Environmental Modelling & Software 80 (2016) 54–65

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Mapping agriculture's impact by combining farm management handbooks, life-cycle assessment and search engine science

Javier Navarro [a, d, *], Brett A. Bryan [b], Oswald Marinoni [a], Sandra Eady [c], Anthony Halog [d]

[a] CSIRO Land and Water, EcoSciences Precinct, 41 Boggo Road, Dutton Park, QLD 4102, Australia
[b] CSIRO Land and Water, Waite Campus, SA 5064, Australia
[c] CSIRO Agriculture, FD McMaster Laboratory, Armidale, NSW 2350, Australia
[d] The University of Queensland, School of Geography, Planning and Environmental Management, Brisbane, QLD 4072, Australia

A R T I C L E I N F O

Article history:
Received 23 March 2015
Received in revised form
12 February 2016
Accepted 12 February 2016
Available online xxx

Keywords:
Sustainable intensification
Life-cycle inventories
Agricultural knowledge systems
Spatial modelling
Information retrieval
Farm management handbooks

A B S T R A C T

Despite great advancements in recent years, the availability of detailed and regionalised farm practice data at national scale remains an obstacle for spatially-detailed research on sustainable intensification. Parsing and information retrieval techniques were applied to 385 farm management handbooks to estimate farm practices (use of fertilisers, pesticides, water, fuel) of 72 commodities grown in 42 regions. Life-cycle inventories were used to derive GHG emissions and energy use from farm practice data. Practices and impacts were mapped at 1.1 km^2 resolution using agricultural census data and a remote-sensing-based land use map. Existing data was linearly extrapolated using a rule-based approach to fill spatial gaps. Estimates were, in aggregate, comparable to the best available data at national and local scales. Our method contributes to the push to create more spatially-detailed assessments of agricultural impacts at a national scale by focusing on the production of basic data at the farm level.

Data availability

A spatial database of agricultural inputs and practices from 385 farm management handbooks covering 72 commodities grown in 42 regions across Australia will be available from CSIRO's Data Access Portal at https://data.csiro.au/dap/landingpage?pid=csiro:12823.

1. Introduction

As global agricultural production strives to keep up with demand, increasing the accuracy and the spatial resolution of agricultural production data (e.g. yields and farm−level activities such as use of fertilisers and pesticides) will be instrumental for informing sustainable intensification (Garnett and Godfray, 2012; Smith et al., 2008; Van Ittersum et al., 2008). In the last few years several studies have gained insights into the environmental impacts of agriculture by combining techniques such as Remote Sensing, Life Cycle Assessment (LCA), or Integrated Assessment. Remote sensing studies have produced continental to global scale maps of key aspects of agricultural production such as fertiliser inputs or Net Biome Production (Ciais et al., 2010; Potter et al., 2010). Some state-of-the-art life-cycle inventories (LCI) at the national scale now contain data at the regional level (about 20–30% of total data) (Colomb et al., 2014; Eady et al., 2014). Other environmental assessments have also drawn from regional farm statistics to produce better analyses (Dolman et al., 2012; Thomassen et al., 2009; Van der Meulen et al., 2014). Van Ittersum et al. (2008) pioneered the integrated assessment of agriculture at a continental scale; the SEAMLESS integrated framework modelled 12 European-produced crops at EU-15 scale and average resolution of 16,310 km^2 (NUTS-2 level). The studies above are representative examples of the push to produce increasingly more spatially-detailed assessments.

Despite the noteworthy ways in which past studies have circumvented the scarcity of farm practice data (use of inputs, timing of use, crop rotations), this remains an issue (Ciais et al., 2010; Folberth et al., 2012; Ridoutt et al., 2014; Van Ittersum et al., 2013). For example, in Europe the Farm Accountancy Data Network (FADN) has been used to inform environmental research (Dolman et al., 2012; Samson et al., 2012; Thomassen et al., 2009;

* Corresponding author. CSIRO Land and Water, EcoSciences Precinct, 41 Boggo Road, Dutton Park, QLD 4102, Australia.
E-mail address: javi.navarro@csiro.au (J. Navarro).

http://dx.doi.org/10.1016/j.envsoft.2016.02.020
1364-8152/

Van der Meulen et al., 2014), but FADN does not contain data on farm practices or environmental impacts (Janssen et al., 2009; Zander et al., 2009). Spatially-detailed and quantitative farm practice data is needed to estimate environmental impacts of agriculture in a precise way (Borkowski et al., 2007; Bryan et al., 2014; Janssen et al., 2009; Ridoutt et al., 2014). However, collecting observed regional farm practice data at national to global scale with a high level of detail is very skill- and time-demanding; it requires the engagement of numerous regional experts with knowledge in a broad range of fields (Zander et al., 2009). The use of farm management handbooks (a.k.a. gross margin handbooks, farm enterprise budgets, or crop budgets) is helpful in reducing those skill and time demands to manageable levels (Janssen et al., 2009). These documents contain a rich set of farm management information and have been used to quantify agricultural management and farm economics (Bryan et al., 2009). However, in covering large areas, multiple documents spanning multiple regions and crops are required. These are often created by different agricultural agencies, using different templates and processes and formats, and extracting consistent information about agricultural practices remains a challenge.

Here a new method to assemble detailed farm management data from a large number of heterogeneous farm management handbooks (hereafter simply *handbooks*) is presented. The method uses query expansion with background knowledge (an information retrieval algorithm popularly used in web search engines) and Natural Language Processing (NLP) to process the handbooks. The goal of this study is to describe the method and to test it as part of a broader system that maps the cradle-to-gate impacts of agriculture at national scale and farm level resolution (1.1 km grid cells) for Australia. The impacts mapped are currently limited to greenhouse gas (GHG) emissions and energy use, but further work would allow us to estimate and map other impacts. The results are divided in two main sections. First the quality of the farm practice data produced is evaluated by comparing it to data available in other published sources at both national and local scales. Second illustrative results from the mapping system are presented. These depend on the farm practice data, so their comparison with other sources of data (where possible) serves as a further quality check of the farm practice data. The results presented are: Fertiliser use at national scale (million tonnes of N, P and K applied) and local scale (kg of product per hectare), pesticide use at national and local scale (kg active ingredient (a.i.) per hectare), water use for irrigation at national scale (total Gigalitres and Megalitres per hectare), fuel use per state (litres per hectare), greenhouse gas emissions and energy use per production process (million tonnes of CO_2-equivalent and PetaJoules respectively), average water use intensity (Megalitres per hectare) and total water use (Gigalitres) per agricultural land use type, and a map of pesticide use intensity (kg a.i. per hectare) across Australian croplands at 1.1 km^2 resolution.

2. Methods

2.1. Overview

Here a method to process one or more handbooks into estimates of farm practices and environmental impacts is described (at this stage limited to GHG emissions and energy use). The method consists of five main steps: 1) Creation of a dictionary which links terms appearing in handbooks to reference tables containing knowledge about technical aspects of crop production (e.g. characteristics of fertilisers, pesticides, farm machinery, etc.). This complementary information is referred to as "background knowledge". The dictionary was produced by applying NLP techniques over a comprehensive collection of farm management handbooks published in Australia between 1998 and 2010 (Fig. 1a). 2) A set of parsing modules were developed to navigate each farm management handbook and use the dictionary to express information in highly-structured and accessible form (Fig. 1b). 3) The processed handbook data was stored in a database together with life-cycle inventory data, spatial statistics on historical yields and livestock numbers, spatial statistics on expenditure in livestock enterprises, statistics of global fertiliser and pesticide trade, modelled regional estimates of livestock greenhouse gas emissions per head, and modelled regional estimates of herd structures (Fig. 1c, left to right). 4) Farm practices and environmental impacts were geo-referenced in a Geographic Information System using publicly available spatial layers of local government boundaries, growing regions, statistical regions, and agro-ecological regions (Williams et al., 2002) (Fig. 1d). Spatial gaps in the geo-referenced data were resolved using a rule-based approach which draws information from the most agronomically-similar region possible (which is in effect a linear extrapolation of existing handbook data guided by agronomic constraints). 5) Finally, the farm practice data was linked to a map of land use at 1.1 km^2 (Fig. 1e). These processes are described in detail below.

Additional modelling of livestock enterprises was required to achieve a complete picture of agricultural impacts at the national scale because there were few livestock farm management handbooks available and they did not contain data on inputs or machinery use. The focus was on beef, dairy and sheep enterprises as these are the most widespread livestock industries in Australia.

2.2. Study area

Our study area is the entire Australian crop and pasture land. Australian agricultural production covers over 400 million hectares of land. Crop and livestock production systems have adapted to the highly diverse growing conditions in tropical, subtropical, grassland, and temperate climate zones (Stern et al., 2000). Agriculture supplies 2.4% of national GDP (ABS, 2012a) and is the economic backbone for 2.5 million rural residents (World DataBank, 2015). The agriculture and forestry sector accounts for approximately 24% of total greenhouse gas (GHG) emissions and 3% of total energy use in Australia (AGEIS, 2013; ABS, 2012b). Livestock occupies 95% of total agricultural land and generates 46% of total revenue (Marinoni et al., 2012). Cereals, legumes and oilseeds (canola) occupy 92% of total crop area and generate 34% of total crop revenue (Marinoni et al., 2012). Fruit, vegetables, sugar and cotton occupy 4% of croplands and generate 45% of total crop revenue (Marinoni et al., 2012).

2.3. Datasets used

Farm management handbooks are agricultural extension documents created to help farmers plan the technical aspects of crop production and their expenses. They describe all on-farm activities up to harvesting, some post-harvest activities such as packing and transport, and provide an estimated yield based on the described practices. Table 1 shows part of a farm management handbook as an example.

It is debated whether the handbooks represent best practices or average practices. Anecdotal evidence argues for both possibilities, published sources acquainted with their production seem to indicate the handbooks reflect typical practices in a region and year (Brock et al., 2012; NSW DPI, 2007). Other studies support the idea that the handbooks provide realistic information (Ridoutt et al., 2014).

A total of 385 crop- and-region-specific handbooks covering the study area were assembled (a complete list of them can be found in