



Manaaki Whenua
Landcare Research

The NZ Land Use Classifier

Prepared for: Ministry of Business, Innovation & Employment as an output from the Innovative Data Analysis (IDA) research programme

July 2018



The NZ Land Use Classifier

Contract Report: LC3335

Andrew Manderson, Ben Jolly, Anne-Gaelle Ausseil

Manaaki Whenua – Landcare Research

Reviewed by:

Alex Herzig
Geospatial Modeller
Manaaki Whenua – Landcare Research

Approved for release by:

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

Disclaimer

This information may be copied and distributed to others without limitation, provided Landcare Research New Zealand Ltd are acknowledged. Under no circumstances may a charge be made for this information without the written permission of Landcare Research.

Contents

Summary.....	v
1 Introduction	1
2 Relevant national datasets	1
2.1 Land Use Cover Database (LCDB).....	1
2.2 LUCAS land use map.....	2
2.3 AgriBase.....	2
2.4 FarmsOnline land use.....	4
2.5 LINZ cadastral data.....	4
2.6 Valuation land use (Corelogic).....	5
2.7 Agricultural Production Census property matching project.....	5
3 Agribase data pre-processing.....	6
3.1 Completeness (missing farms).....	6
3.2 Currency.....	6
3.3 Logical consistency.....	7
3.4 Managing overlapping records	8
3.5 Farm stocking rate	10
3.6 Tools for semi-automating the Agribase preparation process.....	11
4 National Land Use classifications and methodologies.....	11
4.1 Primary Land Use Types (PLUTs) for soil quality monitoring.....	11
4.2 Land use classification for soil cadmium monitoring (proposed).....	13
4.3 Alternative land use classification for cadmium risk (proposed)	14
4.4 MOTU Land Use in Rural New Zealand (LURNZ).....	17
4.5 Land Use New Zealand (LUNZ).....	20
5 Automation – The Land Use Classifier	28
5.1 Background	28
5.2 The ‘pyluc’ framework	28
5.3 Testing and examples.....	30
5.4 Results	31
5.5 Discussion	35
6 Summary and conclusions.....	36
7 Recommendations.....	36
8 Acknowledgements.....	36
9 References	37

Appendix 1 – LURNZ classification rules example	41
Appendix 2 – pyluc LURNZ classification definition.....	42
Appendix 3 – LURNZ documentation.....	44
Appendix 4 – LURNZ provenance information	45

Summary

Project and Client

- The Innovative Data Analysis (IDA) programme is an MBIE-funded research project currently being undertaken by Manaaki Whenua – Landcare Research (MWLR). A sub-objective of the programme is to rationalise New Zealand land use (LU) classifications into a transparent and repeatable system that can be used to provide data and information that is up-to-date, reliable, and of practical relevance to end users.
- Towards this end, we aim to create a land use classifier tool that can consistently and automatically reproduce several NZ LU classifications, drawing from a range of public and commercially available datasets.

Objectives

- Review characteristics of public and commercial datasets relevant to national land use classification, and discuss availability and suitability.
- Evaluate the Agribase land use dataset, and develop procedures for cleaning and enhancing for LU classification purposes.
- Deconstruct and document the methodological steps of key NZ LU classifications, and then reconstruct each as GIS-based models.
- Develop and test a NZ Land Use Classifier tool with automated generation of dataset provenance and documentation.

Methods

The method involved two key processes:

- First, LU classifications were identified, deconstructed, and then rebuilt using conventional GIS techniques and systems. Methods for key NZ LU classifications were sourced from a combination of published methods, systems diagrams, and GIS-deconstruction in cases where both the classification and the underlying input data were available in GIS format. Method rebuilding was undertaken within vector space. Workflow complexity was minimised by developing a pre-classification system that can be applied to all LU classifications (a non-complex workflow is helpful for the coding and automation component). Workflows are captured as ArcModels and scripts to ensure transparency and ease of replication.
- The second step focused on automating key LU classifications. Precise replication of LU classifications requires exact and complete definitions of classification rules as well as access to original data sources and high-quality technical documentation. A software framework called 'pyluc' was constructed that ingests python scripts containing all classification rules as well as links (URLs) to original data sources stored on the LRIS portal. Operating within raster-space, pyluc is able to acquire the inputs and apply the classification rules automatically to exactly recreate original LU classifications. An important feature is pyluc's ability to automatically generate provenance information and detailed technical documentation designed to

accompany the LUC classification. Finally, pyluc is capable of running on either a desktop or in a high-performance computing (HPC) environment that can significantly reduce the time it takes to generate an LU classification while leaving a researcher's computer free for other duties.

Results

- LU classification methods: Three key NZ land use classifications were deconstructed and rebuilt, including Primary Land Use Types (PLUTS, used for indicator monitoring), Motu's Land Use Rural NZ (LURNZ), and MWLR's Land Use NZ (LUNZ). These three were selected as they represent an increasing plane of classification complexity.
- LU classifier tool: The reconstructed LURNZ and LUNZ LU classifications were successfully redefined as python scripts within the pyluc framework. Pyluc was able to produce almost exact matches to the original LU classification with small differences due to raster vs vector methods used to produce each. Pyluc was also able to produce technical documentation and provenance information, detailing the exact steps of the workflow as well as who was responsible for authoring and running each step along with their organisational affiliation and the date/time it occurred.

Conclusions and recommendations

- We have successfully reconstructed three national land use classifications for PLUTs, LUNRZ, and LUNZ, and automated the methodology through the Land Use Classifier tool.
- Pyluc has proved to be a useful tool capable of exactly replicating the LU classifications reconstructed above using a single human-readable text file (python script). The provenance data produced automatically by pyluc is useful for end users as it provides a record of exactly what happened during the processing workflow and when original data sources (which must now be hosted on the LRIS portal) were accessed. The automated technical documentation and additional graphics derived from the provenance data are useful for understanding complex LU classifications.
- While good at producing final 'official' outputs, pyluc is not always the best tool for researchers used to GIS programs during the initial definition or refinement of LU classification rulesets. We suggest that pyluc be incorporated at the final stages of LU classification development, with the pyluc output used for final delivery. This would formalise a given version of a classification, allowing easy recreation later on or by third parties (with appropriate access controls to the data hosted on LRIS).
- Our key result is developing and demonstrating a national land use classifier tool, with the functionality to reproduce LU classifications on-demand in a transparent and consistent manner.
- We recommend, therefore, that any new LU classifications of national relevance, including those now being developed as part of the Our Land and Water National Science Challenge, be integrated with the Land Use Classifier to ensure easy future reproduction (as land use changes) according to the pyluc framework.

1 Introduction

Reliable and up-to-date land use (LU) information is important for the increased use of spatial modelling, for analysing and reporting trends (State of Environment reporting), and for the development of fair and consistent land use policies.

Several independent LU classifications have been developed for NZ, all of which tend to draw on the same public and commercial datasets. Classification diversity reflects differences in end-user requirements. However, classification methodologies are only ever occasionally documented to a standard that permits exact repeat classification at a future date as land use (and land use data) change. Unclear classification methodologies also obstruct the potential for critiquing the method (limited transparency). Further, few classifications attempt to evaluate the accuracy of their classifications, but rather tend to rely on an assumed quality of underlying datasets despite known and widely recognised inaccuracies.

This report describes the process used to develop a NZ Land Use Classifier tool. The purpose of the tool is to be able to automatically produce several possible LU classifications using the latest source data, to a high standard of reliability, on demand according to end user needs.

Key principles:

- Development involves two stages. The first stage focuses on the development of methodologies. Stage 2 concerns the automation and coding component. Documented procedural steps and basic GIS models are Stage 1 outputs, while the replicator itself is the Stage 2 output.
- Complete methodology descriptions are not always available for all LU classifications, thus requiring the development of bridging steps/methods.
- Emphasis is on practical LU classifications, and particularly those orientated towards rural and agricultural uses.

2 Relevant national datasets

The purpose of this section is to briefly describe national land cover and use datasets, and comment on their usefulness, quality, and availability.

2.1 Land Use Cover Database (LCDB)

The New Zealand Land Cover Database (LCDB) is a digital map of the cover of New Zealand created using satellite imagery. It contains detailed information on 33 classes of land cover, and the most recent release (LCDB4.1) contains land cover classifications representing four time steps: summer 1996/97, summer 2001/02, summer 2008/09, and summer 2012/13.

The LCDB is a particularly useful dataset as it includes a classification of grassland. Grassland that intersects with pastoral land use types from other databases (e.g. AgriBase) provides a representation of effective (actual) farmed pastoral land. The LCDB is available under a public license.

Accuracy of the LCDB is periodically evaluated using methods developed in the original pilot project (i.e. Pilaar et al. 1995). The most recent evaluation was for LCDB3, which achieved an overall map accuracy of 96%, which was an improvement on the 93% accuracy for LCDB1 (Landcare Research 2012).

2.2 LUCAS land use map

The Land Use and Carbon Analysis System (LUCAS) Land Use Map (LUM) is a cover classification designed for Kyoto Protocol reporting responsibilities (Newsome et al. 2013). It includes 12 land-use classes and three temporal classifications nominally dated at 1 January 1990, 1 January 2008, and 31 December 2012 (known as '1990', '2008' and '2012'). Classifications are tailored for carbon accounting, and use four key land use classes of Natural forest, Pre-1990 Planted Forest, Post-1989 Forest, and Grassland with woody biomass.

LUM also contains classifications for high- and low-producing grassland which would be particularly useful for land use classification. However, the method used to differentiate the two was not particularly robust (high and low derived from the vegetation association factor in the NZLRI).

Overall map accuracy of the 2012 version of LUM was assessed to be 95% (Poyry 2014). Manaaki Whenua Landcare Research are currently compiling the 2016 version of LUM.

2.3 AgriBase

The following discussion is based on a Landcare Research report to the Ministry for the Environment (Ausseil et al. 2015).

The AgriBase is a national spatial farms database first launched in 1993 (Sanson & Pearson 1997; Sanson 2005, 2013), originally as biosecurity response tool. Today, it contains approximately 135,000 farms, with data describing farm contacts, predominant farm type, size, animal numbers by stock class, and planted areas. The AgriBase is developed and maintained byASUREQuality, and the data are commercially available under license for most uses according to principles of the Privacy Act 1993. Before the development of FarmsOnline, the AgriBase was endorsed as NZ's sole farm database by the former Ministry of Agriculture and Forestry (MAF). Revenue to maintain the AgriBase comes entirely from paying users.

"Farms" are defined as business management units that utilise a defined area of NZ's land and in-land or in-shore water resources in the production of food or other primary products (Sanson 2005). Spatial representation is based on LINZ cadastral parcels that are aggregated to farms based on ownership detail, and information provided by farmers themselves. Maintenance is achieved through regular monthly updates of LINZ cadastral

parcels (identifies sub-divisions and change of ownership), and an online voluntary questionnaire is available to farmers. Valuation data were regularly used to update the AgriBase before 2004, but a law change reduced access and valuation-based updates are now implemented less frequently. Updates are also achieved through client- and industry-partnerships such as surveys.

Pearless (2005) reviewed the AgriBase and concluded: “while the AgriBase has shortcomings it is clearly the most advanced and suitable data set on which to build NZ’s Farm Register. The level of coverage and completeness of farm enterprise data is impressive for a purely voluntary system, updated opportunistically” (p. 4).

Few studies have been undertaken to evaluate the quality of the AgriBase. Sanson et al. (2004) surveyed small holders and compared the results to what was then held within the AgriBase. They concluded that between 47 and 78% of all small holders were captured in the database.

The database came under close scrutiny after the 2005 Waiheke Island foot and mouth disease scare (Pearless 2005; Heinz 2011; LINZ 2012). Before this, MAF relied on the AgriBase as the tool for identifying farms during a biosecurity incursion. However, for Waiheke Island, MAF’s response was impeded by lack of accurate data and gaps in the data (e.g. lifestyle blocks). Information was available for only 50% of the properties, extending the response time from an estimated 2 days to a full week.

MAF commissioned an analysis of the current state of rural property information soon after the Waiheke operation, focusing on the coverage, currency and accuracy of the AgriBase (Pearless 2005). A set of quality assurance protocols were developed to improve and monitor database quality. Further, phone surveys of approximately 700 South Wairarapa farmers were undertaken to check data accuracy. The following results are based on a high response rate of 90%:

- 26% of AGB phone numbers provided direct contact with the Key Decision Maker
- Farm Type was correct 87% of the time
- 16% of livestock farms had a mismatch between stock types and Farm Type. 8% of non-livestock farms had livestock types and numbers not recorded in the AGB
- “A very low level of accuracy regarding stock numbers was experienced” (p. 16). Actual differences were not reported.

In several instances, MAF also reported that the AgriBase included coverage of only 57% of rural properties (MAF 2011b).

In its current form, MAF considered the AgriBase inadequate as a rural property database suitable for managing biosecurity and adverse events. They then evaluated three options for developing an accurate, up to date source of information about rural properties including farm location, ownership, management and stock and crop information (as part of building the FOL business case), two of which were based on the AgriBase. MAF concluded that enhancing the AgriBase would be both more expensive and no faster than building a new application (MAF 2011a). In 2009 Cabinet approved the development of the third option – an entirely new system known as FarmsOnline (MAF 2011b).

The AgriBase is still a useful and comprehensive register of NZ farms, but, as noted previously, it does have some shortcomings. These are discussed more fully in Section 3, along with the presentation of methodology to clean and prepare the Agribase for modelling applications.

2.4 FarmsOnline land use

FarmsonLine (FOL) was developed by Biosecurity New Zealand, the Ministry for Primary Industries (MPI). The aim was to develop a system to gather contact details from property owners and managers and hence improve emergency contact in case of biosecurity outbreak or severe flood. A farm is defined as all rural properties, including lifestyle blocks.

In principle, the FOL database can be made available for research purposes. However, access to FOL for this project could not be obtained. A request was made, initial feedback was positive, but no further feedback was forthcoming after the request went to a higher level of decision-making in MPI.

Only broad descriptions are available regarding the type of data stored in FOL. At the very least, it records address and ownership details, and uses the valuation land use classification. This level of detail for individual farms can be sourced online at <https://farmsonline.mpi.govt.nz> by entering a farm address, and may have some validation utility when valuation data are not available. Property owners can access additional detail, including stock type, crop type, and there is an attribute for 'industry codes'. FOL farm ids are also linked to the National Animal Identification and Tracing (NAIT) database, which contains considerable detail on stock and stock movements.

2.5 LINZ cadastral data

Land Information New Zealand (LINZ) is responsible for maintaining NZ's land title and cadastral survey records (data about property ownership and boundaries). This is referred to the digital cadastre, which is used as the foundation for all other NZ land use datasets (e.g. Agribase, Corelogic's valuation data).

Key parts of the digital cadastre are regularly made available for public use. One of these is the Primary Parcels layer, which represents ownership parcels, and contains certain attributes (e.g. parcel intent) that can be used to class non-agricultural land use types (e.g. roads, hydro parcels, land gazetted for public or special purpose use). Further, freehold or DCDB type parcels can be used to infill some types of missing parcels/data from other databases with missing records (e.g. Agribase). Non-agricultural classifications have been used in part to build the Protected Areas Network – New Zealand (PAN-NZ) dataset (<http://www.landcareresearch.co.nz/resources/maps-satellites/pannz>).

Spatial accuracy of the digital cadastre is variable, ranging from 0.2 to 5 m for urban areas, and from 0.5 to 20 m for rural areas. Particularly remote rural areas can have inaccuracies of $\pm 100\text{m}$ (LINZ 2017).

2.6 Valuation land use (Corelogic)

Each property in NZ has been evaluated for valuation purposes. Part of the valuation process involves tagging the property (via a property id) with a valuation land use classification. In principle, this classification should be updated each time a property is revalued as part of council rating programmes, and when new parcels are created or aggregated as part of the planning and consent process.

Corelogic manages NZ's valuation data. Valuation data can be tagged back to LINZ property parcels. For a fee, Corelogic will undertake this data matching and supply valuation land use as a spatial layer. A detailed land use classification is used, but the terminology and definition of classes has only weak metadata and explanation, and a degree of interpretation is required.

Quality of the valuation land use classification is unknown. We have used it in the past on a catchment basis, and found inconsistencies. We suspect the classification is not regularly updated in tandem with rates reviews, although this may vary between council areas.

2.7 Agricultural Production Census property matching project

New Zealand's richest source of land use data for agricultural properties is from the Agricultural Production Survey (APS). The title 'Agricultural Production Survey' includes both the Agricultural Production Census and the Agricultural Production Survey. The Census is undertaken every 5 years, drawing from a population of approximately 80,000 farm businesses, while the Survey is undertaken annually between Census years using a representative stratified sample of approximately 30,000 farm businesses. Statistics NZ collects and maintains APS data on behalf of MPI.

"Farm businesses" include all units identified on Statistics New Zealand's Business Frame as having agricultural activity (Statistics New Zealand 2015a). This includes individuals or farming enterprises involved in livestock farming, arable farming, horticulture or forestry. The Business Frame is a list of businesses in New Zealand, based on their registration for goods and services tax (GST) with Inland Revenue. Since the compulsory registration level for GST is \$60,000, there is an unknown proportion of units below this level that are excluded from the APS population (e.g. lifestyle blocks and other small farming endeavours paying <\$60,000 in GST per year).

Farmers are obligated to complete APS questionnaires – completing and signing a questionnaire is a compulsory requirement under the Statistics Act 1975. Response rates thus tend to be high (~80% or more), which contributes to APS data being recognised as of high quality, suitable for Tier 1 statistics (Ausseil et al. 2015). Based on the 2012 Census questionnaire, farmers are required to provide detail on location, ownership, land use, production, and management activities such as irrigation, fertiliser use, effluent application, and nutrient budgeting. Statistics NZ record these data in an aspatial form.

In 2015, the Ministry for the Environment (MfE) initiated an 'agricultural land database matching and data integration' project, whereby records from the APS were spatially matched to parcels by farm ownership and address (Ausseil et al. 2015, Statistics NZ

2015b). Where possible, matches were made for APS census data for the period 1996–2014.

Access to APS data is restricted under the Statistics Act 1975. Only those authorised under the Act are allowed to see APS data, and the data can only be used for statistical purposes. Consequently, spatial APS data are not generally available except under special circumstances.

3 Agribase data pre-processing

The Agribase is a particularly useful dataset in that there are few readily available alternatives that capture agricultural land use and enterprises in a spatial and publicly available form. However, it is also an imperfect database in that relies on voluntary input from farmers, is incomplete in terms of coverage, and contains data configurations that capture real-world land use situations that can be difficult to process from an analysis perspective (e.g. overlapping polygons).

This section reports on techniques that are useful for tidying, identifying, and flagging known inconsistencies with the Agribase. These routines and analyses were performed using the July 2015 version of the Agribase.

3.1 Completeness (missing farms)

Agricultural land not currently included with the Agribase has been added where missing Agribase coverage corresponds with significant areas of LCDB4 grassland and horticulture (horticulture is defined to include both annual and perennial crops). Added units are based on LINZ Primary Parcels, and Agribase-equivalent farm types are interpreted through a combination of LCDB4 cover and an assumption that surrounding farms are likely to be of the same farm type (farm type was spatially inherited from farm type dominance within a 2-km radius). Stocking rates have been estimated from district averages for livestock farm types.

A further 1.1M hectares of otherwise omitted agricultural land is added using this method. All new records are flagged (Fig. 1).

3.2 Currency

The Agribase contains a succession of records that extend back to 1993. Source dates have been converted to years and classified by currency (Table 1). Approximately 66% of records are considered current (i.e. less than 5 years old) representing 75% of database area (Fig. 1). A classification of currency was added (Table 1).

Table 1 Currency of the AgriBase (July 2015 version)

Currency classes	No. farms	% of farms	Area (ha)	Area (%)
1. Very recent	2,991	2%	1,416,744	6%
2. Current	85,850	64%	17,770,335	69%
3. Old	45,027	33%	6,533,097	25%
4. Very old	628	0%	33,751	0%
Total	134,496		25,753,927	
0. New parcels	89,775		1,125,916	

3.3 Logical consistency

Logical consistency describes how well related data within the database agree or compare.

3.3.1 Area agreement

Polygon area (geographical area) was compared against attribute area (enterprise area). There appears to be a poor match (Table 2). Only 9% of the farms (19% of total area), qualified as a match (i.e. were within $\pm 10\%$). Approximately 68% of records had no enterprise hectares recorded (Fig. 1). This indicates that enterprise areas should be used with caution if at all. The degree of unreliability suggested here, means that all subsequent analysis performed for this report is undertaken using geographical hectares.

Table 2 Degree of agreement between enterprise and geographical hectares (July 2015 version)

Area agreement class	Difference ($\pm\%$)	No. farms	% of farms	Geo hectares	% of geo ha
1. Very high	<5%	8,804	7%	4,719,664	18%
2. High	5–10%	2,521	2%	347,544	1%
3. Risky	10–20%	2,968	2%	460,201	2%
4. Not recommended	>20%	28,909	21%	9,954,731	39%
5. Quality unknown	nodata	91,294	68%	10,271,787	40%
Totals		134,496		25,753,927	
(new parcels)		8,9775		1,125,916	

3.3.2 Farm type

It is not uncommon to encounter questionable farm type classifications that either align poorly with the dominant enterprise, and/or exhibit a questionable match with LCDB land cover. The intent is to flag these records. We do not recommend modifying the farm type classifications, however, as we have yet to develop a sound method of determining which data are true, and there may be non-transparent reasons for farmers assigning particular farm types (e.g. they may be based on income or financial contribution to the business rather than the extent of a particular farm use).

3.4 Managing overlapping records

The Agribase contains overlapping polygons. Overlaps may represent leased land (both the lessee and lessor have submitted information to the Agribase), new records from recent subdivision but the Agribase record for the original farm has not been updated, or complicated ownership/use arrangements such as Māori land. While all potentially valid scenarios, it is difficult to undertake analysis with overlapping records, and extreme care is required when making area summaries. We suggest five methods for resolving overlaps when the data are to be used for modelling purposes:

- 1 Flatten the data and the top record (most recently added) is kept. This is readily achieved through rasterizing, or through a more complicated process of intersection, erasing and merging in vector space. This is a common method suitable if the attribute data can be discarded, but not suitable if the attribute data are to be retained. This is because attribute data of the bottom record become distorted (i.e. the underlying farm is no longer of the same area and is therefore likely to carry fewer stock, etc.).
- 2 Assimilate overlapping polygons and sum attribute data as 'super farms'. The disadvantage is the risk of 'double counting', for example, stock numbers from both the upper and lower records are added together, whereas in reality the upper record should replace the equivalent stock numbers from the underlying record. Further, polygons exist that overlap with dozens if not hundreds of other polygons (e.g. leased grazing of state land).
- 3 Flatten data and update attribute data of the underlying polygons. This is possible, and some work has been done, but it is a particularly complicated process.
- 4 Convert quantitative attribute data to rates (e.g. farm stock numbers to stocking rate per hectares) and flatten. This is a robust method for a single attribute, but other attributes are lost. However, a variable such as stocking rate is particularly useful.
- 5 Undertake modelling, classification, or analysis without initially modifying overlapping polygons, and then flatten the result at the end. This is our preferred method.

Overlaps have been flagged for 8,206 records, representing just less than 3M hectares of total database area. The flag is expressed as a 'percent of farm overlapping with other farms'. Geographical area covered when the overlaps are flattened (i.e. the actual footprint) is considerably less at just over 1M hectares. This represents 4% of Agribase footprint (i.e. 4% of the area covered by the Agribase has overlapping polygons) (Fig. 1).

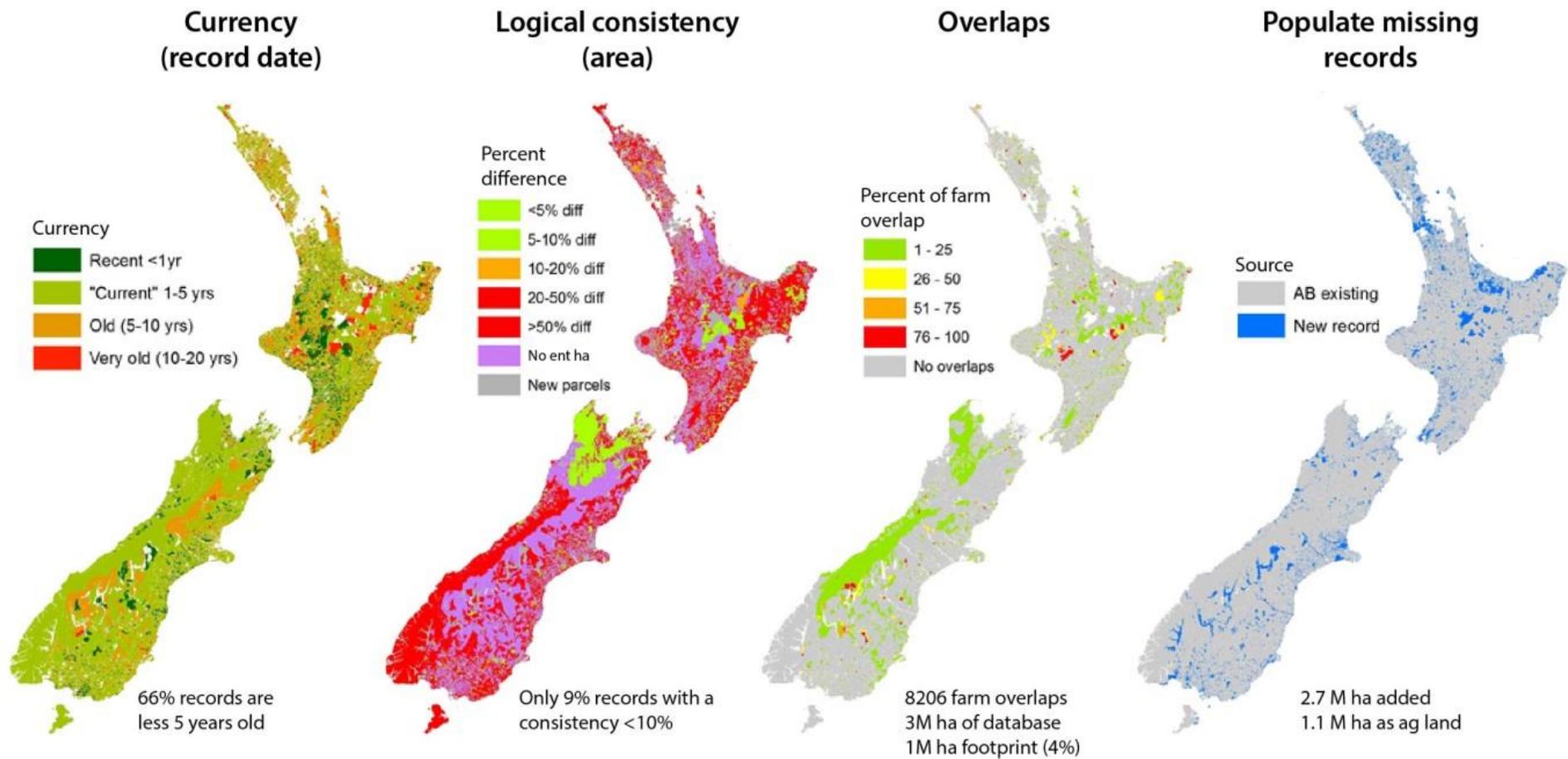


Figure 1 Graphical summaries of results for Agribase July 2015 currency, logical consistency (area), overlaps, and the population of missing records.

3.6 Tools for semi-automating the Agribase preparation process

While not an intended output of the project, the steps and methods described here for cleaning and preparing the Agribase have been captured as a series of Arc models and Python scripts. Among other things this will have value for checking the quality of new versions of the Agribase on delivery. Full automation of the procedures is planned.

4 National Land Use classifications and methodologies

We provide a brief description and discussion of each LU classification, and outline the methods and datasets that have been used to create the classifications. Where published methods and descriptions are lacking, we detail our 'bridging steps' used to replicate each classification.

4.1 Primary Land Use Types (PLUTs) for soil quality monitoring

4.1.1 Description

Primary land use types (PLUTs) is a classification developed to assist with the design of soil quality monitoring frameworks (Hill et al. 2003; Hill & Sparling 2009). It represents a basic single tier classification (Table 3), that differentiates land uses known to explain some of the variance between soil quality measurements.

Table 3 Primary Land Use Types for soil quality monitoring (Hill et al. 2003)

Primary land use types	Code	Description
Cropping and horticulture	CH	Horticulture: orchards, vineyards and berry crops. Typically, in rows, with or without grass cover between rows. Cropping: land cultivated for one or more crops <u>each year</u> . Typically involving tillage for seedbed preparation and harvest.
Plantation forest	PF	Typically, exotic pine or Eucalyptus to produce timber for construction, pulp and paper.
Indigenous vegetation	IV	Typically, beech or broadleaf Podocarp forest with understorey species (ferns, etc.); but could also include wetland and coastal habitat and indigenous dominated scrubland.
Intensive pastoral farming	PI	Permanent grass-legume pastures. Typically used for dairy farming or beef cattle.
Extensive pastoral farming	PE	Permanent pastures or grasslands (including tussock grasslands). Typically used for sheep and beef.
Unclassed	UX	All other uses and covers

4.1.2 Classification method

Most of the PLUT classes relate well to those in the LCDB4 and AgriBase (Table 4). First preference is given to classifications drawn from the LCDB4 on the basis that it is likely to be more accurate given the similarities between the classifications, and that LCDB4 classes represent an actual spatial footprint of a target class rather than a spatial unit that may include more than one classification (i.e. AgriBase enterprises within a farm).

The exception is intensive and extensive pastoral farming, which requires use of stocking rate from the AgriBase. The terms 'intensive' and 'extensive' are qualitative. Here we use a threshold of ≤ 10.5 su/ha to represent land that is farmed less intensively. This captures the greater proportion of farm types commonly considered to be of lower farming intensity.

Table 4 Primary Land Use Types classification criteria

Code	LCDB4	AgriBase
CH	All Short-rotation Cropland (30) All Orchard Vineyard & Other Perennial Crops (33)	
PF	All Exotic forest (71)	
IV	All indigenous/native classes (15, 45, 46, 47, 50, 52, 54, 55, 58, 69, 70)	
PI	All grassland (40, 41, 43, 44) that intersects with...	Rvsd_AB_SR >10.5 su/ha
PE	All grassland (40, 41, 43, 44) that intersects with...	Rvsd_AB_SR \leq 10.5 su/ha
UX	All non-veg (1, 2, 5, 6, 10, 12, 14, 16, 20, 21, 22) All other exotic veg (51, 56, 64, 68)	

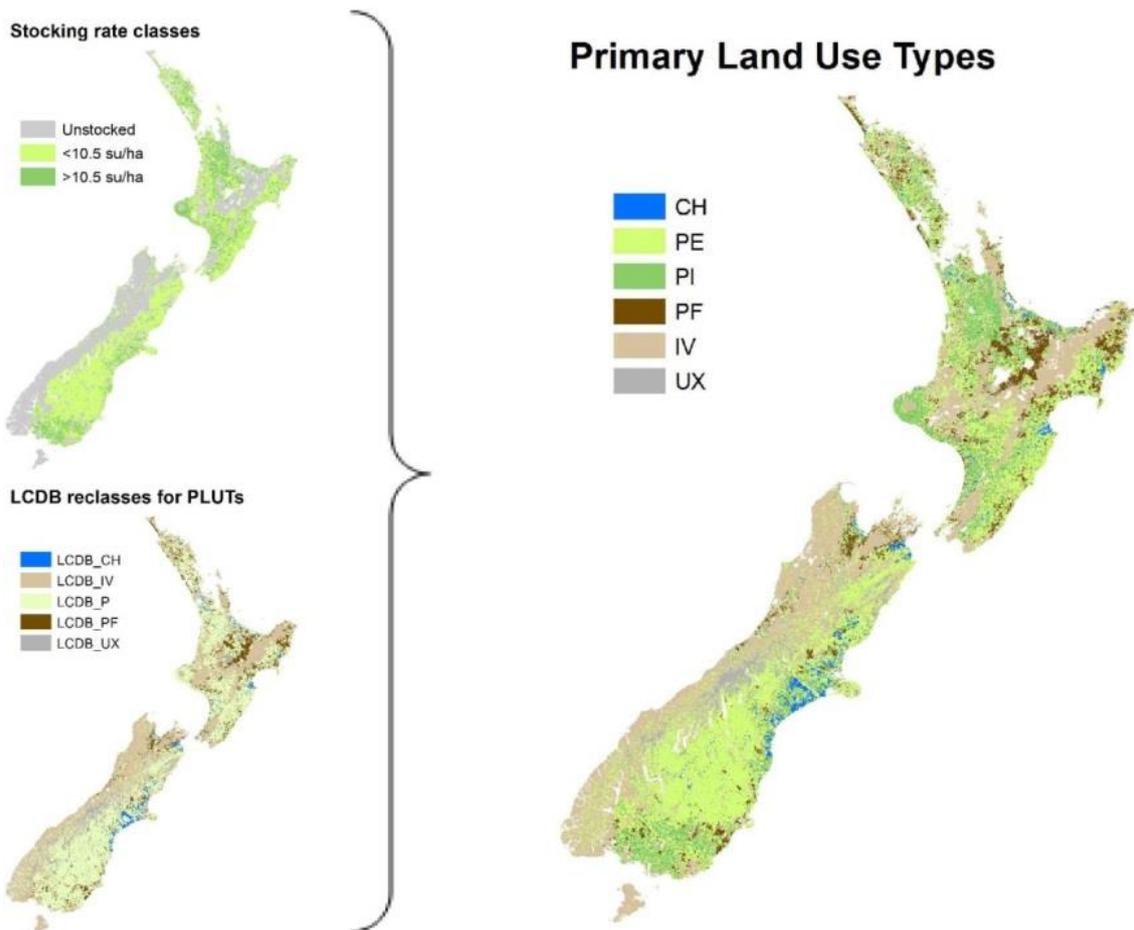


Figure 4 Primary Land Use Types (PLUTs) for soil quality monitoring (after Hill et al. 2003).

4.2 Land use classification for soil cadmium monitoring (proposed)

This classification is differentiated from other soil quality monitoring LU classifications in that it focuses on land uses that are considered to influence cadmium concentrations in soils. It has evolved from a single tier five-class classification for stratifying a national Cd monitoring programme (Cavanagh 2013, 2014), to become a comparatively more comprehensive three-tier, ten-class classification (Table 5). A key feature of the classification is that it is designed to accommodate detailed land use classes that can be recorded during field sampling, in a way that can be aggregated to higher categories for national classification and reporting.

Table 5 Proposed land use classification for soil cadmium monitoring (Cavanagh 2015)

Land use category 1	Land use category 2	Land use category 3	Comments
Pastoral - Dairy	Dairy	Dairy	
Pastoral – other livestock	Sheep and beef	Sheep	Includes sheep and beef, deer, goat and is likely to be a mix of intensive and extensive systems. Ideally intensive and extensive systems could be identified with extensive (low input) systems including lifestyle blocks
		Beef	
	Deer	Deer	
	Other livestock	Other livestock	
Pastoral - other	Unspecified pasture, pasture seed crops	Unspecified pasture, pasture seed crops	
Cropping	Cropping	Crop type	Market gardens, vegetable crops
Arable*	Arable	Crop type	Includes grain crops, hay, fodder crops
Perennial crop*	Orchard	Crop type	Stonefruit, berry fruit, kiwifruit, grapes
	Vineyard		
Forestry	Forestry	Tree type	
Background	Indigenous forest, native scrub	Indigenous forest, native scrub	
	Reserve in non-urban areas	Reserve in non-urban areas	
	Native tussock, not used for grazing	Native tussock, not used for grazing	
Urban	Parks and reserves	Parks and reserves	
	Other	Specify	
Other		Specify	

* There can be overlap as to what might be considered an arable crop vs a horticultural crop, e.g. potatoes can be considered as either.

Unfortunately, we have been unable to secure the type of datasets that would allow us to classify down to the 3rd category of this classification. In lieu, we have developed an alternative 'top down' classification that draws on datasets we have available, and land use principles that influence Cd accumulation.

4.3 Alternative land use classification for cadmium risk (proposed)

An alternative land use classification is constructed to accommodate the potential use of phosphate fertilisers on livestock farms (P-fertiliser is the principle source of Cd accumulation in NZ soils). This is achieved through a potential P-fertiliser demand index (PFDI), which calculates:

- P-development requirements for a given farm according to soil types and minimum optimal Olsen P levels.

- P-maintenance requirements according to current stocking rate, estimated milk solids production (dairy), and soil types.

Actual Olsen P levels will vary widely between farms, and in many cases P-development is likely to have been achieved on intensive farms decades past. For this reason, we use an index as an indicator of potential P-fertiliser demand, and acknowledge that this will not necessarily be representative of actual P-fertiliser use over the years.

The PFDI method draws on values recommended for P-fertiliser use on NZ dairy farms (Roberts & Morton 1999) and sheep/beef farms (Morton et al. 1994). Required spatial inputs include the Agribase (stocking rate and farm type), Land Improvement Corporation (LIC) dairy statistics by district (LIC 2014), LCDB4, and soil information from the Fundamental Soils Layers (FSLs).

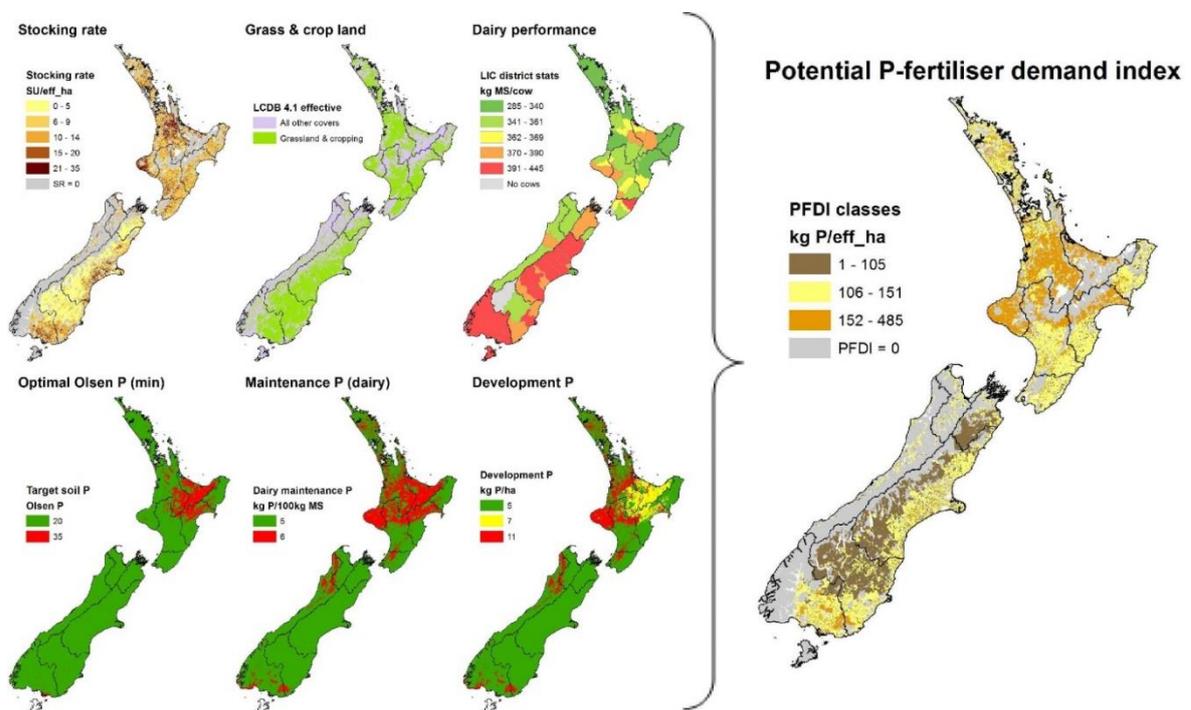


Figure 5 Graphical representation of the Potential Fertiliser Demand Index (PFDI) method.

Non-livestock farms that are likely to include a Cd accumulation risk include all forms of horticulture, defined as the science or art of cultivating fruits, vegetables, flowers, or ornamental plants. This definition is interpreted to include vineyards (grapes), orchards (including nuts), herb farms, nurseries, flowers, and all forms of vegetable growing.

All horticultural farm types, enterprises, and land covers are extracted from the Agribase and LCDB4 databases. Areas from each class and database are cross-tabulated to generate a weighted classification (Table 6), which is used together with the PFDI for grassland to generate an alternative Cd classification for NZ (Fig. 6, Table 7).

Table 6 Horticultural and non-horticultural classification by Agribase vs. LCDB4 cross-tabulation. Refer to Table 7 for a description of codes

	LCDB4 classes						
	Acrop	Pcrop	Grass	Forestry	Other	Oveg	No_LCDB
Arable	ARA_1	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Cropping	Crop_1	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Orchard	AC	ORCH_1	Grassland	Forestry	Other	Other vege	nodata
Viticulture	AC	VIT_1	Grassland	Forestry	Other	Other vege	nodata
Pastoral	ARA_1	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Other uses	AC	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Arable ent.	ARA_2	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Cropping ent.	Crop_2	Pcrop	Grassland	Forestry	Other	Other vege	nodata
Orchard ent.	AC	ORCH_2	Grassland	Forestry	Other	Other vege	nodata
Viticulture ent.	AC	VIT_2	Grassland	Forestry	Other	Other vege	nodata
No_AB	AC	Pcrop	Grassland	Forestry	Other	Other vege	nodata

Table 7 Cadmium accumulation risk classification

Tier 1	Tier 2	Description
Arable	ARA_1	Land on arable and pastoral farms used for annual cropping (grain, fodder, etc.)
	ARA_2	Land on farms with arable enterprises, that is used for annual arable cropping (grain, fodder, etc.)
	AC	Land on other farm types that is used for annual cropping
Cropping	Crop_1	Land on cropping farms used for annual cropping of vegetables, etc.
	Crop_2	Land on farms with cropping enterprises, that is used for annual arable cropping (likely vegetables, etc.)
Orchards	ORCH_1	Land on orchards used for perennial crops (apples, kiwifruit, etc.)
	ORCH_2	Land on farms with orchard enterprises, that is used for perennial crops (likely to be apples, kiwifruit, etc.)
	Pcrop	Land on other farm types that is used for perennial crops
Viticulture	VIT_1	Land on vineyards used for perennial crops (grapes for wine)
	VIT_2	Land on farms with viticultural enterprises, that is used for perennial crops (likely to be grapes for wine)
Grassland	Grass_0	Grassland where the PFDI = 0 (mostly unfarmed grassland)
	Grass_1	Pastoral grassland with a low PFDI (e.g. SI high country grassland)
	Grass_2	Pastoral grassland with a moderate PFDI
	Grass_3	Pastoral grassland with a high PFDI (mostly dairy in Ash/Pumice areas)
Background	Forestry	Forestry on all farm types
	Other	Other non-vegetation land covers (rivers, rocks, etc.) on all farm types
	Other veg	Other non-forestry or pastoral covers (indigenous, scrub, etc.) on all farm types

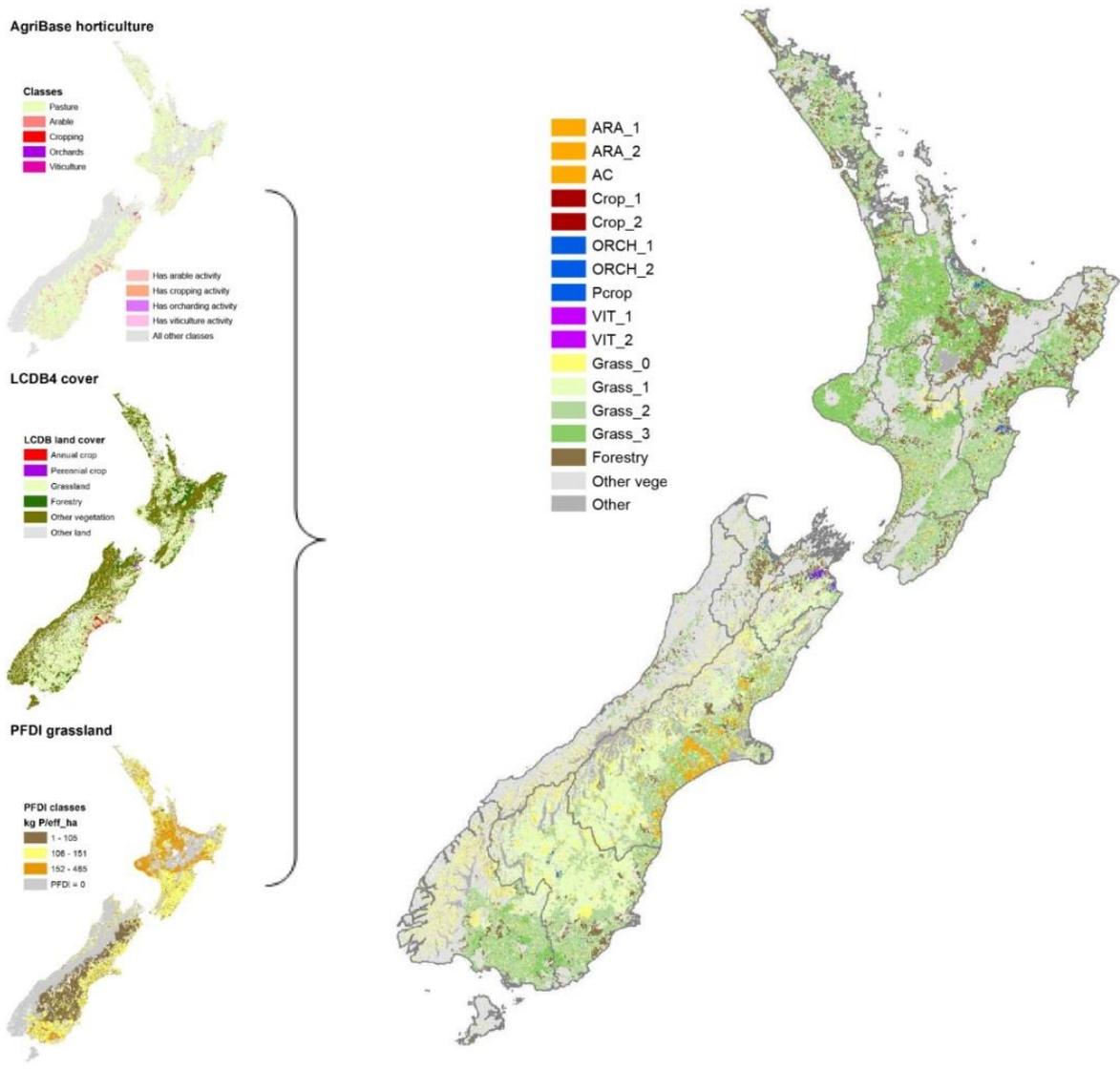


Figure 6 Graphical representation of method used to construct the alternative cadmium risk classification.

This is a proposed risk-based land use classification that requires further work and validation. Among other things, PFDI class thresholds are arbitrary, and the assignment of 'ash' soils is based on high P retention soils other than Pumice (other non-ash soils can also have high P-retentions). Until it is refined, we do not advance it as a land use classification for replication.

4.4 MOTU Land Use in Rural New Zealand (LURNZ)

4.4.1 Description

Land Use in Rural New Zealand model (LURNZ) is a national-level model of land use and land-use change. LURNZ uses a base land-use map, and then uses modelling to predict land use change. Here we are interested in the base land-use map (LURNZ 2002), although the LURNZ modelling framework has been used to generate land use classifications for

several years (Anastasiadis & Kerr 2013; Anastasiadis et al. 2013, 2014; Parshotam et al. 2013; PCE 2013).

Original data layers used to build LURNZ 2002 include the LCDB (version 3), a Land Ownership map created by Landcare Research, the Agribase Enhanced Land Cover Database (ELCDB2), and 'land quality' maps. Both the ELCDB2 and Land Ownership datasets are no longer available, although the underlying methods of construction and classification are straightforward and reproducible.

- The ELCDB2 was originally constructed by sub-classing LCDB2 grassland and horticultural land covers according to land use classes from a 2001 version of the Agribase. In doing so, more detailed classes were produced (e.g. dairy pasture, kiwifruit orchards, etc.). Despite the added detail, the classes are aggregated heavily for LURNZ 2002 (Table 10) which makes them straightforward to reproduce without having to replicate the detail of the ELCDB2.
- The land ownership layer is an early version of the Protected Areas Network NZ (PAN-NZ) dataset, which differentiates private and public land. The latest version of PAN-NZ is 2014.

4.4.2 Classification method

LCDB4.1 2012 and the Agribase (July 2015) are pre-classed into categories relevant to LURNZ (Tables 8 & 9). The results are overlaid together with the public/private ownership layer, and rules are applied to classify the various combinations into LURNZ classes (Table 10). Rules are implemented by priority in an 'else if' framework (example rule code included in Appendix 1).

There is a degree of uncertainty regarding the final rules. Not so much the rules themselves, but the priority of classification regarding ownership. For example, in the original LURNZ, Department of Conservation (DOC) land appears to trump public pasture (even though there can be considerable public pasture in DOC land), and high country leases appear to trump the classification for sheep and beef pasture.

Table 8 LURNZ LCDB pre-classes

LURNZ LCDB preclasses	LCDB codes	Description
LUR_hort	30, 33	Horticulture
LUR_if	69	Indigenous forest
LUR_oth	6, 10, 12, 14, 16, 20, 21, 22, 45, 46, 70	Other/non-productive
LUR_pas	15, 40, 41, 43, 44	Pasture
LUR_pf	64, 68, 71	Plantation forestry
LUR_scrub	47, 50, 51, 52, 54, 55, 56, 58	Scrub
LUR_urb	1, 2, 5	Urban

Table 9 LURNZ Agribase pre-classification criteria

LURNZ pre-class	Agribase farmtypes	Description
m_dai	DAI, DRY, GRA	Dairy
m_hort	ARA, FLO, FRU, NUR, OPL, VEG, VIT	Horticulture
m_lf	LIF, TOU	Lifestyle properties
m_oa	ALA, API, DEE, DOG, EMU, FIS, GOA, HOR, OAN, OST, PAS, PIG, POU, ZOO	Other animals
m_oth	MTW, NEW, OTH, SLY, URB	Other
m_pf	FOR	Plantation forestry
m_sb	BEF, SHP, SNB	Sheep beef
m_scb	NAT, NOF, UNS	Scrub land

Table 10 Overlay combinations used to classify LURNZ land use classes. Order of rule-implementation determines classification priority (i.e. which combinations are classified first)

LCDB preclasses	AB preclasses	Ownership classes	LURNZ classes
Urban (LUR_urb)			Urban
Pasture (LUR_pas)	Dairy (m_dai)		Dairy
Pasture (LUR_pas)	Sheep beef (m_sb)		Sheep & Beef
Pasture (LUR_pas)	Horticulture (m_hort)		Sheep & Beef
Horticulture (LUR_hort)			Horticulture
	Other animals (m_oa)		Other animal
	Lifestyle properties (m_lf)		Lifestyle
Pasture (LUR_pas)		HC lease	Public pasture
Pasture (LUR_pas)		Private	Oth private pasture
		Public	DOC/Public land
Other/non-productive (LUR_oth)			Non-productive
Scrub (LUR_scrub)			Scrub
Plantation forestry (LUR_pf)			Forestry
Indigenous forest (LUR_if)			Indigenous forest

* HC lease = high country lease

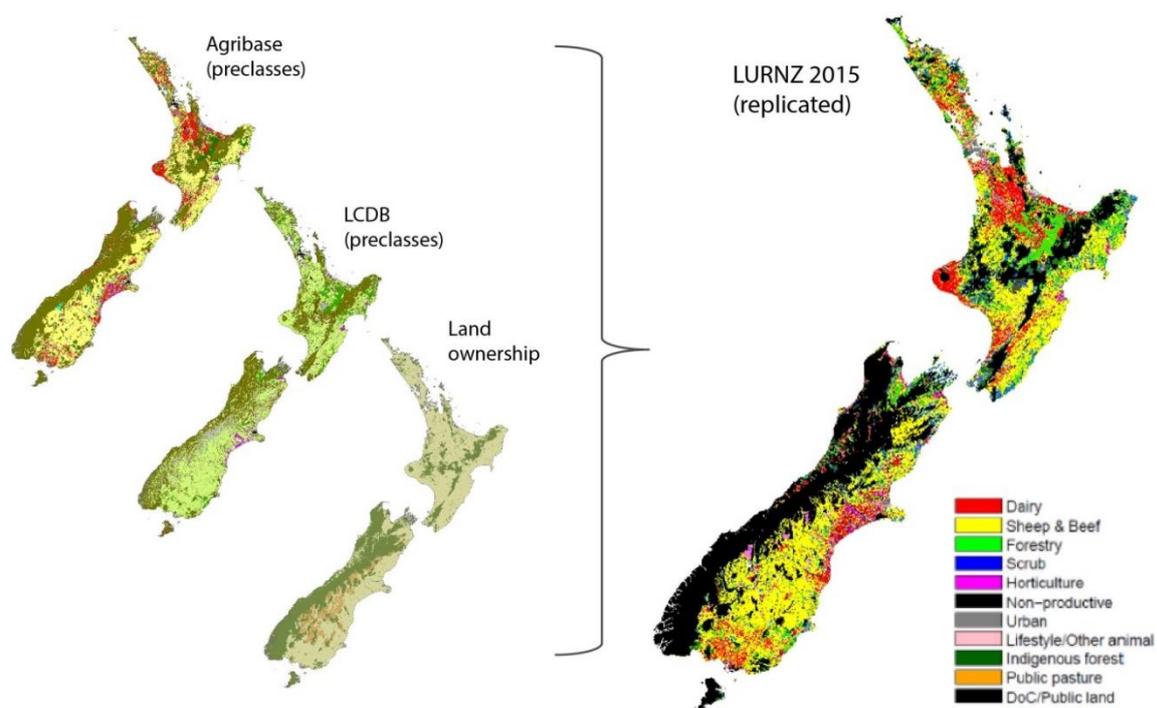


Figure 7 LURNZ 2015 – layers and result of the replication process.

4.5 Land Use New Zealand (LUNZ)

4.5.1 Description

LUNZ is described as a comprehensive land use dataset for New Zealand that was created by Landcare Research using information from AgriBase, LCDB2, LENZ and MAF monitor farm types (Motu 2010), originally to help support the CLUES programme (Woods et al. 2006). The first version was for 2002. The method was later automated into a Java language framework by Harald Waxenegger and Robert Gibb, and a 2011 LUNZ version was produced.

The classification itself has three tiers (Fig. 8). The first tier includes six land use categories (pastoral, arable, horticultural, forestry, native forest, and 'other'). These are subdivided into 25 detailed classes at Tier 2. Tier 3 simply tags a district to the Tier 2 classes.

4.5.2 Classification method

The LUNZ version 2 (2011) Java framework is no longer operational, and was deemed too archaic and complex to reconstruct within the bounds of this project. There is no published methodology regarding LUNZ. We could not obtain a layer of LUNZ for any year to reverse engineer (we did find a `lunz_tier2.shp` file but this did not have the final Tier 2 classification). We have reconstructed the classification to the second Tier, largely by inferring the method from Figure 8, Woods et al. (2006), and a conceptual diagram by Robert Gibb (Gibb 2010).

A pre-classification is firstly applied to the LCDB (Table 11). The earlier version of the LCDB (version 2) had more cover classifications than the current version 4. For example, vineyards and orchards were separate classifications in LCDB2, while they are combined as one in LCDB4. Consequently, we could only initially pre-class the LCDB4 into 10 classes rather than the 11 indicated by Figure 8. Seven of the 10 LCDB pre-classes mapped directly to the final LUNZ classification.

Table 11 LUNZ LCDB pre-classes

LUNZ LCDB preclasses	Description	LCDB codes
LC_arable	Annual cropping land	30
LC_art	Artificial surface	1, 2, 5, 6
LC_BGR	Bare ground	10, 12, 14, 16
LC_for	Plantation forestry	64, 71
LC_hort	Horticulture (vineyards, orchards, etc.)	56
LC_nat	Indigenous forest	68, 69, 70
LC_pastoral	Pastoral	15, 40, 41, 43, 44
LC_scrub	Scrub	50, 51, 52, 54, 55, 56, 58
LC_wat	Water bodies inc rivers	20, 21, 22
LC_wet	Wetlands	45, 46, 47

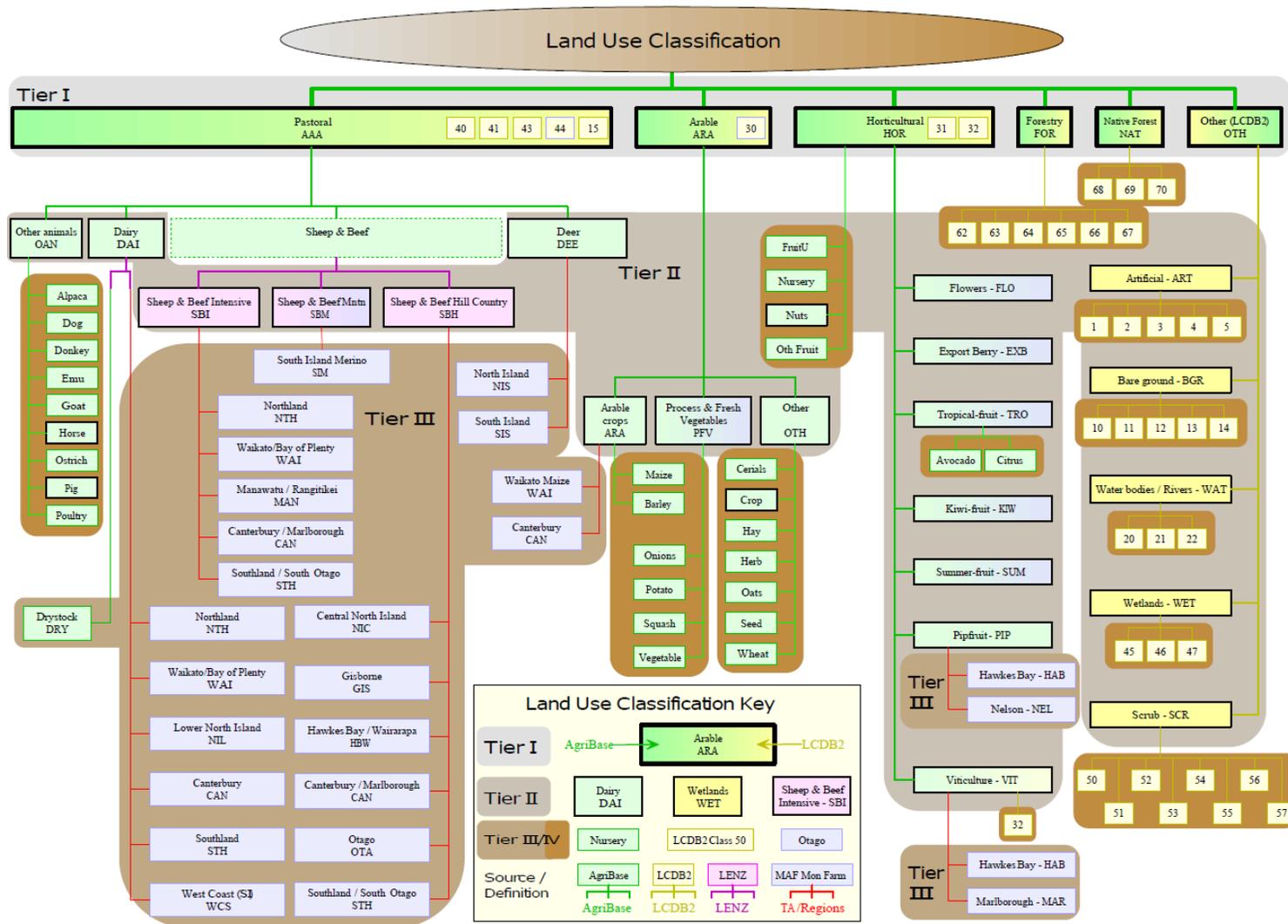


Figure 8 Overview of the LUNZ classification (Robert Gibbs).

The pastoral class (LC_pastoral) was further differentiated using a weighted 'balance of expected probabilities' involving a pre-classification of Agribase farm types (Table 12) and enterprise stock-type dominance (based on total stock units for each stock type within a farm). Rules were used to combine and class combinations of farm types and enterprise stock dominance. For example, if LCDB pre-class = LC_pastoral, and if Agribase pre-class = dairy, and if total_dairy_stock_units = maximum of all stock types, then LUNZ Tier 2 = AAA_DAI.

Table 12 LUNZ Agribase pre-classification criteria

LUNZ Agribase pre-class	Agribase farmtypes	Description
Dairy	DAI, DRY	Dairy farming systems
Deer	DEE	Deer farming systems
SnB	BEF, SHP, SNB, ARA, LIF, GRA	Sheep, beef, or sheep and beef systems
OAN	ALA, DOG, EMU, GOA, HOR, OAN, OST, PAS, PIG, POU	Other animal farming systems
Non-livestock	All other Agribase farm types	Not otherwise farmed with livestock

Sheep and beef farms attract a further sub-classification in LUNZ, based on a landform class derived using pre-classes for LUNZ Level 3 environments (Table 13).

Table 13 LUNZ pre-classifications for LUNZ Level 3 environments

Landform	LUNZ Level 3 environments
Flats	A1.1, A2.1, A3.1, A4.1, A4.2, A5.1, A5.2, A5.3, A5.4, A6.1, A7.1, A7.2, A7.3, B1.1, B1.2, B2.1, B3.1, B3.2, B4.1, B5.1, B5.2, B5.3, B6.1, B7.1, B8.1, B9.1, B9.2, C1.1, C1.2, C1.3, C2.1, C3.1, C3.2, F5.1, F5.2, F5.3, F7.1, G1.1, G2.1, G3.1, G3.2, G3.3, G3.4, G4.1, G5.1, G6.1, H1.1, H1.2, H1.3, H2.1, I3.1, I3.2, I3.3, J1.1, J1.2, J2.1, J2.2, J3.1, J3.2, J4.1, J4.2, J4.3, L1.1, L1.2, L1.3, L2.1, L2.2, L3.1, L3.2, L4.1, L5.1, L5.2, M1.1, M2.1, M2.2, M2.3, M3.1, M3.2, M4.1, N1.1, N1.2, N2.1, N2.2, N3.1, N3.2, N3.3, N6.1, N6.2, N7.1, N8.1, N8.2, O1.3, O1.4, O3.1, O3.2, O3.3, Q4.2
Hills	A6.2, B1.3, D1.1, D1.2, D1.3, D2.1, D2.2, D2.3, D3.1, D3.2, D3.3, D4.1, E1.1, E1.2, E1.3, E1.4, E2.1, E2.2, E3.1, E3.2, E4.1, E4.2, F1.1, F1.2, F1.3, F1.4, F2.1, F3.1, F3.2, F3.3, F4.1, F6.1, F7.2, F7.3, H2.2, H2.3, H3.1, H3.2, H4.1, I1.1, I2.1, I2.2, I4.1, I4.2, I5.1, I5.2, I6.1, L6.1, L6.2, N4.1, N5.1, N5.2, O1.1, O1.2, O2.1, O2.2, O2.3, O4.1, O4.2, O5.1, O5.2, P1.1, P1.2, Q1.1, Q1.2, Q2.1, Q2.2, Q3.2, Q4.1, Q4.3, S1.1, S2.1, S2.2, S3.1
Mountains	F6.2, K1.1, K1.2, K2.1, K2.2, K3.1, K3.2, K3.3, K4.1, K4.2, K5.1, P2.1, P3.1, P3.2, P4.1, P5.1, P5.2, P6.1, P6.2, P7.1, P8.1, P8.2, Q3.1, Q3.3, R1.1, R1.2, R2.1, R2.2, T1.1

Horticulture and arable land uses (LC_arable and LC_hort) use a similar but more comprehensive weighted 'balance of expected probabilities' approach. This involved the creation of several layers, each of which offers some degree of relative likelihood to support the occurrence of certain crop or tree types at particular locations. Horticulture involved an additional pre-step to disaggregate the LCDB4 combined orchards and vineyards class, using the following layers:

- LCDB4 pre-classes for horticulture (LC_hort)
- LCDB2 classifications for vineyards and orchards (i.e. original separate classifications)
- Agribase enterprise dominance (by hectares of orchard or vineyard enterprise types)
- Agribase farm type classifications for vineyards (VIT) or orchards (e.g. FRU).
- Land Use Capability subclasses that most commonly associate with viticulture or orchard based on an intersection of the NZLRI with LCDB2 vineyard and orchard cover classes (Table 14).
- The dominance of either orchards or vineyards within territorial authority districts as determined by an analysis of Agricultural Production Census statistics.

Table 14 Likely land quality for viticulture and orchards based on LUC subclass associations with LCDB2 viticulture and orchard cover classifications (e.g. viticulture occurs more commonly on 3s, 3w, etc. Neither LCDB2 viticulture nor orchard occurs on 1w, 2c, etc.)

LCDB2 class	LUC subclass associations
Viticulture	3s, 3w, 4s, 6s, 7s
Orchards	1c, 1s, 2e, 2s, 2w, 3e, 4e
Neither	1w, 2c, 3c, 4c, 4w, 5c, 5e, 5s, 5w, 6c, 6e, 6w, 7c, 7e, 7w, 8c, 8e, 8s, 8w

Combination matrices were constructed to determine combinations and weightings (Fig. 9). Weightings were allocated based on experience – we do not yet have sufficiently detailed spatial datasets to quantitatively determine actual probabilities – this comment is similar to one made by Robert Gibb in Woods et al. (2006). The combinations and weightings were converted to rules, and implemented to generate intermediary orchard classifications, and directly for the final vineyard classification (HOR_VIT). These steps would not be necessary if the LCDB reinstated the separate classification of orchards and vineyards.

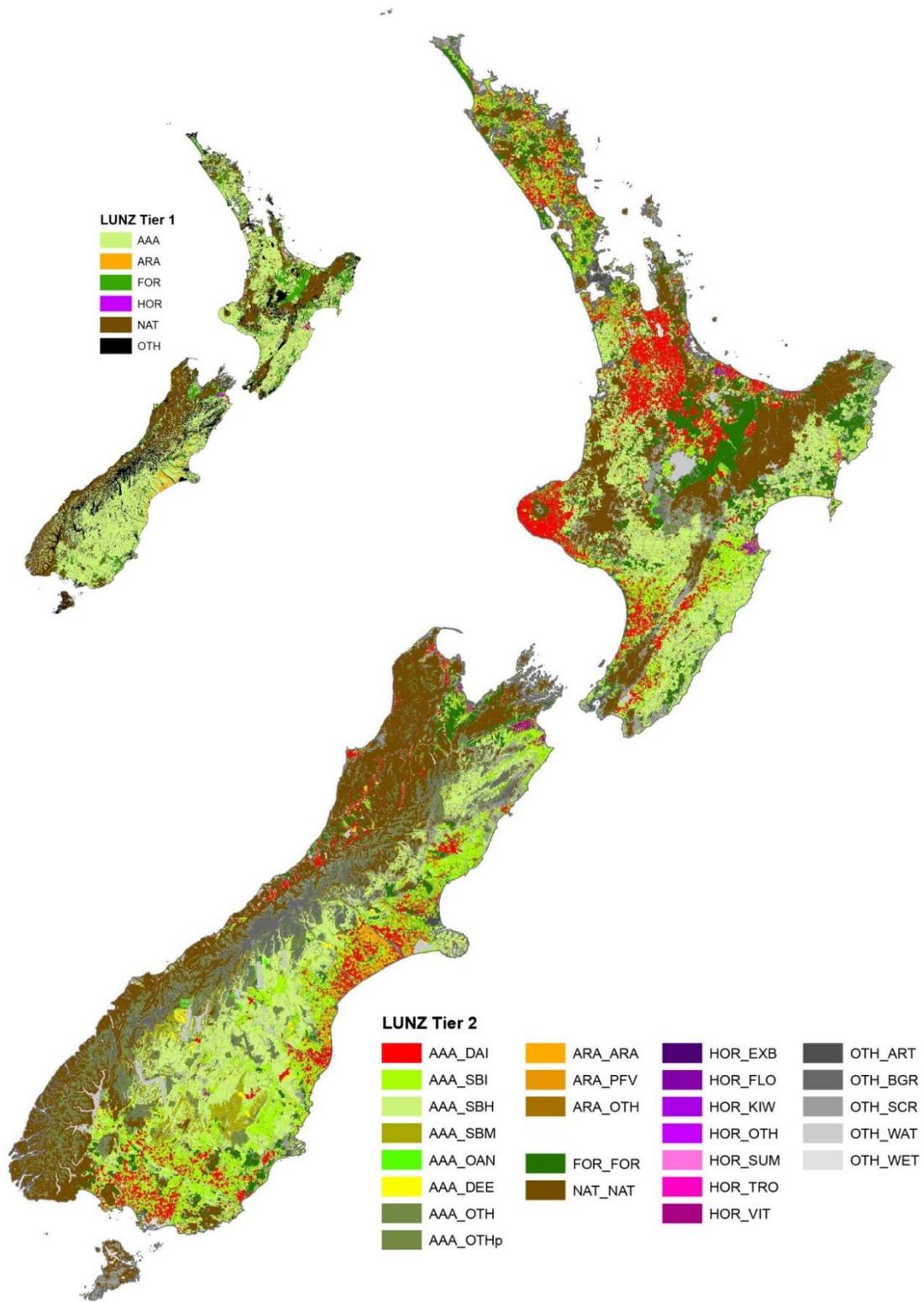


Figure 10 LUNZ Tiers 1 and 2.

Table 15 LUNZ codes and descriptions

Tier 1	Tier 1 description	Tier 2	Hectares	Tier 2 description
ARA	Arable	ARA_ARA	176891	Land used for arable cropping of barley and maize (original LUNZ description)
		ARA_PFV	28485	Land used for commercial vege growing
		ARA_OTH	164311	Land used for other cropping (wheat, oats, cereal, seed, hay, lucerne, brassicas, etc.
HOR	Horticultural	HOR_VIT	41738	Vineyard
		HOR_EXB	2400	Land used for berry fruits
		HOR_OTH	16142	Land used for other perennial hort types
		HOR_SUM	3773	Land used for 'Summer' fruits (stone fruits = plums, peaches, apricots, etc.)
		HOR_TRO	7764	Land used for 'Tropical' fruits (avocados, feijoas, tamarillos, passionfruit, and all citrus fruits)
		HOR_KIW	23330	Land used for kiwifruit
		HOR_FLO	8376	Land used for flower growing
AAA	Pastoral	AAA_DAI	2100689	Dairy farm pasture
		AAA_SBI	3076754	Intensive sheep and beef pasture
		AAA_SBH	5035667	hill country sheep and beef pasture
		AAA_SBM	469810	Mountain land sheep and beef pasture
		AAA_DEE	213903	Deer pasture
		AAA_OAN	228214	Pasture used by other animals
		AAA_OTHp	1438498	Other pasture/grassland on public land (e.g. reserves, Doc conservation estate)
		AAA_OTH	627746	Other pasture on private land
FOR	Forestry	FOR_FOR	2045220	Forestry
NAT	Native forest	NAT_NAT	6432333	Native forest
OTH	Other LCDB	OTH_SCR	2701532	Scrub
		OTH_BGR	1051106	Bare ground
		OTH_WAT	536766	Water
		OTH_WET	165284	Wetlands
		OTH_ART	245648	Artificial surfaces including urban areas
not lcdb	not lcdb	not lcdb	201601	Residual area from other databases that is outside the Topo50 NZ coastline

5 Automation – The Land Use Classifier

5.1 Background

Good quality land use (LU) classifications need to be well-documented and easily repeatable, whether by a third-party or by the original authors, if they are to be trustworthy and dependable sources of information. A LU classification that is well documented should, by definition, be easily replicated using data from primary sources, classification rules that are clear and explicitly stated, and processes that are transparent and easy to repeat. Documentation at this level of detail is lengthy and tedious to update – particularly if several intermediate datasets must be derived from primary sources – which can lead to unintentional discrepancies between the steps that were documented and those that were followed. Ideally, it should be possible to define a LU classification in a way that is self-documenting (in a technical sense) and able to be understood by both a human and a computer such that the classification can be easily and transparently recreated with minimal manual effort.

An increasingly important aspect of dataset documentation is ‘data provenance’, which can be defined as a record of what has happened to some data, where it happened, when it happened, how it happened, who did it, using which tools/instruments, and for what purpose (why). A good record of data provenance demonstrates the quality of a dataset by providing an audit trail and greatly helps third-parties understand processes involved in the creation and evolution of a given dataset; however, it does not replace the need for ‘traditional’ documentation such as this report. Data provenance is typically recorded using a formalised notation and is machine-readable which allows humans to use software to query and create visualisations for what can be extremely complicated and detailed information. These queries and visualisations can then be integrated into more human-readable documentation, while the raw provenance information should remain accessible to third-parties to query themselves.

5.2 The ‘pyluc’ framework

5.2.1 Overview

We have created a software tool called ‘pyluc’ (Python land use classification framework) which ingests a single simplified Python script that defines an entire LU classification from primary sources to final result. Using this script, pyluc obtains the input data specified, applies the classification logic, and produces a dataset with accompanying documentation including provenance. The entire process is automated and does not require user input; no user-interface is provided. The documentation and provenance information are created automatically from the internal model of the LU classification built by pyluc during ingestion – it is possible to quickly recreate just the documentation and/or provenance without re-processing the entire dataset as this step can be time-consuming on large datasets. Provided the primary sources are available to a third-party, the only requirements to exactly recreate the LU classification are the pyluc tool and the single definition script. Users are also able to use the script itself along with the recreated documentation to delve into the logic behind any step of the classification process and

can modify and re-run the script to create new iterations of the dataset with this reflected in the automatically generated provenance information. Pyluc can operate in a Windows or Linux environment on either a desktop machine or a high-performance computer and scales efficiently to handle very large or detailed classification tasks.

Early in the design phase of pyluc it was decided that all input datasets must be located on the LRIS (Land Resource Information Systems) portal¹ or a similar 'Coordinates'-based site such as the Land Information New Zealand data service.² There are two primary reasons for this approach: these sites are remotely accessible with appropriate authorisation so data distribution is easy and secure, and the datasets on them are immutable (will not change once uploaded) – an essential property for the repeatability of a LU classification. Thus, users of pyluc can be confident they are downloading unmodified copies of the original datasets used by the original LU classification authors. Another important design decision was that pyluc would be a raster-based processing engine. This was made for a number of reasons, but predominantly to reduce the complexity present, and processing power required, when many vector layers with potentially different line-work must be combined in an automated system. The resolution of the resulting classification can be adjusted as required; because pyluc scales well with large datasets it can handle very fine resolutions if needed. Pyluc accepts vector datasets as inputs and will automatically rasterise them once downloaded. If vector output is required, pyluc will also vectorise final results on demand.

5.2.2 Definition script

The intention during the design of pyluc was that LU classification authors would continue to use tools familiar to them, such as ArcGIS, to develop and test the logic behind their classifications (potentially on subsets of nation-wide datasets if computation power was an issue). Once the authors were ready to begin formalising the LU classification for delivery, these rules would be transferred into a pyluc definition script and tested over the entire region of interest to ensure the results were correct. The syntax for defining a rule for pyluc is similar to that for ArcPy, such that ArcPy rules can be copied into the definition script with minor modifications. As the script is just a text file containing Python code it integrates extremely well with version-control tools such as SVN or Git, where further tweaks or modifications to the ruleset can easily be tracked, attributed, and rolled back if required. The automated documentation produced by pyluc means that, by the time the 'final' version of the definition script (and therefore the LU classification) is reached, the technical documentation is automatically up to date with no extra effort on the part of the authors to maintain it during the development process.

Each pyluc LU classification definition script may be broken up into four sections: setup, inputs, classification rules, and outputs. The 'setup' section defines the name, geographic extent, and raster resolution of the LU classification as well as providing a place where all those involved with the LU classification (people and organisations) should be defined.

¹ <https://lris.scinfo.org.nz/>

² <https://data.linz.govt.nz/>

This 'inputs' section specifies exactly which LRIS data layers are used by the LU classification and provides an opportunity for these to be attributed to organisations responsible (just because a layer is in LRIS it doesn't mean that layer was created by Manaaki Whenua). The 'classification rules' section is for all rule definitions from lookup tables to more complicated classes requiring multiple inputs, while the 'outputs' section links classification rules with outputs, either intermediate (for consumption by subsequent classification rules) or final.

An example definition script for the Motu LURNZ classification is included in Appendix 2, though most of the 'help' comments have been removed for brevity. Original copies can be requested from the authors.

5.3 Testing and examples

5.3.1 Constructing the definition scripts

To test the pyluc framework we decided to create definition scripts for two of the LU classifications in this document: MOTU LURNZ, and LUNZ. Original data sources used to recreate LU classifications were uploaded to the LRIS portal (as per pyluc requirements) but kept in a private group with access limited only to the IDA team members directly involved in this project. The scripts were constructed by Ben Jolly who developed pyluc but had no hand in recreating the classifications in the first place so had no prior knowledge of what was involved for each and primarily relied on documentation supplemented by occasional discussions. Each definition script was started by extracting ArcPy code from the Arc Toolboxes produced for each classification, with further tweaks or gaps in knowledge filled via other technical documentation.

Several test runs were carried out over smaller regions of the country and compared with the corresponding outputs from Section 4 of this report, with further tweaks to the scripts and framework as various bugs were identified and fixed. A small number of particularly difficult bugs delayed this stage of the project, where minor differences between the way that ArcPy and regular Python dealt with data types and rounding affected some 'edge cases', which resulted in incorrect classifications (e.g. AAA_OTH vs AAA_OAN for the LUNZ classification). Additionally, some errors were only found when testing switched to the entire country.

A raster resolution of 20 m (or 400 m²) per pixel for each national-scale LU classification was picked as a good balance between accuracy and avoiding issues with minor differences in line-work when it came time to compare results.

5.3.2 Processing

Processing at local-scale was relatively fast, with initial runs constrained by the time it took LRIS to export the input layers (around one hour) and subsequent runs taking a matter of seconds or minutes depending on the size of the local area (hundreds to thousands of square kilometres). National-scale processing was moved to the National e-Science Infrastructure (NeSI) High Performance Computer (HPC) in Auckland, also known as 'Pan'.

Pyluc has been designed to take advantage of a cluster-computing environment by computing LU classifications in tiles with an additional 'merge' step to collate the results. Splitting the nation-wide classification runs (at 20-m resolution) resulted in 988 tiles, with the entire process taking 2 hours for the MOTU LURNZ classification and approximately 6 hours for the more complicated LUNZ classification. These figures are rough averages taken over a number of iterations as Pan (like most HPCs) is a shared environment and so performance is impacted by other users.

Pyluc output was compared with original vector-based files by rasterizing those files using the same resolution/extent. These 'thematic rasters' were then compared with a 'difference' raster produced that highlighted and quantified inconsistencies. This approach allowed relatively rapid assessment of each pyluc output on a regular desktop PC, whereas performing an assessment in vector-space would have been prohibitively time-consuming.

5.4 Results

5.4.1 MOTU Land Use in Rural New Zealand (LURNZ)

The definition script input, technical documentation output, and provenance output for the MOTU LURNZ LU classification can be found in Appendices 2, 3 (as the embedded PDF), and 4 respectively. The MOTU classification is relatively simple, with two lookup tables providing 'pre-classes' that are fed into a simple 'if this and/or this then that' classifier. The technical documentation (Appendix 3) outlines the *Organisation* involved in Part 1, then the *People* (authors and operators) in Part 2, where each person is assigned an organisational affiliation. Part 3 lists the *Inputs* from LRIS, including the URL to the layer as well as the field name(s) used and the organisation responsible for producing each layer. Part 4 covers the LU classification *Rules*, first with a table that documents how each rule relates to the inputs available and who wrote it, then with an exact replication of the code used.

The MOTU LURNZ raster produced by pyluc is shown in Figure 11, where the colour scheme has been modified to match that of Figure 7 so the figures may be directly compared. Results are consistent for both the manual recreation and its pyluc implementation, except for the 'Private pasture unclassified' class, which remained in the technical documentation used to create the pyluc definition script but not the final manual recreation in Figure 7.

The provenance output in Appendix 4, in particular Figure 14 and Figure 15, helps visualise the connections between the organisations, people, input layers, rules, and outputs. These figures are taken from the ProvStore website³, specifically the page for the uploaded MOTU LURNZ auto-generated PROV-N data in Figure 13.⁴ Figure 14 is actually a screenshot of an interactive diagram so it is recommended to visit the original version linked in the caption.

³ <https://openprovenance.org/store/>

⁴ <https://openprovenance.org/store/documents/819>

MOTU_LURNZ.Final

- Unclassed
- Indigenous forest
- Forestry
- Scrub
- Non-productive
- DoC/Public land
- Private pasture unclassified
- Public pasture
- Lifestyle/Other animal
- Horticulture
- Sheep and Beef
- Dairy
- Urban

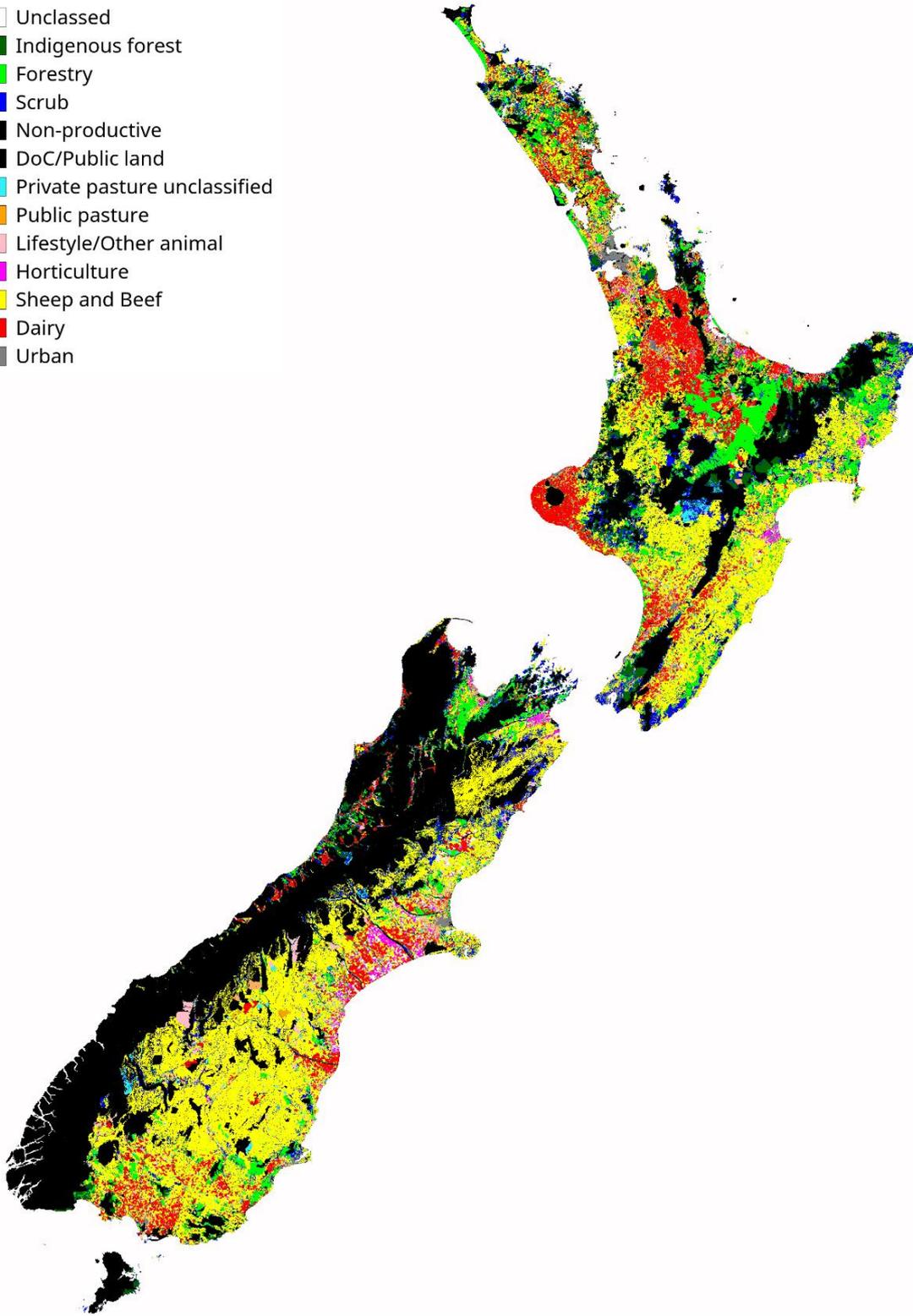


Figure 11 MOTU LURNZ LU classification produced by pyluc (visualised with QGIS).

5.4.2 Land Use New Zealand (LUNZ)

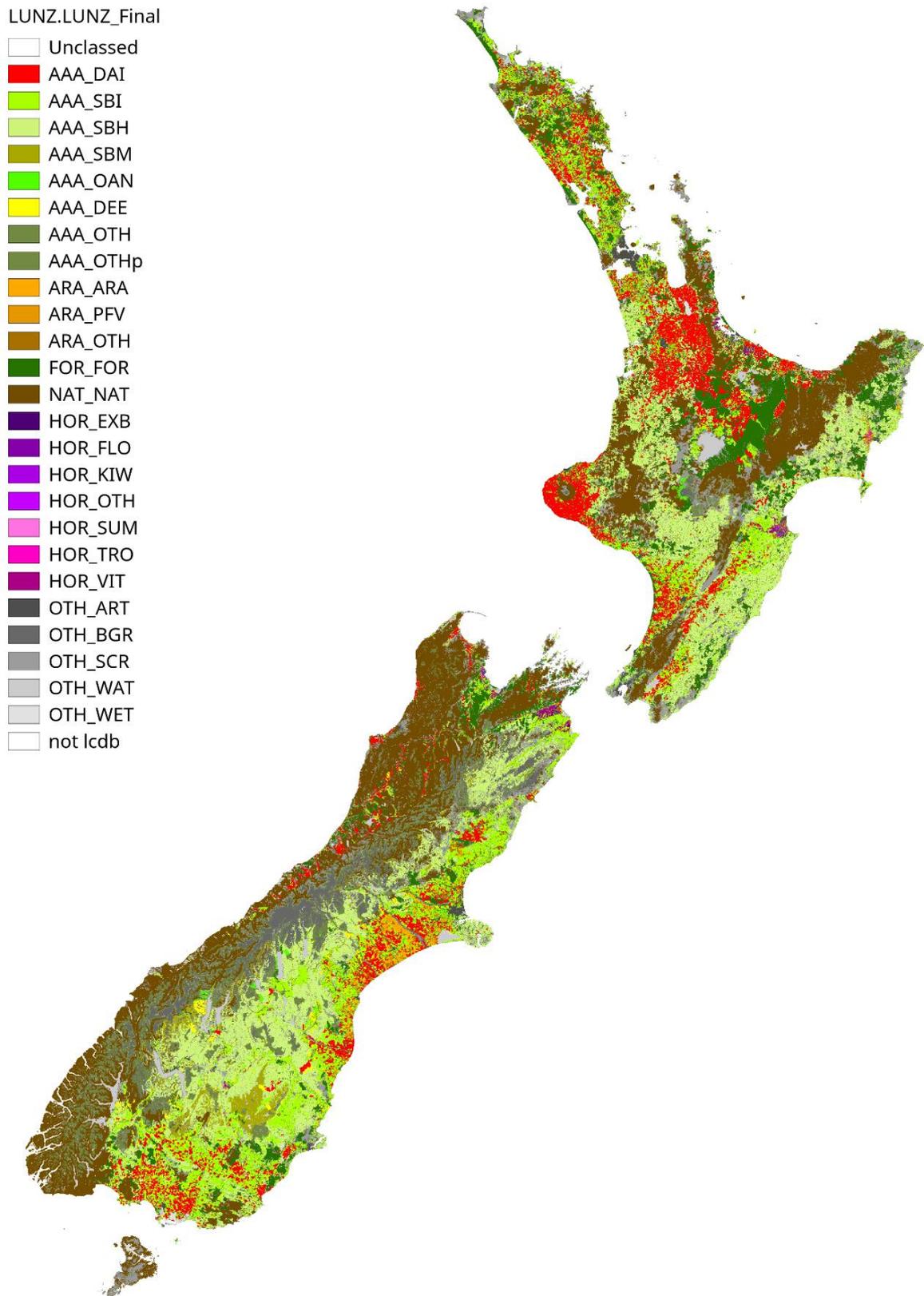


Figure 12 LUNZ LU classification produced by pyluc (visualised with QGIS).

Table 16 Comparison of LUNZ classification areas between the original recreation and the pyluc output. Note this is comparing a sum of vector (polygon) areas with a sum of raster pixel areas so small differences are to be expected

Tier 1	Tier 1 description	Tier 2	Orig. ha	pyluc ha	Diff (ha)	Diff (%)
ARA	Arable	ARA_ARA	176,891	176,933	42	0.02%
		ARA_PFV	28,485	285,98	113	0.40%
		ARA_OTH	164,311	164,215	-96	-0.06%
HOR	Horticultural	HOR_VIT	41,738	41,748	10	0.02%
		HOR_EXB	2,400	2,399	-1	-0.04%
		HOR_OTH	16,142	16,188	46	0.28%
		HOR_SUM	3,773	3,781	8	0.21%
		HOR_TRO	7,764	7,750	-14	-0.18%
		HOR_KIW	23,330	23,341	11	0.05%
		HOR_FLO	8,376	8,383	7	0.08%
		AAA	Pastoral	AAA_DAI	2,100,689	2,100,987
AAA_SBI	3,076,754			3,075,972	-782	-0.03%
AAA_SBH	5,035,667			5,036,475	808	0.02%
AAA_SBM	469,810			469,709	-101	-0.02%
AAA_DEE	213,903			213,757	-146	-0.07%
AAA_OAN	228,214			227,273	-941	-0.41%
AAA_OTHp	1,438,498			1,437,906	-592	-0.04%
AAA_OTH	627,746			629,087	1341	0.21%
FOR	Forestry	FOR_FOR	2,045,220	2,045,102	-118	-0.01%
NAT	Native forest	NAT_NAT	6,432,333	6,432,969	636	0.01%
OTH	Other LCDB	OTH_SCR	2,701,532	2,701,252	-280	-0.01%
		OTH_BGR	1,051,106	1,051,230	124	0.01%
		OTH_WAT	536,766	536,605	-161	-0.03%
		OTH_WET	165,284	165,120	-164	-0.10%
		OTH_ART	245,648	245,638	-10	0.00%
not lcdb	not lcdb	not lcdb	201,601	n/a	n/a	n/a

Table 17 Comparison between the RASTERISED original recreation of the LUNZ classification and the pyluc output to eliminate the differences caused by comparing raster and vector areas (only differences > 1 ha included)

Orig. LUC	Differences		Diff (ha)
		pyluc LUC	
HOR_FLO		HOR_TRO	2
AAA_OAN		AAA_OTH	1,026
AAA_OAN		AAA_OTHp	3
HOR_FLO		HOR_KIW	2
ARA_OTH		ARA_PFV	56
HOR_VIT		HOR_OTH	25
HOR_VIT		HOR_SUM	6
HOR_VIT		HOR_FLO	7

Figure 12 shows the product of the LUNZ pyluc definition script as visualised by QGIS, with colours chosen to allow direct comparison with the Tier 2 map in Figure 10. As these figures look very similar, Table 16 has been provided to show the differences in area for each class. This only represents total areal difference so does not detect individual misclassified areas, but it does show that the areas generally agree. As the caption mentions, a large proportion of these differences are likely due to fact that vector areas are being compared with raster, where many small boundary differences over thousands of classified patches eventually add up to large areal differences. Table 17 supports this by showing a direct pixel-to-pixel comparison between the rasterised original output and the pyluc output, where only a small number of classes are affected over relatively small areas. This comparison also accounts for any misclassifications and shows that, for most LUNZ classes, the pyluc output is identical to the original recreation. The main class causing problems within pyluc is AAA_OAN, with approximately 1026 ha of this misclassified as AAA_OTH, corresponding to 0.45% of the total AAA_OAN area. There are small differences adding up to a few tens of hectares for the HOR and ARA Tier 1 classes.

Technical documentation and provenance examples for LUNZ are lengthy and complicated, reflecting the nature of this LU classification, and for brevity are not included in this report. Please contact the authors for more information.

5.5 Discussion

As the MOTU LURNZ LU classification logic is relatively straight-forward, the majority of the validation effort has been directed at the more complicated LUNZ classification scheme. Output from the pyluc definition script compares very well with the original recreation when both are compared in raster-space; however, there are some minor differences in statistics when comparing the pyluc raster output with the original vector-based dataset (up to 0.41% of total area). Misclassifications are relatively rare, with only some classes affected and total areas of the order of tens of hectares or less. The exception to this is the 'AAA_OAN' ('AAA – other animal') class, where the pyluc definition

script incorrectly assigns the 'AAA_OTH' ('AAA other') to 1026 hectares. This has been traced to a small number of properties that fall into a 'grey-zone' where it is likely calculated stock units are rounded differently between the Arc GIS and Python backends. The affected area is proportionally small (0.45%) and the distinction between the two in the input data can be minor when reported stock units are low, so the decision was made not to pursue the misclassification any further.

6 Summary and conclusions

The Python land use classification framework (pyluc) has been shown to reliably replicate more manually produced LU classifications within a formalised environment. Automatically generated technical documentation and provenance look to be useful additional outputs ready for incorporation into more traditional reports and will greatly help in the later recreation of a given LU classification. Provenance information, combined with the classification definition script, also helps provide an audit-trail for source material and timing of an LU classification. Two potential disadvantages of pyluc are the additional time taken to create the definition scripts, and the fact that processing happens in raster-space so final output, even if vectorised, will not exactly match the line-work of the input datasets. A mitigating factor for the definition script creation is that this helps to formalise the classification logic clearly and explicitly in a way that is concretely tied to the final output so we argue the additional time is worth investing.

7 Recommendations

We recommend that other land use classifications – particularly those developed independently by regional councils, and those currently being developed for national projects such as Our Land and Water National Science Challenge – be evaluated and described according to the approach used in this project.

We also recommend that pyluc be adopted as a 'final step' in the process for creating formal LU classifications for all national applications. The existence of a classification definition script alone will help future reproducibility, while the provenance records can lend more credibility, from a technical sense, to the delivered output.

8 Acknowledgements

This project is undertaken within the Innovative Data Analysis programme, funded by the Ministry of Business and Innovation.

9 References

- Anastasiadis S, Kerr S, Daigneault A, Doole G, Greenhalgh S, Montes de Oca Munguia O, Rutledge D, Turner J 2013. Understanding the practice of land use modelling. Wellington, Motu Economic and Public Policy Research.
- Anastasiadis S, Kerr S, Zhang W, Allan C, Power W 2014. Land use in rural New Zealand. Spatial land use, land-use change, and model validation. Motu Working Paper 14-07. Wellington, Motu Economic and Public Policy Research.
- Anastasiadis S, Kerr S 2013. Land use and farming intensity: for 1996, 2008 and 2020. Wellington, Motu Economic and Public Policy Research Report for the Parliamentary Commissioner for the Environment.
- Ausseil A-G, Manderson A, Rutledge D, Wyman T, Osborne T 2015. Land use mapping review for Ministry for the Environment. Landcare Research contract report LC 2356.
- Cavanagh J 2014. Status of cadmium in New Zealand soils. Lincoln, Landcare Research.
- Gibb R 2010. The perfect world. The real world. Conceptual diagrams of LUNZ. Lincoln, Landcare Research.
- Heinz D 2011. FarmsOnLine Privacy Impact Assessment (Version 3.2). Retrieved from <http://www.mpi.govt.nz/document-vault/2564> (10 June 2015).
- Hill RB & Sparling GP 2009. Soil quality monitoring. Land and soil monitoring: a guide for SoE and regional council reporting. Hamilton, Land Monitoring Forum. Pp. 27–88.
- Hill RB, Sparling G, Frampton C, Cuff J 2003. National soil quality review and programme design. Technical Paper 75, Land. Wellington, Ministry for the Environment.
- Land Information New Zealand 2012. Stronger biosecurity protection using online location-based data. Farmsonline enhances New Zealand's information of the rural sector. Geospatial Case Study. Retrieved from http://www.linz.govt.nz/system/files_force/media/file-attachments/mpifarmsonlinecasestudysml.pdf?download=1&download=1 (10 June 2015).
- Landcare Research 2012. LCDB v3.0 Accuracy assessment [Online]. Available: <http://www.lcdb.scinfo.org.nz/about-lcdb/accuracy-assessment>. Accessed: 10 July 2016.
- Landcare Research 2014. NZ Land cover database. www.lcdb.scinfo.org.nz.
- LINZ 2017. Accuracy of the digital cadastre (webpage). Available: <http://www.linz.govt.nz/data/linz-data/property-ownership-and-boundary-data/accuracy-digital-cadastre> (accessed 20 July 2017).
- Livestock Improvement Corporation (LIC) 2014. New Zealand dairy statistics 2013–14. <http://www.dairynz.co.nz/publications/dairy-industry/new-zealand-dairy-statistics-2013-14/> (accessed 18 February 2015).
- Ministry for the Environment 2010. Soil health environmental snapshot. Wellington, Ministry for the Environment.

- Ministry of Agriculture and Forestry 2011a. National Animal Identification and Tracing Bill - Ministry of Agriculture and Forestry – FarmsOnline Stage 2 Business Case Final Options analysis. http://www.parliament.nz/en-nz/pb/sc/documents/advice/49SCPP_ADV_00DBHOH_BILL10501_1_A187083/ministry-of-agriculture-and-forestry-farmsonline-stage. (retrieved 10 June 2015).
- Ministry of Agriculture and Forestry 2011b. Further Information regarding Farmsonline for the National Identification and Tracing Bill. Departmental Report of the Primary Production Select Committee National Animal Identification and Tracing (NAIT) Bill Annex 2. http://www.parliament.nz/resource/en-nz/49SCPP_ADV_00DBHOH_BILL10501_1_A187081/12db253ea43f7f918328bab5384cea873a3fc4a7 (retrieved 10 June 2015).
- Motu 2010. Land Use New Zealand (LUNZ) Map. Data documentation. Available: <https://motu.nz/our-work/environment-and-resources/lurnz/land-use-new-zealand-lunz-map/> (retrieved 21 Jan 2017).
- Newsome P, Shepherd J, Pairman D 2013. Establishing New Zealand's LUCAS Land Use and Land Use-Change and Forestry 2012 Map. Landcare Research contract report LC1667.
- Parliamentary Commissioner for the Environment. 2013. Water quality in New Zealand: Land use and nutrient pollution. Wellington, PCE.
- Parshotam A, Elliott S, Shankar U, Wadhwa S 2013. National nutrient mapping using the CLUES model. NIWA Client report HAM2013-086.
- Pearless S 2005. Report back on Agribase optimisation project 2. Service level agreement and metrics/. Wellington, HeXaD Consulting Group for Biosecurity NZ.
- Pilaar C, Thompson S, Ching N 1995. The New Zealand land cover database: a pilot scheme and a forestry perspective. New Zealand Forestry Journal, May 1995. http://nzjf.org.nz/free_issues/NZJF40_1_1995/A35A39C7-BD9A-4FE5-9EAF-E1E342BD9B68.pdf
- Poyry Ltd 2014. Accuracy assessment of LUCAS 2012 Land Use Map. Where based, Poyry Management Consulting (NZ) Limited.
- Sanson R 2005. The Agribase™ farm location database. Proceedings of the New Zealand Society of Animal Production 65: 93–96
- Sanson R 2013. AgriBase and its potential for use in NZ traceability and verification. FAB News (official newsletter of the NZVA Food Safety, Animal Welfare & Biosecurity Branch), Issue 16 (December 2013). http://fab.nzva.org.nz/sites/default/files/domain-4/FAB%20newsDec2013_0.pdf (retrieved 10 June 2015).
- Sanson R, Cook A, Fairweather J 2004. A study of smallholdings and their owners. MAF Information Paper No: 53. <http://maxa.maf.govt.nz/mafnet/publications/statistics/forestry/smallholdings/smallholders.pdf>. (retrieved 10 June 2015).
- Sanson R, Pearson A 1997. Agribase – A national spatial farm database. Proceedings of 8th International Symposium on Veterinary Epidemiology and Economics. *Epidémiologie et Santé Animale* Paris, France (31–32): 12.16.11–12-16.13

Statistics New Zealand 2015b. Agricultural Land database matching and data integration: Privacy impact assessment. www.stats.govt.nz.

Statistics New Zealand 2015a. Agricultural production statistics: June 2014 (final). http://www.stats.govt.nz/browse_for_stats/industry_sectors/agriculture-horticulture-forestry/AgriculturalProduction_final_HOTPJun14final/Data%20Quality.aspx. (retrieved 23 June 2015)

Waikato Regional Council (WRC) 2015. Stock density indicator. <http://www.waikatoregion.govt.nz/Environment/Environmental-information/Environmental-indicators/Land-and-soil/Land/riv9-technical-information/> (accessed 18 September 2014).

Woods R, Bidwell V, Clothier B, Elliott S, Harris S, Hewitt A, Wheeler D 2006. The CLUES project: predicting the effects of land-use on water quality – Stage II. NIWA Client Report: HAM2006-096 July 2006. Available <http://maxa.maf.govt.nz/mafnet/rural-nz/sustainable-resource-use/clues/stage-2/index.htm>

Appendix 1 – LURNZ classification rules example

Example code (python via Arc field calculator)

```
Motu(!MOTU12!,!Motu1!,!Ownshp!)
```

```
def Motu(MOTU12,Motu1,Ownshp):
    if(MOTU12 == "LUR_urb"):
        return "Urban"
    elif (MOTU12 == "LUR_pas" and Motu1 == "m_dai"):
        return "Dairy"
    elif (MOTU12 == "LUR_pas" and Motu1 == "m_sb"):
        return "Sheep and beef"
    elif (MOTU12 == "LUR_pas" and Motu1 == "m_hort"):
        return "Sheep and beef"
    elif (MOTU12 == "LUR_hort" and Motu1 == "m_hort"):
        return "Horticulture"
    elif(Motu1 == "m_oa"):
        return "Other animal"
    elif(Motu1 == "m_lf"):
        return "Lifestyle"
    elif (MOTU12 == "LUR_pas" and Ownshp == "HC lease"):
        return "Public pasture"
    elif (MOTU12 == "LUR_pas" and Ownshp == "Private"):
        return "Private pasture unclassified"
    elif(Ownshp == "Public"):
        return "DoC/Public land"
    elif(MOTU12 == "LUR_oth"):
        return "Non-productive"
    elif(MOTU12 == "LUR_scrub"):
        return "Scrub"
    elif(MOTU12 == "LUR_pf"):
        return "Forestry"
    elif(MOTU12 == "LUR_if"):
        return "Indigenous forest"
    else:
        return "Unclassified"
```

Appendix 2 – pyluc LURNZ classification definition

```
from pyluc import LUC

"""
SETUP
"""
#preamble
name = 'MOTU_LURNZ'
extent_nztm = 1760000, 5828000, 1805000, 5808000#1083000, 6207000, 2096000, 4745000
resolution_m = 20
classification = LUC.Classification(name, extent_nztm, resolution_m)
#organisations
classification.doc.add_luc_owner(
    'lr',
    'Landcare Research',
    'http://www.landcareresearch.co.nz/'
)
classification.doc.add_other_organization(
    'linz',
    'Land Information New Zealand',
    'http://www.linz.govt.nz/'
)
classification.doc.add_other_organization(
    'aq',
    'Asure Quality',
    'https://www.asurequality.com'
)
#author/operator
classification.doc.add_luc_author(
    'mandersona',
    'Andrew Manderson',
    'lr'
)
classification.doc.add_luc_operator(
    'ben.jolly',
    'Ben Jolly',
    'lr',
    delegators=['mandersona']
)

"""
INPUTS
"""
classification.add_input_layer(
    'agribase',
    'Frmtyp2',
    'https://iris.scinfo.org.nz/layer/461-agribase-enhanced-001/',
    attributed_to='aq'
)
classification.add_input_layer(
    'lcdb',
    'Class_2012',
    'https://iris.scinfo.org.nz/layer/423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand/',
    attributed_to='lr'
)
classification.add_input_layer(
    'Ownshp',
    'Ownshp',
    'https://iris.scinfo.org.nz/layer/460-publicprivatenzland01b/',
    attributed_to='linz'
)

"""
CLASSIFICATION RULES
"""
#Lookup tables
lcdb_preclass_lut = {
    1: 'LUR_urb', 2: 'LUR_urb', 5: 'LUR_urb', 6: 'LUR_oth',
    10: 'LUR_oth', 12: 'LUR_oth', 14: 'LUR_oth', 15: 'LUR_pas',
    16: 'LUR_oth', 20: 'LUR_oth', 21: 'LUR_oth', 22: 'LUR_oth',
    30: 'LUR_hort', 33: 'LUR_hort', 40: 'LUR_pas', 41: 'LUR_pas',
    43: 'LUR_pas', 44: 'LUR_pas', 45: 'LUR_oth', 47: 'LUR_scrub',
```

```

50: 'LUR_scrub', 51: 'LUR_scrub', 52: 'LUR_scrub', 54: 'LUR_scrub',
55: 'LUR_scrub', 56: 'LUR_scrub', 58: 'LUR_scrub', 64: 'LUR_pf',
68: 'LUR_pf', 69: 'LUR_if', 70: 'LUR_oth', 71: 'LUR_pf'
}
agribase_preclass_lut = {
'ALA': 'm_oa', 'API' : 'm_oa', 'ARA': 'm_hort', 'BEF': 'm_sb',
'DAI': 'm_dai', 'DEE' : 'm_oa', 'DOG': 'm_oa', 'DRY': 'm_dai',
'EMU': 'm_oa', 'FIS' : 'm_oa', 'FLO': 'm_hort', 'FOR': 'm_pf',
'FRU': 'm_hort', 'GOA' : 'm_oa', 'GRA': 'm_dai', 'HOR': 'm_oa',
'LIF': 'm_lf', 'MTW' : 'm_oth', 'NAT': 'm_scb', 'NEW': 'm_oth',
'NOF': 'm_scb', 'NUR' : 'm_hort', 'OAN': 'm_oa', 'OPL': 'm_hort',
'OST': 'm_oa', 'OTH' : 'm_oth', 'PAS': 'm_oa', 'PIG': 'm_oa',
'POU': 'm_oa', 'Pcrop': 'm_hort', 'SHP': 'm_sb', 'SLY': 'm_oth',
'SNB': 'm_sb', 'TOU' : 'm_lf', 'UNS': 'm_scb', 'URB': 'm_oth',
'VEG': 'm_hort', 'VIT' : 'm_hort', 'ZOO': 'm_oa'
}
}
#classification rules
def motu_classification(inputs):
    return [
        ('Unclassed', inputs.default),
        ('Indigenous forest', inputs.MOTU12 == 'LUR_if'),
        ('Forestry', inputs.MOTU12 == 'LUR_pf'),
        ('Scrub', inputs.MOTU12 == 'LUR_scrub'),
        ('Non-productive', inputs.MOTU12 == 'LUR_oth'),
        ('DoC/Public land', inputs.Ownshp == 'Public'),
        ('Private pasture unclassified', (inputs.MOTU12 == 'LUR_pas') & (inputs.Ownshp == 'Private')),
        ('Public pasture', (inputs.MOTU12 == 'LUR_pas') & (inputs.Ownshp == 'HC lease')),
        ('Lifestyle', inputs.Motu1 == 'm_lf'),
        ('Other animal', inputs.Motu1 == 'm_oa'),
        ('Horticulture', (inputs.Motu1 == 'm_hort') & (inputs.MOTU12 == 'LUR_hort')),
        ('Sheep and beef', ((inputs.Motu1 == 'm_sb') | (inputs.Motu1 == 'm_hort')) & (inputs.MOTU12 ==
'LUR_pas')),
        ('Dairy', (inputs.Motu1 == 'm_dai') & (inputs.MOTU12 == 'LUR_pas')),
        ('Urban', inputs.MOTU12 == 'LUR_urb')
    ]
}
"""
OUTPUTS (defining classification steps)
"""
classification.add_classification_step('mandersona', 'Motu1', agribase_preclass_lut, 'agribase')
classification.add_classification_step('mandersona', 'MOTU12', lcdb_preclass_lut, 'lcdb')
classification.add_classification_step('mandersona', 'Final', motu_classification)

```

Appendix 3 – LURNZ documentation



motu-luc-automate
d.pdf

Appendix 4 – LURNZ provenance information

```
document
  prefix foaf <http://xmlns.com/foaf/0.1/>
  prefix inputs <http://www.landcareresearch.co.nz/lucs/MOTU_LURNZ/inputs/>
  prefix linz <http://www.linz.govt.nz/>
  prefix rules <http://www.landcareresearch.co.nz/lucs/MOTU_LURNZ/rules/>
  prefix results <http://www.landcareresearch.co.nz/lucs/MOTU_LURNZ/results/>
  prefix aq <https://www.asurequality.com>
  prefix lr <http://www.landcareresearch.co.nz/>
  prefix lris <https://lris.scinfo.org.nz/layer/>

  wasGeneratedBy(results:MOTU12, rules:lcdb_preclass_lut, 2018-03-01T10:36:51.934000+00:00)
  wasGeneratedBy(results:Final, rules:motu_classification, 2018-03-01T10:36:51.934000+00:00)
  wasGeneratedBy(results:Motu1, rules:agribase_preclass_lut, 2018-03-01T10:36:51.934000+00:00)
  activity(rules:lcdb_preclass_lut, -, -)
  activity(rules:motu_classification, -, -)
  activity(rules:agribase_preclass_lut, -, -)
  entity(lris:48460-publicprivatenzland01b)
  entity(lris:Ownshp, [foaf:name="Ownshp"])
  entity(lris:48423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand)
  entity(lris:48461-agribase-enhanced-001)
  entity(lris:lcdb, [foaf:name="Class_2012"])
  entity(results:Motu1)
  entity(results:MOTU12)
  entity(lris:agribase, [foaf:name="Frmtyp2"])
  entity(results:Final)
  agent(lr:ben.jolly, [prov:type='prov:Person', foaf:name="Ben Jolly"])
  agent(linz:linz, [prov:type='prov:Organization', foaf:name="Land Information New Zealand"])
  agent(aq:aq, [prov:type='prov:Organization', foaf:name="Asure Quality"])
  agent(lr:mandersona, [prov:type='prov:Person', foaf:name="Andrew Manderson"])
  agent(lr:lr, [prov:type='prov:Organization', foaf:name="Landcare Research"])
  wasAttributedTo(lris:48423-lcdb-v41-land-cover-database-version-41-mainland-new-zealand, lr:lr)
  wasAttributedTo(lris:48460-publicprivatenzland01b, linz:linz)
  wasAttributedTo(lris:48461-agribase-enhanced-001, aq:aq)
  wasAttributedTo(results:Final, lr:mandersona)
  wasAttributedTo(results:Motu1, lr:mandersona)
  wasAttributedTo(results:MOTU12, lr:mandersona)
  actedOnBehalfOf(lr:ben.jolly, lr:lr, -)
  actedOnBehalfOf(lr:mandersona, lr:lr, -)
  used(rules:motu_classification, lris:Ownshp, -)
  used(rules:lcdb_preclass_lut, lris:lcdb, -)
  used(rules:agribase_preclass_lut, lris:agribase, -)
  used(rules:motu_classification, results:Motu1, -)
  used(rules:motu_classification, results:MOTU12, -)
  wasAssociatedWith(rules:motu_classification, lr:mandersona, -, [prov:role="author"])
  wasAssociatedWith(rules:motu_classification, lr:ben.jolly, -, [prov:role="operator"])
  wasAssociatedWith(rules:lcdb_preclass_lut, lr:mandersona, -, [prov:role="author"])
  wasAssociatedWith(rules:lcdb_preclass_lut, lr:ben.jolly, -, [prov:role="operator"])
  wasAssociatedWith(rules:agribase_preclass_lut, lr:mandersona, -, [prov:role="author"])
  wasAssociatedWith(rules:agribase_preclass_lut, lr:ben.jolly, -, [prov:role="operator"])
  wasDerivedFrom(lris:agribase, lris:48461-agribase-enhanced-001, -, -, -)
  wasDerivedFrom(lris:Ownshp, lris:48460-publicprivatenzland01b, -, -, -)
  wasDerivedFrom(lris:lcdb, lris:48423-lcdb-v41-land-cover-database-version-41-mainland-new-zeala
endDocument
```

Figure 13 Syntax-highlighted PROV-N code uploaded to [ProvStore](#) (hosted [here](#)).

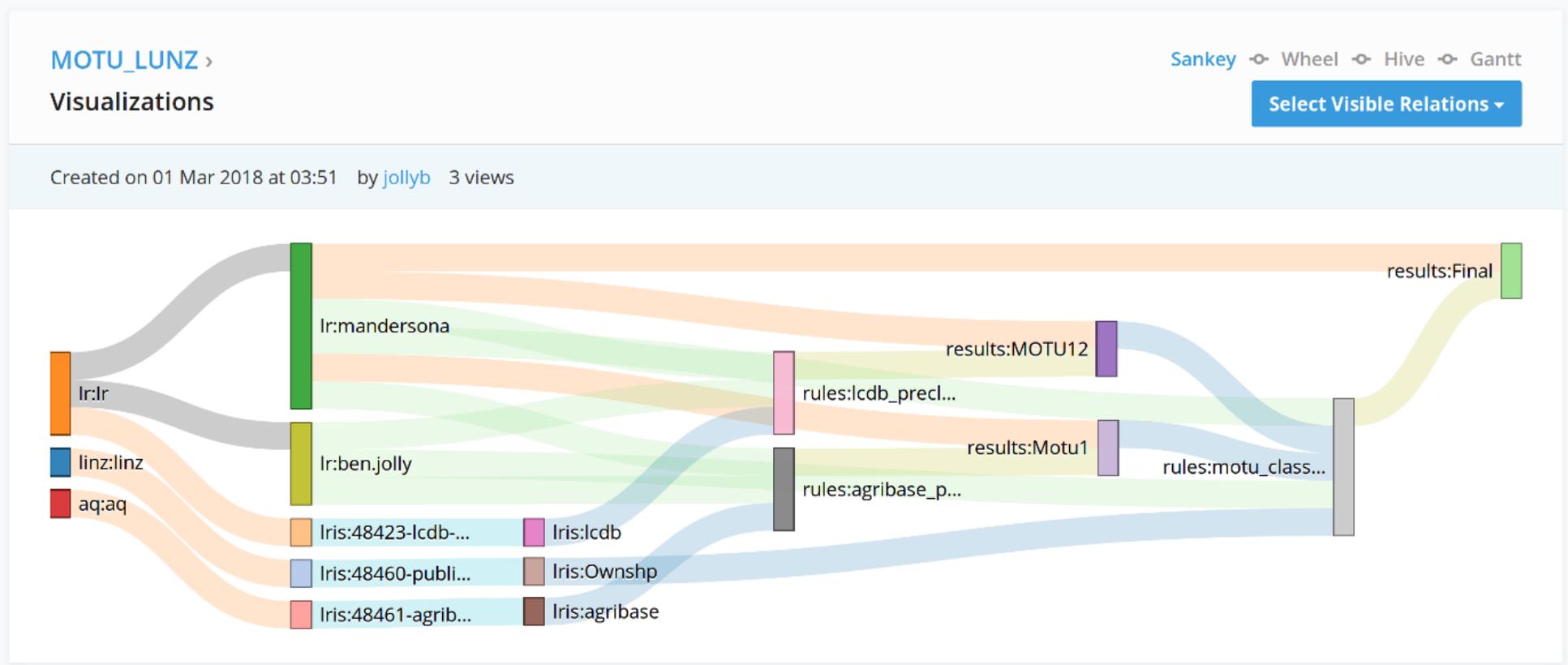


Figure 14 Screenshot of interactive 'Sankey' diagram ([original](#)).

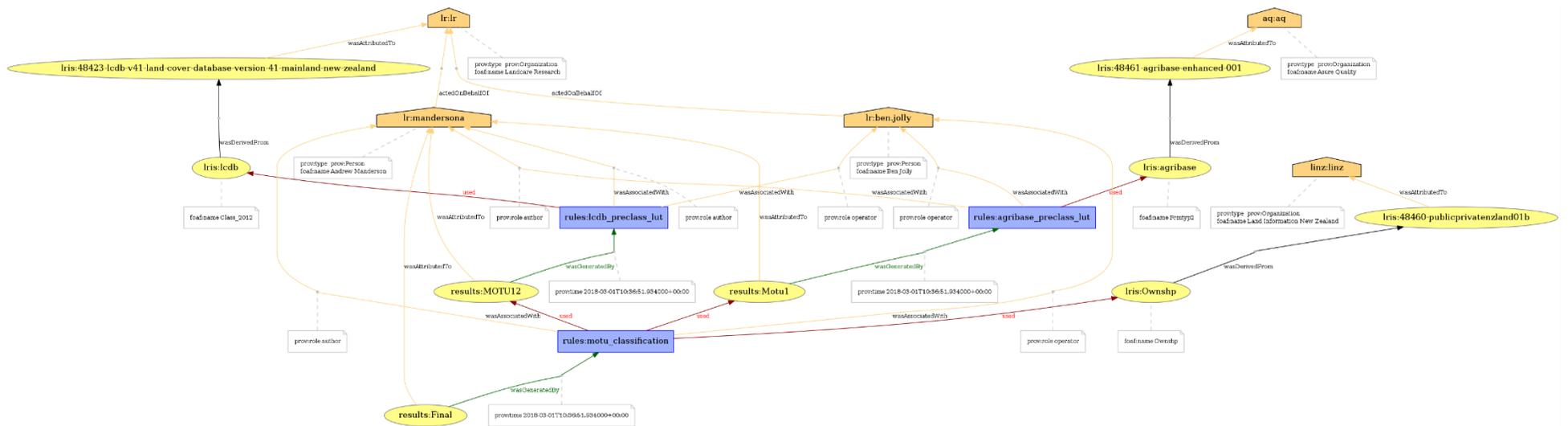


Figure 15 Visual graph of provenance information for the Motu LURNZ LU classification ([original](#)).