, & Ruru J. (2021). Te Pūtahitanga: A Tiriti-led science-policy approach for Aotearoa New Zealand. Auckland: Ngā Pae o te Māramatanga.

Lesslie R, Barson M, Smith J 2006. Land use information for integrated natural resources management: a coordinated national mapping program for Australia. Journal of Land Use Science 1(1): 45–62. https://doi.org/10.1080/17474230600605244

Lindsay JB 2016a. The practice of DEM stream burning revisited. Earth Surface Processes and Landforms 41(5): 658–668. https://doi.org/10.1002/esp.3888

Lindsay JB 2016b. Whitebox GAT: A case study in geomorphometric analysis. Computers & Geosciences 95: 75–84. https://dx.doi.org/10.1016/j.cageo.2016.07.003

Ministry for the Environment 2021. Hazardous Activities and Industries List. https://environment.govt.nz/publications/hazardous-activities-and-industries-list-hail/#what-is-the-hail

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, et al. 2021. Sustainable data analysis with Snakemake. F1000Research 10: 33. https://doi.org/10.12688/f1000research.29032.2

Newsome P, Shepherd J, Pairman D, Belliss S, Manderson A 2018. Establishing New Zealand's LUCAS 2016 Land Use Map. Wellington, Ministry for the Environment.

North H, Amies A, Dymond J, Belliss S, Pairman D, Drewry J, Schindler J, Shepherd J 2022. Mapping bare ground in New Zealand hill-country agriculture and forestry for soil erosion risk assessment: an automated satellite remote-sensing method. Journal of Environmental Management 301: 113812. https://doi.org/10.1016/j.jenvman.2021.113812

Northland Regional Council, Toitū Te Whenua Land Information New Zealand (LINZ) (2022). Northland, New Zealand 2018-2020. Collected by RPS Group, distributed by OpenTopography and LINZ. https://doi.org/10.5069/G9BR8QDQ.

Office of the Valuer-General 2010. Rating valuations rules 2008. (LINZS30300).

OpenTopography 2022. Northland, New Zealand 2018-2020. https://doi.org/10.5069/G9BR8QDQ

Pearson L, Couldrey M 2016. Methodology for GIS-based land use maps for Southland. Environment Southland.

Purss MBJ, Peterson PR, Strobl P, Dow C, Sabeur ZA, Gibb RG, Ben J 2019. Datacubes: a discrete global grid systems perspective. Cartographica: The International Journal for Geographic Information and Geovisualization 54(1): 63–71. https://doi.org/10.3138/cart.54.1.2018-0017

Rennó CD, Nobre AD, Cuartas LA, Soares JV, Hodnett MG, Tomasella J, Waterloo MJ 2008. HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment 112(9): 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018

Rutledge DT, Price R, Briggs C, Cowell S 2009. Geospatial land-use classification for New Zealand: review and recommendations. Statistics New Zealand.

Sahr K, White D, Kimerling AJ 2003. Geodesic discrete global grid systems. Cartography and Geographic Information Science 30(2): 121–134. https://doi.org/10.1559/152304003100011090

Tarboton DG 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research 33(2): 309–319. https://doi.org/10.1029/96WR03137

Uber Engineering 2023. H3: a hexagonal hierarchical geospatial indexing system (4.1.0) [Computer software]. h3geo.org.

Wang L, Liu H 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science 20(2): 193–213. https://doi.org/10.1080/13658810500433453