Motu

New Zealand Monitor Farm Data

Loïc Henry, Edmund Lou and David Fleming

A Technical Paper

Motu Economic and Public Policy Research

November 2017

Document information

Author contact details

Loïc Henry

Motu Economic and Public Policy Research

loic.henry@agroparistech.fr

Edmund Lou

Motu Economic and Public Policy Research

edmund.lou@motu.org.nz

David Fleming

Motu Economic and Public Policy Research

david.fleming@motu.org.nz

Acknowledgements and Disclaimer

This work has been conducted as part of the project "Barriers to Adoption of 'No-cost' Mitigation Options" funded by the New Zealand Government under 'The Global Research Alliance on Agricultural Greenhouse Gases' (MPI research project contract #RFP17526). All errors, omissions, and opinions expressed are the responsibility of the authors.

Motu Economic and Public Policy Research

PO Box 24390 info@motu.org.nz +64 4 9394250

Wellington www.motu.org.nz

New Zealand

 ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

Table of Contents

Tables

1 Introduction

The New Zealand Monitor Farm Data (NZMFD) is a merged dataset of two sources. The first source, called financial data, contains information on the financial status of farms. These data were collected by the Ministry of Agriculture and Forestry (MAF)[1] under the Farm Monitoring Programme, which was designed to provide an annual aggregated overview of a range of farm types throughout New Zealand. The second source documents information about each farm's physical production inputs alongside their economic and environmental outcomes. As with the financial data, most of the input information was collected by MAF, while the environmental outcomes were calculated using OVERSEER® (Overseer) version 6.2.1, which is an agricultural decision support tool, developed by AgResearch and aimed at assisting farmers in examining the on-farm impact of nutrient flows.[2,3] For this reason, we call the second source the Overseer data.

Structurally, NZMFD is an unbalanced panel. Every farm in this dataset is uniquely identified by a Farm ID (e.g. CANDY1). In total, the NZMFD includes 407 observations, with data ranging from 2009 to 2012. These include farm data from most regions of the country, subdivided into dairy farms (223 farms), sheep and beef farms (165), and deer farms (19).

In the next section we describe in detail how this merged dataset was constructed and the main characteristics of the data. The dataset may be available for research purposes, although confidentiality agreements will first need to be established with the Ministry for Primary Industries (MPI).[4]

2 Data sources

2.1 The financial data

The financial data are farm-level data collected by MAF under the Farm Monitoring Programme.[5] The original raw financial data are made up of 75 CSV files, each containing several spreadsheets. Each file is categorised by region and farm type in a given year between 2006 and 2012. There are three farm types under consideration: dairy, sheep and beef, and deer. In total, there are 28 files for dairy farms, 36 for sheep and beef farms, and 11 for deer farms.[6] MAF used all the individual CSV files to create a representative model farm for each farm type, which was the main focus of their farm monitoring overview report (e.g. MAF 2011).[7] However, these model farms might give

[1] MAF was dissolved and became part of the then-newly formed Ministry for Primary Industries (MPI) in 2012.
[2] The Overseer model is based on nutrient budgeting. A nutrient budget, like a financial budget, offers information about inputs and outputs within a farm, but in terms of nutrients rather than finances. See Wheeler et al. (2003) for more information.
[3] Overseer is now jointly owned by MPI, AgResearch and the Fertiliser Association of New Zealand.
[4] Please contact Motu if you would like to have access to the database.
[5] We received the raw data from Stephen Murray at MPI (stephen.murray@mpi.govt.nz).
[6] The file name of every CSV file contains information about the farm type, the year of monitoring and the region.
[7] A representative farm model is a simple average of all the monitor farms within a given farm type.

biased information as the MAF monitor farms were not necessarily chosen in an empirically random way.

To carry out some analysis beyond the descriptive information of representative farms across regions, as provided in the MAF reports, we combined the CSV files and obtained an unbalanced panel dataset, which we called the financial dataset. This dataset encompasses 1,605 observations, with, in some cases, more than one observation per farm across different years.[8] To merge these CSV files, we followed the steps below:

1. We cleaned and transposed every spreadsheet in every CSV file, such that each row stands for one particular farm and all the financial variables are listed in columns. We then saved these as a new CSV file.[9] If there was more than one spreadsheet in a CSV file before transposition, we merged these spreadsheets using Farm ID and eliminated duplicate variables afterwards.
2. We took one CSV file of a farm type and defined it as our 'base file', then created two new variables: year (of monitoring) and region.[10] We documented all the variables of this base file in a dictionary.
3. We took the remaining CSV files, with year and region variables added, and proceeded to match their variables. If the file had exactly the same variables as the base file, we simply closed it. However, if it did not, we updated the dictionary by adding the new variables.[11] If a new variable had a similar name to that of one of the variables in the base file, it was replaced by the existing name in the dictionary. We repeated this procedure iteratively until all the CSV files of the given farm type were checked. The outcome was 75 CSV files with three accompanying dictionaries, one for each farm type.
4. We imported all the CSV files into RStudio and applied the '`data.table::rbindlist`' command to create a combined dataset.[12]
5. We conducted consistency checks, using descriptive statistics, to see if there were unusual values. These values pointed to some possible mistakes that occurred during the procedure; these were fixed accordingly.

It is important to highlight here that, even though these data management steps were carefully followed, the financial dataset may still not be error-free. If a user does observe some unusual outliers, it is therefore suggested they go back to the raw financial data. The three dictionaries were combined to create a glossary in order to facilitate this procedure.

[8] Specifically, the data include observations of 300 dairy farms, 300 sheep and beef farms, and 84 deer farms.
[9] It is worth mentioning that all the derived variables were discarded in the data cleaning process.
[10] The region variable here was originally called 'monitoring model'. See Appendix for details.
[11] This means that not all the spreadsheets available have the same set of variables.
[12] This command helps to make a dataset by matching column names. Missing values are represented by 'NA'.

2.2 The Overseer data

All the information collected by MAF was originally stored in 1,313 Overseer files. These Overseer files are indicated by the .ovp file extension, and can be run by the Overseer application.[13] We gave all the Overseer files to AgResearch, and received back an Excel version of the data.[14] In total, there are 1,166 observations, in some cases encompassing more than one observation per farm between the years 2009 and 2012.[15] The 147 missing observations were originally created in an old version of Overseer (earlier than 6.2.1), which could not be run by AgResearch.

As with the financial data, we turned the Overseer data into a manageable dataset (called the Overseer dataset). Both the data cleaning and the final compilation (there are nine spreadsheets altogether in the Overseer data) were carried out using Stata software. Below, we explain some of the important steps in the data cleaning process:

1. We removed the 'dairy replacement' category in the variable 'enterprise type' because there are no data for it.[16]
2. We added the monthly totals of irrigation water to give an annual reading for every farm. In the 'Block' spreadsheet, we removed farms with missing or zero values for annual temperature or annual rainfall.[17] However, since the Overseer model measures these two variables, this might explain some of the mistakes that became apparent in the process of extracting information from the Overseer files.
3. We calculated the total block area when 'block type' = 'pastoral'.[18] Note that in some cases there are different annual rainfall measures for the same farm. For example, monitor farm CANDY1 has an annual rainfall of 650 mm for its dairy block but 620 mm for its effluent block.[19] To create a single annual rainfall reading for farms like CANDY 1, we applied a block-area-weighted approach, i.e.

$$\text{rainfall} = \sum_{i} \frac{\text{area of block i}}{\text{total block area}} \times \text{rainfall}_{i}$$

4. All the nine spreadsheets were combined according to their index number.[20]

[13] We received all the Overseer files from Aaron Carson on 14 December 2015; Aaron left MPI in October 2016.
[14] We received the Overseer data from Mike Rollo at AgResearch on 30 June 2016.
[15] Specifically, there are 273 dairy farms, 296 sheep and beef farms, and 19 deer farms.
[16] The variable enterprise type takes a type t ∈ {dairy, dairy replacement, sheep, beef, deer}.
[17] The 'Block' spreadsheet contains mainly information about farm blocks, such as block area, topography, rainfall, temperature, soil type, etc.
[18] Block type ∈ {fodder crop, house, riparian, pastoral, tree}.
[19] This could happen if the farm size is sufficiently large, otherwise there might be a problem either in the model or in the data extraction process.
[20] In the Overseer data, every farm that corresponds to an Overseer file has an index number.

3 Overview of the NZMFD data

The NZMFD dataset is a merging of the two datasets described above. The process was achieved by matching farm ID and year. Unfortunately, however, the final NZMFD has only 407 observations. The reasons for such a large drop in observations are mainly because:

- the financial dataset ranges from 2006 to 2012, but the Overseer data cover only the period from 2009 to 2012;
- all the financial files of beef and sheep farms from 2009 to 2010 are lost except for one from Northland in 2010;[21] and
- all the financial files of dairy farms in 2012 are lost except for the same one from Northland.

The following tables summarise the number of farms of each farm type in total and in each year.

Table 1. Number of unique farms for each farm type[22]

Farm type	Number of unique farms	Percentage
Dairy	145	55.98
Sheep and beef	101	39.00
Deer	13	5.02
Total	259	100.00

Table 2. Number of unique farms for each year

Year	Number of unique farms	Percentage
2009	90	34.75
2010	43	16.60
2011	98	37.84
2012	28	10.81
Total	259	100.00

[21] Stephen Murray at MPI told us that these files were locked but they could not find a technician to break the passwords.
[22] As already mentioned, the sample of farms under the Farm Monitoring Programme was not consistent across years, and some farms appeared in several years.

Table 3. Number of unique farm types for each year

Year	Dairy farm	Sheep & beef farm	Deer farm
2009	89	0	1
2010	19	24	0
2011	34	52	0
2012	3	25	12
Total	145	101	13

The main areas of interest in the NZMFD can be summed up in the five following categories:

1. Geophysical information, e.g. total effective area, rainfall, temperature, topography and soil type.
2. Production status, e.g. milk solids for dairy farms, and revised stock units (RSU)[23] for both sheep and beef and deer farms.
3. Financial results, e.g. milk solids sales, cattle sales, sheep sales, deer and velvet sales, farm profit before tax and farm profit after tax.
4. Management practices, e.g. irrigation, use of dicyandiamide (DCD), use of feed pad and use of wintering pad.
5. Emissions and leaching, e.g. greenhouse gas (GHG) emissions (in particular, methane and nitrous oxide emissions), nitrogen leaching and phosphorus leaching.[24]

The following table shows the summary statistics for the above-mentioned variables. It is worth noting that some of the variables have fewer observations but may be important nonetheless for potential empirical analyses, and hence are included in the table.

[23] RSU is defined in Coop (1965) and Nicol and Brookes (2007).
[24] Unfortunately, we do not have *Escherichia coli* information.

Table 4. Summary statistics

Variable		Obs	Mean	SD	Min.	Max.
1	Total effective area (ha)	407	646.7	1,682.5	40	21,910
	Rainfall (mm)	407	1,165.8	366.2	400	2,500
	Temperature (°C)	407	12.6	2.1	8	16
2	Milk solids (kg)	223	148,879	99,382.1	31,000	645,000
	RSUs	162	3,365.6	2,496.4	25	17,000
3	Milk solids sales[25] ($)	107	1,021,030	718,944.3	227,740	4,772,220
	Cattle sales ($)	237	72,569.5	61,952.6	10,080	445,434
	Sheep sales ($)	162	540,921.9	339,138.2	18,296	1,662,125
	Deer and velvet sales ($)	29	426,403.9	328,465.6	63,844	1,410,269
	Net cash income ($)	407	803,441.9	603,154.1	47,210	4,814,077
	Farm working expenses ($)	407	438,417.8	356,115.7	34,072.2	2,423,022
	Cash operating surplus ($)[26]	407	352,057.9	296,567.2	-45,776	2,391,055
	Economic farm surplus ($)	134	414,807.5	370,794.4	-27,102.2	2,361,055
	Farm profit before tax ($)	407	181,455.8	232,692.9	-949,336	1,698,184
	Farm profit after tax ($)	407	153,786.4	216,677.7	-949,336	1,475,771
4	Irrigation*	407	0.13	0.33	0	1
	DCD use*	198	0.07	0.26	0	1
	Feed pad*	407	0.08	0.27	0	1
	Wintering pad*	407	0.005	0.07	0	1
5	Methane emissions (kg CO_2-eq/ha)	407	5,315.26	2,970.32	176	12,748
	Nitrous oxide emissions (kg CO_2-eq/ha)	407	1,938.90	1,279.54	60	5,934
	Carbon dioxide emissions (kg CO_2-eq/ha)[27]	407	945.03	857.05	19	4,536
	Total GHG emissions (kg CO_2-eq/ha)[28]	407	8,199.21	4,961.24	256	21,250
	Nitrogen leaching (kg N/ha)	407	32.83	25.88	2	132
	Phosphorus leaching (kg P/ha)	407	1.46	1.72	0	18.3

Notes: For cattle sales, sheep sales and deer and velvet sales, we treat 0 as no observation. All the variables with per hectare measure are calculated by dividing total effective area. *Denotes dummy variables.

[25] These figures are calculated using the formula shown in the spreadsheets: milk solid sales = advanced payment + deferred payment + dividend on wet sales + capacity adjustment.
[26] Cash operating surplus is defined as net cash income less farm working expenses.
[27] Carbon dioxide emissions are from fuel and electricity.
[28] Total GHG emissions is the sum of methane, nitrous oxide and carbon dioxide emissions.

Table 5. Distribution of farm topography

Topography	Number of farms	Percentage
Flat	279	68.55
Easy hill	62	15.23
Rolling hill	56	13.76
Steep hill	10	2.46

Table 6. Distribution of farm soil type

Soil type	Number of farms	Percentage
Peat	3	0.74
Podzol	7	1.72
Pumice	23	5.65
Recent/YGE/BGE	80	19.66
Sand (high P loss)	9	2.21
Sedimentary	189	46.44
Volcanic	96	23.59

4 Conclusion

This dataset fills an important gap in New Zealand agricultural economics at a practical level. It generates different fields for researchers to use in interrogating agricultural production, nutrient and greenhouse gases emissions and financial outcomes across farms.

Even though the NZMFD data is not large is restricted to only four years and is already five years old, it still presents useful information for analysis of the agricultural industry across New Zealand. We hope the dataset will be used by researchers interested in better understanding farm dynamics and their financial constraints.

References

Coop, I. E. 1965. 'A review of the ewe equivalent system'. *New Zealand Agricultural Science* 1(3): 13–18.

Ministry of Agriculture and Forestry (MAF). 2011. 'Farm monitoring overview'.

Nicol, A. M., and I. M. Brookes. 2007. 'The metabolisable energy requirements of grazing livestock, in pasture and supplements for grazing animals'. *In*: Rattray, P. V., A. M. Nicol and I. M. Brookes (*eds*). *Pasture and supplements for grazing dairy cattle*. Occasional Publication 14. New Zealand Society of Animal Production.

Wheeler, D. M., S. F. Ledgard, C. A. M. DeKlein, R. M. Monaghan, P. L. Carey, R. W. McDowell and K. L. Johns. 2003. 'OVERSEER® nutrient budgets – moving towards on-farm resource accounting'. *Proceedings of the New Zealand Grassland Association* 65: 191–194.

Appendix

Distribution of farm monitoring models

The following tables illustrate the distribution of farm monitoring models for different farm types in NZMFD.

Appendix Table 1. Distribution of monitoring models for dairy farms

Model	Number of farms	Percentage
Canterbury (CANDY)	38	17.04
Northland (NHLDY)	18	8.07
Southland (SHLDY)	25	11.21
South North Island (SNIDY)	30	13.45
Taranaki (TARDY)	46	20.63
Waikato/Bay of Plenty (WSADY)	66	29.60

Note: Model codes are in parentheses.

Appendix Table 2. Distribution of monitoring models for sheep and beef farms

Model	**Number of farms**	**Percentage**
Canterbury/Marlborough breeding and finishing (CMBF)	28	16.97
Canterbury/Marlborough hill country (CMHC)	14	8.48
Central North Island hill country (CNIHC)	13	7.88
East Coast hill country (ECHC)	7	4.24
Hawke's Bay/Wairarapa hill country (HBW)	16	9.70
Northland (NLD)	26	15.76
Otago dry hill (ODH)	15	9.09
South Island high country (SIHC)	9	5.45
Southland/South Otago hill country (SSOH)	17	10.30
Southland/South Otago intensive (SSOIF)	8	4.85
Waikato/Bay of Plenty intensive (WBOP)	12	7.27

Note: Model codes are in parentheses.

Appendix Table 3. Distribution of monitoring models for deer farms

Model	**Number of farms**	**Percentage**
North Island (NIDEER)	6	31.58
South Island (SIDEER)	13	68.42

Note: Model codes are in parentheses.