READING SOILS

Using easily obtainable soil information to assess soil functioning

Maricke M.W.J. van Leeuwen

Propositions

1. Visual soil evaluation should use quantitative rather than qualitative visual observations.
   (this thesis)

2. Visual soil evaluation should in particular be used to assess soil functions that are driven by soil physical properties.
   (this thesis)

3. Moving from lowland deltas to higher land will become an essential measure to deal with sea level rise.

4. Environmental impact assessment of a product should include effects of production on human health and life expectancy of labourers and neighbouring population.

5. Universities should provide arts and dance rooms on campus.

6. Premarital coaching should be subsidized.

Propositions belonging to the thesis, entitled:

READING SOILS
Using easily obtainable soil information
to assess soil functioning

Maricke M.W.J van Leeuwen
Wageningen, 2 July 2019

Reading soils – Using easily obtainable soil information to assess soil functioning

Maricke Maria Wilhelmina Johanna
van Leeuwen

Reading soils – Using easily obtainable soil information to assess soil functioning

Maricke Maria Wilhelmina Johanna
van Leeuwen

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 2 July 2019
at 1.30 p.m. in the Aula.

Maricke Maria Wilhelmina Johanna van Leeuwen

Reading soils – Using easily obtainable soil information to assess soil functioning
170 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)
With references, with summaries in English and Dutch

Table of contents

Dankwoord

Ik ben enorm blij dat het gelukt is om de PhD uit te voeren en succesvol af te ronden, en ik ben enorm dankbaar dat ik er niet alleen voor stond, maar dat er van alle kanten hulp kwam.

Allereerst wil ik Marthijn Sonneveld bedanken. Hij was de initiator van het project, maar is helaas in december 2013 overleden. Marthijn had een enorme passie voor bodems en landschappen. Ik ben dankbaar dat ik in ieder geval tijdens mijn studie van Marthijn heb les gehad, en nog een stukje heb kunnen samenwerken aan de BodemConditieScore. De andere begeleiders van mijn PhD wil ik zeker ook bedanken: Gerard (voor ongeveer de eerste zeven maanden mijn dagelijkse begeleider als vervanger voor Marthijn), Cathelijne, Jakob, Imke, Jos, Everhard, Frank en Simon. Jullie zijn een goed team geweest, met visie, geduld, en humor. Dankjewel voor jullie ondersteuning tijdens de PhD! I am also very thankful for the fantastic office mates I have had: Chantal, Femke, Jetse (you came in so often that you count as office mate), Luc, Kasia, Marcos, Rockie, Rafa, Cindy, Nienke, Simona, Alexandre, Aukjen and Yinxia. Thanks for the friendship and beautiful office moments, like showing interest by talking and listening, crying, laughing, seriously working, singing, or decorating each other's desk. I want to thank Gert for the opportunities I got to join field practicals in Drenthe. I learned more about soils and landscapes than I did in my own student time. I am also thankful that you stimulated me to lead and to teach the entire group of students. Mieke and Henny, thank you for always being stand-by, and for helping with practical or financial questions. You felt like the mothers of the group, what I really appreciated. Thank you for your great care! Also, I want to thank the other SGL colleagues: without you (and the abovementioned colleagues) there were for example no Sexy Soil Soldiers at the WeDay (at least during my PhD contract time ;-)), SGL barbecues, Cake Fridays, Lunch woks, Lunch walks or group outings. Thanks for your help with practical questions, and thanks for your presence in boring and gezellige moments. Very valuable moments were those where we were open and honest, and where vulnerable topics could be shared (e.g. for me, difficulty with scientific writing or causes of stress like fear of failure). Especially in an environment where I sometimes felt competition, I want to encourage you to stay open and honest to yourself and to others, and to listen to each other. For me, these moments helped me realizing that I was not the only one that was stressing out. Thanks for your support, and thanks for laughing, which helped to be not too serious at some

Chapter 1

Introduction

1.1 General introduction

To meet the food demand needs of the growing world population, crop production has to increase while soil quality of agricultural land should be maintained or improved and pollution from agricultural land should be minimized (United Nations, 2019). Soil properties affect the availability of water and nutrients for crop growth, they filter water by binding nutrients and solutes, they regulate the amount of greenhouse gas emissions and act as a habitat for all kinds of soil microbes (Karlen et al., 1997). When soil is disturbed, also its functionality can be disturbed. Soil quality assessment, therefore, is needed to evaluate the status of all different kind of soil functions, and to be able to give recommendations for land management. Soil quality and environmental sustainability can be assessed using a wide range of techniques that vary in complexity, availability, and ease of use. The aim of this thesis is to evaluate how easily obtainable soil information can be used to assess soil functioning, in order to improve the environmental performance of agricultural land. Fieldwork, laboratory analyses and environmental modelling were performed to study the associations between easily obtainable soil information (soil maps and visual soil evaluation) and quantified indicators for soil functioning. I particularly focussed on the use of soil maps and visual soil evaluation to assess soil functions, because soil maps and visual soil evaluations are likely available to, or obtainable by, many users around the world. This thesis thereby contributes to the assessment of soil quality in agricultural land, fostering environmentally sustainable crop production.

1.2 Multifunctional soils and their visual assessment

When assessing soil quality in agricultural land, it is important to understand that soils are not only used to sustain crop growth, but soils also have other functions such as cycling of water and nutrients, serving as a carbon pool and preservation of biodiversity (CEC, 2006). Soil quality assessments, therefore, ideally focus on multiple soil functions, instead of focussing on one soil function and thereby missing possible synergies or trade-offs with other soil functions (Schröder et al., 2016a). This multi-functionality of soils is reflected in the definition of soil quality by Karlen et al. (1997): 'the capacity of a specific kind of soil to function, within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality, and support human health and habitation'. According to this definition, the 'capacity of a specific kind of soil to function' implies that the current level of soil functioning has to be compared with the desired value (or a reference level) of soil functioning, in order to be able to evaluate soil functioning. The desired value of a specific soil function could be a situation without any limiting factors for that soil function, and therefore the soil function is at its maximum value. In contrast, limiting factors could create a gap between the desired value and the actual value of soil quality and soil functioning. For example, poor management could result in nutrient and soil

organic matter depletion, which could negatively affect crop growth and the capacity of the soil to hold water and nutrients. When soil quality of different soil types is evaluated in relation to the desired value for each specific soil, soil functioning of different soils can be compared with each other, which gives insight into the soils that have relatively most room for improvement ('window of opportunity', Bouma, 1994).

This thesis focusses on the soil functions crop production, and storing, filtering, and transforming nutrients and water (Table 1.1), on Dutch dairy farms. I assessed the crop production function under wet and dry conditions. Relevant indicators are crop yield gap under wet and dry conditions, crop oxygen and drought stress, and plant available water. Selected indicators for the soil function 'storing, filtering, and transforming nutrients and water' are water storage capacity, nitrate and phosphate concentrations in drain water and groundwater, and soil nutrient balances (Table 1.1).

1.3 Easily obtainable soil information
This thesis assesses how easily obtainable soil information can be used to assess soil functioning. Examples of easily obtainable soil information are the use of soil series that are available through soil maps by indicated soil mapping units, and the use of visual soil evaluation to assess soil quality.

Table 1.1. Selected soil functions (white boxes) and their selected indicators in this thesis.

Soil functions (CEC, 2006)	Soil function indicator	Easily obtainable soil information	Chapter
Crop production	Crop yield gap under wet and dry conditions	VSE[1]	4
Storing, filtering, and transforming nutrients and water	Crop sensitivity to dry and wet periods: oxygen and drought stress	VSE[1]	4
	Plant available water	VSE[1]	4
Maintaining biodiversity	Water storage capacity	VSE[1]	4
Serving as carbon pool	Nitrate and phosphate concentrations in drain water and groundwater	Soil series	2
Serving as archive for geological and cultural heritage		VSE[1]	4
	Soil nutrient balance	Soil series	2

[1] Visual soil evaluation (VSE).

13

This thesis focusses on easily obtainable soil information, because only basic soil knowledge or training is required to use or obtain this soil information. It is expected that easily obtainable soil information is available to many users, such as farmers or scientists, in many countries worldwide. The use of easily obtainable soil information to eventually increase soil quality and the sustainability of agricultural land management is thus available to many people, which increases the importance to test how this information can be used to assess soil functions.

1.3.1   *The use of soil series to improve nutrient balance calculations*

Currently, most nutrient balances that are used in practice are at farm level or at a coarser scale. For policy makers this level of detail is sufficient to have an indicator of overall farm sustainability, and to evaluate the environmental pressure of agriculture in a region (Van Beek et al., 2003). Farmers and land managers, however, also make land management decisions at field level, and therefore farm-level sustainability indicators are not informative for decision-making at field level (Van Beek et al., 2003). Nutrient balances at field level, instead, could help farmers in making correct decisions with regards to environmental sustainability.

Soil maps provide information about the degree of soil variation within a farm, and therefore soil maps may be used to decide whether nutrient balances should be performed at farm level (in case of low soil variation) or at field level (in case of high soil variation). Contrasting soil series can have contrasting soil properties (Batjes, 2016) that can affect water and nutrient availability to plants. As such, soil series indicated on a soil map could be used to estimate nutrient balances at field level. Nutrient balances, in turn, do not only reflect the amount of external nutrient inputs, but also the soils' capacity to provide water and nutrients to crops, and the ability to buffer water and nutrients and to prevent leaching (Bindraban et al., 2000). My hypothesis is that if soil variation within a farm is low, field level nutrient balances are similar across fields due to similar soils, and therefore a nutrient balance at crop or farm level is sufficient. In contrast, when a soil map indicates high soil variation within a farm, it is expected that nutrient balances across fields vary, and therefore the need to have nutrient balances at field level increases. In that case, nutrient balance at the crop or farm level is then no longer a good representation for nutrient balances at field level.

1.3.2   *Visual soil evaluation to assess soil quality*

Visual soil evaluation (VSE) assesses soil quality based on a number of readily observable soil properties, which can be assessed using an extracted soil block or using a soil profile (e.g. Peerlkamp, 1959; Roger-estrade et al., 2016; Shepherd, 2009). The various VSEs that exist have in common that the assessment is rapid and cost-effective, assessment uses visual observations; VSEs were originally developed to

evaluate soil properties in relation to crop growth (e.g. McKenzie, 1998; Peerlkamp, 1959; Shepherd, 2009); the intended users are farmers or land managers, scientists, government agencies and agricultural consultants (Bünemann et al., 2018); and the method descriptions include reference pictures of soil and/or detailed instructions, which make the methods easy to use. The main differences between the various VSEs are the used set of visual observations (Bünemann et al., 2018) and the depth of assessment (topsoil, subsoil, or entire profile). For more details about the main developed VSEs, please refer to Emmet-Booth et al. (2016) and Bünemann et al. (2018). In this thesis a broad set of visual observations is used, which is based on Visual Soil Assessment (Shepherd, 2009; Sonneveld et al., 2014).

There are three key concerns with regards to VSE. First, it is unclear whether the background and the use of qualitative or semi-quantitative visual observations helps in evaluating soil functions. Currently, scores are assigned to evaluate a soil observation, with classes referring to a good, moderate or poor soil condition. Scores are often based on expert knowledge and they are region-specific (e.g. McKenzie, 2013; Shepherd, 2009; Sonneveld et al., 2014). If soil quality scores are based on qualitative assessment (e.g. comparing the abundance of roots with reference pictures), there is no indication whether the distance between the soil quality classes is linear or nonlinear, and whether the correlation with soil functions is linear or nonlinear. Also, the use of scores may impair quantitative analyses and the development of soil function evaluations because many statistical tests require numerical data. An alternative would be to make observations as quantitative as possible, and if it is desired, to assign a soil quality score only in the final evaluation of soil functioning. Quantitative visual observations are more useful than soil quality scores for the development of soil function evaluations. Several studies proved that visual observations corresponded with standard field or laboratory measured soil properties (e.g. Ball et al., 2007; Johannes et al., 2017; McKenzie, 2001; Newell-Price et al., 2013; Sonneveld et al., 2014). For the broad range of quantitative visual soil observations, however, the correspondence with standard field or laboratory measurements is unknown and needs to be investigated.

Second, the subjectivity of VSE is questioned (Guimarães et al., 2011). VSEs consist of several steps, starting with soil excavation and soil preparation, followed by visual observations and interpretation. Each step could be influenced by the observers' interpretation, though the manuals with reference pictures and descriptions are as clear as possible to reduce subjectivity (Guimarães et al., 2017). To my knowledge, only Guimarães et al. (2011) studied the reproducibility of the soil block excavation, whereas Ball et al. (2007) studied the reproducibility of the visual evaluation of soil structure. Both studies found that the visual evaluation of soil structure was reproducible. For other visual observations that can be used in VSEs (e.g. grass cover, number of

earthworms, soil compaction) this reproducibility has not previously been investigated.

Third, soil texture is not always included in the evaluation of soil quality (Bünemann et al., 2018), although texture is associated with visual soil properties (Ball et al., 2017), with soil hydraulic characteristics (Wösten et al., 2001) and with soil functions (Guimarães et al., 2017; Peerlkamp, 1959). Based on expert knowledge, Shepherd (2009) assigns the highest soil quality scores in relation to grass productivity to loamy soils and the lowest scores to sandy and clayey soils. Newell-Price et al. (2013) found that higher sand contents increased the overall Visual Soil Assessment score. Johannes et al. (2017), in contrast, found that higher clay contents (higher than 20% clay) increased visual soil structure scores, applied on excavated soil cores. For the same soil cores, the relation between sand contents and visual soil structure scores showed an optimum with the best scores (moderate scores) for sand contents higher than 55%. As soil texture influences soil hydrological properties (Wösten et al., 2001), it can be expected that the correlation between visual observations and soil functions is also influenced by soil texture, and therefore it is important to include soil texture in the VSE manuals when evaluating soil functions. Until now, VSEs did not include measured texture effects in the evaluation of visual soil observations in relation to soil functions.

Despite the abovementioned concerns about VSE, it is recognized that VSE shows potential in the evaluation to assess various soil functions in order to assess the environmental performance of agricultural crop production (Ball et al., 2017; Shepherd, 2009). Only a few studies, however, investigated whether visual soil observations are a good indicator for several soil functions (da Silva et al., 2014; Giarola et al., 2013; Mueller et al., 2013, 2009). Most studies that used VSE to assess soil functioning focussed on crop growth (e.g. Ball et al., 2015; Batey and McKenzie, 2006; McKenzie, 2013; Shepherd, 2009). As far as I know, VSEs that evaluate visual properties in relation to crop production, nutrient loss, greenhouse gas emissions and carbon sequestration are based on expert knowledge (Shepherd, 2009). Improved understanding of the quantitative relationships between visual observations and soil functions (including texture effects) can improve VSEs by adding those relationships in the evaluation. This may help farmers and land managers to assess the various aspects of soil quality, and to improve the environmental performance of agricultural land.

1.4 Knowledge gaps

As outlined above, easily obtainable soil information (soil series and visual soil evaluation) can be used to assess soil functions and the environmental performance of agricultural land. Yet, considerable knowledge gaps remain, and these are the focus of this thesis:

1. It is unknown whether nutrient balances at farm level represent nutrient balances at field level, given the within-farm diversity of soil series.
2. Visual soil evaluations (VSEs) make use of scores to evaluate a soil property. The use of quantitative instead of qualitative visual observations in VSE, enables a quantitative assessment of the relation between visual observations and soil functions. For a broad set of quantitative visual observations, however, the reproducibility by different observers and the correlation with standard field or laboratory measurements is unknown.
3. Until now VSE methods are only developed to evaluate visual soil properties in relation to the soil function crop production. The correlation between VSE and other soil functions is not yet studied in a quantitative way. I expect that texture affects the correlation between visual soil properties and soil functions, but this is not yet quantified.

1.5 Thesis aim and research questions

The knowledge gaps are summarized in the main aim of this thesis: to evaluate how easily obtainable soil information can be used to assess soil functioning.

Each knowledge gap is translated into one research question (Figure 1.1):

1. To what extent does the required spatial scale in nutrient balances depend on the level of soil variation? (Chapter 2)
2. To what extent are quantitative visual soil observations reproducible, and do they correlate with standard field or laboratory measurements? (Chapter 3)
3. Can quantitative visual soil observations be used to assess soil functioning? (Chapter 4)

The overall aim will be addressed in the synthesis (Chapter 5).
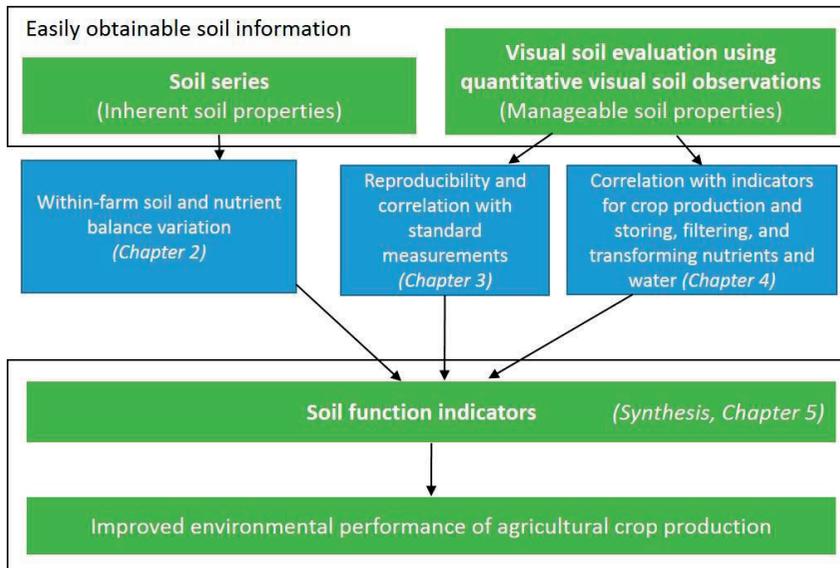
Figure 1.1. Thesis outline. The aim of this thesis is to evaluate how easily obtainable soil information can be used to assess soil functioning. This research contributes to understanding of soil quality and the environmental performance of agricultural crop production. The blue boxes represent the chapters that will, or have been, published in peer-reviewed scientific journals. The green boxes represent the main themes studied.

1.6 Thesis outline

Each knowledge gap and research question is discussed in one chapter (Figure 1.1).

Chapter 2 addresses the role of spatial scales in nutrient balances on dairy farms. To assess whether nutrient balances at farm level represent nutrient balances at field level given the within-farm soil variation, field-level nutrient balances are associated with soil series. Also, this chapter assesses the effect of the used spatial scale of soil-dependent leaching factors on farm-level nitrate leaching.

Chapter 3 and 4 address whether visual soil evaluation (VSE) can be used to assess various soil functions. Chapter 3 evaluates the VSE 'building blocks' by focussing on the quantitative visual observations. First, the reproducibility of the visual observations is evaluated, when assessed by a group of farmers and soil scientists on various soil types. Second, the correlation between quantitative visual observations and standard field or laboratory measurements is assessed. Chapter 4 subsequently discusses whether a set of quantitative visual observations can be used to assess various soil functions. Also, the effect of soil texture on the relation between quantitative visual observations and soil functions is assessed.

Finally, Chapter 5 is the synthesis where I discuss what knowledge is created in this thesis, and how this knowledge can be used in today's sustainable management of agricultural land and in ongoing soil scientific research.

1.7 Study area

This study is carried out on dairy farms located on sandy, peaty and clayey soils in the Netherlands (Figure 1.2). The dairy sector covers approximately half of the total agricultural land in the Netherlands (CBS, 2017; WEnR, 2016).

For Chapter 2 and 4, five dairy farms were selected that participated in the ongoing project 'Cows and Opportunities' (in Dutch: 'Koeien & Kansen'). This project monitors nutrient inputs and outputs at field, crop and farm level (Oenema et al., 2011), which enabled us to calculate nutrient balances at various spatial scales. These five farms were located on sandy and clayey soils, and selected based on the number of different soil series within the farm. The range of soil textures enabled us to assess the effect of soil texture on the relation between quantitative visual observations and soil functions.

For Chapter 3, 26 Dutch dairy farmed sites were selected in the North Friesian Woodlands. Within the North Friesian Woodlands, sandy, peaty and clayey soils can be found, which enabled us to include soil type in the analyses of the reproducibility and validation of quantitative visual observations.

Climate in the Netherlands is temperate. Between the years 2001 and 2016 meteorological station 'De Bilt' recorded mean monthly temperatures between 3.7 °C and 18.2 °C. The mean annual precipitation was 862 mm (Royal Netherlands Meteorological Institute, 2018). Further details about the study areas can be found in Chapter 2, 3 and 4.

Clay soils (marine)

Peaty soils
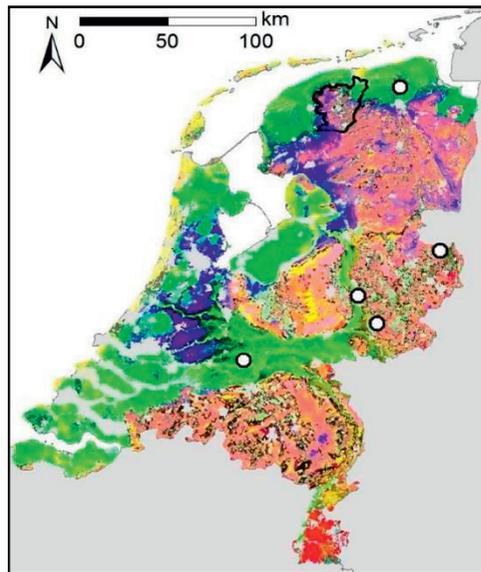
Podzol soils

Clay soils (fluvial)

'Enk' earth soils

Figure 1.2. Location of the study areas in the Netherlands, and an impression of the diverse dairy farmed landscapes with typical soil profiles. Colours in the map of the Netherlands indicate the soil series. In this map the irregular hollow shape in the North of the Netherlands represents the North Friesian Woodlands area, which was used in Chapter 3. The white circles represent the locations of the five dairy farms used in Chapter 2 and 4.

Chapter 2

The relevance of spatial scales in nutrient balances on dairy farms
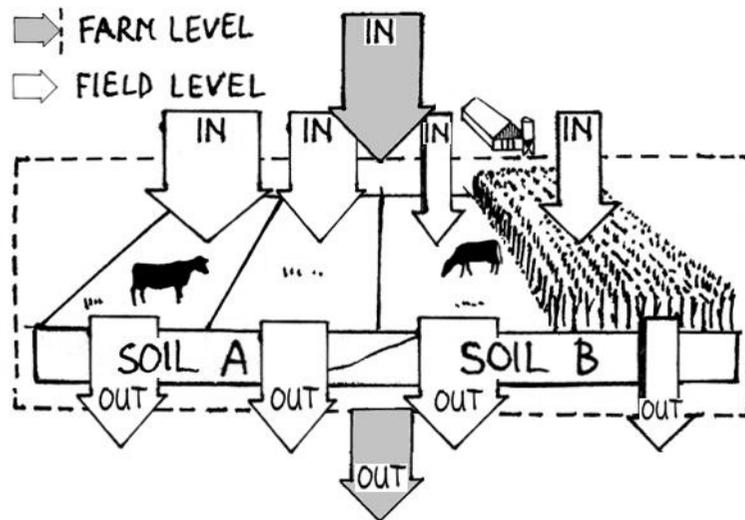
Highlights

- Large variations in N and P balances were found between fields at five dairy farms.
- The variation was likely caused by management factors and soil types.
- This calls for the use of field-level rather than farm-level nutrient balances.

Abstract

Policy makers and farmers use tools, such as a nutrient balance, to gain insight into the environmental impact of agricultural practices. A discrepancy, however, exists between the needs of policy makers and farmers, about the use and the spatial scale of such tools. Farm balances calculate nutrient balances across all agricultural fields within a farm without distinguishing separate fields, whereas field balances calculate a nutrient balance on a delineated field. For farmers, a nutrient balance at field level is more useful than at crop or farm level, because decision making and fine-tuning management occurs at the field level. A field balance, however, requires more detailed data than a farm balance and therefore is less easy to implement. As soil types influence nutrient balances, we hypothesize that if within-farm variation in soil types is low, there is no need to replace a farm balance by a field balance. To test this hypothesis, we computed nutrient balances at farm and field level on five Dutch dairy farms (three on sand, two on clay), varying in degree of within-farm variation in soil series. A full year of soil nitrogen (N) and phosphorus (P) input and output data on farm and field level were provided by farmers, while soil variation was determined using the Dutch 1:50,000 soil map. The Annual farm Nutrient Cycle Assessment (ANCA) was used to calculate soil N and P surpluses, and soil nutrient fluxes such as nitrate leaching and nitrous oxide emission at farm and field level. Even on farms with few soil series, a considerable variation in N and P inputs, outputs and balances across fields was found, due to management differences and soil properties not represented by the soil map. Furthermore, field-level balances better represented nitrogen leaching than farm-level balances on farms with diverse soils (reflected by different leaching factors) and negative nitrogen field balances (deficits). Also, using field balances, for one case study farm the highest soil N surplus (kg ha^{-1}) was found on grass fields with the highest risk of N leaching. A field balance, therefore, provides more meaningful information than a farm balance when variation in soil types and/or management factors is found within the farm, because soil types and management factors affect N and P balances, N leaching and N emissions. For farms with the highest variation in soil types and/or management, we recommend using field-level nutrient balances in order to detect extreme surpluses, deficits, leaching and/or emissions, to improve management decisions.

2.1 Introduction

The sustainable development goals (United Nations, 2017) address the dual-challenge to produce enough food to feed a growing and more prosperous population, and to produce this food in an environmentally friendly way. The current food production system, however, has a major impact on the environment. Livestock production in Europe, for example, is responsible for about 80% of soil acidification and air pollution (via emission of mainly ammonia and nitrogen oxides), and for 73% of the water pollution (via leaching of nitrate or phosphate) (Leip et al., 2015). To reduce emissions to soil, air and water, the European Union introduced the National Emission Ceilings Directive to reduce air pollutants, and the European Nitrates Directive to reduce ground- and surface water nitrogen (N) and phosphorus (P) concentrations (European Commission, 1991; European Environment Agency, 2017). To reach the targets, policy makers and other actors need tools to monitor the environmental impact of agricultural practices at farm level.

Nutrient balances can be used as a policy tool (Sassenrath et al., 2013). For farmers, who manage the land field by field, nutrient balances at field level are more useful than at crop level (all fields within a farm with the same crop) or farm level (the whole farm: land, housing and animals), because decision making and fine-tuning management occurs at the field level (Van Beek et al., 2003). A nutrient balance at field level, however, would require more data than a balance at crop level or farm level and, therefore, is less easy to implement (Öborn et al., 2003). Defining a nutrient balance at crop or farm level assumes that nutrient balances and associated losses are equal across fields. For example, the Overseer model (Thomas et al., 2005) used leaching and emission factors specific for farming systems, but on the national level, and the Annual farm Nutrient Cycle Assessment (Aarts et al., 2015) used field-specific leaching and emission factors that were aggregated to the crop and farm level. Nutrient balances vary between fields due to differences in soil characteristics (soil compaction or the depth and the soil organic matter content of the A horizon), hydrological conditions, grazing regimes, fertilizer applications and crop yields (Lipiec and Stępniewski, 1995; Oenema et al., 2010; Van Es et al., 2006). A discrepancy, therefore, exists between the needs of policy makers and farmers, about the use and the spatial scale of the tools that are used to quantify the environmental impact of agricultural practices.

Furthermore, the effect of soil characteristics (such as soil compaction, depth of the A horizon, and groundwater tables) on the nutrient balance is often excluded despite the fact that soil processes, such as denitrification and the build-up and decline of soil organic matter, are included in nutrient balances (e.g., Van Beek et al. (2003) and Watson and Atkinson (1999)). Van Beek et al. (2009) assessed soil nutrient balances

on three dairy farms located on well-drained sand, and poorly-drained clay and peat in the Netherlands. They found that average denitrification rates were highest for peat and clay and lowest for sand. Average N leaching was highest for sand and lowest for clay. P surplus, however, was highest for peat and lowest for sand and clay. The presence of soil organic carbon and water content stimulate denitrification (Van Beek et al., 2003). Soil texture, groundwater depth and precipitation regulate N leaching, and texture, groundwater depth, the presence of iron and aluminum (hydr)oxides and the P content of the soil regulate the soil P adsorption capacity and P leaching (Freese et al., 1992; Oenema et al., 2004; Schoumans et al., 2013). Even within the soil type classes 'sand', 'clay' and 'peat', soils vary considerably (e.g., in texture, soil organic matter and groundwater depth), thereby affecting local nutrient balances. This illustrates that more detailed descriptions of soils are important to consider in nutrient balances, rather than using these broad soil type classes.

We hypothesize that if within-farm variation in soil types (combination of texture, groundwater table, depth of the A horizon) is low (Figure 2.1A), there is no need to replace a farm balance by a field balance. Unless manure application rates vary highly across fields within a crop type on a specific soil series, we expect similar biomass yields, N and P balances, N leaching and N emissions on those fields. Yet if within-farm soil variation is high (Figure 2.1B), a nutrient balance at crop or farm level may not be a good representation of the various nutrient balances at field level. The objective of this paper is to test the above hypothesis and to assess whether it is relevant to calculate a nutrient balance at field rather than at crop or farm level, given any soil variation within a farm. Since January 2017, dairy farmers in the Netherlands are required to use Annual farm Nutrient Cycle Assessment (ANCA, in Dutch: KringloopWijzer), in an attempt to improve nutrient use efficiency at their farm and
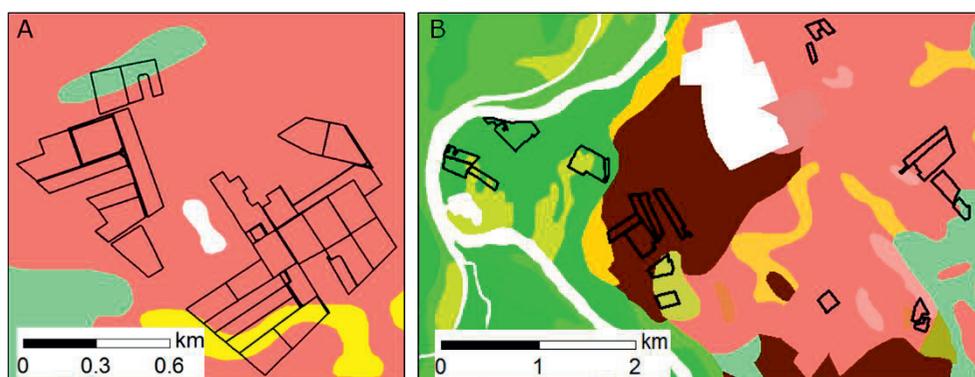
Figure 2.1. Example of two farms on a 1:50,000 soil map, each colour depicts a soil series. Black lines delineate fields belonging to one farm. A) variation in soil series is low: 'homogeneous'; B) variation in soil series is high: 'heterogeneous' (See Section 2.2.1 for more information of about the farm and soil characteristics).

to reduce nutrient losses to the environment (Aarts et al., 2015). We therefore used ANCA to calculate N and P balances at field, crop and farm level, and estimate N leaching and N emissions, using five case study dairy farms.

2.2 Materials and methods

2.2.1 Farm and soil characteristics

We selected five dairy farms that take part in the ongoing project 'Cows and Opportunities' (in Dutch: 'Koeien & Kansen'). This project monitors soil nutrient input and output data at field level, which is used in the present study to calculate nutrient balances at field, crop and farm level (Oenema et al., 2001). We used nutrient balances at the field level for 2014.

The selected five dairy farms in the Netherlands were experimental Farm SH0 (Farm 'De Marke', sand, homogeneous), and four commercial pilot farms which will be referred to as Farm SHe1 (sand, heterogeneous 1), Farm SHe2 (sand, heterogeneous 2), Farm CHe (clay, heterogeneous) and Farm CH0 (clay, homogeneous; Figure 2.2). Selection criteria included soil texture (either mainly on sand or clay) and within-farm spatial variation in soil series (the number of soil series on the Dutch 1:50,000 soil map). Farm SH0 and CH0 had the lowest number of soil series (four each) and were classified as homogeneous (Figure 2.2, Table 2.1). For both farms, 80% of the farmland was located on two main soil series. Farm SHe1, She2 and CHe were classified as heterogeneous farms in terms of soil series, as Farm She1 and CHe had eight soil series, and Farm SHe2 had ten soil series (Figure 2.2, Table 2.1). Land of farm SHe2 was most equally distributed across the various soil series; each soil series had a surface area of 1-17% of the land. About 43% of the land of Farm SHe1 was located on two main soil series, and only 2-16% on the remaining six soil series. About 63% of the farmland of Farm CHe was located on the two main soil series, and between 4-14% on the remaining six soil series.

Farm SH0 was located on aeolian cover sands. The dominant soil series of the four soil series present within this farm was the '*veld*' podzol soil (ordinary hydropodzol, Table 2.1), which covered 46.4 ha of the total 54.5 ha (Figure 2.2, Table 2.1). About 80% of the farmland was in grass-maize rotation in 2014. As Farm SH0 is an experimental farm (Knowledge Transfer Centre De Marke, 2017), most data were based on measurements and not on estimates, whereas for the other farms most data were based on estimates and not on measurements (Oenema et al., 2015). Also Farm SHe1 and SHe2 were located on aeolian cover sands (Figure 2.2, Table 2.1).
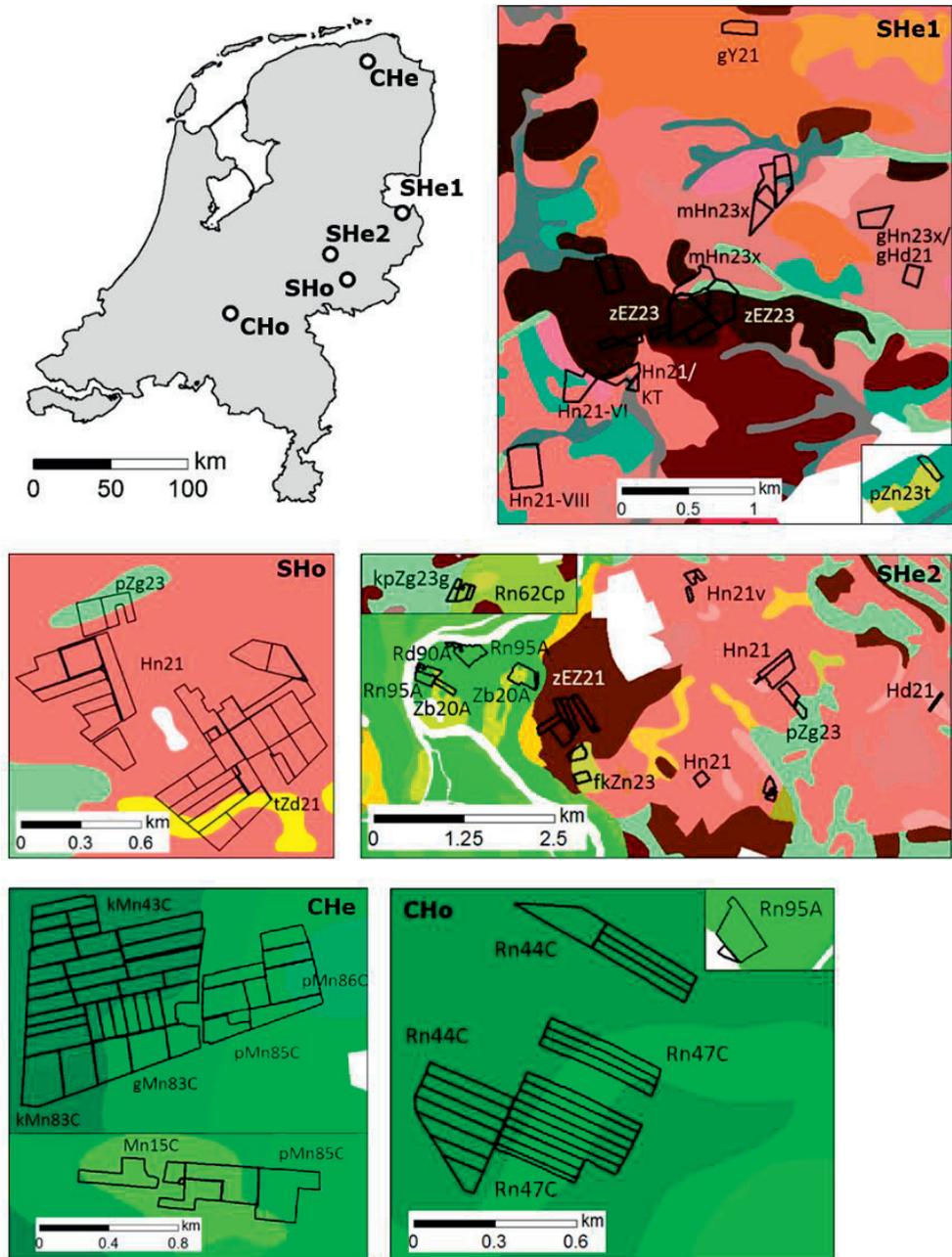
Figure 2.2. Location and soil series of Farms SHo (sand, homogeneous), Farm SHe1 (sand, heterogeneous), Farm SHe2 (sand, heterogeneous), Farm CHe (clay, heterogeneous) and Farm CHo (clay, homogeneous), in the Netherlands. Smaller inserted boxes present the soil series for fields far away from the farm household and are at the same scale as the rest of the farm. Please refer to Table 2.1 for soil classification and codes.

Farm SHe1, however, had glacial till in the subsoil: ~30% of the farmland had glacial till within 1.2 m depth, for the remainder of the farmland the glacial till was located deeper. The main soil series were the *'veld'* podzol soils and manmade black *'enk'* earth soils (Table 2.1.). The black *'enk'* earth soils have a ploughed Aap-horizon of at least 50 cm thick. Farm SHe2 had besides *'veld'* podzol soils and manmade black *'enk'* earth soils also clayey xero- and hydrovague soils in riverine sediments. The fields that were located in the river floodplains (27 ha) were extensively managed; no manure was applied. Farm CHe was located on marine clay (Figure 2.2). The main soil series were the *'polder'* vague soils (Table 2.1). Farm management was organic since the year 2009. Farm CHo was located on river clays (Figure 2.2). Similar to Farm CHe, the main soil series were the *'polder'* vague soils (Table 2.1). More information about soil characteristics for these typical Dutch soils is provided by De Bakker and Schelling (1989).

Farm SHe1 was the least intensively producing farm, with a milk production of 11450 kg ha^{-1} in 2015, and about 85 cows. Farm CHo was the most intensively producing farm, with a milk production of 23500 kg ha^{-1} in 2013, and about 110 cows (Table 2.2, Koeien & Kansen, 2017).

Climate in the Netherlands is temperate. Average daily temperatures ranged from 3.2°C in winter to 17.0°C in summer, and mean annual precipitation was 809 mm in the period 2004-2014. The year 2014 was a warmer, dryer and sunnier year than normal, with monthly temperatures ranging between 5.1°C in winter, and 17.1°C in summer, and with 771 mm precipitation (Royal Netherlands Meteorological Institute, 2017).

2.2.2 Nutrient balance concepts

Nutrient balances, also referred to as 'element balances' or 'nutrient budgets', are used to evaluate nutrient inputs and outputs over defined system boundaries and over a defined time period (Cherry et al., 2008). Depending on the definition used, nutrient balances can vary in system boundaries and consequently in the accounted inputs, outputs and balances. In general, there are two categories of nutrient balances: full balances and partial balances. The first category, 'full balance', include *all* nutrient inputs and outputs of the system. The second category, 'partial balance', do *not* include *all* inputs and outputs of the system. For example, Van den Bosch et al. (1998) only included the 'easy to quantify' inputs and outputs (e.g., fertilizers, animals, feed, meat, milk, manure), leaving out the 'difficult to quantify' flows (inputs through atmospheric deposition, fixation, and sedimentation; and outputs through gaseous losses, leaching, runoff and erosion). In this case, the partial balance refers to those nutrient flows that are directly managed or that have some economic value. Soil surface balances, as introduced by Oenema et al. (2003), are also partial balances.

Table 2.1. Farm and soil characteristics from the dominant soil series occurring within a field. International soil classification from De Bakker and Schelling (1989). [1] Groundwater table classes retrieved from Alterra (2005). Note that in the texture class, 'L' means 'loamy'.

Farm (total cropland area)	Crop type	Number of fields	Total area (ha)	Soil map units	Soil group	Soil classification	Ground-water table class	Texture class	Leaching factor
SHo (54.4 ha)	Grass	14	26.6	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VII	Sand	0.37
		2	3.8	pZg23	Sandy hydroearth soils	'Beek' earth soils	VII	Sand	0.37
		2	4.2	tZd21	Sandy xeroearth soils	'Kant' earths soils	VII	Sand	0.37
	Arable	9	19.8	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VII	Sand	0.75
SHe1 (55.6 ha)	Grass	1	4.1	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VI	Sand	0.29
		1	7.2	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VIII	Sand	0.37
		2	2.7	Hn21/ KT	Ordinary hydropodzols/ Tertiary clays	'Veld' podzol soils	VI	Sand	0.29
		5	11.8	mHn23x	Ordinary hydropodzols	'Veld' podzol soils	VI	Sand	0.29
		1	1.4	pZn23t	Sandy hydroearth soils	'Goor' earth soils	V	Sand	0.19
		2	12.4	zEZ23	Thick 'enk' earth soils	Black 'enk' earth soils	VIII	Sand	0.37
	Arable	2	5.1	gHn23x/ gHd21	Ordinary hydropodzols/ xeropodzol soils	'Veld' podzol soils/ 'Haar' podzol soils	VII	Sand	0.75
		1	2.4	gY21	Moder podzol soils	'Holt' podzol soils	VIII	Sand	0.75
		5	8.7	zEZ23	Thick 'enk' earth soils	Black 'enk' earth soils	VIII	Sand	0.75
SHe2 (90.5 ha)	Grass	2	6.2	fkZn23	Sandy hydrovague soils	'Vlak' vague soils	VII	Sand	0.37
		2	11.7	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VIII	Sand	0.37
		1	1.0	kpZg23g	Sandy hydroearth soils	'Beek' earth soils	II*	Sand	0.19

Group	Use	N	Area	Code	Soil description	Soil type	Class	Texture	Value
		1	3.8	Rd90A	Clayey xerovague soils	'Ooi' vague soils	VIII [1]	Clay	0.11
		3	3.6	Rn62Cp	Clayey hydrovague soils	'Polder' vague soils	IV	Clay (L)	0.11
		2	12.1	Rn95A	Clayey hydrovague soils	'Polder' vague soils	VIII [1]	Clay (L)	0.11
		2	11.4	Zb20A	Sandy xero vague soils	'Vorst' vague soils	VIII [1]	Sand	0.37
		3	15.1	zEZ21	Thick 'enk' earth soils	Black 'enk' earth soils	VIII	Sand	0.37
	Arable	1	0.9	Hd21	Xeropodzol soils	'Haar' podzol soils	VII	Sand	0.75
		6	15.4	Hn21	Ordinary hydropodzols	'Veld' podzol soils	VIII	Sand	0.75
		1	2.1	pZg23	Sandy hydroearth soils	'Beek' earth soils	VI	Sand	0.59
		2	7.3	zEZ21	Thick 'enk' earth soils	Black 'enk' earth soils	VIII	Sand	0.75
CHe (140.6 ha)	Grass	12	32.0	gMn83C	Clayey hydrovague soils	'Polder' vague soils	VI	Clay	0.11
		22	55.9	kMn43C	Clayey hydrovague soils	'Polder' vague soils	VI	Clay	0.11
		2	12.0	kMn63C	Clayey hydrovague soils	'Polder' vague soils	III	Clay (L)	0.11
		4	15.0	Mn15C	Clayey hydrovague soils	'Polder' vague soils	VI	Clay (L)	0.11
		5	19.8	pMn85C	Clayey hydroearth soils	'Woud'/ 'Leek' earth soils	VI	Clay	0.11
		1	6.0	pMn86C	Clayey hydroearth soils	'Woud'/ 'Leek' earth soils	VI	Clay	0.11
CHo (41.6 ha)	Grass	12	17.7	Rn44C	Clayey hydrovague soils	'Polder' vague soils	IV	Clay	0.11
		11	16.1	Rn47C	Clayey hydrovague soils	'Polder' vague soils	IV	Clay	0.11
	Arable	1	2.8	Rn95A	Clayey hydrovague soils	'Polder' vague soils	VII	Clay (L)	0.34
		2	5.1	Un-known	Unknown	Unknown	Unknown	Clay	0.34

Table 2.2. General farm and soil quality characteristics. N_total: total N in soil. P_AL: plant available P in soil. N_total and P_AL were only reported for grassland soils, and were sampled from 0-10 cm. For Farm SH0 samples were taken from 0-20 cm depth.

Farm	Total cropland area (ha)	Number of cows	Number of young stock	N_total (mg N in 100 g soil)		P_AL (mg P_2O_5 in 100 g soil)	
				Average	Standard deviation	Average	Standard deviation
SH0	54.4	86	58	1471	256	64	21
SHe1	55.6	84	74	2263	228	50	15
SHe2	90.5	160	108	3080	1015	68	19
CHe	140.6	182	138	4659	1918	83	59
CH0	41.6	109	73	5259	737	20	11

These balances calculate the difference between inputs (nutrients that enter the soil) and outputs (crop uptake or harvested and grazed crops) at the field surface. At the farm level, the literature often refers to 'farm-gate balances' or 'input-output accounting systems' for the partial balances, and 'whole-farm balances' for the full balances. Internal flows within the system are not defined in a partial balance, therefore partial balances can be considered as a black box (Cherry et al., 2008; Öborn et al., 2003). The difference between outputs and inputs for the full balance is automatically a change in the nutrient stock. The partial balance is a proxy for the sustainability of the system, but more detailed analysis is needed to know the fate of the nutrient losses or surpluses.

In the present study, we deal with partial balances. Nutrient flows from the soil, such as nitrous oxide (N_2O-N) emission and nitrate (NO_3^--N) leaching were quantified, but nitrogen gas (N_2) or nitrogen oxides (NO_x-N) emissions were not. Therefore estimated surpluses included the nutrient losses and processes that were not accounted for (Cherry et al., 2008).

2.2.3 Nutrient balance at crop and farm level

ANCA determines the nutrient flow through a dairy farm and through its four major components: livestock, manure, soil and crops (Figure 2.3, Oenema et al., 2011; Aarts et al., 2015). For each component a separate balance is calculated. This study focusses on the soil component within ANCA (Figure 2.3). ANCA calculates nutrient balances at the crop level, and subsequently aggregates the balances to farm level (Schröder et al., 2016b).Concerning the soil component, 'farm level' in the present study means all agricultural fields within a farm, excluding housing, animals, etc. N and P inputs of the soil component at crop and farm level included fertilizer (chemical fertilizer and within-farm produced organic manure; for the N balance corrected for NH_3-N emission), excreta during grazing (for the N balance corrected for NH_3-N emission), N deposition and N fixation. N and P outputs included harvested crops, grazed and mowed grass. Additional outputs for the N balance were NO_3^--N leaching and N_2O-N emission. Also, to account for N balance when grass and maize rotated in the past three years, mineralization (input) and accumulation (output) of soil organic matter
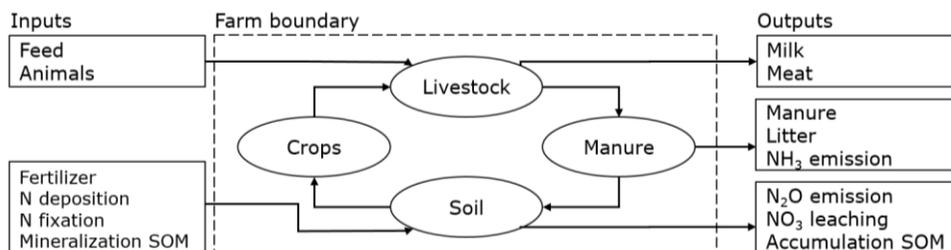
Figure 2.3. ANCA determines N and P flows for a dairy farm, through the components livestock, manure, crops and soil (adapted from Gourley et al., 2007).

were included. ANCA also estimates NH_3-N emission as a fraction of manure and fertilizer input, but it was excluded in the present study because NH_3-N emission is not affected by soil type (Schröder et al., 2016b). For the P balance ANCA calculates a surplus or deficit, but not the fate of P (Schröder et al., 2016b).

ANCA calculates soil N_2O-N emission and NO_3^--N leaching using leaching and emission factors (Fraters et al., 2015; Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2012; Schröder et al., 2016b). In the present study, for the calculation of N_2O-N emission only direct emissions were considered. Indirect N_2O-N emission mainly occurs outside the farm. It can be an intermediate product of denitrification of leached nitrate, a by-product of nitrification of ammonium, or a by-product of redeposited soil-N emissions (IPCC, 2006). In ANCA, a crop-level emission factor is multiplied with the amount of applied fertilizer at crop level. For chemical fertilizers an emission factor of 0.008 was used, and for excreta during grazing an emission factor of 0.024 was used. For organic fertilizers an emission factor of 0.003 was used for grasslands, and an emission factor of 0.013 was used for arable land (Schröder et al., 2016b). Most soil N_2O emission originates from applied fertilizers and excreta. Therefore, N_2O emission on unfertilized land was considered as a background emission and estimated as 1 kg N_2O-N ha^{-1} y^{-1} (IPCC, 2006; Schröder et al., 2016b).

ANCA calculates NO_3^--N leaching by multiplying the crop level partial N balance (surplus or deficit, corrected for soil N_2O-N emission but not yet for leaching) with a crop level leaching factor. N leaching factors were based on crop type (arable or grass); soil type classes sand, peat and clay; and for the sandy soils also the groundwater table classes wet, moist, or dry (Fraters et al., 2015; Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2012; Table 2.1). Currently, in ANCA farmers manually enter the soil type classes that can be found within their farm. If more than one soil type occurs within a crop type, ANCA applies a weighted mean leaching factor to the soil class.

2.2.4 Quantification of nutrient balances at field level

In this study, N and P inputs and outputs (mentioned in Section 2.2.3) were available at field level. Field-level data were obtained from the Cows and Opportunities project (Oenema et al., 2015). NO_3^--N leaching and N_2O-N emission were calculated at the field level by applying the leaching and emission factors at field level, in contrast to using leaching and emission factors at the crop and farm level previously shown in Section 2.2.3 (Schröder et al., 2016). Field-specific leaching and emission factors were determined by using the field-specific N and P inputs registered by Cows and Opportunities and by using field-specific soil texture classes obtained from the 1:50,000 soil map of the Netherlands (Alterra, 2006) and using field-specific groundwater table classes (Hoogland et al., 2014; Wageningen Environmental Research, 2017; Table 2.1). For four fields of Farm SHe2, groundwater tables were not

indicated at the soil map; for those fields groundwater tables were obtained from an earlier soil survey (Alterra, 2005). For each field we selected the prevailing soil series and groundwater table to represent the entire field, in case more than one soil series and groundwater table classes were present within a field. To determine N leaching factors, ANCA uses general textural classes 'sand' and 'clay', and has no provision for loamy soils. We therefore classified fields with loamy soils (8-25% clay), and fields with higher clay contents as 'clay'; all other soils were classified as 'sand'. N and P inputs (corrected for NH_3 emission), outputs (including N_2O emission) and balances were calculated for each farm based on the field balances. NO_3^--N leaching (calculated as a fraction of an N surplus), N_2O-N emission, and the soil N and P partial balances were assessed at field level as a function of soil series (soil mapping unit on the 1:50,000 soil map).

2.2.5 Comparing N leaching calculations at different spatial scales

To assess whether it is relevant to calculate nutrient balances at the field level rather than at crop or farm level, two methods were used.

1. The ANCA method. Although the operational ANCA uses nutrient input and output data at crop and farm level, in the present study those data were obtained by aggregating field balances to crop and farm level. Also, the leaching factors were determined at the field level first, and then aggregated to the crop level. This modification was performed to guaranty that the difference in results between ANCA and a field balance method was fully explained by a difference in aggregation level and not by a difference in leaching factors. Subsequently, according to the operational ANCA, the crop-level leaching factor was multiplied with the crop-level soil N balance to come up with N leaching at crop level and at farm level (Section 2.2.3). The operational ANCA tool that is used by farmers has no information of N inputs, outputs and balances at the field level. In theory fields can exist that have a negative balance.
2. The field balance method. N leaching was calculated at the field level first, using the field-specific leaching factors and soil N surpluses. In case of an N deficit, no leaching was calculated. N leaching and N deficits at the field level were separately aggregated to the crop level (Section 2.2.4).

Both methods used the same soil nutrient balance (surplus or deficit) data and the same leaching factors at the field level. To compare the calculation methods of the two approaches, nutrient balances of both approaches were aggregated to crop and farm level. A calculation example of ANCA (Method 1) and the field balance method (Method 2) is given in Figure 2.4.

1. N balance (surplus or deficit) * Leaching Factor (LF) = N leaching
2. N surplus * Leaching Factor (LF) = N leaching
   and/or N deficit

Example data
N balance ~ LF

1. ANCA method

2. Field balance method

grass

SOIL X
-20 ~ 0.5
DEFICIT

SOIL X
120 ~ 0.5
SURPLUS

FIELD AGGREGATION
(-20* ½ + 120* ½)
*
FIELD AGGREGATION
(0.5* ½ + 0.5* ½)
=
50 * 0.5 = **25**

Deficit = **20**

Leaching:
120 * 0.5 = **60**

maize

SOIL Y
-20 ~ 0.75
DEFICIT

SOIL X
120 ~ 0.5
SURPLUS

FIELD AGGREGATION
(-20* ½ + 120* ½)
*
FIELD AGGREGATION
(0.75* ½ + 0.5* ½)
=
50*0.625 = **31.25**

Deficit = **20**

Leaching:
120 * 0.5 = **60**

Total at farm level:
25+31.25 = **56.25** leaching.

Total at farm level:
20+20 = **40** deficit,
60+60 = **120** leaching.

Figure 2.4. Illustration of the ANCA method (Method 1) and the field balance method (Method 2). The size of four fields (2 grass, 2 maize) are equal. Both grass fields have soil type SOIL X (with leaching factor 0.5), while one maize field has SOIL X and the other field has SOIL Y (with leaching factor 0.75). While the ANCA method calculates N leaching at the crop level before aggregating to farm level, the field balance method calculates N leaching at the field level before aggregating to farm level.

2.3 Results

2.3.1 N and P field balances

Average N inputs on grassland fields were 265-445 kg N ha⁻¹, whereas average N inputs on arable fields were 150-320 kg N ha⁻¹. Similar to soil N inputs, average soil P inputs were higher for grassland than for arable land, averaging 35-44 kg P ha⁻¹ and 11-34 kg P ha⁻¹ respectively. On average, most farms had a positive N balance and a negative P balance. Exceptions are Farm CH0, which had a negative N balance for arable land (-51 kg N ha⁻¹), Farm SHe2, which had a positive P balance for arable land (4 kg P ha⁻¹),
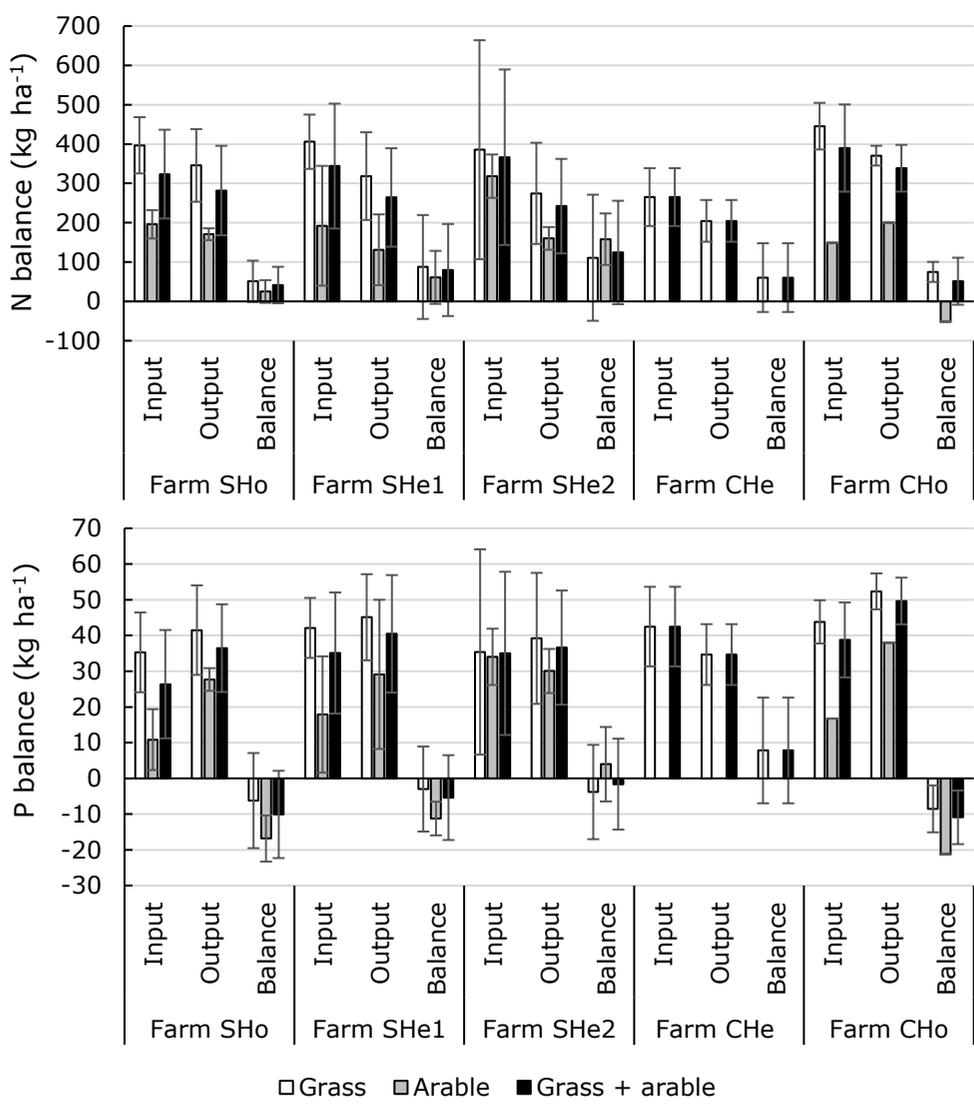
Figure 2.5. Average soil N and P inputs, outputs (including N_2O-N emission, but not yet NO_3^--N leaching) and partial balances at crop level and farm level (grassland and arable land together) for the year 2014. Error bars represent the standard deviations of N and P balances across the fields within a crop type and within the farm.

and Farm CHe, which had a positive P balance (only grassland present: 8 kg P ha^{-1}). Farm SHe2 had on average the highest N surplus of all farms (124 kg N ha^{-1}, grassland and arable land together), while Farm SHo had the lowest average N surplus (42 kg N

ha^{-1}, grassland and arable land together). Farms SHo and CHo had on average the most negative P balance of -10 kg P ha^{-1} and -11 kg P ha^{-1} respectively (Figure 2.5).

The greatest variation in N and P inputs and outputs across fields within a farm was found for the grass fields of SHe2 and arable fields of SHe1, and the least variation was found for farm CHo, followed by Farm CHe and SHo. The standard deviations on grassland of Farm SHe2 equalled 278 kg N ha^{-1} and 29 kg P ha^{-1} for total inputs, and 129 kg N ha^{-1} and 18 kg P ha^{-1} for total outputs. The standard deviations on arable land of Farm SHe1 equalled 152 kg N ha^{-1} and 16 kg P ha^{-1} for total inputs, and 90 kg N ha^{-1} and 21 kg P ha^{-1} for total outputs (Figure 2.5). The least variation in N and P inputs and outputs across fields within a farm was found for Farm CHo. The standard deviations on grassland equalled 59 kg N ha^{-1} and 6.1 kg P ha^{-1} for total inputs, and 25 kg N ha^{-1} and 5 kg P ha^{-1} for total outputs. Farm CHe had no arable land. Low standard deviations of N (53 kg ha^{-1}) and P (9 kg ha^{-1}) outputs were found (Figure 2.5). Nevertheless, across grassland fields the highest variation of total N and plant available P in soil were found for Farm CHe (Table 2.2).

2.3.2 N and P balance per soil series

When assessing N and P balances per soil series (Figure 2.6, 2.7 and Appendix 2), it was found that for some soil series a relatively narrow range of N balances, NO$_3$$^-$-N leaching, N$_2$O-N emissions and P balances across the fields was found, while for other soil series broad ranges were found. To illustrate this in further detail we take Farm SHo as an example: from the farms located on sandy soils, the average nutrient inputs and outputs at field level showed the least variation for Farm SHo (a homogeneous farm); the standard deviations of the inputs and outputs for grass and arable land were smaller than for the other farms (all heterogeneous) located on sand (Figure 2.5). N leaching and N$_2$O-N emission on Farm SHo were highest for soil series tZd21 (two fields), and lowest for soil series pZg23-VII (two fields). On the main soil unit, Hn21-VII, the widest ranges of the N balance, N leaching, and N emissions were found for both grassland (14 fields) and arable land (nine fields, Figure 2.6), compared to the other soil series within this farm.

The P balance was negative for most fields on Farm SHo (Figure 2.7). P balances were close to zero for soil series tZd21 (two fields), and most negative for soil series pZg23-VII (two fields). A wide range of P balances was found on the main soil series Hn21-VII (23 fields).
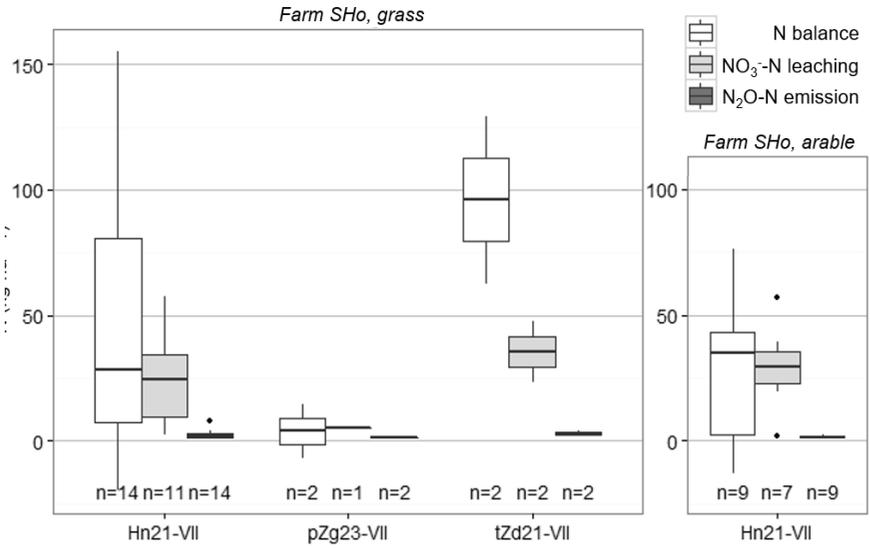
Figure 2.6. Box and whisker plots of partial N balance (corrected for N_2O-N emissions but not yet for leaching), NO_3^--N leaching and N_2O-N emission as function of soil and crop type, for Farm SHo (sand, homogeneous). Note that a part of the N balance ends up in NO_3^--N leaching. 'n' denotes the number of fields per soil type. Please refer to Table 2.1 **Table**for soil classification and codes.
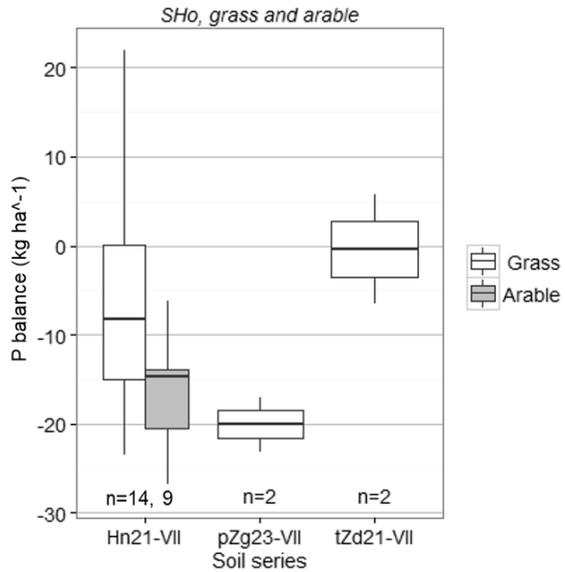
Figure 2.7. Box and whisker plots of P balance as function of soil type for dairy farm SHo (sand, homogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.

2.3.3 Comparing nutrient balance methods

The farms on sandy soils (SH0, SHe1, SHe2) had higher NO_3^--N leaching on both grassland and arable land, than the farms on clay soils (CHe and CH0, Figure 2.8). When using the ANCA method (Method 1), Farm CH0 even had a net N depletion at farm level (grassland and arable land) of 3 kg N ha^{-1} (Figure 2.8C), while the field balance method (Method 2) showed 8.2 kg N ha^{-1} leaching on 34 ha of farmland (of the 41.7 ha total farmland). For the farms on sandy soils, NO_3^--N leaching was higher on arable land than on grassland (Figure 2.8), while in general nitrogen surpluses on arable land were lower than on grassland (Figure 2.5). The largest differences in NO_3^--N leaching, between the ANCA method (Method 1) and the field balance method (Method 2), were found for the calculations on grassland, especially at Farms SHe1 (sand, heterogeneous) and SHe2 (sand, heterogeneous) (Figure 2.8A). The ANCA method (Method 1) resulted in N leaching of 29 kg N ha^{-1} for Farm SHe1 and 32 kg N ha^{-1} for Farm SHe2, whereas the field balance method (Method 2) resulted in N leaching of 50 kg N ha^{-1} for Farm SHe1 and 90 kg N ha^{-1} for Farm SHe2. For the same two farms, aggregating N leaching from crop level to farm level (Figure 2.8C) resulted in smaller differences between the two methods, than when assessing N leaching at the crop level. For Farm SH0, CHe and CH0, the calculated N leaching at crop level was slightly higher for the field balance method (Method 2) than for the ANCA method (Method 1). Furthermore, the calculated N deficits for Farm SH0, CHe and CH0 were smaller than for Farm SHe1 and SHe2, for grassland (Figure 2.8A) and for grass and arable land together (Figure 2.8C).

2.4 Discussion

2.4.1 N and P field balances

2.4.1.1 Nitrogen

The Dutch Manure and Fertilisers Act made rules about the maximum amount of nitrogen from animal manure that may be applied on agricultural fields, according to the European Union Nitrates Directive (part of the Water Framework Directive; Henkens and Van Keulen, 2001; Schröder and Neeteson, 2008). In 2014 a maximum amount of animal manure of 170 kg N ha^{-1} could be applied on dairy farmed fields. These rules were implemented to avoid groundwater nitrate concentrations above 50 mg N L^{-1} (Henkens and Van Keulen, 2001; PBL, 2017). For farms that received derogation the maximum amount of nitrogen was increased to 230 kg N ha^{-1} if located on sandy soils in de East and South of the Netherlands, or to a maximum of 250 kg N ha^{-1} if located on other soils and regions. In the present study, the five farms had average animal manure inputs of 190-230 kg N ha^{-1}. All farms, except Farm CHe, applied for derogation.
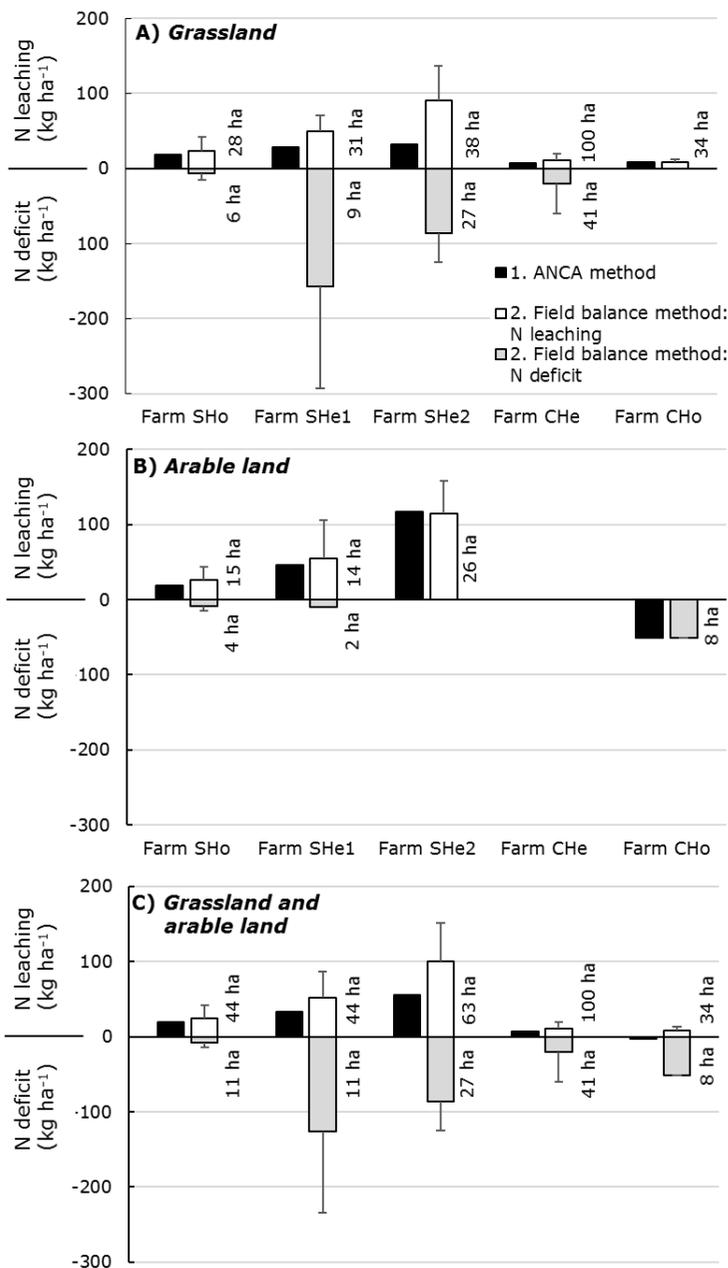
Figure 2.8. Average N leaching at crop level and farm level, obtained with the ANCA method (Method 1) and with the field balance method (Method 2). A: grassland; B: arable land; C: grassland and arable land together. Note that N leaching obtained with the field balance method is only calculated for fields with a surplus (positive Y-axis), and N deficits are represented by the negative Y-axis. For the field balance method (Method 2): the numbers at the bars denote the surface area of land (hectare) included in the calculations; error bars represent the standard deviations of N leaching and N deficit across the fields.

The average N surpluses of the five farms in the present study ranged between 40 and 125 kg N ha^{-1} (Figure 2.5). These average N surpluses were comparable with the lowest surpluses found in other studies. For example, Daatselaar et al. (2015) assessed partial balances (using similar inputs and outputs as in the present study) on the soil level, on 275 dairy farms in the Netherlands, located on sand, clay and peat soils (year 2009-2011). They found N balances of 111-150, 121-188, 111-127 kg N ha^{-1} respectively. Other studies assessed N balances on the farm level, across the four components livestock, manure, soil and crops (Buckley et al., 2015; Gourley et al., 2012; Mihailescu et al., 2014). In these studies N leaching and N emissions were not estimated. Gourley et al. (2012) assessed partial balances on 44 dairy farms in Australia on 'various' soil types. N inputs included fixation and deposition. They found farm level N surpluses between 47 and 600 kg N ha^{-1}. Buckley et al. (2015) assessed partial balances on 242 dairy farms in Ireland and found farm level N surpluses between 25 and 310 kg N ha^{-1} (mean N balance of 145 kg N ha^{-1}). Mihailescu et al. (2014) assessed partial balances on 21 dairy farms in South-West Ireland (year 2009-2011) and found farm level N surpluses between 69 and 239 kg N ha^{-1} (mean N balance of 175 kg N ha^{-1}).

One reason that lower N surpluses were found in the present study than in other studies, may be the difference in system boundaries defined when assessing nutrient balances at different spatial scales (see also Section 2.2.2). In contrast to the present study, the previously mentioned studies (Buckley et al., 2015; Gourley et al., 2012; Mihailescu et al., 2014) computed partial balances at the farm level and not at the field level. At farm level, the components livestock, manure, soil and crops are assessed together (Figure 2.3), while a nutrient balance at the field level focusses on the soil component. Some nutrient outputs from housing and feed storage (e.g., emissions) are included in a farm-level balance, but not in a field-level balance. Van Beek et al. (2003) computed partial balances at farm level (farm-gate balance) and at field level (soil surface balance, including atmospheric deposition) on six dairy farms on a peat soil in the Netherlands (1999-2001). The farm-level balances showed average N surpluses between 213 and 271 kg N ha^{-1}, while the field-level balances showed average N surpluses between 110 and 155 kg N ha^{-1}. Field-level balances included, for example, N inputs through slush application from dredges and atmospheric deposition, while they were not included in the partial balance at farm level. Van Beek et al. (2003) also showed that for some farms the partial balances at farm level was positively correlated with the partial balance at field level, while for other farms they were negatively correlated. This illustrates that it is worth assessing nutrient balances at field level and at farm level if the environmental performance of a farm is studied, as the correlation between the two is not always the same.

2.4.1.2 Phosphorus

The Dutch Manure and Fertilisers Act (PBL, 2017) includes rules for the maximum amount of phosphorus that can be applied on farmland. In 2014, depending on the phosphate status of the soil, the maximum amount of P that could be applied was 37.1-43.6 kg P ha^{-1} (85-100 kg phosphate ha^{-1}) for grasslands, and 24.0-34.9 kg P ha^{-1} (55-80 kg phosphate ha^{-1}) for arable land (RVO, 2017). On the five dairy farms in the present study the P input was 35-44 kg P ha^{-1} on grasslands. On arable land, P inputs were 10-18 kg P ha^{-1}, except for Farm SHe2 which applied 34 kg P ha^{-1} (Figure 2.5). Nevertheless, the farms almost reached the P guidelines. P balances at the five dairy farms in the present study were negative (between -2 and -11 kg P ha^{-1}), except for Farm CHe which had a P surplus of 8 kg P ha^{-1}. PBL (2017) reported that the year 2014 was more productive (higher outputs, lower surpluses) than normal, due to more favourable weather conditions for crop growth than other years. Since 2000, 2014 was the only year with negative soil P balances on the national level for sand, clay and peat soils. For the Netherlands, in the period 2011-2014 the average P surplus across all dairy farms was about 2.2 kg P ha^{-1} (5 kg phosphate ha^{-1}; PBL, 2017). Comparing P balances with other studies, for instance, Gourley et al. (2012) found P balances between -7 and +133 kg P ha^{-1}; Buckley et al. (2015) found a mean P balance of 6.2 kg ha^{-1} (about -100 to +42 kg P ha^{-1}); Van Beek et al. (2009) found soil surface balances at field level of 20 kg P ha^{-1} on sand, 21 kg P ha^{-1} on clay and 14 kg P ha^{-1} on peat; and Van Beek et al. (2003) found soil surface balances at field level between 3 and 7 kg P ha^{-1}. Our results (Figure 2.5) are comparable with P balance found in the previously mentioned studies, even though different system boundaries may be used. While for N balances it matters whether a balance is made at the farm level (including the components livestock, manure, soil and crops) or at the field level (soil component), in the case of a P balance, a farm balance will have the same result as a P balance at the field level, because P cannot be emitted for example, from housing or feed storage.

2.4.1.3 Variation in soil series and N and P balances

The greatest variation in within-farm N and P inputs and outputs was found for the grasslands of Farm SHe2 (Figure 2.5), which was classified as heterogeneous in terms of soil series, like Farms SHe1 and CHe. This may suggest that soil series heterogeneity is responsible for within-farm variation, but even on homogeneous farms a wide range of N and P balances, N leaching and N emissions was found within the same soil and crop type (Figure 2.6, 2.7, A.2.13, A.2.17). This is not surprising because additional factors besides soil and crop type influence N and P balances, such as management factors (type and amount of fertilizer or manure) and soil properties not represented by soil maps (for instance soil structure, rooting depth, soil organic matter contents, as discussed previously by Lipiec and Stępniewski (1995) and Oenema et al. (2010)). When nutrient balances are assessed at the field level, extremes in N and P balances

(extreme surpluses or deficits) become visible (Figure 2.5-2.7), which could help farmers improve their management. The factors that affect N and P balances in ANCA, are variations in nutrient inputs (management factors) and outputs (management factors, crop growth), as well as N emission and N leaching factors, which will be discussed below.

Variation in nutrient inputs

The variation in N and P inputs results from farm management decisions on grazing (inputs through excreta), fertilizer and manure application (Lanyon, 1994). If fields are grazed, N and P inputs through manure and fertilizer should be compensated to avoid over-fertilization. In practice, only Farm SH0 and SHe1 compensated most excreta on grazed fields by applying less manure and fertilizers (data not shown). The other farms, however, did not compensate the excreta on grazed fields, which means that grazed fields received more N and P inputs than fields that were not grazed.

Another factor that could explain variation in N and P inputs, is the distance from the farmhouse to the fields (Powell et al, 2007). Fields that are nearby or easily accessible, typically receive more fertilizer and excreta than fields that are distant or poorly accessible. Gourley et al. (2015) found the highest concentrations of plant available P in soil for fields close to the farmhouse, where grazing intensity, manure and fertilizer inputs were the highest. Sometimes management decisions may be related to soil type but it can be a coincidence, if for example, the fields that are easily accessible are on the same soil type and receive more nutrients by excreta during grazing and additional manure application. In our study, this was likely the case for Farm SHe2 and its fields around its farmhouse on soil series zEz21 (Figure 2.2, Table 2.1), which received the highest nutrient inputs through fertilizer and excreta during grazing (Figure A.2.11).

Another factor that explains variation in N and P inputs is that farmers may increase fertilization in the current growing season if fields had relatively low crop yields in the previous growing season (Lanyon, 1994; Van Beek et al., 2003). It was expected that crop yield increases with increasing nutrient inputs. If this relation is poor (in the present study the poorest correlations were found for all maize land, and for grasslands of Farm SHe1, CHe and CH0; data not shown), then the farmer did not consider field characteristics that influence yield, such as soil condition (e.g. soil N supply, water availability, and compaction), type of grass species, and the optimal amount of manure of fertilizer for crops to grow (Bramley, 2009). Instead, all the earlier mentioned factors in this section, could be an explaining factor for the amount of nutrient inputs.

Variations in N and P inputs almost certainly occur within a farm and are invisible in a farm level nutrient balance. If high nutrient inputs occur locally, this may lead to high nutrient losses to the environment that are invisible in a farm level nutrient balance.

Variation in nutrient outputs

At field level, crop yield can be affected by many factors, such as N and P availability (management factor or soil factor) and soil conditions (Bramley, 2009). An example of a soil condition that can influence the N and P balance in ANCA, is soil compaction. Extreme soil compaction, for instance, negatively affects root growth and soil aeration, thereby decreasing yield Furthermore, when water logging occurs on a compacted layer, denitrification rates increase while crop growth decreases (Ball et al., 2017; Lipiec and Stępniewski, 1995). Effects of, for instance, compaction on crop yield would be measurable in ANCA, while effects of compaction on altered denitrification rates are not included in ANCA.

For N_2O emissions, the same emission factors were used for all the fields within a crop type, because all soils were mineral soils and not organic soils. The N_2O emissions, therefore, were the result of the amount of (chemical and/or organic) fertilizer inputs, which is a management factor. In fact, the amount of N_2O emission varies strongly in space and time, because it depends on many factors, such as soil texture, soil moisture, crop type and weather conditions (Mosier et al., 1998; Van Groenigen et al., 2010). For the five farms, grassland received on average more manure than arable land (Figure 2.5), and hence N_2O emissions on grassland were higher than on arable land (Figure 2.6). Although the emission factors used could give the impression that the relationship between fertilizer application rate and N_2O emission is linear, other studies show a non-linear relationship. For example, Van Groenigen et al. (2010) show for arable crops that at low fertilizer rates (below 200 kg N ha^{-1}) N_2O emission is stable between 1 and 2 kg N_2O-N ha^{-1}, and at fertilizer rates above 300 kg N ha^{-1} N_2O emission increases rapidly, which could be the result of fertilizer inputs that exceed crop demands.

Variation in nutrient balances

In ANCA, variation in the N and P balance (surplus or deficit) resulted from the combined effects of the amount of N and P applied via manure, fertilizer and excreta (corrected for NH_3 emission), the amount of N_2O emission, and the N and P in harvested crops. We calculated N leaching based on the N balance (surplus or deficit) and leaching factors (based on soil texture class, groundwater table class and crop type). Therefore we expected the resulting N leaching (corrected for NH_3 and N_2O

emission) to vary with soil series and crop type. The leaching factors were determined using the 1:50,000 soil map. In ANCA land use and the number of soil classes within the farm (texture and groundwater tables) are filled in manually by the farmers, which is sensitive to classification errors and can introduce uncertainty in the nutrient balances. These errors may be reduced by digitally linking the Dutch 1:50,000 soil map to ANCA.

2.4.2 Comparing N leaching calculations at different spatial scales

2.4.2.1 General N leaching observations

Farms on sandy soils had on average a higher N leaching than the farms on clayey soils (Figure 2.8), which was a result of the higher leaching factors for sandy soils than for clayey soils used by ANCA (Schröder and Neeteson, 2008), while the N balances for the farms on clay soil were comparable to the farms on sandy soil (Figure 2.5). From the farms on sandy soils, Farm SHo had the lowest N leaching (Figure 2.8) and on average the lowest soil N surplus (Figure 2.5). Farm SHo is an experimental farm, which has the facilities to monitor field inputs and outputs more accurately than regular farms, and to make precise management decisions at the field level (Oenema et al., 2015), which could have resulted in lower N leaching. While soil N surpluses for the farms on sandy soils were lower for arable land than for grassland (Figure 2.8), leaching on arable land was higher than on grassland (Figure 2.8) due to higher leaching factors.

For farm SHe1 a positive correlation was found between N surplus (kg ha^{-1}) and leaching factors for grassland (Figure 2.9: r^2 of 0.71), whereas for Farm SHe2 a weak correlation was found (r^2 of 0.06). This means that for Farm SHe1 on average the highest N surplus was found at fields with the highest leaching factor (0.37). The driest sandy soils have the highest leaching factors (Schröder and Neeteson, 2008). For grass
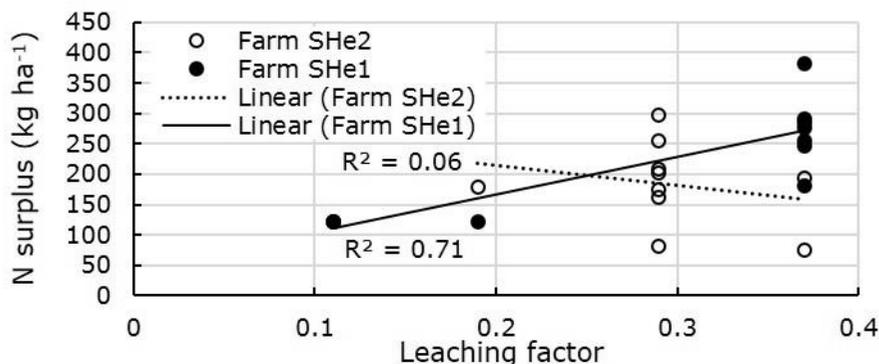
Figure 2.9. Correlation between N balance (surplus or deficit) and leaching factors on grassland.

production on these soils, water availability is probably more limiting than nutrients (Aarts et al., 1999), hence, N inputs should be decreased on these fields to avoid N losses to the environment.

For the farms on sandy soils, a lower N leaching at farm level was found with the ANCA method (Method 1) than for the field balance method (Method 2), because ANCA aggregated the N balances (positive and negative balances together) of all fields to farm level before multiplying with a leaching factor. In the field balance method (Method 2), however, N leaching was only calculated for fields with an N surplus.

2.4.2.2 Use of leaching factors at different spatial scales

The estimated amount of N leaching depends on the calculation method (ANCA method or field balance method). Especially for farms that have fields with varying leaching factors (due to different soil types) within the same crop type (Table 2.1), the ANCA method and the field balance method calculate a different amount of N leaching. Van Beek et al. (2003) concluded that 'first averaging, then calculating' (which is comparable to the ANCA method) does not give the same results as 'first calculating, then averaging' (comparable to the field balance method), when they assessed the correlation between nutrient balances and N leaching, at farm and field level, across six dairy farms in the Netherlands. Our results show a similar mathematical consideration. The exact difference between the two calculation methods depends on the degree of correlation between the N balances (kg ha^{-1}) and leaching factors. Farm SHe1 in the present study, for example, has the highest N surplus (kg ha^{-1}) on grasslands that had the highest leaching factors (Figure 2.9, Section 2.4.2.1). When the correlation between the N surplus (kg ha^{-1}) and leaching factors is positive (Farm SHe1, Figure 2.9), then ANCA calculates a lower N leaching then the field balance method at the crop or farm level.

Homogeneous farms (in terms of soil series) on sand, and all farms on clay had the same leaching factors within a crop type (Table 2.1). When there are no N deficits, N leaching is the same when calculating at the crop or farm level (ANCA method) or when first calculating at the field level (for example, Farm CH0, Figure 2.8). Therefore, a nutrient balance that uses leaching factors based on soil type should be calculated on the field level when different soil types are present within a farm.

2.4.3 Uncertainty in leaching factors

Although estimated N leaching highly depends on the N surplus and the leaching factors used, ANCA gives a good representation of the nutrient balance at both crop and farm level (Oenema et al., 2017), and the leaching factors were based on long-term measurements of N soil surpluses and NO_3^- concentrations in ground and surface waters (Fraters et al., 2015; Rijksinstituut voor Volksgezondheid en Milieu (RIVM),

2012). Nevertheless, the nature of N inputs and the accuracy of the leaching factors should be considered when applying at field level. The leaching factors link the N balance (surplus or deficit) to N leaching. For the calculation of the leaching factors 18 years (year 1991-2009) of nitrogen measurements in groundwater and surface water, and of soil N balance (surplus or deficit) measurements were used across dairy farms in the Netherlands. The leaching factors, therefore, are usually applied at large temporal and spatial scales, i.e. for several years and across all Dutch croplands and grasslands for two soil texture classes (Fraters et al., 2015; Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2007), because spatial scales of management and soil types were aggregated to determine the degree of correlation between the N balance and N leaching. A reasonably reliable correlation between soil N balance and N concentrations in ground- and surface water has been reported for farms on sandy soils. For clay soils however, this correlation was poor (Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2007; Schröder et al., 2016b). This is an indication that even within crop types and within clay soils, a high variation in N leaching can be found across farms that is caused by management factors or by soil characteristics that were not described. The N leaching values in the present study can be considered as a mean representation of the farms, when these would have been measured for several years (Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2007). In the present study, however, leaching factors were applied to one year of data. Therefore, it is not likely that estimated N leaching represents multi-year N leaching, and hence absolute N leaching values should not be used for comparison with other studies. Also, using one year of data, N deficits or N surpluses could occur on certain fields because of enhanced mineralization rates or enhanced nitrogen immobilization because of crop rotation schemes. When considering a longer period, e.g., three years, mineralization and immobilization of nitrogen as a result of crop rotation would come closer to an average nitrogen balance than when using one year of data. Reijneveld et al. (2009) found that soil organic carbon (SOC) in the top 25 cm soil was stable for a 16-year period under grassland in the Netherlands, but SOC contents fluctuated approximately 5 g kg^{-1} (for average SOC contents of 40-45 g kg^{-1}) from year to year. The assumption of stable soil stocks is only acceptable when considering a longer timer period than a period of one year.

Also, within a year, timing of manure application and grazing is crucial for the N balance, but this is not included in ANCA. For example, manure application and grazing in fall causes higher leaching than in spring, because plant uptake in fall is lower than in spring (Cuttle and Bourne, 1993; Van Es et al., 2006).

Consequently, as the leaching factors were designed to determine the correlation between N balances and N leaching for crop types at the national level, calculations of N leaching across crop types at the national level are more accurate than
48

calculations at the farm level or field level. Uncertainty of N leaching calculations at field level is increased by varying mineralization rates, varying availability of nitrogen in the organic or chemical fertilizers, and varying precipitation surpluses (Vellinga and André, 1999).

2.5 Conclusions and recommendations

Nitrogen and phosphorus field balances were determined for five Dutch dairy farms, located on sand and clay soils. Results showed that even homogeneous farms (with few different soil types) had large variation in field-level nutrient balances (grassland: -20 to +150 kg N ha^{-1} and -24 to +22 kg P ha^{-1}; arable land: -13 to +76 kg N ha^{-1} and -27 to -6 kg P ha^{-1}) within the same soil and crop type. This variation was likely caused by management factors or yield variations.

Also, we concluded that field-level balances better represent nitrogen leaching than crop or farm-level balances when:

1. Many negative nitrogen balances (deficits) at field level can be expected (in case of highly productive soils). If negative field balances are together with positive field balances aggregated to farm level, this results in lower N leaching at farm level than when using a field balance method where leaching is only calculated as a fraction of positive field balances (surpluses).
2. Within a farm, and within a crop type, different leaching factors are unequally present due to variation in soil texture and groundwater tables. Aggregation of leaching factors first before multiplying with the N balance gives different N leaching estimates, than when multiplying the leaching factor with the N balance first at field level before aggregating to crop or farm level (field balance method).

In conclusion, given the within-farm variability of management decisions, soil characteristics and/or soil series (and hence varying leaching factors), field balances are preferred above a farm balance. Yet, field-level nutrient balance data is not widely available due to financial constraints and labour availability. Therefore, only for farms with the highest variation in soil types and/or management it is recommended to use field-level nutrient balances. This could eventually lead to improved management decisions and reduced nutrient losses to the environment while maintaining or increasing yield.
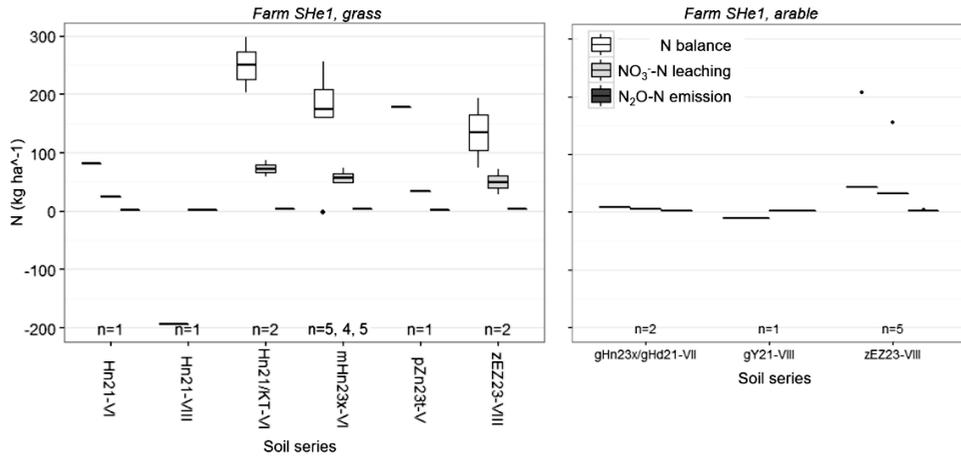
Figure A.2.10. Box and whisker plots of partial N balance (corrected for N_2O-N emissions but not yet for leaching), N leaching and N_2O emission as function of soil and crop type, for Farm SHe1 (sand, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
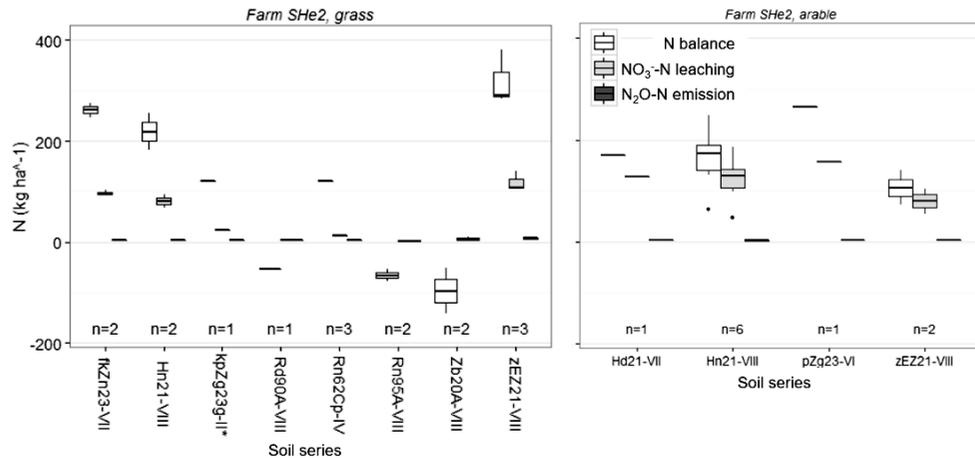
Figure A.2.11. Box and whisker plots of partial N balance (corrected for N_2O-N emissions but not yet for leaching), N leaching and N_2O emission as function of soil and crop type, for Farm SHe2 (sand, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
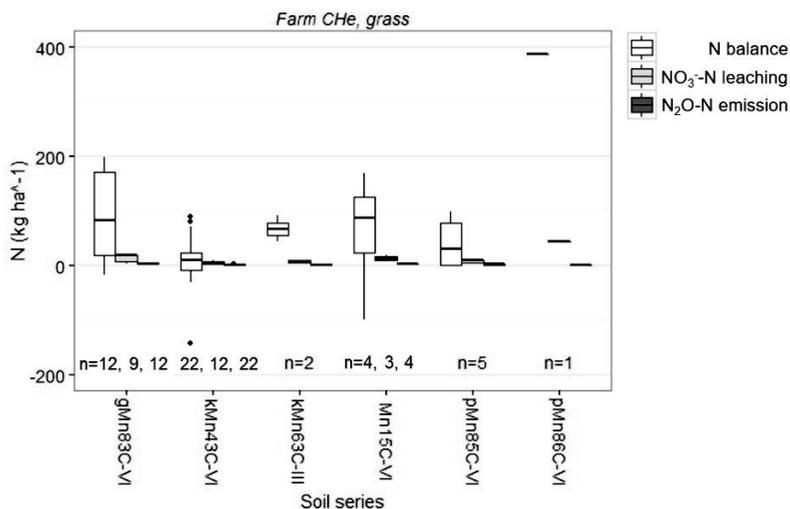
Figure A.2.12. Box and whisker plots of partial N balance (corrected for N_2O-N emissions but not yet for leaching), N leaching and N_2O emission as function of soil and crop type, for Farm CHe (clay, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
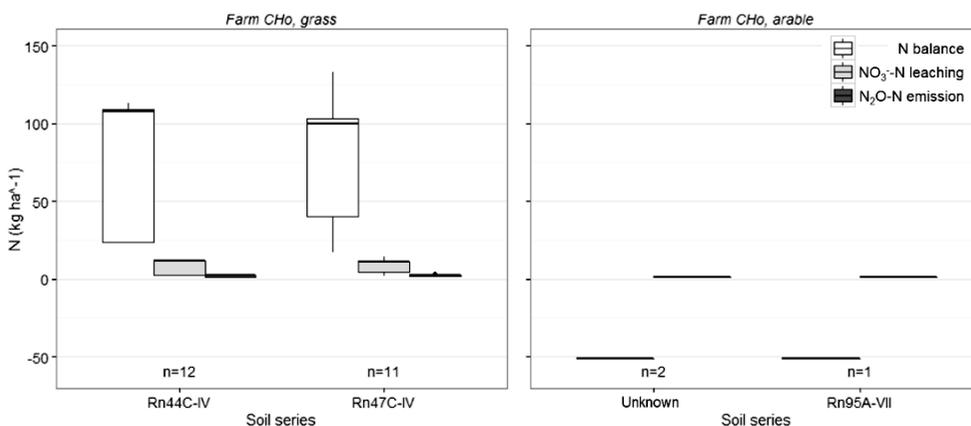
Figure A.2.13. Box and whisker plots of partial N balance (corrected for N_2O-N emissions but not yet for leaching), N leaching and N_2O emission as function of soil and crop type, for Farm CHo (clay, homogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
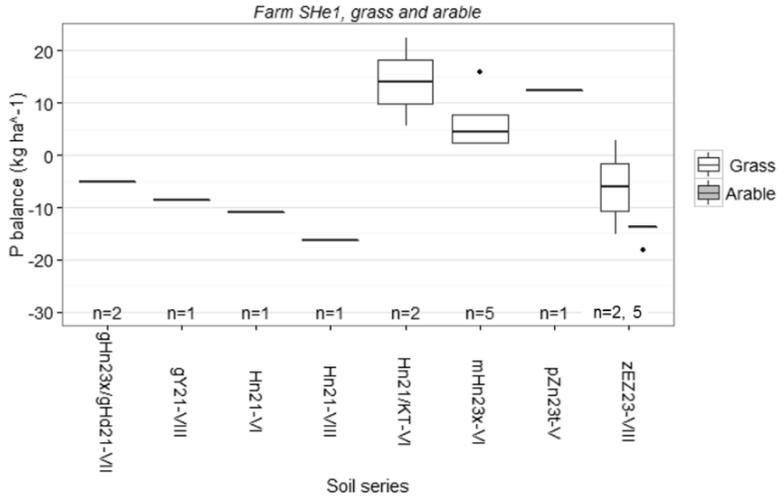
Figure A.2.14. Box and whisker plots of P balance as function of soil type for dairy farm SHe1 (sand, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
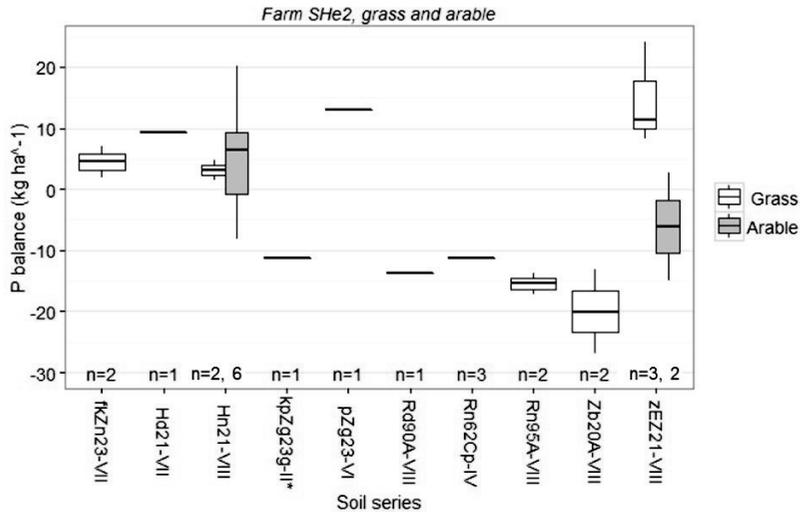
Figure A.2.15. Box and whisker plots of P balance as function of soil type for dairy farm SHe2 (sand, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
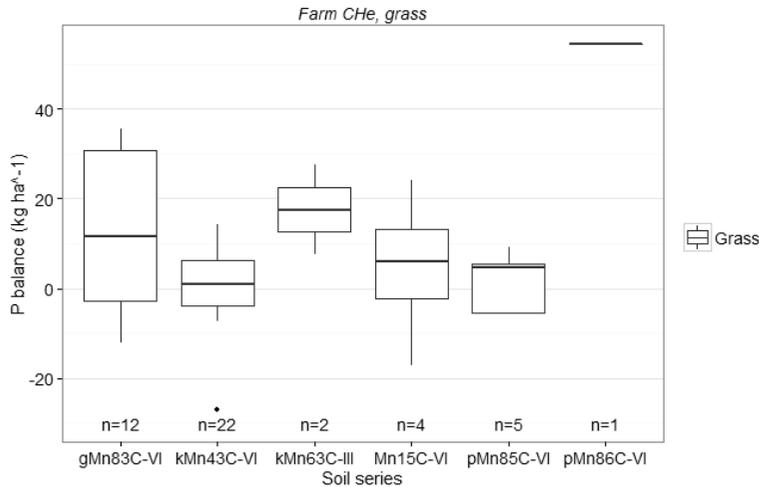
Figure A.2.16. Box and whisker plots of P balance as function of soil type for dairy farm CHe (clay, heterogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.
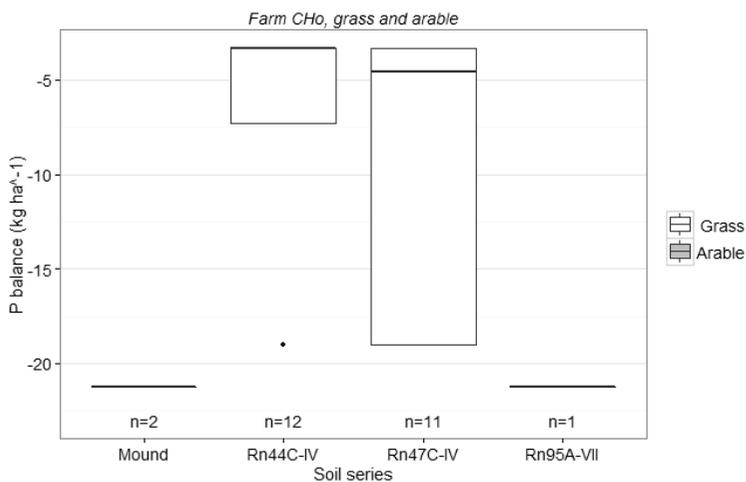
Figure A.2.17. Box and whisker plots of P balance as function of soil type for dairy farm CHo (clay, homogeneous). 'n' denotes the number of fields per soil type. Please refer to Table 2.1 for soil classification and codes.

Visual soil evaluation: reproducibility and correlation with standard measurements
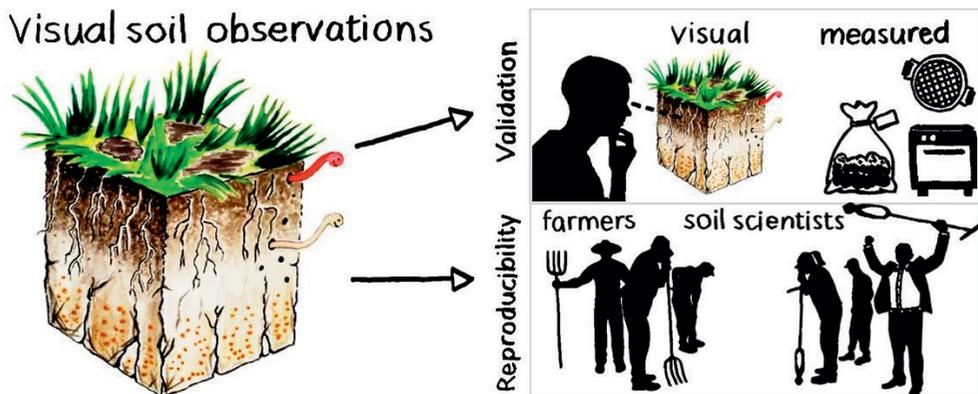
Highlights

- Accuracy of visual soil observations was studied as part of visual soil evaluation.
- Reproducibility of visual observations was studied among farmers and soil scientists.
- Visual observations were correlated with quantitative field or laboratory data.
- Farmers and soil scientists gave similar results for six out of seven observations.
- Visual soil evaluation procedures should be soil type specific.

Abstract

Visual soil evaluation (VSE) is a simple and fast method to assess soil quality in situ, and is becoming increasingly popular. Besides soil structure assessment, also other soil properties can be assessed such as grass cover, roots and earthworms. Yet, the full set of visual observations has not been properly evaluated for reproducibility and correlation with standard field or laboratory measurements, for several soil types. The objectives of this study were therefore to evaluate the reproducibility and the correlation of visual observations with closely related field or laboratory measurements. We used quantitative visual observations where possible, to enhance objectivity of VSE. The reproducibility and correlation of visual observations with standard measurements was evaluated for three soil types (sand, peat and clay) in the North Friesian Woodlands, The Netherlands. Reproducibility of quantitative visual observations was tested by comparing observations made by farmers and soil scientists, on the same soils. A linear mixed-effect model indicated that for all quantitative visual observations except for the depth of soil compaction, subjectivity due to the observers' background (farmer or soil scientist) had no significant effect on the observations. For assessment of relative soil quality differences between sites, the results suggested that a single observer can make the visual observations, when assessing the fraction largest soil structural elements, earthworms, gley mottles and the depth of soil compaction. Spearman's rank correlation coefficients indicated that visual observations of grass cover, root count, maximum rooting depth and the fraction largest soil structural elements correlated significantly with closely related field or laboratory measurements regardless of soil type. Maximum rooting depth, root count, soil colour, the fraction largest soil structural elements, and the degree of soil compaction only significantly correlated with field or laboratory measurements for specific soil types. Analyses showed that the correlation of visual observations with standard measurements were soil type dependent, suggesting that the evaluation of soil quality should also be soil type dependent.

3.1 Introduction

Visual soil evaluation (VSE) methods are becoming increasingly popular among farmers, organisations and companies that focus on soil management and environmental sustainability (Ball et al., 2016). A VSE determines soil quality through several soil quality characteristics that are observable by eye (e.g. Ball et al., 2007; McKenzie, 2013; Shepherd, 2009). After visual observations of soil quality characteristics, weight factors are assigned to indicate the relative importance of each soil quality characteristic, and soil quality is evaluated using a grading system. Visual soil evaluations can be used to monitor soil quality, to identify constraints for soil functioning, and to identify soils that are in an early stage of degradation (McGarry, 2004). Visual soil evaluation is cost-effective and rapid, e.g., the visual soil assessment (VSA) of Shepherd (2009) takes approximately 45 minutes. Visual soil evaluation is furthermore a valuable addition to soil chemical and physical analyses for the interpretation of land degradation issues (McKenzie, 2013). Because of the increased use of VSE, it is essential that the method is reproducible and the made observations are correct. We therefore focus on the first step in VSE: the visual assessment of soil quality characteristics.

The visual soil assessment (VSA) of Shepherd (2009) uses one of the broadest sets of visual soil quality characteristics among all VSEs. However the relationship between each of the visual soil quality characteristics and soil physical measurements is only assessed for clay soils (Sonneveld et al., 2014) and not for other soil types. As relationships between visual soil quality characteristics and laboratory-measured soil parameters likely vary between soil types, use of VSE methods developed for a single soil type may lead to poor accuracy when it is applied on other soil types. Other VSEs that significantly correlate with soil physical measurements for various soil types are the visual evaluation of soil structure (such as such as SoilPAK (McKenzie, 2001), the Peerlkamp test (Ball et al., 2007; Mueller et al., 2009), Visual Evaluation of Soil Structure (Guimarães et al., 2013; Newell-Price et al., 2013; Pulido Moncada et al., 2014), CoreVESS (Johannes et al., 2017), and VSA soil structure (Mueller et al., 2009)), as well as the visual assessment of soil compaction using the French *profil cultural* method (Peigné et al., 2013). To the best of our knowledge no other VSEs have been related to soil physical measurements for several soil types.

An additional challenge of VSE is that its usefulness is often questioned by critics because of the potential subjectivity in the visual observations, although VSE protocols are easy to use and self-explaining (Guimarães et al., 2017). While the reproducibility of visual assessment of vegetation cover and the visual evaluation of soil structure have been studied (Klimeš, 2003; Ball et al., 2007), the reproducibility of the full range of visual soil quality characteristics has not yet been evaluated for

potential users (agricultural land managers and environmental scientists) and contrasting soil types. Klimeš (2003) found that visual grass cover observations were not reproducible among five observers, on seven sites. Ball et al. (2007) in contrast found the visual evaluation of soil structure to be reproducible, assessed by two experts and two non-expert users at two sites. It is relevant to know whether farmers can assess soil quality on their own, or whether a specialist should be hired to assess soil quality.

Aside from the benefit of assessing the correlation between visual observations and standard field or laboratory measurements for contrasting soil types and having insight into its reproducibility, the quality of VSE may improve if a more quantitative approach is taken. VSE is usually based on qualitative or semi-quantitative information, where visual observations are reported as scores rather than numeric quantitative observations (Ball et al. 2007; Peerlkamp, 1959; Shepherd, 2009). However, a quantitative assessment may give a better representation of soil quality and allows VSE methods to be universally applicable (Emmet-Booth et al., 2016).

The objectives of this study were to evaluate the reproducibility of visual observations and to evaluate the correlation between visual observations and standard field or laboratory measurements. The reproducibility of visual observations was tested by comparing visual observations made by farmers and soil scientists at the same sites. We used quantitative visual observations where possible, rather than semi-quantitative or qualitative visual observations as an attempt to make VSE more objective. In this study we use the broad set of visual soil quality characteristics proposed by Shepherd (2009, VSA) as it covers the most used visual indicators of soil quality.

3.2 Materials and methods

3.2.1 Study area

The study area is the North Friesian Woodlands in the North of The Netherlands (Figure 3.1A). The North Friesian Woodlands were selected because of the various soils present (Figure 3.1FB). Dominating soil types are cultivated hydromorphic podzols ('veldpodzols' or 'laarpodzols') developed in Pleistocene aeolian cover sand, histosols ('vlierveengronden'), and fluvisols ('poldervaaggronden') developed in Holocene marine clay (Table A.3.3). The cultivated hydromorphic podzols have a dark coloured plough layer (Sonneveld et al., 2002), of around 30 cm deep and with a Munsell colour value < 3. Glacial till can be found within 120 cm depth from the surface. Groundwater
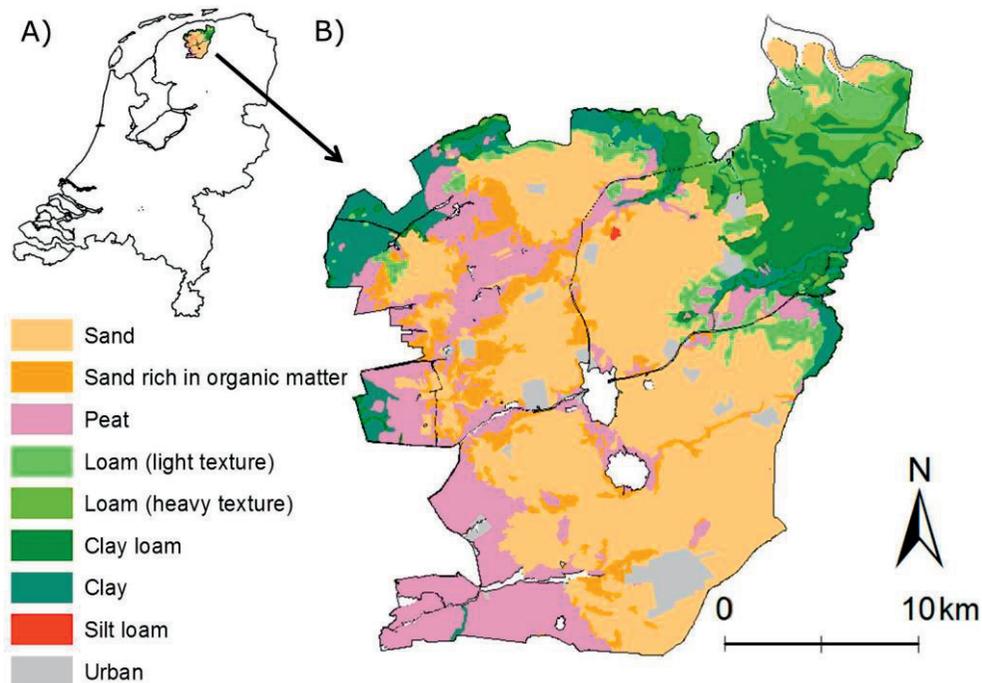
Figure 3.1. Location of the North Friesian Woodlands in The Netherlands (A) and prevailing soil textures (B) (data retrieved from Alterra, 2006).

often perches on the glacial till, and can be found between 25 and >120 cm depth (Sonneveld et al., 2006). Histosols have groundwater tables between 0 and 100 cm depth and fluvisols have groundwater tables between 0 and 120 cm depth. The dominant land use in the North Friesian Woodlands is grassland for dairy farming, with approximately 80% of dairy farmers being member of cooperative 'Noardlike Fryske Wâlden' and using sustainable agricultural practices (Noardlike Fryske Wâlden, 2016). Climate in the region is temperate. Temperatures range from 0.3-5.3°C in winter to 13.2-21.6°C in summer. Mean annual precipitation was 861 mm in the period 2004-2014.The field study year of 2014 was a warmer, dryer and sunnier year than normal, with temperatures ranging between 2.5-7.3°C in winter, and 12.4-21.3°C in summer, and with 671 mm precipitation (Royal Netherlands Meteorological Institute, 2017).

For the reproducibility study, five fields under grassland were selected that were located on sand (n=2), peat (n=1), and clay soils (n=2). The fields were located on three dairy farms that were used in the correlation study (see next paragraph). The fields were homogeneous in terms of topography, grass cover and soil profiles and located

close to each other (within a radius of 4 km) so that the observers could analyse all five fields within the same day.

To correlate visual observations with standard measurements, we selected 26 farms in the North Friesian Woodlands. The farms were more or less equally distributed within a radius of 13 km. At each farm we randomly selected one site (a field) under grassland to carry out visual soil observations and standard field or laboratory measurements. These 26 sites were located on sand (n=11), peat (n=7), and clay soils (n=8). Four of the sites had been renewed within the last three years, but most sites were between 10 and 50 years old (Table A.3.3). From the 26 farms, 22 farms had dairy cattle and four farms had meat cattle.

3.2.2 Procedure of visual observations
From the range of soil parameters in the VSA of Shepherd (2009), we selected grass cover, porosity, root length and root density, soil colour, soil structure, earthworms, gley mottles and soil compaction (Table 3.1). Except for grass cover, root length and root density, we only considered those indicators that directly assess soil quality characteristics, rather than plant quality characteristics that only indirectly assess soil quality. Soil smell was not considered as this was beyond the scope of the study. In contrast to VSA we assessed most visual observations quantitatively rather than using soil quality scores.

First, grass cover was assessed within 1 m² from the place where the soil block would be extracted. Grass cover was observed as the percentage of grass base covering the soil surface. Grass was pulled apart by hand to make bare soil visible, facilitating the estimation of grass cover. We did not cut the grass before assessment, because we wanted farmers to be able to assess grass cover any time without cutting it first. Subsequently, a soil block of 20×20×20 cm was extracted from the topsoil with a spade. Three parameters were quantified on the bottom of the block: 1) earthworm burrows larger than 2 mm were counted over the entire 20×20 cm surface area (biopore count); 2) all roots (living and dead) were counted over a surface area of 10×10 cm (root count); and 3) the soil organic matter content was quantified using the colour value of field moist soil (Wills et al., 2007), with the Munsell Soil Color Charts (Munsell Color, 1975). The bottom half of the soil block (the 10-20 cm depth layer) was subsequently used for soil structure assessment as soil structure was often more distinct in this lower part of the extracted block. For soil structure assessment, the soil had to be broken up first. The drop-shatter method in VSA proved only functional across a narrow soil moisture range (not too moist, not too dry). We therefore broke up the soil by hand along natural cracks following Guimarães et al. (2011), who found that this produces similar results as the drop-shatter test. We then described two structural properties: 1) the

Table 3.1. Set of visual observations with the corresponding standard field or laboratory measurements.

Visual observation	Unit	Corresponding measurement	Unit
Grass cover	% grass cover	Grass cover classification	% pixels grass
Biopores, 20 cm depth	Count, over 20×20 cm	Bulk density, 10-20 cm depth	g/cm³
Roots, 20 cm depth	Count, over 10×10 cm	Root dry weight, 10-20 cm depth	g/L
Maximum rooting depth	cm	Root dry weight, 10-20 cm depth	g/L
Soil colour, 20 cm depth	Munsell value	Soil organic matter, 10-20 cm depth	%
Soil structure, 10-20 cm depth			
Fraction largest elements	%	Mean weight diameter, 10-20 cm depth	-
Shape of largest elements	Score		
Shape of smallest elements	Score		
Earthworms 0-20 cm depth	Count, in 20×20×20 cm	Mean weight diameter, 0-20 cm depth	-
Gley mottles	%	Mean lowest groundwater table	cm below surface
Compaction depth	cm below surface	Penetration resistance >2 MPa	cm below surface
Degree of compaction	Score	Bulk density, 10-20 cm depth	g/cm³

fraction of largest soil structural elements, and 2) the shape of the largest and smallest structural elements, both obtained from VSA soil structure assessment. On a plastic bag the largest soil structural elements (>1.5 cm) were separated from the smallest elements (<1.5 cm). Instead of using score classes for soil structure as in VSA, the fraction of the largest soil structural elements covering the bag was visually estimated. The shape of soil structural elements was assessed and scores were assigned as there was no quantitative alternative: 2 = rounded shape; 1 = sub-angular shape; 0 = angular shape. Finally, the entire 20×20×20 cm soil volume was carefully searched for earthworms, and the earthworm number was recorded.

After extracting and visually analysing the soil block as described above, the sides of the pit were used for visual assessment of gley mottles. If gley mottles were present, the percentage of surface that was covered was reported. The soil pit was then extended to 50 cm depth and the sides of the pit were used to assess soil compaction. If soil compaction was present, the degree of compaction and the depth of the compacted layer were recorded. Following VSA, initially we used the change of soil structure and the presence and absence of roots to identify a potentially compacted layer. Yet given that dry soil conditions influenced the soil structure on clay soils, we then used the presence and absence of roots to identify compacted layers. The degree of compaction was assessed by penetrating the soil with a knife. For this there was no quantitative alternative, so we used VSA score classes: 2 = no compaction; 1 = moderate compaction; 0 = strong compaction. Maximum rooting depth was subsequently assessed by visually identifying the presence of roots in a single 0-1.2 m deep profile extracted using an Edelman soil auger.

Finally, the soil at each site was categorized as peat, sand, or clay according to site's parent material. Soils having a peat layer starting within 40 cm depth of at least 10 cm thick (folic layer, IUSS Working Group WRB, 2007) were in this study named as peat. If the soil was not classified as a peat soil, the classification into cover sand (in this study named as 'sand') or marine clay (in this study named as 'clay') was done by assessing soil type at 20 cm depth. The site was classified as a sand soil when loose sand grains were clearly visible and when the soil could not be smeared between the fingers when moist; it was classified as a clay soil when loose sand grains were mostly absent and when the soil could be smeared between the fingers when moist.

3.2.3 Reproducibility of quantitative visual observations
The five reproducibility sites, having sand (n=2), peat (n=1), or clay (n=2), were independently assessed by eight local dairy farmers from the North Friesian Woodlands (all male, 30-55 years old) and eleven Dutch, academically trained soil scientists (male and female, 25-60 years old). All observers received a 1-h training in the field in which the quantitative visual observations (Section 3.2.2) were
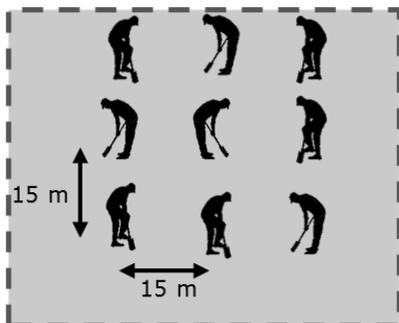
Figure 3.2. The position of the observers in a field during the reproducibility experiment.

demonstrated by the lead author. After the training, the observers used this procedure to carry out the quantitative visual observations. The observers were subsequently asked to report the most reliable and unreliable visual observations according to their experience. The observers were not allowed to communicate with each other during the assessment. The distance between the observers was therefore set to 15 m (Figure 3.2).

3.2.4 Correlation of visual observations with standard measurements
To prevent potential observer bias affecting the results, one single observer made the visual soil observations. Fieldwork was conducted in September and October 2014. Visual observations were correlated with a standard field or laboratory measurement that represented the same or a closely related soil property (Table 3.1).

The following visual observations were correlated with a closely related field or laboratory measurement: grass cover was correlated with digital image analysis of a photograph taken from the same surface area; the number of roots was correlated with root dry weight; the fraction of the largest soil structural elements was correlated with the mean weight diameter (an index to describe the dry aggregate distribution, Van Bavel, 1949) of soil structural elements, the depth of a compacted layer in the soil was correlated with a profile of penetration resistance and the degree of soil compaction was correlated with bulk density.

The other visual observations were correlated with an indirectly related field or laboratory measurement. The number of biopores was correlated with bulk density, as Głab and Kulig (2008) found high bulk density to correspond with low total soil porosity, and the number of biopores to be positively correlated with total porosity. Moist soil colour value was correlated with soil organic matter content, as moist colour value was previously found to be a valid single-factor predictor of soil organic carbon content for agricultural soils (Wills et al., 2007), and soil organic carbon content is closely related to soil organic matter content (Ball, 1964). Maximum rooting

depth was correlated with root dry weight at 10-20 cm depth, as the root biomass in the topsoil has been reported to be inversely related with maximum rooting depth (Jackson et al., 1996). The shape of the soil structural elements was correlated with mean weight diameter, as was previously reported by Sonneveld et al. (2014) for clay soils in The Netherlands. The number of earthworms was correlated with the mean weight diameter, because some earthworm species are known to improve soil aggregate stability and thus increase the dry aggregate distribution index of the soil (Six et al., 2004). We finally correlated the number of gley mottles with the mean lowest groundwater table that was derived from groundwater classes indicated at the 1:50,000 soil map (Alterra, 2006) according to Van der Sluijs and De Gruijter (1985), as the presence of gley mottles is a hydromorphic feature indicating the mean highest and mean lowest groundwater levels.

3.2.4.1 Field procedure

At each site visual observations were carried out as described in Section 3.2.2. Field measurements and soil samples were taken from the same soil pit. After the visual assessment of grass cover and before extracting the soil block for visual assessment, grass was cut to a length of 2 cm over a surface of 38×38 cm, and a photograph of the surface was taken with a digital camera (Canon PowerShot A720 IS). To ensure comparable light conditions between sites, these pictures were taken under shaded conditions. After the soil pit was dug, a bulk density sample was taken between 10 and 18 cm depth, using a core with a height of 7.9 cm and a diameter of 7 cm. The soil from each bulk density sample was used for analysis of soil organic matter content. For the analyses of root dry weight and mean weight diameter, a separate soil block of at least 15×15×20 cm was carefully excavated from an undisturbed side of the soil pit, and transported to the laboratory in a plastic box (17×17×25 cm) so that the block would remain intact. Finally the penetration resistance was measured five times within 50 cm distance from the soil pit, using an electronic penetrometer with a cone angle of 60°, a base area of 1 cm², and penetration recordings with a centimetre interval (Penetrologger, hardware and software version 1.0, Eijkelkamp Soil & Water, Giesbeek, The Netherlands). Soil samples were stored in a cold room at 2°C and analysed within three months.

3.2.4.2 Processing of soil samples and field data

Bulk density and volumetric water content was determined by drying soil samples (48 h, 105°C), and then weighing them. Subsequently, the dried soils were sieved at 2 mm (for soils with low cohesion, including some peat samples) or crushed with a soil crushing machine (for soils having high cohesion, including some peat samples). From each sample a subsample of about 10 g was taken for analysis of soil organic matter

content by loss on ignition (4 h at 550°C) without correcting for water bound to clay and iron minerals.

For the analysis of root dry weight and mean weight diameter, a 1-L subsample was taken from the bottom half of the sampled soil block (10-20 cm, root dry weight), and subsamples of about 500 g were additionally taken from 0-10 and 10-20 cm (mean weight diameter). For root dry weight, soil samples were rinsed with water and all living and dead roots were collected, after which the roots were dried (20 h, 105°C) and weighed. To be able to express root dry weight per volume of soil, the exact sampled volume of soil was calculated using the weight of the field-moist soil sample, volumetric water content and bulk density.

For dry aggregate size distribution, the mean weight diameter (Van Bavel, 1949) was determined following a standard Wageningen University & Research protocol (Alterra, 2014). Prior to analysis, soil samples were carefully passed through a 12-mm sieve by gently breaking up the soil when needed, and dried (7 days, 40°C). Sieves were used with a mesh size of 8, 4, 2, 1, 0.5 and 0.25 mm.

The image of the surface cover that was taken in the field was digitally analysed for grass cover using ArcMap 10.2.1. The image was classified as grass or bare soil, after manually selecting five training samples representing grass and bare soil (ten in total). From the penetrometer data, an average penetration resistance profile was calculated from the five replicate measurements at each site. As root growth is restricted when penetration resistance exceeds 2 MPa (Gugino et al., 2009; Leao et al., 2005), we recorded the depth at which the penetration resistance exceeded this threshold value.

3.2.5 Statistical analyses

3.2.5.1 Reproducibility of quantitative visual observations
The reproducibility of the quantitative visual observations was evaluated using a linear mixed-effect modelling approach (Pinheiro and Bates, 2000). A linear mixed-effect model takes into account the sample size and number of observations. Moreover, quantitative visual observation results can be affected by a number of factors acting at the same time, such as the background of the observers, systematic or random errors by the individual observers, as well as site characteristics, which are taken into account in this modelling approach. In our analysis, observer background was considered as a fixed effect as there were no other potential user types than farmers and soil scientists. Sites and individual observers were taken as random effects as these five sites and 19 observers are merely samples from a much larger population and our interest is in this larger population (i.e., the soils of North Friesian Woodlands, all farmers in the region and all qualified soil scientists). Count data and the observations of gley mottles were

strongly skewed towards the 0-boundary, and hence not normally distributed. Those data were log-transformed to be used in the linear mixed-effect model.

For each quantitative visual observation that was assessed, the following linear mixed-effect model was used to assess the reproducibility:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

where Y_{ijk} is the response (observed quantitative visual observation) at site i ($i = 1, ...,$ 5) for observer type j ($j = 1, 2$) and for individual observer k ($k = 1, ..., 11$ if $j = 1$ and $k = 1, ..., 8$ if $j = 2$). μ is the overall mean, α_i the deviation from the overall mean that represents site effect (random effect), β_j the deviation from the overall mean that represents observer type (fixed effect), γ_k the deviation from the overall mean that represents individual observer bias (random effect), and ε_{ijk} is a random effect representing deviations due to factors not described by the other terms, e.g. random fluctuations in assessments by individual observers and within-site spatial variation, although we selected the sites based on homogeneous topography, grass cover and soil profiles. We assumed that the α_i's, γ_k's and ε_{ijk}'s were mutually independent, normally distributed random variables, had zero mean and constant variances. The linear mixed-effect model was implemented using R (R Core Team, 2014) version 3.1.2 and the lme4 package (Bates et al., 2015). We used the likelihood-ratio test to evaluate the significance of the fixed effects and random effects (Pinheiro and Bates, 2000).

3.2.5.2 Correlation of visual observations with standard measurements

Descriptive statistics (mean and standard deviation) were calculated for all quantitative visual observations and for all field and laboratory measurements grouped per soil type. For categorical data, variation within a soil parameter was expressed as dispersion index 'I^2', where 0 represents maximum variation and 1 represents minimum variation (Blair and Lacy, 2000). Finally, Spearman's rank correlation coefficients (Spearman's ρ) were used to correlate visual observations with standard field or laboratory measurements. Continuous variables were also analysed with Spearman's rank correlation coefficients, to allow comparison between all correlated visual observations, and to account for non-normally distributed data. Significant and valid correlations were considered those with P<0.05 and Spearman's ρ>0.4. The analyses were done in R (R Core Team, 2014) version 3.1.2 using the Hmisc package (Harrel Jr et al., 2015).

3.3 Results and discussion

3.3.1 Reproducibility of quantitative visual observations

3.3.1.1 Grass cover

The observations of grass cover were significantly affected by site (P=0.01, Table 3.2) but not by observer type or individual observers. From the random effects, the standard deviation of the residuals term was relatively high compared to the standard deviation of the individual observers and site effects (Table 3.2). This means that the main error in grass cover assessment was caused by inconsistent observations of a substantial number of individual observers over all sites or by within-site spatial variation, which was not described by the fixed (observers' background) and random effects (Site and individual observers). The significant site effect suggests that the total error was small enough and the group of observers was able to detect differences between sites; hence the mean observed value of a group of observers can be used to get an accurate estimate of grass cover rather than the value of a single observer.

3.3.1.2 Biopore count

The biopore count was affected by site (P=0.00, Table 3.2). Similar as for grass cover, the standard deviation of the residuals term was relatively high compared to the standard deviation of the individual observers and the site effects. This suggests that

Table 3.2. Regression coefficients for the linear mixed-effect models for each quantitative visual observation. Levels of statistical significance: **P<0.01, *P<0.05. Fixed effect represents the background of the observer; values are the difference between soil scientists and farmers (soil scientist mean minus farmer mean).

Quantitative visual observation	Overall mean	Fixed effect observer type	Random effects		
			St.dev. Individual observers	St.dev. Site	St.dev. Residuals
Grass cover (%)	79.3	3.5	5.7	6.2 **	13.1
Biopore count	1.4 [1]	0.2 [1]	0.3 [1]	0.4 [1, **]	0.7 [1]
Root count	3.7 [1]	0.2 [1]	0.4 [1, **]	0.2 [1]	0.5 [1]
Soil structure fraction largest elements (%)	31.4	8.5	7.2 *	21.7 **	14.4
Earthworm count	2.0 [1]	0.1 [1]	0.4 [1, **]	0.7 [1, **]	0.5 [1]
Gley mottles (%)	1.8 [1]	-0.5 [1]	0.4 [1]	1.0 [1, **]	0.9 [1]
Compaction depth (cm below surface)	24.6	-8.0 *	5.8 *	5.0 *	7.3

[1] Log-transformed value.
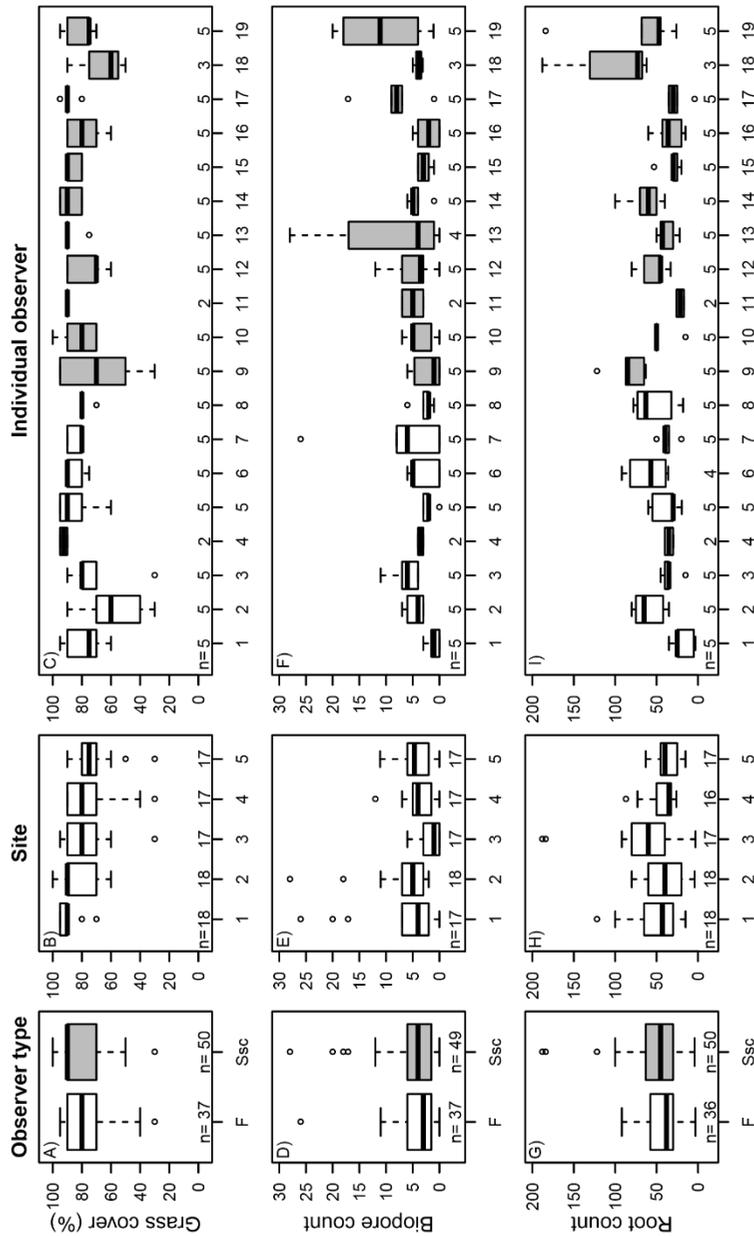
Figure 3.3. Quantitative visual observations grouped by observer type, site, and individual observer. Observer type: farmers ('F', white boxes) and soil scientists ('Ssc', grey boxes); Sites: 1 (clay), 2 (clay), 3 (peat), 4 (sand), 5 (sand); individual observers: observers 1-8 are farmers (white boxes); observers 9-19 are soil scientists (grey boxes). Observer 19 is the lead author and expert in VSA. Observer 4 and 11 were only present at Site 1 and Site 2. Observer 18 was only present at Site 3, Site 4, and Site 5. Note that horizontal lines in the boxes represent median values.

the main error was caused by inconsistent observations by the individual observers or by within-site spatial variation. It is known that biopore numbers can vary over sites, due to local variation in soil moisture content and soil microbial activity (Nakamoto, 1997). Nevertheless, as site was significant, the total error was small enough to be able to detect differences between sites. It is interesting to note that the observers evaluated biopore count as the most unreliable visual observation at three sites: Site 3 (peat), Site 4 and 5 (both sand). For Site 3 this could be explained by the low number of earthworms and thus the low number of biopores (on average each observer counted 2 biopores and 2 earthworms, Figure 3.3E), and for Site 4 and 5 observers reported that it was difficult to recognize biopores, and they reported that biopores easily collapsed. As for grass cover, the significant site effect suggests that instead of using observations by single observers, the mean observed value from a group of observers should be used.

3.3.1.3 Root count

Root count was significantly affected by the individual observers (P=0.00, Table 3.2). Some observers systematically observed more roots than others (Figure 3.3I), irrespective of the site, as site effect could not explain variation in root count (P=0.11). Subjectivity by individual observers was high and therefore visual observations of root count cannot be considered to be reproducible. The observers reported that when having many fine roots on the cut surface of the block, they easily lost their counts (this was especially the case for Site 1-3, which had clay and peat). This resulted in a poor within-group agreement, as the standard deviations were more than 50% of the mean observed value, which is reflected in a larger spread of values for Site 1-3 than for Site 4 and 5 (Figure 3.3H). To increase reproducibility an improvement in the instruction could be to divide the 10 cm-squared surface in e.g. four or nine smaller squares, in order to be more systematic in counting and to avoid losing counts when having many (fine) roots.

3.3.1.4 Fraction largest soil structural elements

The observed fraction of largest soil structural elements was significantly affected by the individual observers (P=0.04) and site (P=0.00, Table 3.2). As for root count, this means that subjectivity of the assessment of the largest soil structural elements by individual observers was high. As the site effect could explain the largest part of the variation in soil structural elements (Table 3.2), this means that individual observers were consistent over the sites and could detect relative differences between the sites. Besides the soil type, the largest difference between the sites that may have influenced the visual assessment of soil structure was the soil moisture condition. Especially Site 1 and 2 (clay) were dry, and hence soil below 10 cm was hard. This hampered the extraction of the soil block, and also breaking open the soil for the assessment of soil

structure. At Site 2 the average soil block depth was 15.8 cm (five observations) while for the other sites the soil block depths were on average between 18.5 and 20.8 cm. Any deviation from the desired soil block depth of 20 cm could have resulted in a deviation on the observed fraction of largest soil structural elements. We observed that the dry conditions moreover resulted in hard soil structural elements on clay. This hampered the breaking of the soil block along natural cracks, resulting in a larger fraction of the largest soil structural elements compared to the case when soil would be moist. The observers found it difficult to judge when to stop breaking open the soil and to move on to the quantification of the fraction of large soil structural elements. From all visual observations, for Site 1 and 2 (clay) the observers indicated that the soil structure assessment was the most unreliable quantitative visual observation.

Ball et al. (2007) assessed the reproducibility of the Peerlkamp soil structure test among four observers on sandy loam soils, and also found that differences between sites were larger than the effect of subjectivity from the individual observers. They also found that individual observers were influenced by site effects (interaction). Interactions were not taken into account in the present study, as this was beyond the scope of this research. The interaction between observer type and site could be studied in future research.

3.3.1.5 Earthworm count
The observed number of earthworms was affected by the individual observers (P=0.00) and site (P=0.00, Table 3.2). As for the observations of soil structural elements, this means that subjectivity by individual observers was high, but the observations were consistent over the sites. The observers reported that if they would have had more time to search the soil for earthworms, they probably would have found more. For the clay sites, which were dry and hard, the observers reported difficulties in breaking open the soil to look for earthworms.

3.3.1.6 Gley mottles
The observations of gley mottles were significantly affected by site (P=0.00, Table 3.2). In contrast to the observations of grass cover and biopores, the standard deviation of the residuals term was not high compared to the standard deviations of the individual observers and the site effects. Therefore we can conclude that the individual observers made consistent observations and they agreed with each other.

3.3.1.7 Depth of soil compaction
The observed depth of soil compaction was affected by observer type (P=0.04), individual observers (P=0.04) and site (P=0.00, Table 3.2). Soil scientists recorded the soil compaction depths on average 8 cm shallower than farmers did. Subjectivity by observer type (farmer or soil scientist) and individual observers was high. However, as

the largest part of the variation was explained by the site effect, this means that given their background being farmer or soil scientist, each observer was consistent over the sites and was able to detect relative differences between the sites.

The depth of a compacted layer was only reported if compaction was present; therefore there were fewer observations for all sites than for other visual observations (Figure 3.3T). This especially resulted in fewer observations at Site 3 (peat; five in total), Site 4 and 5 (sand; both three observations at each site) than for Site 1 and 2 (clay).

3.3.1.8 Implications for the applicability of quantitative visual observations

For all quantitative visual soil observations, except for root count, differences in soil quality characteristics between sites were detected, even though the total number of observations was small and individuals had substantial systematic and random observation errors (Table 3.2). Also, there was always some within-site spatial variation. The fact that site effect was significant for most visual observations indicates that the total error (i.e., the combined effect of random errors by observers and within-site variation) was small enough to be able to detect differences between sites. As such, it can be concluded that a group of observers is able to detect systematic differences between sites. If one is interested in the absolute soil quality at a particular site, then the average value from a VSE observation from a group of observers would produce an accurate representation (Ball et al., 2007; Klimeš, 2003). To increase agreement between observers, those observers could cross-check their findings regularly in order to become more consistent in the way of observing (Guimarães et al., 2011; Ball et al., 2015). For sites having more similar conditions it could be more difficult to detect differences in soil quality characteristics, as subjectivity errors will become more evident (Ball et al. 2007). In that case more observers at each site are needed to visually describe soil quality characteristics (Ball et al. 2007).

For all quantitative visual observations, except for the assessment of compaction depth, the observers' background (farmer or soil scientist) did not influence the quantitative visual observations (Table 3.2). This indicates that quantitative visual observations can be used by farmers or soil scientists without correcting for subjectivity.

Individual observers, however, were subjective. Sometimes individual observers had systematically higher or lower estimates than the group mean (systematic error). This hinders comparison of VSE results across individuals and sites. If one is interested in relative differences of soil quality in space (between sites) or in time (for a given site), however, one single observer could carry out VSE. This holds for the observations of the fraction of large soil structural elements, earthworms, gley mottles and the depth

of soil compaction. Observers found it difficult to break open the soil when the soil was very dry and hard. Avoiding too wet and too dry soil conditions may therefore increase reproducibility. The observers also found it difficult to judge when to stop breaking open the soil and to move on to the quantification of the fraction of large soil structural elements. For root count, individual observers always had estimates having the same order of magnitude, irrespectively the site conditions (random error). Observers especially found it difficult to count roots when many roots were present or when the roots were very fine. Therefore care should be taken when counting root on peat and sand. For the observation of grass cover and biopore count, where observers tend to be inconsistent but differences between sites were detected, several replicates at a given site may increase the robustness of the estimate, as previously discussed by Shepherd (2009). Also it is expected that more training will increase the reproducibility.

3.3.2 Correlation of visual observations with standard measurements
Detailed information about the relation between visual observations and measured properties is provided in Appendix 3.1.

3.3.2.1 Grass cover
The visual assessment of grass cover was strongly and significantly correlated with quantified grass cover through digital image analysis for all soil types together (Spearman's ρ=0.76; P=0.00) and when the data were separated by soil type (sand: Spearman's ρ=0.67 and P=0.02; peat: Spearman's ρ=0.77 and P=0.04; clay: Spearman's ρ=0.75 and P=0.03, Figure 3.5A). However, grass cover values assessed by digital image analysis exceeded those determined by visual observation. For observations close to 100% grass cover, the visual observations corresponded best with the grass cover by
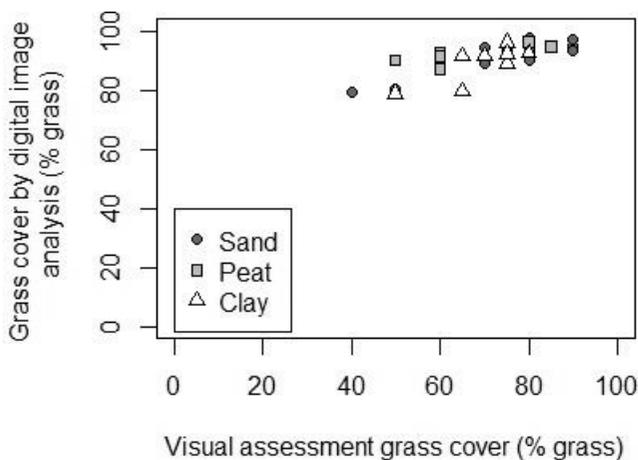
Figure 3.4. Relation between visual assessment of grass cover and grass cover measured by digital image classification.

digital image analysis, more than for lower grass cover values (Figure 3.4). This was also found by Sykes (1983) and Killourhy et al. (2016), who concluded that observed values close to 0 or 100% can be visually estimated more precisely than values around 50% cover.

3.3.2.2 Biopore count

The correlation between biopore count and bulk density was poor, regardless of whether data were separated by soil type (Spearman's ρ between -0.05 and 0.35, P>0.05, Figure 3.5B). This does not mean that the abundance of biopores is not a valuable property to assess in the field, as previous studies found that biopores are associated with water and air flow in soil (Shipitalo et al., 2000), and together with bulk density it is associated with root growth (Gaiser et al., 2013). Głab and Kulig (2008) found a relationship between biopore number and bulk density in a mulched reduced tillage system on luvic chernozem planted with wheat, but they concluded that biopores abundance was influenced by tillage effects and not by earthworm activity. This relationship between bulk density and biopore abundance therefore cannot be used for our sites, which were unploughed grasslands. Instead of using bulk
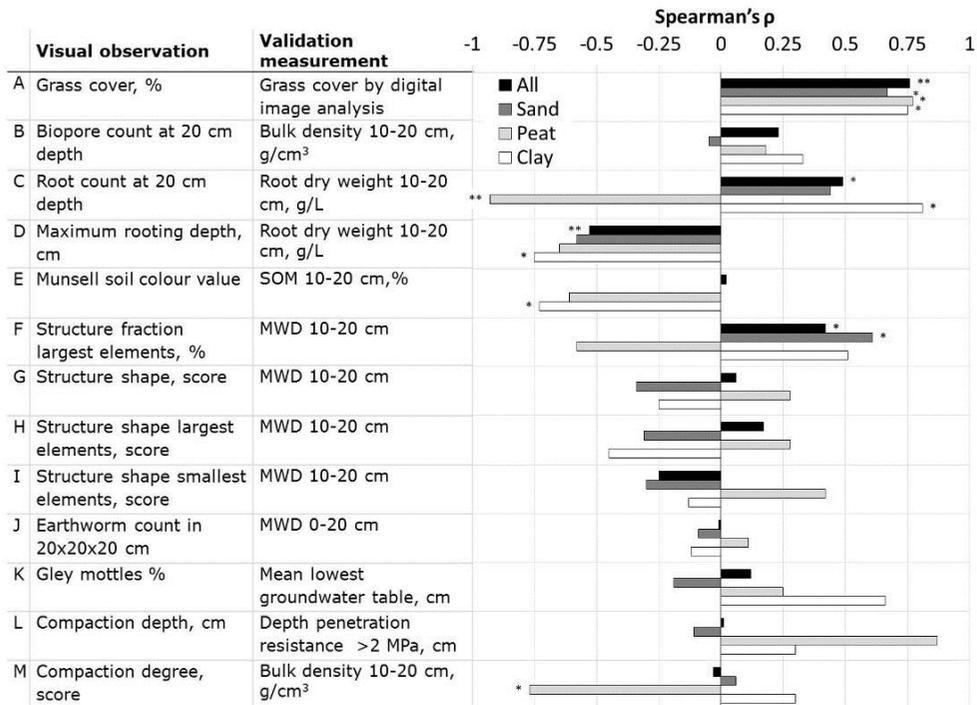
Figure 3.5. Spearmans' ρ correlation coefficients indicate the correlation between visual observations and standard field or laboratory measurements. Values close to 0 indicate poor correlations. Levels of statistical significance: **P<0.01, *P<0.05. MWD: mean weight diameter.

74

density to correlate with the abundance of biopores, medical X-ray computed tomography of a core sample (Katuwal et al., 2015) or digital image analyses of a soil surface could be used (Nakamoto, 1997), but the biopore classification in those methods are time consuming. Nevertheless, we experienced difficulty in the field assessment of biopore count, especially when the surface of the soil block was crumbly, or in the case of dry sand where biopores could easily collapse (the same caused uncertainty in the reproducibility study; see Section 3.3.2.2). These difficulties could have caused an underestimation of the biopore count.

3.3.2.3 Root count

Although root count was significantly correlated to root dry weight for all soil types together (Spearman's ρ=0.49 and P=0.01), separating the data by soil type showed a stronger correlation for clay (Spearman's ρ=0.81 and P=0.01, Figure 3.5C and 3.6). Separating the data by soil type also showed an insignificant correlation between root count and root dry weight for sand (P=0.17). While for sand and clay an expected positive correlation between root count and root dry weight was found, for peat surprisingly a negative correlation was found that was statistically significant (P=0.00). We could not explain why this trend was negative, but there were several factors that made the root count difficult. First, the dense root system of grass species on peat sites (Figure 3.7A) made counting of roots in the field challenging. Also, at the cut surface loose sand grains with equal or larger size than fine roots reduced the visibility of finer roots (Figure 3.7B), which additionally may have obscured them from being counted and thereby may have both induced errors in the root count in field.
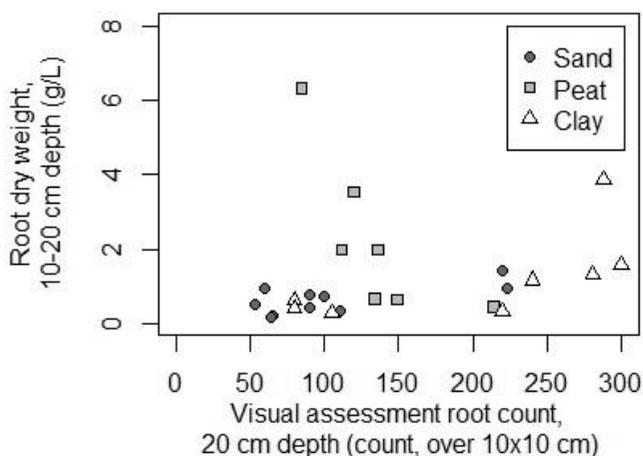
Figure 3.6. Relation between visual assessment of root count and root dry weight.
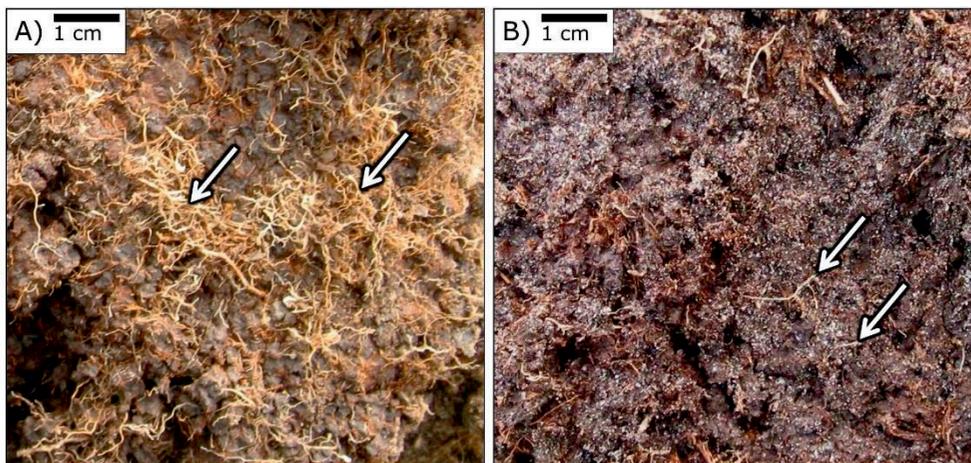
Figure 3.7. Close-ups from the cut surfaces of soil blocks extracted from a peat site, at 20 cm depth. True size of both photos ca. 7×7 cm. A: It was difficult to count roots when the root system was dense. Arrows indicate dense root systems. B: Sand grains are visible as white dots and were sometimes bigger than fine roots. Arrows indicate roots that are bigger than the white sand grains.

Given the fact that root count was not reproducible (Section 3.3.1.3) and root count showed poor correlations with standard measurements for peat and sand, one must be careful when counting roots on peat and sand.

3.3.2.4 Maximum rooting depth

The correlation between maximum rooting depth and root dry weight between 10 and 20 cm depth was significant, for all soil types together (Spearman's ρ=-0.53; P=0.01). When the data were separated by soil type, the maximum rooting depth was only correlated with root dry weight for clay (Spearman's ρ=-0.75; P=0.03); and not for sand (P=0.10) and peat (P=0.11, Figure 3.5D and 3.8A).

3.3.2.5 Soil colour value

The correlation between Munsell soil colour value and soil organic matter (SOM) content was strong and significant for clay (Spearman's ρ=-0.73; P=0.04), but not for sand, peat, and for all soil types together (Figure 3.5E). In the case of clay (6-22% SOM) the observations had a soil colour value between 2 and 4; while in the case of sand (4-17% SOM) and peat (8-33% SOM) all observations had a soil colour value of 2 (the darkest chip possible) except for one observation with a soil colour value of 3.5 on peat (Figure 3.8B). This suggests that while Munsell soil colour value may be a useful quantitative visual observation to estimate soil organic matter content (Wills et al., 2007), this proxy is only effective in regions with a distinct variation in soil darkness (colour value) within a given soil type. Sonneveld et al. (2014) also correlated Munsell
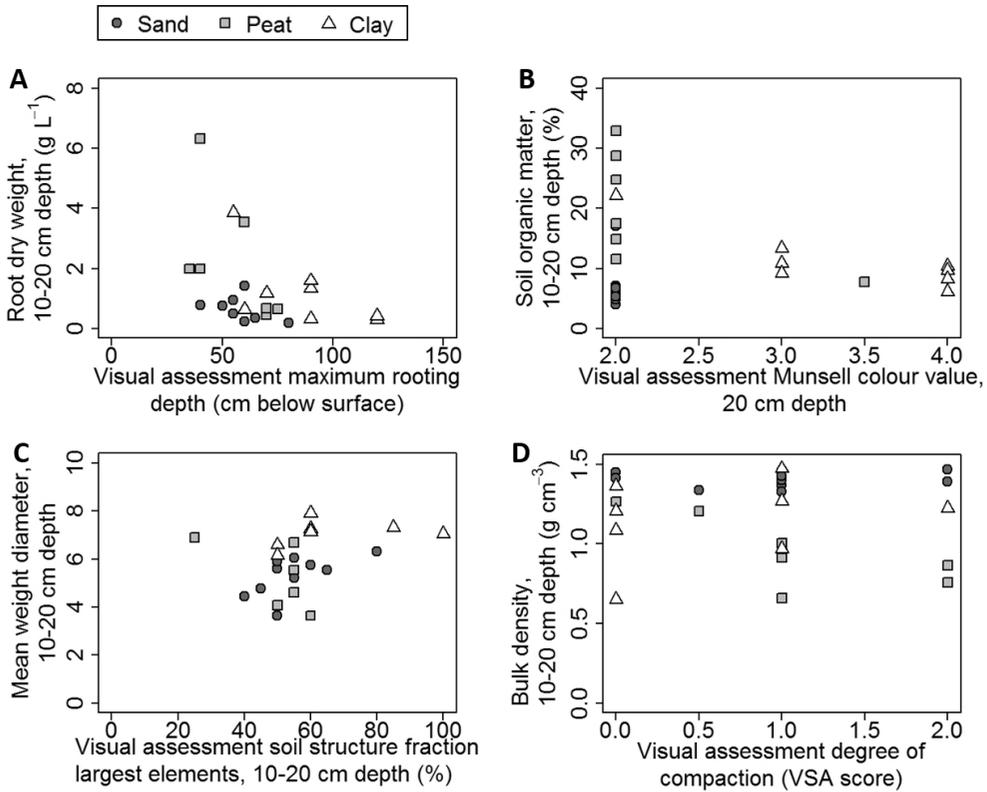
Figure 3.8. Relation between maximum rooting depth and root dry weight (A); between Munsell soil colour value and soil organic matter content (B); between fraction largest soil structural elements and mean weight diameter (C); and between the degree of soil compaction and bulk density (D).

soil colour value with soil organic matter content for clay soils but they found a lower correlation coefficient (Spearman's ρ=0.26) than in the present study (Spearman's ρ=-0.73), as most of their observations (n=65 out of 71) received a soil quality score of 2 (12% SOM) and the other observations a soil quality score of 1 (10.1% SOM). The high SOM content in sand (17%) was expected. Large areas of the cover sands were covered with peat before reclamation. Peat was excavated but remnants were ploughed and mixed in the topsoil after reclamation, resulting in high SOM contents and dark soil colours.

3.3.2.6 Fraction largest soil structural elements
We found a significant correlation between the fraction of largest soil structural elements and the mean weight diameter for all soil types together (Spearman's ρ=0.42, P=0.04) as well as for sand (Spearman's ρ=0.61 and P=0.046, Figure 3.5F and 3.8C). We faced difficulty in breaking open the soil, because the clay soils were generally dry and hard, and the peat soils were generally moist. This resulted in the loss of two data

points where breaking open the clods was impossible. The reduced number of observations likely negatively affected the correlation of this property with mean weight diameter for clay and peat soils. Yet, while Sonneveld et al. (2014) found a significant correlation between VSA's soil structure assessment and mean weight diameter on clay soils, we did not find the same for the fraction largest soil structural elements on clay soils in the present study. A potential explanation for this difference may be that Sonneveld et al. (2014) assessed soil when it was not too wet and not too dry, and therefore it is likely that they did not face problems with breaking up the soil, in contrast to our study. Also Sonneveld et al. (2014) followed VSA, which uses a combined soil quality score for the fraction largest soil structural elements and the assessment of the shape of the soil structural elements (angular or granular), whereas we separated those properties. Another explanation could be that in the case of a limited number of score classes (three to five in VSA), the use of soil quality scores can reduce variation in observations compared to the situation when using quantitative data, and therefore the use of soil quality scores (as done in Sonneveld et al., 2014) could lead to better fits compared to when using quantitative data.

3.3.2.7 Shape of soil structural elements

The correlation between the shape of soil structural elements and mean weight diameter was poor, regardless of whether data were separated by soil type (Spearman's ρ between -0.45 and +0.07, P>0.05, Figure 3.5G). Similar patterns were found by Pulido Moncada et al. (2014) who also found insignificant correlations between the shape of soil structural elements and the mean weight diameter for sandy loam and silt loam soils. In the present study the correlation between the score that represented the soil structural elements and the mean weight diameter did not improve when distinction was made between the largest and the smallest soil structural elements (Figure 3.5H and I).

3.3.2.8 Earthworm count

The correlation between earthworm count and the mean weight diameter of the top 20 cm of soil was poor, regardless of whether data were separated by soil type (Spearman's ρ between -0.09 and -0.12, P>0.05, Figure 3.5J). Using soil quality scores instead of exact earthworm numbers, Sonneveld et al., (2014) found a significant correlation between earthworms and mean weight diameter on peat soils (histosols, Spearman's ρ=0.90, P<0.01), but not on clay soils (fluvisols, Spearman's ρ=0.16, P>0.05), concluding that earthworms contributed to macro-aggregation on peat soils. Mueller et al. (2009) found an insignificant relationship between earthworm count and several visual soil structure assessments (e.g. Peerlkamp score and VSA) on sandy loam to loamy sand. A potential explanation for the poor correlation between earthworm count and mean weight diameter (in the present study) or with other soil

structural characteristics (in the case of Mueller et al., 2009) could be that the appearance of earthworms is highly dependent on local soil moisture conditions (Curry, 2004; Mueller et al., 2009). In our case, especially the clay sites and some of the sand sites were dry, which could have resulted in lower earthworm numbers in the topsoil. This demonstrates the importance of applying VSE when soil is not too wet and not too dry (Guimarães et al., 2017).

3.3.2.9 Gley mottles

The correlation between surface covered by gley mottles and the mean lowest groundwater table was poor, regardless of whether data were separated by soil type (Spearman's ρ between 0.12 and 0.66, P>0.05, Figure 3.5K). The reason could be that the mean lowest groundwater table classes have wide depth ranges where the groundwater table can be found (the widest range is 50 cm, Figure A.3.9F, Van der Sluijs and De Gruijter, 1985). Also, the mean lowest groundwater tables were available at a much coarser spatial scale (each delineated area on a 1.50.000 soil map is bigger than 500 m²) than the visual observations made at the point level. Hence spatial variation of groundwater tables within and between fields is not visible at the 1:50,000 soil map that was used, and may have resulted in the poor correlation between gley mottles and the soil map-derived mean lowest groundwater tables.

3.3.2.10   Depth of soil compaction

The correlation between the depth of the layer where soil compaction visually started and the depth where penetration resistance exceeded 2 MPa was poor, regardless of whether data were separated by soil type (Spearman's ρ between 0.01 and 0.87, P>0.05, Figure 3.5L and Figure A.3.9G). At the clay sites we mainly used the presence of roots to identify compacted layers, to avoid errors in the identification of compacted layers as a result of varying soil moisture contents. Yet, penetration resistance measured by a penetrometer is influenced by soil moisture content as well as by soil texture and soil organic matter content (Costantini, 1996). Especially at sites with a dry and hard topsoil, penetrometer values tend to overestimate resistance to root growth and thus the depth of soil compaction (Bengough et al., 2001). While some sites were dry and dense with high penetration resistance, on other sites the penetration resistance did not exceed 2 MPa, resulting in a reduction of observation pairs from a total of 26 to 21. For peat the total number of sites remaining was three, for which the correlation between penetration resistance and the visual depth at which compaction started was high (Spearman's ρ=0.87) yet insignificant (P=0.33). For sand and peat we had six respectively five pairs of observation left. The poor correlations do not necessarily mean that the depth of soil compaction is an invalid visual observation, but the site conditions were too dry to be assessed with a penetrometer.

3.3.2.11 Degree of soil compaction

We found a strong and significant correlation between the degree of soil compaction and bulk density for peat (Spearman's ρ=-0.77, P=0.04), but not for sand (P=0.85) or clay (P=0.47), nor for all soil types together (P=0.90, Figure 3.5M and 3.8D). At the time of field assessment, especially the clay sites and some of the sand sites were dry (Table 3.6) and therefore very hard. Compaction in the field due to dry conditions could be confused with compaction due to human impact (Guimarães et al., 2017), and therefore the visual observations of the degree of soil compaction for clay and sand sites were invalid. Nevertheless it should be noted that bulk density is soil type dependent (Da Silva et al., 1997), which was also clear in the present study where bulk density was significantly higher for sand (1.37 g cm^{-3}) than for clay (1.15 g cm^{-3}) and peat (0.95 g cm^{-3}), at P=0.05 (Table 3.6). This indicates the necessity of separating the data by soil type when the degree of soil compaction is correlated with bulk density.

3.3.2.12 Implications for the applicability of visual observations

We found statistically significant correlations for grass cover, root count, maximum rooting depth and the fraction largest soil structural elements with a field or laboratory measurement for all soil types together (Figure 3.5). This suggests that these visual observations are valid and sensitive enough to be applied in the North Friesian Woodlands, as long as a broad range of soil conditions can be expected. If visual observations are carried out on a single soil type, then grass cover, root count on clay, maximum rooting depth on clay, Munsell soil colour value on clay, fraction largest soil structural elements on sand and the degree of soil compaction on peat could be used (Figure 3.5). On the other hand, we found insignificant correlations between the visual observations and the standard field or laboratory measurements for biopore count, soil structure shape, earthworms count, gley mottles and compaction depth. Although some visual observations were affected by dry soil conditions, the poor correlations do not necessarily mean that the visual observations were invalid as some standard field or laboratory measurements were not closely related to the visual properties. Therefore these visual observations should not be omitted from VSE, but require further validation research. From the abovementioned correlations between visual soil observations and standard field or laboratory measurements, we can conclude that the correlations were affected by soil type. Therefore VSE guidelines should be soil type specific as well.

To improve the accuracy of the visual observations made, too wet or too dry field conditions should be avoided when applying VSE because it increases the observation error. We experienced that the soil structure assessment was influenced by dry conditions on Site 1 and 2 (clay) and wet conditions on Site 3 (peat). The same dry conditions caused difficulties in the identification of compacted soil layers. Guimarães

et al. (2017) concluded that the range of optimal soil water contents for VSE should be studied in future research, but until then the friability of the soil (ease of soil to crumble) could be used as indicator for optimal water contents. For this, the optimal range of soil water contents could be approached by determining the soils' lower plastic limit (wet end) and the energy it costs to fragment soil (dry end).

3.4 Conclusions

Analysis of reproducibility of VSE soil observations indicated that subjectivity due to the observers' background (farmers and soil scientists) was only significant in the assessment of the depth of soil compaction. We also showed that relative differences between sites can be detected by a single observer, for all properties apart from grass cover, biopore and root count. However, accurate evaluation of absolute soil quality characteristics at a particular site requires the mean observed value from a group of observers.

The correlation between visual observations and standard field or laboratory measurements indicated that grass cover, root count, maximum rooting depth and the fraction largest soil structural elements showed significant correlations, even when data were not separated by soil type (sand, peat or clay). When separating data by soil type some additional visual observations showed a significant correlation with standard field or laboratory measurements.

Soil moisture conditions affected the reproducibility of the visual soil structure assessment, as well as the correlation of visual soil structure assessment and soil compaction assessment with laboratory and field measurements.

The correlation between visual observations and field or laboratory measurements were often soil type dependent, indicating that the procedures for visually evaluating soil quality should also be soil type dependent.

Appendix 3.1
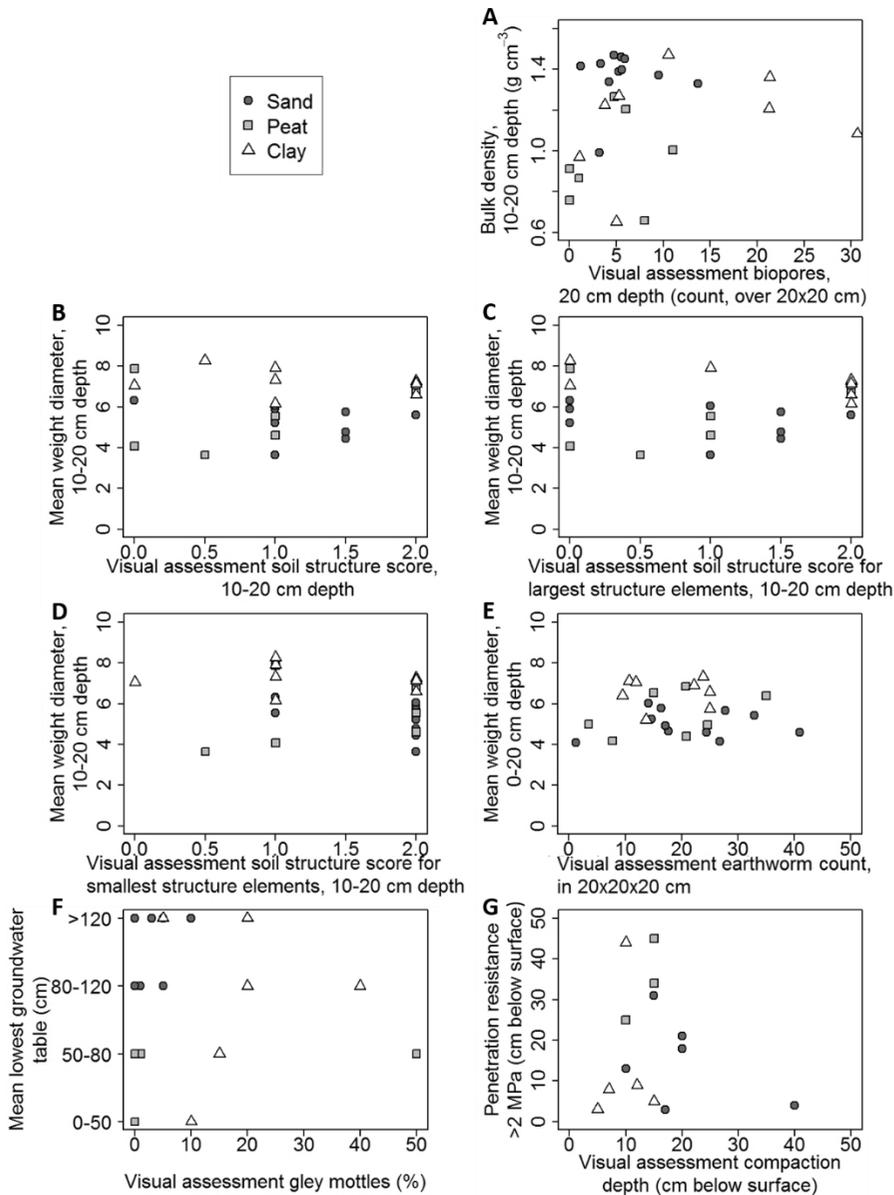
Figure A.3.9. Relation between biopore count and bulk density (A); between shape of soil structural elements (0=angular, 1=sub-angular, 2=granular) and mean weight diameter (B); between shape of the largest soil structural elements (0=angular, 1=sub-angular, 2= granular) and mean weight diameter (C); between shape of the smallest soil structural elements (0=angular, 1=sub-angular, 2= granular) and mean weight diameter (D); between earthworm count and mean weight diameter (E); between gley mottles and mean lowest groundwater table indicated at the soil map (F); and between soil compaction depth and depth where penetration resistance >2 MPa (G).

Table A.3.3. Soil characteristics at the sites. Average SOM content over the layers 0-10, 10-20 and 20-30 cm; number of years since last grassland renewal (due to land use conversion, plowing or reseeding); visual observations of texture according to FAO (2006); soil groups and Dutch soil classification according to the Dutch soil map.

Soil characteristics	Sand		Peat		Clay	
SOM 0-30 cm (%)	7.3	(3.5)	19.0	(7.2)	11.2	(2.9)
Number of years since grassland renewal	26	(20)	27	(33)	41	(50)
Visual observation texture (at 20 cm depth)	Fine sand (n=9) Sand (n=1) Sandy loam (n=1)		Fine sand in material rich in organic matter (n=6) Sandy clay loam (n=1)		Heavy clay (n=4) Clay (n=3) Sandy clay loam (n=1)	
Soil groups	Ordinary hydropodzol soils (n=7) Peaty earth soils (n=2) Sandy hydroearth soils (n=1) Peaty podzol soils (n=1)		Ordinary raw peat soils (n=3) Peaty earth soil (n=2) Peaty podzol soils (n=1) Clayey hydrovague soils (n=1)		Clayey hydrovague soils (n=5) Ordinary raw peat soils (n=2) Ordinary hydropodzol soils (n=1)	
Dutch soil classification	'Laar' podzol soil (n=5) 'Veld' podzol soil (n=2) Peaty earth soil (n=2) 'Goor' earth soil (n=1) 'Moer' podzol soil (n=1)		'Vlier' peat soil (n=3) Peaty earth soil (n=2) 'Moer' podzol soil (n=1) 'Polder' vague soil (n=1)		'Polder' vague soil (n=4) 'Drecht' vague soil (n=1) 'Waard' peat soil (n=2) 'Veld' podzol soil (n=1)	

Deviating number of observations years since grassland renewal: sand: n=10, peat: n=5, clay: n=7.

Chapter 3

Table A.3.4. Means and standard deviations (between brackets) for quantitative visual observations, for all observations together, and for sand (n=11), peat (n=7), and clay (n=8).

Visual observations	Unit	All observations		Sand		Peat		Clay	
Grass cover	%	71	(14)	74	(16)	69	(14)	69	(9)
Biopores	Count, over 20x20 cm, at 20 cm depth	7	(7)	6	(3)	4	(4)	12	(11)
Roots	Count, over 10x10 cm, at 20 cm depth	142	(78)	104	(61)	136	(40)	199	(96)
Maximum rooting depth	cm	68	(22)	59	(12)	56	(17)	87	(25)
Soil structure fraction of largest elements	%	57	(15)	55	(11)	50	(13)	66	(19)
Earthworms	Count, in 20x20x20 cm	19	(9)	21	(11)	18	(11)	18	(7)
Gley mottles	%	9	(13)	3	(3)	7	(19)	18	(10)
Compaction depth	cm below surface	14	(8)	18	(10)	11	(4)	10	(4)

Deviating number of observations: maximum rooting depth sand: n=9; soil structure fraction of largest elements peat: n=6, clay: n=7; depth of soil compaction sand: n=9, peat: n=5, clay: n=6.

Table A.3.5. Descriptive statistics for visual observations having ordinal values: minimum, maximum and median value, and I², for all observations together (n=26), and for sand (n=11), peat (n=7), and clay (n=8).

Visual observation	All observations				Sand				Peat				Clay			
	Min	Max	Median	I²	Min	Max	Median	I²	Min	Max	Median	I²	Min	Max	Median	I²
Munsell soil colour (value)	2	4	2	0.7	2	2	2	1.0	2	3.5	2	0.9	2	4	3.5	0.7
Soil structure shape (score)	0	2	1	0.6	0	2	1	0.5	0	2	1	0.3	0	2	1	0.5
Soil structure shape largest elements (score)	0	2	1	0.5	0	2	1	0.3	0	2	1	0.3	0	2	2	0.6
Soil structure shape smallest Elements (score)	0	2	2	0.7	1	2	2	0.7	0.5	2	2	0.4	0	2	1	0.5
Degree soil compaction (score)	0	2	1	0.6	0	2	1	0.4	0	2	1	0.3	0	2	0.5	0.5

85

Table A.3.6. Descriptive statistics for standard field or laboratory measurements: mean value and standard deviation, for all observations together, and for sand (n=11), peat (n=7), and clay (n=8).

Measured	Unit	All observations		Sand		Peat		Clay	
Grass cover by digital image analysis	% grass	90.9	(5.6)	91.2	(6.2)	92.4	(3.1)	89.0	(6.3)
Bulk density 10-20 cm	g cm^{-3}	1.19	(0.26)	1.37	(0.13)	0.95	(0.22)	1.15	(0.26)
Volumetric water content 0-20 cm	cm^3 cm^{-3}	0.36	(0.08)	0.30	(0.06)	0.42	(0.06)	0.39	(0.06)
Root dry weight 10-20 cm	g L^{-1}	1.1	(1.2)	0.6	(0.6)	2.1	(2.0)	1.0	(0.7)
SOM 10-20 cm	% mass	11.7	(7.8)	6.9	(3.5)	19.8	(9.3)	11.2	(4.9)
MWD 0-10 cm	-	5.2	(1.0)	4.7	(0.8)	5.3	(0.7)	5.9	(1.2)
MWD 10-20 cm	-	6.0	(1.3)	5.3	(0.8)	5.6	(1.6)	7.2	(0.7)
Mean lowest groundwater table	cm below surface	80-120 [1]		>120 [1]		50-80 [1]		80-120 [1]	
Penetration resistance: depth at which 2 MPa is exceeded	cm	15.3	(13.5)	14.3	(8.8)	24.2	(16.2)	10.3	(15.5)

[1] Median values, as the data type is ordinal. SOM: soil organic matter, MWD: mean weight diameter

Chapter 4

Quantitative visual soil examination to evaluate soil functions on dairy farms

Highlights

- A set of quantitative visual soil observations was collected at sand and clay soils.
- Quantitative visual observations associated with crop growth and water quality.
- Those associations were influenced by soil texture.
- Visual soil evaluation can be used to assess various soil function indicators.

Abstract

With the growing pressure on agricultural land, it is important to assess and evaluate soil quality and soil functions to ensure environmental sustainability of food production. Visual soil evaluation provides an easily obtainable means to assess soil quality and several soil functions, but it is often only used to assess soil quality in relation to crop growth. The aim of the present study was to assess the association between a set of quantitative visual observations and several soil function indicators, and to include soil texture effects. A broad set of quantitative visual observations was collected on 25 dairy-farmed sites in the Netherlands. We used laboratory-measured soil physical properties and an ecohydrological model to determine the following soil function indicators: plant available water and water storage capacity; yield gap, oxygen and drought stress for a wet year (2001), a dry year (2003), and a 'normal' year (2016). We also used measured nitrate and phosphate concentrations in drain and groundwater. Stepwise linear regression models showed that, except for drought stress in a dry year, soil function indicators correlated with a set of quantitative visual soil observations. Clay content was found to influence the associations between visual observations and soil function indicators, which shows the importance of evaluating soil quality in relation to the soils' potential to function, which is site-specific. We suggest, therefore, to include soil texture in future visual soil evaluations. This study showed that visual soil evaluations can assess several soil functions at the same time, which contributes to the evaluation of soil quality in relation to environmental performance of agricultural land.

4.1 Introduction

Healthy soils form the basis for a sustainable production of food. The assessment and evaluation of soil quality, therefore, is of utmost importance. Visual soil evaluations show great potential in assessing soil quality and various soil functions (Shepherd, 2009), but studies that associate a set of visual observations with various soil functions are scarce. Such studies, however, could give insight into the relative importance of each visual soil observation in relation to each soil function, which could contribute to the design of proper scoring functions to evaluate soil functioning and soil quality.

Soil quality is "the capacity of a soil to function, within managed or natural ecosystem boundaries, to sustain plant and animal productivity, maintain and enhance water and air quality, support human health and habitation" (Karlen et al., 1997). Soils in agroecosystems have many functions, the main functions are (CEC, 2006): crop production; storing, filtering, and transforming nutrients and water; maintaining biodiversity; serving as carbon pool, and serving as archive for geological and cultural heritage. Because of the multi-functionality of soil, it is important to assess several soil functions at the same time. This is because the delivery of one soil function may go at the expense of another soil function. In agricultural soils, for instance, the focus is often on crop production, but trade-offs with other soil function can be overlooked (Schröder et al., 2016). Therefore, in the present study we focussed on crop production (selected indicators are modelled crop yield under wet and dry conditions, and plant available water) and storing, filtering, and transforming nutrients and water (selected indicators are water storage capacity and nitrate and phosphate concentrations in ground- and drain water, Table 4.1).

The term 'soil function' may suggest that the five soil functions (Table 4.1) only depend on soil itself. This is true according to the definition: "the potential of soil to function...". But the quantified soil functions also depend on aboveground factors, such as climate and land management (Bünemann et al., 2018; Schröder et al., 2016, Figure 4.1). For instance, the soil function 'crop production' is a function of soil biological, physical and chemical properties (Gregorich et al., 1994), climatic conditions, and management factors (e.g. nutrient availability, crop characteristics) (Mueller et al., 2013). Visual soil evaluation considers mostly soil physical properties and a few biological properties. It can be used to identify potential limiting soil physical (and a few biological) properties for soil functioning (Guimarães et al., 2017), but it also depends on the other factors if these factors will be limiting soil functioning.

Chapter 4

Table 4.1. Quantification of soil functions. White boxes are the focus in the present paper, the soil functions in the grey boxes are not considered.

Soil functions	Quantified by
Crop production	Yield gap, as function of actual and potential crop (dry matter) yield
Storing, filtering, and transforming nutrients and water	Crop sensitivity to dry and wet periods: plant transpiration 'gap', as function of actual and potential plant transpiration
Maintaining biodiversity	Plant available water
Serving as carbon pool	Water storage capacity
Serving as archive for geological and cultural heritage	Nitrate and phosphate concentrations in ground- and drain water
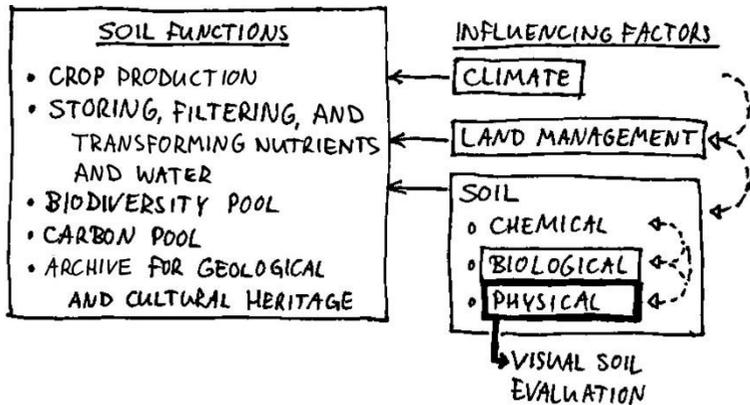
Figure 4.1. Factors that influence soil functions under a given crop. Note that those factors interact with each other (dashed arrows).

There are two main approaches for soil quality assessment. In the first approach, soil functioning is assessed by describing soil characteristics of various soil aspects (Gregorich et al., 1994). This could be, for example, routine soil chemical analyses or visual soil evaluation of (mainly) soil physical properties. In the second approach, soil functioning is assessed in relation to the soils' full potential, given the crop type, climate and parent material (Karlen et al., 1997). In the present paper we adopted the second approach where possible, because it allows a fair comparison of soil functions of agricultural fields across soil types and weather patterns, which is important for designing scoring functions. Bünemann et al. (2018) reviewed soil quality monitoring programs that are used internationally. Visual soil evaluation showed potential to assess soil functioning (Ball et al., 2017; Shepherd, 2009).

Many studies report visual soil evaluations (VSEs) to correspond with standard or laboratory measured soil physical properties (e.g. Ball et al., 2007; Johannes et al., 2017; McKenzie, 2001; Newell-Price et al., 2013; Sonneveld et al., 2014). But fewer studies reported the association between visual soil observations and soil functions, and these studies have in common that they used single correlations between visual assessment or soil quality index and 'measured' soil properties or soil functions. First of all, Mueller et al. (2013) reported a significant correlation between visual evaluation of soil structure and crop production, while other studies showed insignificant correlations between visual evaluation of soil structure and crop production (Giarola et al., 2013; Mueller et al., 2009). Van Groenigen et al. (2014) showed that the presence of earthworms can have a positive effect on crop growth, but they also showed that this relationship is dependent on many biophysical factors. Furthermore, Pulido Moncada et al. (2014) reported that the soil quality index of Visual Soil Assessment and the Visual Evaluation of Soil Structure significantly correlated with plant available water. Lastly, visual evaluation of soil structure was insignificantly correlated with CO_2 and

N$_2$O emission in clayey oxisols planted with soybean, although the soil structure in interrows was poorer than within rows of soybeans (da Silva et al., 2014). These studies show the potential of VSE in assessing soil functioning, but as soil functioning depends on many factors, it is important to include various soil quality indicators that preferably describe soil physical, chemical and biological properties (Karlen et al., 2001).

It is furthermore possible that associations between visual soil observations and soil functions are soil type dependent, because both soil functioning (McBratney et al., 2014) and visual soil observations are soil type dependent (Johannes et al., 2017; Newell-Price et al., 2013; Van Leeuwen et al., 2018), and because of possible interactions with soil texture (Bünemann et al., 2018). Taking into account interactions with soil texture could yield surprising results. For example, for most soils soil compaction is unfavourable for crop growth (e.g. Peigné et al., 2013). In coarse-textured soils, however, some soil compaction could increase water storage capacity and capillary rise of groundwater to the plant roots, thereby having a positive effect on crop growth: Douglas (1997) showed that the second harvest (in a dry growing season) had higher yields on compacted soil (not visually assessed) than on non-compacted soil, because the water reserves for crop growth were probably higher in compacted soils than on non-compacted soils. The effect of soil compaction on crop growth is affected by the available water and soil texture (Batey, 2009). These studies show that it is important to consider soil texture in VSE when assessing soil functioning.

The aim of this work is to assess the associations between a set of quantitative visual soil observations (Shepherd, 2009; Van Leeuwen et al., 2018) and several soil functions, which are crop production and storing, filtering, and transforming nutrients and water. The influence of soil texture on these associations is assessed. This contributes to a better understanding of visual soil evaluations in relation to soil functions, which may contribute to sustainable food production.

4.2 Materials and methods

4.2.1 Experimental design

We selected five Dutch dairy farms that were located on sand and clay soils. In 2016 (September 12 to October 5), five quantitative visual soil assessments (Van Leeuwen et al., 2018) were carried out on each farm, and additional soil samples were taken for soil physical analyses (Section 4.2.3). The indicators used as a proxy for crop production were plant available water, water storage capacity, potential and actual crop yield (and its difference, known as the yield gap), plant oxygen and drought stress. Indicators used as a proxy for 'storing, filtering, transforming nutrients, solutes

and water' were nitrate (NO_3) and phosphate (PO_4) concentrations in ground- and drain water. Plant available water and water storage capacity were determined from soil moisture retention characteristics (Section 4.2.6). Potential and actual crop yield, yield gaps, plant oxygen and drought stress were simulated in the ecohydrological model SWAP (Soil-Water-Atmosphere-Plant, Kroes et al., 2017) coupled to crop growth model WOFOST (World Food Studies, De Wit et al., 2019) (Section 4.2.7), for the years 2001 (wet year), 2003 (dry year), and 2016 ('normal' year). Nitrate and phosphate concentrations in groundwater and drain water were measured by RIVM in 2016 (Rijksinstituut voor Volksgezondheid en Milieu, 2016). The quantified indicators for soil functions were related to quantitative visual observations using stepwise linear regression (Section 4.2.8).

4.2.2 Farm and soil characteristics

Five Dutch dairy farms were selected which were located on sand and clay soils in the Netherlands (Figure 4.2). Those farms were part of the ongoing project 'Cows and Opportunities' (in Dutch: 'Koeien & Kansen'; Oenema et al., 2011). The project monitors nutrient inputs and outputs at field, crop and farm level, and measured soil texture at field level. The same farms were used and described in Van Leeuwen et al. (2019).

Farm 1, 2 and 3 were predominantly located on aeolian cover sands. Dominant soil series for these farms were well-drained hydropodzols ('*veld*' podzol soils), and thick black '*enk*' earth soils (Figure 4.2, Alterra, 2006; De Bakker and Schelling, 1989). Farm 1 was an experimental farm, which had higher nutrient use efficiencies and lower nutrient losses to the environment than typical commercial farms (Oenema et al., 2001). Farm 2 was located on the top of an ice pushed ridge consisting of Pleistocene (sand and gravel) and Tertiary (clay, loam and sand) formations. On the top of the ridge some glacial till of the ground moraine was found. Locally the ice pushed materials were covered with aeolian cover sands with varying depths up to 2 m thick (Van den Berg and Den Otter, 1993). Loamy or clayey layers in the subsoil can cause stagnation of rainwater. Because these materials were pushed, the subsoil is complex, and within-field soil variation can be large. For example, soils can locally be poorly or well drained as a result of the presence or absence of stagnating layers. Farm 3 had in addition to the sandy soils also clayey to loamy '*polder*' vague soils (Figure 4.2, Alterra, 2006; De Bakker and Schelling, 1989) for the fields that were located close to the river IJssel. Farm 4 was located on marine clay and Farm 5 was located on fluvial clay. Dominant soil series for Farm 4 and 5 were '*polder*' vague soils. Farm 4 was an organic farm, and Farm 5 was the only farm that kept the cows continuously inside.

Climate in the Netherlands is temperate. Meteorological station 'De Bilt' recorded for the period 2000-2016 average monthly temperatures between 3.7 °C and 18.2 °C, and a
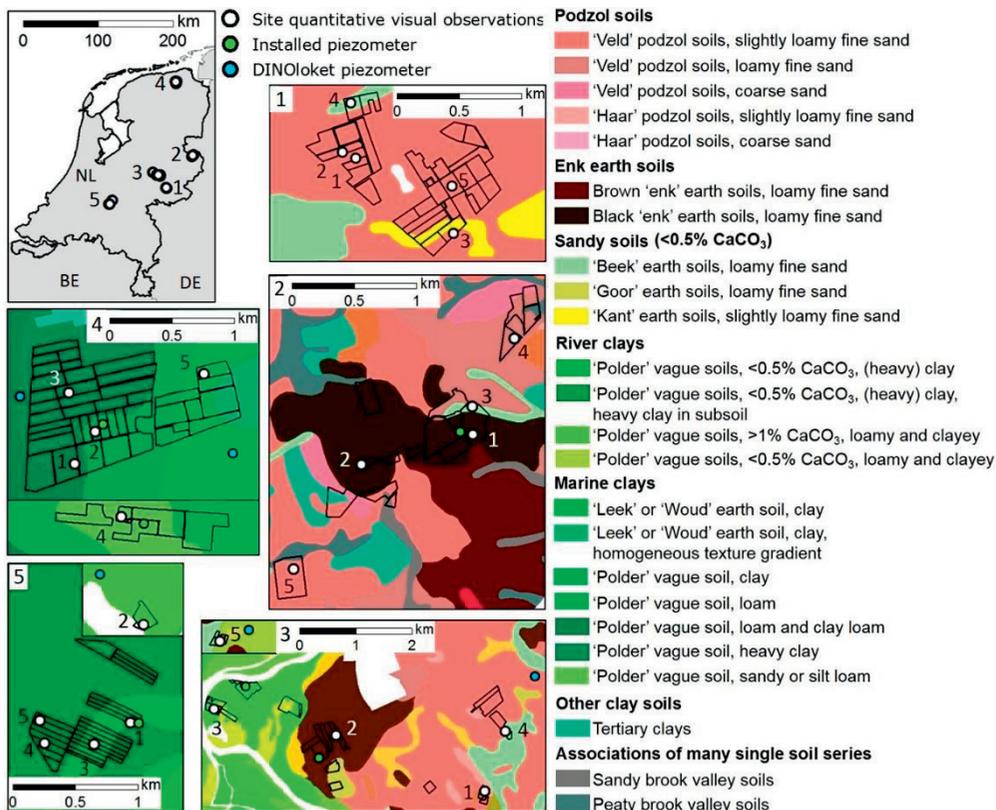
Figure 4.2. Locations of the dairy farms (Farm 1-5), sampled fields, piezometers that we installed, and groundwater table data available through DINOloket (TNO Geologische Dienst Nederland, 2018). The black lines indicate the main fields of the dairy farms, and the underlying colours represent the soil series from the 1:50,000 soil map of the Netherlands (Wageningen Environmental Research (Alterra), 2015).

mean annual precipitation of 862 mm (Royal Netherlands Meteorological Institute, 2018). Please refer to Section 4.2.7.1 for weather details in the years 2001 and 2003.

4.2.3 Soil and water sampling

To select the fields for quantitative visual soil observations and soil sampling, we divided each farm into strata based on the combination of soil series and land use. Soil series were obtained from the 1:50,000 soil map of the Netherlands (Figure 4.2, Alterra, 2006). We only took those fields into account for which crop type (grass or maize) remained unchanged between 2015 and 2016, to exclude effects of recent land use change on soil hydraulic properties. Crop type data were provided by the project 'Cows and Opportunities' and/or by the farmers. Due to the highly time consuming nature of the soil hydrological analyses, the maximum number of fields that could be sampled was 25. For each farm, therefore, the five largest strata were selected for sampling. For

each stratum one field was selected that had a single soil series and an average nitrogen use efficiency compared to other fields within a stratum (nitrogen field balances were provided by 'Cows and Opportunities'). On the 25 fields (20 grass fields and five maize fields), quantitative visual soil observations were done and soil samples were taken. Quantitative visual soil observations were done according to Van Leeuwen et al. (2018), which is based on the Visual Soil Assessment of Shepherd (2009) (Table 4.2). Before digging a pit, grass cover was assessed. Then a pit was dug, and a block of approximately 20x20x20 cm was extracted with a spade. The actual block size was measured after excavation (the number of biopores per surface area, and the number of earthworms per volume could be corrected if the block size was not 20x20x20 cm). On the bottom of the soil block, the number of roots (in a square of 10x10 cm), biopores (on the 20x20 cm surface area) and gley mottles, and the Munsell soil colour value were determined. Then the 10-20 cm deep layer of the soil block was gently crumbled by hand, to assess the fraction largest soil structural elements. The shape of all structural elements was assessed according to Shepherd (2009). The entire soil block was crumbled to count the earthworms. From the soil pit, if soil compaction was present, the depth of the compacted layer, and the degree of soil compaction were determined. Furthermore from the soil pit the depth of approximately 85% of the plant roots, and the depth of the A horizon were assessed. An Edelman soil auger was used to obtain a soil profile up to a depth of 120 cm. From the augered soil profile, the mean highest groundwater table was assessed by assessing the presence of gley mottles. Finally, the augered soil profile was used to assess the maximum rooting depth. If roots were still visible in 110-120 cm depth the maximum rooting depth was set to 130 cm. Next to the set of quantitative visual observations, additional soil characteristics were described and soil samples were taken for use in the SWAP model to quantify several soil function indicators that related to biomass production (Section 4.2.7). Soil horizons were identified and for each horizon soil texture was estimated according to FAO (FAO, 2006) down to 120 cm depth, using the augered soil profile. For analysis of (un)saturated hydraulic conductivity, undisturbed soil cores (10.3 cm diameter, 8 cm high) were collected between 10 and 20 cm depth from the same soil pit where quantitative visual observations were done. The core samples were wrapped in cling film and stored in a cold room (2°C) until they were saturated with water (within 0.5 to 3.5 weeks). For analysis of soil organic matter content, a composite topsoil sample was taken in a W-pattern, consisting of 10 to 15 subsamples (dependent on the field size) from 0-20 cm depth. These samples were stored in a cold room for one night, and analysed the next day by Eurofins Agro (Wageningen, the Netherlands) using near-infrared spectroscopy.

Table 4.2. Quantitative visual soil observations analysed in the field (except for depth of A horizon and mean highest groundwater table): according to Van Leeuwen et al., (2018). If a visual observation was not considered in stepwise regression analysis, the reason is given.

Visual observation	Unit	Considered in stepwise regression analysis
Grass cover	%	No: not observed in maize fields
Root count at 20 cm depth	Count, on a 10x10 cm surface	No: missing observations
Biopore count at 20 cm depth	Count, on a 20x20 cm surface	No: missing observations
Number of gley mottles at 20 cm depth	%	No: mean highest groundwater table (indicated by the shallowest depth where gley mottles appear) is expected to be a more direct indicator for various soil functions.
Munsell soil colour value at 20 cm depth	-	Yes
Soil structure, 10-20 cm depth:		
Fraction largest elements	%	Yes
Shape of structure elements	VSA score[1]	Yes
Earthworm count 0-20 cm depth	Count, in a 20×20×20 cm volume	Yes
Depth where soil compaction starts	cm below surface	Yes
Degree of compaction	VSA score[1]	Yes
Depth of 85% of roots	cm below surface	Yes
Depth of A horizon	cm below surface	Yes
Mean highest groundwater table	cm below surface	Yes
Maximum rooting depth	cm	Yes

[1] According to Shepherd (2009).

Nitrate and phosphate concentrations in groundwater were measured in summer 2016 (Farm 1, 2 and 3) and for drain water measured in winter 2016 (Farm 5). Data was thus available for 14 of the 25 fields, and provided by RIVM (Rijksinstituut voor Volksgezondheid en Milieu, 2016).

4.2.4 Groundwater table monitoring

Minimum and maximum groundwater tables were used to define the drainage condition in the SWAP model (Section 4.2.7). Much groundwater table monitoring data were available for the studied sites in the online database DINOloket (TNO Geologische Dienst Nederland, 2018). These data were included when the following conditions were met if: 1) data was collected within 3 km of one of our 25 fields; 2) data was collected from 2001 to 2003 and/or in the year 2016; 3) the phreatic groundwater table was monitored; and 4) when soil type and topography were comparable with the studied sites. For fields with deep groundwater tables (between 1.5 and 3 m deep, observed from our own measurements or from DINOloket, see Section 4.2.7.5), free drainage was assumed hence no groundwater table measurements were needed in SWAP. For fields where the DINOloket data were not useful and where free drainage could not be assumed, we installed piezometers for additional data collection. On Farm 3 and 4 two piezometers were installed, and on Farm 2 and 5 one piezometer was installed (Figure 4.2, Table A.4.7). These piezometers had depths between 160 and 280 cm and measured the phreatic groundwater table. Groundwater tables were monitored at 15-min intervals using TD-Divers (Eijkelkamp Soil & Water, Giesbeek, The Netherlands) that measure the pressure of the water column, corrected for air pressure. The divers were calibrated before being installed in the piezometer. Groundwater monitoring started at the 16th of August 2016, and continued for a period of six to eight months. For more details, please refer to Table A.4.7.

4.2.5 Determination of soil moisture retention and (un)saturated hydraulic conductivity

The undisturbed core samples were gradually saturated with water (1 cm per day, until water reached a level of 1.5 cm below the edge of the core). The saturated hydrological conductivity was determined using the constant head method (Stolte, 1997), after which the samples were weighed to determine the saturated water content. The soil cores were then kept saturated (between two weeks and two months), until application of Wind's evaporation method (Wind, 1968). This method is used to determine the soil hydraulic functions (soil moisture retention characteristic and unsaturated hydraulic conductivity function, see Section 4.2.7.3) in the pressure head range of 0 to -800 cm. The hydraulic heads were measured with four deaerated tensiometers (4 mm diameter; length ceramic cup 3 cm), which were installed at 1, 3, 5 and 7 cm depth. Pressure heads were recorded at 1 to 2-minute intervals.

Simultaneously, the weight of the sample was recorded (Sartorius CPA2202S) to calculate the evaporation rates in time. Measurements stopped when three tensiometers stopped working because of air entry at dry conditions. If three tensiometers failed before reaching a pressure head of -500 cm, then the sample was gradually saturated again and the evaporation method was repeated. In that case, the tensiometers were replaced by larger tensiometers (6 mm diameter; length ceramic cup 5.3 cm) to ensure good contact with the soil. At the end of the experiment the remaining water content was determined by immediately weighing the soil core, and reweighing after oven drying (24 h at 105°C).

4.2.6   Parameter estimation of the soil hydraulic functions

Parameter estimation for the soil hydraulic functions was performed with the Hydrus1D software package (version 4.16.0110, Šimůnek et al., n.d.). Input data were the saturated water content, and the measured evaporation rates and pressure heads at each time step (converted to a 2-hour time interval). Hydrus1D estimated the residual water content, saturated hydraulic conductivity, and shape parameters α and n, by inverse parameter estimation (Wendroth et al., 1993). The estimated parameters were used in the analytical soil hydraulic functions of Mualem-Van Genuchten (see Section 4.2.7.3). The resulting soil moisture retention curve was used to determine the soil water storage capacity and plant available water content. The soil water storage capacity was estimated by subtracting the soil water content at pF=2 (field capacity) from the saturated soil water content. The plant available water content was estimated by subtracting the soil water content at pF=4.2 (plant wilting point) from the water content at pF=2.

4.2.7   SWAP model

The ecohydrological model SWAP (Soil-Water-Atmosphere-Plant) simulates the vertical water flow in a soil column between the soil surface and shallow groundwater or subsoil, in relation to vegetation growth (Kroes et al., 2017). Grass growth is simulated by the crop growth module WOFOST (WOrld FOod Studies, De Wit et al., 2019).

SWAP version 4.0.1 was run, to obtain the soil function indictors potential crop yield, actual crop yield, yield gap, crop oxygen stress under wet conditions and crop drought stress. For all SWAP inputs please refer to Van Leeuwen et al. (2019b, Dataset).

4.2.7.1 Meteorology

Daily mean weather records (radiation, minimum and maximum temperatures, humidity wind speed and precipitation) were obtained from the KNMI weather station and precipitation station that was closest to each respective farm (Table A.4.8). The year 2001 was warm, sunny and wet compared to other years. At weather station

'De Bilt' monthly temperatures ranged from 2.6 °C in January to 18.5 °C in July and August, and 1039 mm precipitation was measured. The year 2003 was warm, sunny and dry. On weather station 'De Bilt' monthly temperatures ranged from 1.8 °C in February to 19.3 °C in August, and 613 mm precipitation was measured. The year 2016 was sunnier and warmer than normal, with monthly temperatures ranging from 4.6 °C to 18.4 °C, and with 838 mm precipitation (Royal Netherlands Meteorological Institute, 2018).

4.2.7.2 Crop section

All fields, including the maize fields, were modelled as grass to allow comparison of crop yields (potential and actual) and evapotranspiration rates over a growing season of equal duration. Input parameters for grass growth were obtained from WaterVision Agriculture (Hack-ten Broeke et al., 2016). The main model settings were that all fields were mowed and not grazed, and limitations due to nutrient deficiency, pests, diseases and weeds were included in the relative management factor which was set to 0.9 (with 1.0 being optimal growth without limitations). This management factor affects both the potential and actual yield, proportionally.

4.2.7.3 Soil hydraulic functions

SWAP uses the analytical soil hydraulic functions of Mualem-Van Genuchten (Mualem, 1976; Van Genuchten, 1980), which read:

$$\theta = \theta_{\text{res}} + (\theta_{\text{sat}} - \theta_{\text{res}})(1 + |\alpha h|^n)^{-m}$$

$$K = K_{\text{sat}} S_{\text{e}}^{\lambda} \left[1 - \left(1 - S_{\text{e}}^{\frac{1}{m}} \right)^m \right]^2 ; m = 1 - \frac{1}{n}$$

where, θ_{sat} is the saturated water content (cm³ cm⁻³), θ_{res} is the residual water content (cm³ cm⁻³), h is the pressure head (cm), α (cm⁻¹), n and m are shape parameters, K_{sat} is the saturated hydraulic conductivity (cm d⁻¹) and S_{e} is the relative saturation that can be determined using:

$$S_{\text{e}} = \frac{\theta - \theta_{\text{res}}}{\theta_{\text{sat}} - \theta_{\text{res}}}$$

Measured root zone soil hydraulic parameters (residual water content, saturated hydraulic conductivity and shape parameters α and n, Section 4.2.6) and the measured saturated hydraulic conductivity (Section 4.2.5) were used in SWAP. Subsoil parameters were obtained from the Staring Series soil physical database (Wösten et al., 2013), using the observed C-horizon texture (up to 120 cm depth, Section 4.2.3).

4.2.7.4 Heat flow

Soil temperatures affect root activity and therefore crop growth. Soil heat flow is simulated by SWAP with soil heat capacity and soil thermal conductivity, and requires the input of air temperature (from meteorological data), soil texture and organic matter content. For the root zone, field-level soil texture was obtained from the project 'Cows and Opportunities' (measured from composite sample at the field level, sampled at 0-10 or 0-25 cm depths, obtained with near-infrared spectroscopy), as well as field-level topsoil organic matter contents (0-20 cm, Section 4.2.3). For the subsoil, estimated C-horizon textures were used (Section 4.2.3) and subsoil organic matter contents were set to 0.005 g g^{-1} dry soil.

4.2.7.5 Drainage

The drainage flux of the soil column was simulated as a function of the drainage level and drainage resistance, using:

$$q_{drain} = \frac{\Phi_{GWT} - \Phi_{drain}}{\gamma_{drain}}$$

where, q_{drain}= drainage flux (cm d^{-1}), Φ_{GWT}= phreatic groundwater table midway between drains and ditches (cm), Φ_{drain}= drainage level (cm), and γ_{drain}= drainage resistance (d). The parameters for drainage resistance and drainage level were calibrated until simulated groundwater tables reflected measured groundwater tables to the nearest 15 cm (Section 4.2.4). The obtained drainage resistance and drainage level were subsequently used in the SWAP simulations of years 2001 and 2003. In this way groundwater tables were not imposed but followed from the modelled soil water fluxes in the actual weather period. On two fields on Farm 3, and one field on Farm 5, groundwater table data were only available for the years 2001-2003, hence the drainage resistance and drainage level were based on simulated groundwater tables for those years (Table A.4.7). On some fields, for example all fields on Farm 1, groundwater tables were generally deep (between 1.5 and 3 m), and therefore free drainage was assumed (Table A.4.7). In case of free drainage, the soil water flux at the bottom of the soil column only depends on gravity, and is equal to the unsaturated hydraulic conductivity of the subsoil.

Three fields (two on Farm 1 and one on Farm 3) applied sprinkler irrigation, with records obtained from the farmers. The least irrigation was applied in 2001: only one field of Farm 1 was irrigated with 56 mm water. Most irrigation was applied in 2003: the three fields were irrigated with 66-86 mm water.

4.2.8 Data analyses

From the simulated crop growths in SWAP (all simulated as grass), the yield gap was calculated as a fraction of the potential yield. Oxygen stress and drought stress were simulated as reduced transpiration rates (potential transpiration – actual transpiration, cm y^{-1}) due to too wet or too dry soil conditions. In this study the reduced transpiration was expressed as a fraction of the potential transpiration.

Statistical analyses were performed in R Studio version 3.4.3 (R Core Team, 2014). Pearson correlation coefficients were obtained between all quantitative visual observations and all indicators for soil functions, using the Hmisc package (Harrell et al., 2018).

Each soil function indicator that was assessed in the present study was associated with quantitative visual observations. Dutch farmers usually know clay contents from routine soil analyses at field level, therefore clay content was considered as an additional predictor for soil functioning. For each combination of a quantitative visual observation and a soil function indicator, the interaction with clay content (indicated with '×') was assessed using the following multiple linear regression model structure:

Soil function indicator =
visual observation + clay content + visual observation × clay content

Stepwise linear regressions were used to select the quantitative visual observations that contributed to a soil function indicator. Those quantitative visual observations that had a reduced number of observations and/or those that showed overlap with another quantitative visual observation, were omitted from the stepwise linear regressions (Table 4.2). Furthermore, clay content was included in the stepwise linear regressions to account for the soil types. Interactions with clay, however, were not included in the stepwise linear regression models due to the low number of observations (25). Stepwise linear regressions were performed in the MASS package (Venables and Ripley, 2002), using both directions following the AIC criterion to select variables. Subsequently, Ramsey's regression equation specification error test (RESET, Ramsey, 1969) was performed (using the lmtest R package of Zeileis and Hothorn, 2002) to assess whether nonlinearity should be taken into account, by adding interaction and/or quadratic terms.

4.3 Results

Summary statistics of all quantitative visual observations and the indicators for soil functions are given in Table A.4.9 and A.4.10. Please refer to Van Leeuwen et al. (2019b) to view the data.

4.3.1 Associations between quantitative visual observations and indicators for soil functions

When all soil types (sand and clay) were assessed together, all indicators for soil functions (Table 4.1) except for the relative drought stress in 2003 were associated with one of the quantitative visual observations (Table 4.3). For each soil function indicator, the strongest correlations were found for plant available water and grass cover (r= -0.55); water storage capacity and the fraction largest soil structural elements (r= -0.68); yield gap in 2016, 2001 and 2003 and root count (r= 0.63, 0.69 and 0.69 respectively); relative oxygen stress in 2001 and root count (r= 0.75); nitrate concentration in ground- or drain water and Munsell soil colour value (r= -0.86); and phosphate concentration in ground- or drain water and gley mottles (r= 0.86; Table 4.3). The biopore count and the degree of soil compaction did not correlate with any of the soil function indicators (Table 4.3). When interactions with clay content were included, it turned out that the correlation between the degree of soil compaction and yield gap in a wet year (2001) was significantly affected by clay content (Figure 4.3, Table 4.4), but biopore count correlated poorly with any of the soil function indicators when interaction with clay was included. Furthermore, significant interactions with clay content were found for gley mottles and the fraction of largest soil structural elements (Figure 4.3, Table 4.4). The soil function indicators associated more with the quantitative visual observations when interaction with clay content was included, than when the interaction with clay content was not included.

4.3.2 Relation between multiple quantitative visual observations and indicators for soil functions

For all soil functions indicators (except for relative drought stress in 2003), stepwise linear regression models resulted in a better prediction of the soil function indicators (Table 4.5), than when using a single quantitative visual observation (Table 4.3) or/and when interaction with clay content was included (Table 4.4). From the soil function indicators, nitrate and phosphate concentrations were the best predicted soil function indicators using a set of quantitative visual observations (adjusted R^2 =0.86, and 0.78 respectively, Table 4.5). The visual observation of soil compaction depth was most often significant in the stepwise linear regression models (Table 4.5). In contrast, earthworm count was never significant. Ramsey's RESET test was not significant for any of the stepwise linear regression models, meaning that quadratic terms and/or interactions terms would not improve the models significantly (Table A.4.11).

Table 4.3. Pearson correlation coefficients (r) between quantitative visual observations (first row) and soil function indicators (first column). Only those correlation coefficients are shown that were significant at P=0.05. Note that biopore count, degree of soil compaction and relative drought stress in 2003 are not shown in the table, because none of the correlations were significant. Between brackets: R^2. This table continues on the next page.

Soil function indicator	Grass cover (%)	Root (count)	Munsell soil colour value (dark=2, lighter >2)	Gley mottles (%)	Soil structure (VSA score: 2=good; 0=bad)	Fraction largest soil structural elements (%)	Earth-worm (count)
Plant available water (cm³ cm³)	-0.55 (0.30)			0.48			0.49
Water storage capacity (cm³ cm⁻³)		-0.5	-0.47		0.49	-0.68 (0.46)	
Yield gap 2016 (%)	0.51	0.63 (0.4)	0.61			0.45	
Yield gap 2001 (%)	0.45	0.69 (0.48)	0.66			0.53	
Yield gap 2003 (%)	0.5	0.69 (0.48)	0.59			0.46	
Relative oxygen stress 2001 (%)	0.46	0.75 (0.56)	0.66			0.52	
Nitrate concentration (mg L⁻¹)		-0.69	-0.86 (0.74)	-0.53	0.7	-0.8	-0.61
Phosphate concentration (mg L⁻¹)		0.62	0.55	0.86 (0.74)		0.65	0.53

Soil function indicator	Depth compacted layer (cm below surface)	Depth of 85% of roots (cm below surface)	Depth A horizon (cm below surface)	Mean highest ground-water table (cm below surface)	Maximum rooting depth (cm below surface)	Soil organic matter content (fraction)	Clay (fraction)
Plant available water (cm³ cm³)	-0.44						
Water storage capacity (cm³ cm⁻³)				-0.42			-0.48
Yield gap 2016 (%)			-0.42			0.6	0.58
Yield gap 2001 (%)		-0.43	-0.4			0.65	0.66
Yield gap 2003 (%)			-0.4			0.58	0.61
Relative oxygen stress 2001 (%)		-0.46			0.40	0.65	0.65
Nitrate concentration (mg L⁻¹)						-0.66	-0.72
Phosphate concentration (mg L⁻¹)			-0.57	-0.6		0.77	0.73
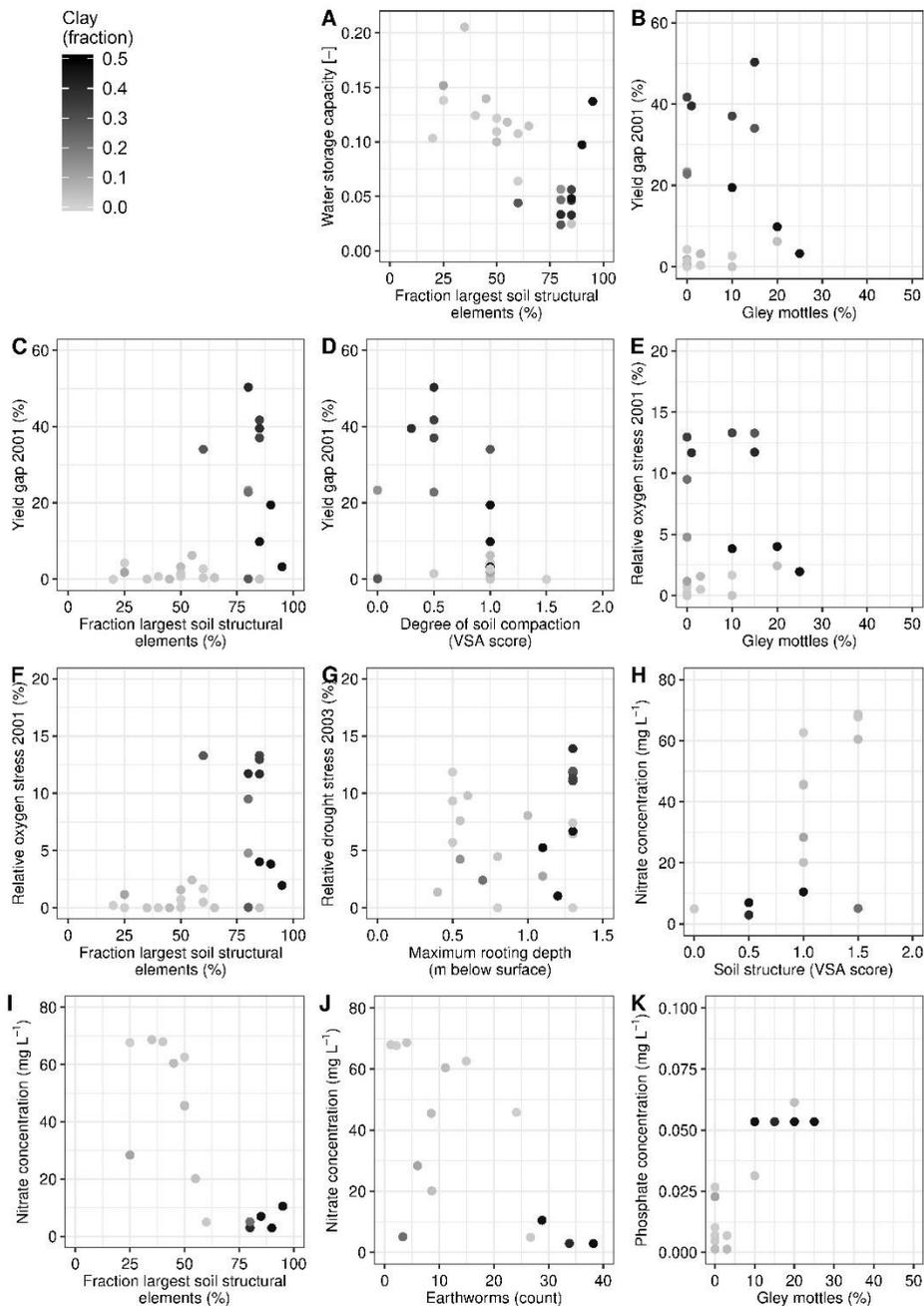
Figure 4.3. If clay content significantly affected the correlation between a soil function indicator and a quantitative visual observation, the relationship between a soil function indicator (Y-axis), quantitative visual observation (X-axis) and clay content (colour of the points: the darker, the higher the clay content) is plotted. For the corresponding regression models that include interaction with clay content, please refer to Table 4.4.

106

Table 4.4. Regression coefficients and the adjusted R^2 of regression models are shown that predict a soil function indicator based on a quantitative visual observation and the interaction with clay content. Regression models are shown when the interaction term (visual observation x clay content) was significantly associated with a soil function at $P=0.05$. For all other combinations of soil function indicators and quantitative visual observations, interaction with clay was not significant.

Label	Soil function indicator	Visual observation	Adjusted R^2	Regression coefficients			
				Intercept	Visual observation	Clay content (fraction)	Visual observation x clay content (interaction)
A	Water storage capacity [-]	Fraction largest soil structural elements (%)	0.55	0.2	-0.002	-0.8	0.01
B	Yield gap 2001 (%)	Gley mottles (%)	0.54	-2.3	0.7	106.4	-4.4
C	Yield gap 2001 (%)	Fraction largest soil structural elements (%)	0.50	-13.1	0.2	299.5	-2.9
D	Yield gap 2001 (%)	Degree of soil compaction (VSA score)	0.52	-2.7	4.2	128.5	-87.8
E	Relative oxygen stress 2001 (%)	Gley mottles (%)	0.51	-0.7	0.3	31.8	-1.4
F	Relative oxygen stress 2001 (%)	Fraction largest soil structural elements (%)	0.53	-4.6	0.1	106.5	-1.1
G	Relative drought stress 2003 (%)	Maximum rooting depth (m below surface)	0.17	11.3	-6.2	-77.4	71.2
H	Nitrate concentration (mg L⁻¹)	Soil structure (VSA score)	0.79	2.8	43.8	41.5	-161.2
I	Nitrate concentration (mg L⁻¹)	Fraction largest soil structural elements (%)	0.70	113.9	-1.3	-504.7	5.9
J	Nitrate concentration (mg L⁻¹)	Earthworm (count)	0.60	68.2	-1.5	-191.3	4.5
K	Phosphate concentration (mg L⁻¹)	Gley mottles (%)	0.84	0.01	0.003	0.1	-0.01

Table 4.5. Results of the stepwise linear regression models. The regression coefficients of the intercept and the selected quantitative visual observations are presented, as well as the adjusted R2 of the model. ** P<0.01, * P<0.05. This table continues on the next page.

Soil function	R2-adj	Intercept	Crop type (maize =1, grass =0)	Munsell soil colour value (dark =2, lighter >2)	Soil structure (VSA score: 2=good; 0=bad)	Fraction largest soil structural elements (%)	Earthworm (count)	Depth compacted layer (m below surface)	Compaction degree (VSA score)	Depth of 85% of roots (m below surface)	Depth A horizon (m below surface)	Mean highest groundwater table (m below surface)	Maximum rooting depth (m below surface)	Clay content (fraction)
Plant available water (cm³ cm⁻³)	0.66	0.2	-0.08*	0.04		0.001*	0.001	0.36*	0.07	0.41*	0.11	-0.09**		-0.41**
Water storage capacity (cm³ cm⁻³)	0.75	-0.05	-0.07*					0.32*	0.08**	0.34**	0.15*	-0.04	-0.08**	
Yield gap 2016 (%)	0.77	50.3*			12.5*	-0.2		-50.6*	-29.4**	-38.8	-41.8*		13.8*	23.0
Yield gap 2001 (%)	0.83	153.1**	9.5	-8.6		-0.8**	-0.2	-140.6*	-39.7**	-134.6**	-62.2*	16.4	10.7	134.0**
Yield gap 2003 (%)	0.75	123.3**	10.0	-8.5		-0.6*	-0.2	-115.3	-28.8**	-90.4*	-39.9*	13.3		118.3**
Relative oxygen stress 2001 (%)	0.83	17.2*			6.1*	-0.1		-28.4**	-11.5**	-20.7*	-10.8		5.9*	10.9

Soil function	R2-adj	Inter-cept	Crop type (maize =1, grass =0)	Munsell soil colour value (dark =2, lighter >2)	Soil structure (VSA score: 2=good; 0=bad)	Fraction largest soil structural elements (%)	Earthworm (count)	Depth compacted layer (m below surface)	Compaction degree (VSA score)	Depth of 85% of roots (m below surface)	Depth A horizon (m below surface)	Mean highest groundwater table (m below surface)	Maximum rooting depth (m below surface)	Clay content (fraction)
Relative drought stress 2003 (%)	0	6.2												
Nitrate concentration (mg L^{-1})	0.88	148.9**	19.1	-27.9**		-0.75*				-73.2		13.72*		87.47*
Phosphate concentration (mg L^{-1})	0.93	0.32*	-0.04	-0.05*	-0.06	-0.001	-0.001	0.13		-0.09		0.02	-0.08*	0.48*

4.3.3   Correlations between soil functions

The general trends over the sites suggested that there were synergies and trade-offs between soil function indicators; some soil function indicators can be optimal while, concurrently, others may or may not be optimal (Table 4.6). For example, synergies were found between plant available water and nitrate concentration and relative drought stress in 2003, and between water storage capacity and yield gaps in 2001, 2003 and 2016. Not surprisingly, yield gaps of the various years correlated positively with each other, as well as with relative oxygen or drought stress. Trade-offs were found for nitrate and phosphate concentrations in groundwater or drain water, which were significantly and negatively correlated with each other. This indicates that low nitrate concentrations in drain or groundwater were associated with high phosphate concentrations, and vice versa. Nitrate concentrations were also significantly and positively correlated with soil water storage capacity, indicating that low nitrate concentrations are associated with low soil water storage capacities. Lastly, nitrate concentrations were negatively correlated with yield gap in 2001 and with relative oxygen stress in 2001, meaning that low yield gaps in a wet year associated with high nitrate concentrations, and vice versa.

4.4 Discussion

4.4.1   Associations between a single quantitative visual observation and indicators for soil functions

Most soil function indicators could be predicted using a single quantitative visual observation (Table 4.3), which could be explained by the fact that both visual soil evaluation and most soil function indicators were quantified based on soil physical properties. Especially the quantified soil function indicators plant available water, water storage capacity, modelled yield gaps, modelled relative oxygen and drought stress are mainly affected by soil texture and soil structure (Vereecken et al., 2010). Nitrate and phosphate concentrations in ground- and drain water were also strongly correlated with quantitative visual observations. This was unexpected, because nitrate and phosphate concentrations were likely not only dependent on soil physical properties, but also on soil chemical properties, climatic conditions and management factors (Freese et al., 1992; Lipiec and Stępniewski, 1995; Oenema et al., 2010, 2004; Schoumans et al., 2013). A reason could be that there are many internal correlations between soil physical, chemical, and biological aspects (Karlen et al., 2001; Pulido Moncada et al., 2014; Sonneveld et al., 2014), and land management and weather patterns (Figure 4.1).

Table 4.6. Pearson correlation coefficients (r) between soil functions, only presented when significant at P=0.05. Bolt font: not optimal combination (trade-off) between soil function indicators. Note that values were only shown below the diagonal, to avoid repetition.

	Plant available water (cm³ cm⁻³)	Water storage capacity (cm³ cm⁻³)	Yield gap 2016 (%)	Yield gap 2001 (%)	Yield gap 2003 (%)	Nitrate concentration (mg L⁻¹)	Phosphate concentration (mg L⁻¹)	Relative oxygen stress 2001 (%)	Relative drought stress 2003 (%)
Plant available water (cm³ cm⁻³)									
Water storage capacity (cm³ cm⁻³)									
Yield gap 2016 (%)		-0.53							
Yield gap 2001 (%)		-0.6	0.95						
Yield gap 2003 (%)		-0.45	0.93	0.94					
Nitrate concentration (mg L⁻¹)	-0.52	**0.64**		**-0.57**					
Phosphate concentration (mg L⁻¹)						**-0.62**			
Relative oxygen stress 2001 (%)		-0.6	0.91	0.96	0.89	**-0.67**			
Relative drought stress 2003 (%)	-0.5		0.68	0.56	0.77			0.5	

Previous studies reported contrasting results about the association between measured crop yields and visual evaluation of soil structure (e.g. Ball et al., 2007; Guimarães et al., 2013) with. For example, visual soil evaluation of soil structure (VESS) significantly correlated with crop yield (Mueller et al., 2013), while other studies found insignificant correlations between VESS and crop yield (Giarola et al., 2013 and Mueller et al., 2009). In these studies it is likely that land management and climatic conditions affected measured crop yields, while VESS focusses on soil physical properties (Mueller et al., 2013, Figure 4.1). VESS may be therefore not a good predictor for measured crop yield, especially when the limiting factors are others than soil physical factors (e.g. above-ground factors or soil chemical factors). Instead, a broad set of visual observations likely better represent soil function indicators, as was found in the present study.

Besides using a broad set of visual observations to assess soil functions, including clay content could improve the assessment of soil functions. For some soil function indicators in the present study (water storage capacity, yield gap 2001, nitrate and phosphate concentrations in drain- or groundwater) the association with a quantitative visual observation became stronger when an interaction with clay content was included (Table 4.4) than when those interactions were not included (Table 4.3). Previous studies showed that texture influenced the VSE scores. Nevertheless, in the present study Ramsey's RESET test indicated that including interactions and/or quadratic terms (indicating nonlinear relationships) would not improve the correlation between a set of visual observations and soil function indicators. The reason could be that we did not observe the full range of possible quantitative visual soil observations and soil function indicators, and that the effect of other (linear) terms overruled the effect of interactions and/or quadratic terms. To conclude, it was found that soil function indicators can be best predicted based on a combination of visual observations and clay content.

4.4.2   *Relation between multiple quantitative visual observations and indicators for soil functions*

The stepwise linear regression models resulted in a better prediction of soil function indicators (i.e. high adjusted R^2, Table 4.5) than when using a single quantitative visual observation (Table 4.3) or when adding an interaction term with clay content to a single quantitative visual observation (Table 4.4). Modelled yield gaps in a wet year (2001) and a dry year (2003) were associated with the same quantitative visual observations, which were soil structure (fraction largest soil structural elements), soil compaction, root depth (85% of roots) and A horizon depth (Table 4.5). It should be noted that these modelled yield gaps were mainly the result of a combination of modelled plant oxygen and drought stress, and not of nutrient limitations, and therefore it was found that visual soil physical properties correlated with modelled yield gaps. As for modelled yield gaps in a wet and a dry year, the crop sensitivity to

oxygen and drought stress is affected by similar soil properties. Under wet conditions, oxygen stress for crop growth is mainly governed by the absence of air filled soil pores, which is affected by texture, soil compaction, soil structure and the presence of continuous soil pores (Håkansson and Lipiec, 2000). Whereas in dry conditions, drought stress is mainly governed by the absence of plant available water, deep groundwater tables and a limited capillary rise of water to the root zone (Kroes, 2018), which is affected by texture, soil compaction and soil structure. Also, deep rooting plants are generally less sensitive to drought than shallow rooting plants. Nevertheless it depends on the water and nutrient availability in deeper soil layers whether plants can actually grow and increase biomass in dry periods, or whether they just remain green without growing (Kemp and Culvenor, 1994). Unexpectedly, although the yield gap in 2003 could be predicted, the drought stress in the dry year (2003) could not be predicted using (a set of) quantitative visual observations. The best model that associated with drought stress was based on maximum rooting depth and an interaction with clay content (Table 4.4). The prediction of drought stress in SWAP might improve if a dynamic root model is included (Kroes, 2018).

It turned out that for each quantified soil function indicator the selected set of quantitative visual observations was different (Table 4.5). This suggests that not all quantitative visual observations need to be assessed when assessing a specific soil function, which reduces assessment time in the field. In practice, farmers often ask which visual soil property to improve to improve crop growth. Another common question is how much a soil property should change to reach a satisfying soil functioning level. The stepwise linear regression modelling results suggest that a soil function can be changed by just changing one of the quantitative visual observations (Table 4.5). However, there are many associations between (visual) soil properties (e.g. Ball et al., 2017; Sonneveld et al., 2014). It is likely that by changing one of the soil properties other soil properties change as well, and therefore it cannot be easily predicted how a change in one soil property will affect a soil function of interest. Instead, after implementing a soil management strategy, stepwise linear regression models could be used to monitor how visual properties change, and how the soil functions change. This only holds for visual observations that fall in similar ranges as the data used to develop the stepwise linear regression models (Table A.4.9 and A.4.10).

4.4.3   *Most optimal combination of soil functions*
Assessing several soil functions at the same time allows assessment of whether soil management practices improve all soil functions, or whether one of the soil functions improves at the expense of another soil function (Schröder et al., 2016a). Within agroecosystems, the preferred combination of soil functions highly depends on the

climatic zone, crop types, soil types and vulnerability of surrounding ecosystems (O'Sullivan et al., 2015; Schulte et al., 2014). In the present study we found a positive correlation between nitrate concentration in ground- and drain water and water storage capacity, meaning that improving water storage capacity (often associated with improved crop yield) is associated with higher nitrate concentrations. For nine of the 15 fields the nitrate concentrations were below the EU threshold of 50 mg L^{-1} (Table A.4.10), which means that for those fields there could be space to increase crop production without exceeding the nitrate concentration threshold.

4.4.4   Future research

The approach used in the present study to associate a set of visual observations to the soils' potential to function, enabled fair comparison of soil quality between sites with different soil properties. The stepwise linear regression approach showed potential to design scoring functions to evaluate various soil functions based on (quantitative) visual observations, which can be easily repeated to other geographic areas if data is available. Another reproducible approach was presented by Pulido Moncada et al. (2014), who used classification and regression trees to identify associations between soil chemical and physical properties. Data in classification and regression trees, however, is grouped based on the distribution of the data, rather than that those identified groups of data correlate with soil functions. A stepwise linear regression approach is than preferred, as it results in direct correlations between visual observations and soil functions.

Despite the fact that the stepwise linear regression approach showed potential to assess various soil functions, the stepwise linear regression models found in this study are not readily applicable to other areas, because they are based on rather small set of observations with site-specific soil properties (Table A.4.9 and A.4.10). For use of these methods elsewhere, we recommend to collect more visual soil observations together with quantified soil function indicators, to make the regression models more robust. A major challenge of the method used in the present paper was that the laboratory measurements of the soil moisture retention characteristics (needed to quantify the modelled soil function indicators) were time consuming. A solution could be to go to sites where data of soil hydraulic properties and/or soil moisture retention characteristics are available, to be used to quantify soil function indicators. Databases are available, such as BOFEK in the Netherlands (Wösten et al., 2013), or Hypres in Europe (Wösten, 2000). Also, it is possible to use less complex versions of SWAP, or to use other soil hydrological models than SWAP, such as Hydrus1D (Šimůnek et al., n.d.) or MACRO (Larsbo and Jarvis, 2003). More quantitative visual soil observations allow analyses of optima and nonlinear relationships between quantitative visual observations and soil function indicators, as well as interactions with soil texture.

Furthermore, the addition of soil chemical data next to collected quantitative visual soil observations could be helpful to understand the broader context of soil functioning and the environmental performance of agricultural land.

4.5 Conclusions

Quantitative visual soil evaluation were used to assess a range of soil functions for dairy farmed soils in the Netherlands. Soil function indicators showed stronger correlations with a set of quantitative visual soil observations, than with single quantitative visual observation. Stepwise linear regression analysis showed that quantitative visual observations associated with the following soil function indicators: measured plant available water and water storage capacity; modelled yield gaps for a wet year (2001), a dry year (2003), and a 'normal' year (2016); and oxygen stress for a wet year (2001). Those quantified soil function indicators were based on soil physical properties, and therefore they could be associated with a set of quantitative visual soil observations that also focus on soil physical properties. Measured nitrate and phosphate concentrations in ground- and surface water also associated with a set of quantitative visual observations. This is attributed to correlations between soil physical properties and soil chemical properties, climate and management. For several soil function indicators, the association with a set of quantitative visual observations was affected by soil texture. This shows the importance of evaluating soil quality in relation to the soils' potential to function, which can vary between soil types. Also, in the present study not all soil function indicators were optimal at the same time. For instance, a high water storage capacity was associated with high nitrate concentrations, and low yield gaps in a wet year were also associated with high nitrate concentrations. This study showed that visual soil evaluations can assess several soil functions at the same time, to be able to assess and evaluate soil quality in relation to environmental performance of agricultural land.

Appendix 4

Table A.4.7. Overview of used groundwater table data or free drainage assumptions at the sampled fields. This table continues on the next page.

Farm	Fields	Free drainage assumption	Groundwater table monthly means (cm below surface)		Data origin	End measurements
			Highest	Lowest		
1	1-5	Yes	150	300	DINOloket B34C0044; B34C0150; B34C0159; B34C0213	
2	1, 3	No	105 (Feb. 2017) [1]	153 (Oct. 2016) [1]	Measured [1]	27/02/17
	2, 4, 5	Yes				
3	2	Yes	214 (Aug. 2016)	271 - piezometer often dry	Measured	27/02/17
	3	Yes	193 (Aug. 2016)	266 - piezometer often dry	Measured	27/02/17
	5	No	99 (Feb. 2016)	158 (Oct. 2016)	DINOloket B33E1281	
	1, 4	No	133 (Jan./Feb. 2001-2003)	232 (Jul./Sep.-Oct. 2001-2003)	DINOloket 33F0247 (2001-2003)	
4	1, 3	No	42 (Mar. 2017)	159 (Oct. 2016)	Measured	10/04/17
	4	No	49 (Mar. 2017)	145 (Oct. 2016)	Measured	7/04/17
	3	No	56 (Feb. 2016)	147 (Oct. 2016)	DINOloket B07B0141 (2016-2017)	
	5	No	66 (Feb. 2016)	144 (Oct. 2016)	DINOloket B07B1166 (2016-2017)	

| 5 | 1, 3, 4, 5 | No | 87 (Mar. 2017) | 112 (Dec. 2016) | Measured | 7/04/17 |
| | 2 | No | 112 (Jan./Feb. 2001-2003) | 195 (Jul./Aug./Sep. 2001-2003) | DINOloket B39C0211 (2001-2003) | |

[1] On Farm 2 the diver that measured the air pressure, inadvertently stopped after one month of measurements, and daily mean air pressure records were used from a nearby weather station (17 km, Table A.4.8), and calibrated to replace and complement the diver records (estimated local air pressure = 0.918 * Twenthe air pressure + 93.242, R^2=0.86 for the 30-day overlap).

Table A.4.8. Daily mean weather records obtained from KNMI weather and precipitation stations.

Farm	Precipitation station name (KNMI reference number)	Distance from farm (km)	Weather station	Distance from farm (km)
1	Hengelo (654)	3	Hupsel (283)	10
2	Tubbergen (361)	7	Twenthe (290)	17
3	Almen (678)	9	Deelen (275)	25
	Apeldoorn (543)	9	Deelen (275)	20
4	Ezinge (136)	9	Eelde (280)	18.5
5	Geldermalsen (584)	5	Herwijnen (356)	9

Chapter 4

Table A.49. Average and standard deviations of all quantitative visual observations.

	Grass cover (%)	Biopore (count)	Root (count)	Munsell soil colour value (dark=2, lighter>2)	Gley (%)	Soil structure (VSA score: 2=good; 0=bad)	Fraction largest soil structural elements (%)	Earth-worm (count)	Depth compacted layer (cm below surface)
Number of observations	20	23	22	25	25	25	25	25	25
Average	59	5	187	2.6	6	0.9	63	14	14
Standard deviation	22	5	134	0.7	8	0.4	23	16	11

	Degree of soil compaction (VSA score: 2=good; 0=bad)	Depth of 85% of roots (cm below surface)	Depth A horizon (cm below surface)	Mean highest groundwater table (cm below surface)	Maximum rooting depth (cm below surface)	Soil organic matter content (fraction)
Number of observations	25	25	25	25	25	25
Average	0.7	29	35	46	98	0.068
Standard deviation	0.4	9	13	51	35	0.033

Table A.4.10. Average and standard deviations of soil functions.

	Plant available water ($cm^3\ cm^{-3}$)	Water storage capacity ($cm^3\ cm^{-3}$)	Yield gap 2016 (%)	Yield gap 2001 (%)	Yield gap 2003 (%)	Relative oxygen stress 2001 (%)	Relative drought stress 2003 (%)	Nitrate concentration ($mg\ L^{-1}$)	Phosphate concentration ($mg\ L^{-1}$)
Number of observations	25	25	25	25	25	25	25	15	15
Average	0.18	0.09	0.10	0.12	0.12	0.04	0.06	33.4	0.03
Standard deviation	0.04	0.05	0.12	0.16	0.11	0.05	0.04	27.3	0.02

Table A.4.11. Ramsey's RESET test shows that quadratic and/or interaction terms would not improve the regression models, because P>0.05.

Soil function	P-value
Plant available water	0.82
Water storage capacity	0.97
Yield gap 2016	0.84
Yield gap 2001	0.87
Yield gap 2003	0.83
Relative oxygen stress 2001	0.60
Relative drought stress 2003	NA
Nitrate concentration	0.94
Phosphate concentration	NA

Chapter 5

Synthesis

5.1 General discussion

In order to evaluate how easily obtainable soil information can be used to assess soil functioning, the three research questions (Section 1.5) are discussed in the following three subsections. Figure 5.1 provides the thesis outline, complemented with the main conclusions.
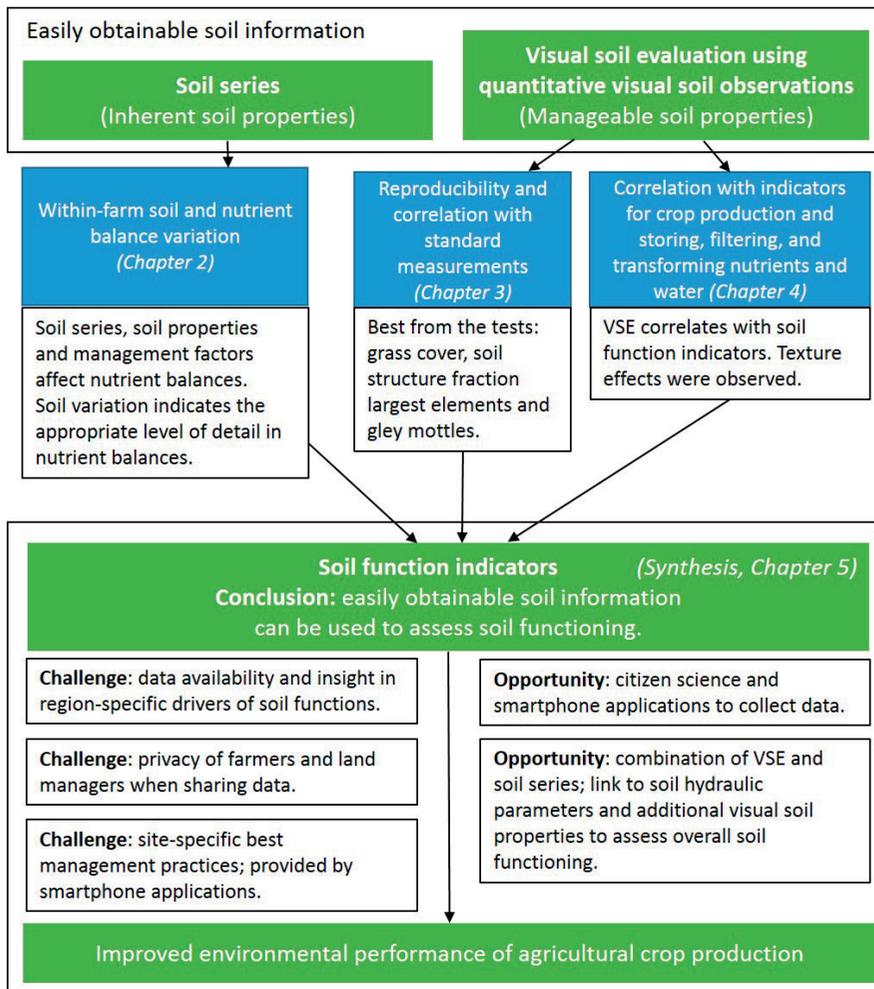
Figure 5.1. Thesis outline (Figure 1.1) complemented with the main findings from each Chapter.

5.1.1 To what extent does the required spatial scale in nutrient balances depend on the level of soil variation?

Chapter 2 assessed whether nutrient balances at farm level represented nutrient balances at field level, given the within-farm diversity of soil series. For dairy farms,

nutrient balances at field level are affected by several factors: soil properties (which can be obtained from soil series, e.g. texture and soil organic matter contents; Oenema et al., 2010; Van Es et al., 2006), soil quality characteristics (e.g. soil compaction; Lipiec and Stępniewski, 1995) and land management factors (e.g. total nutrient inputs; Oenema et al., 2010; Watson and Atkinson, 1999). The required detail in farm-level nutrient balances does therefore not only depend on the level of soil variation within a farm, but also on the 'management variation' within a farm, such as field-specific amount of irrigation, amount of fertilizer applications, grazing and mowing. At field level extremes or 'hotspots' in nutrient balances can be visible (such as soil nitrogen (N) or phosphorus (P) surpluses or deficits), whereas at crop or farm level these hotspots may not be visible because they disappear by averaging (Chapter 2). Chapter 2 demonstrated the effect of the used spatial scale on the minimum and maximum nutrient balances found for the five studied dairy farms assessed in this thesis. Field-level soil N balances varied between -190 and +385 kg N ha^{-1} y^{-1} (Figure A.2.11), whereas field-level soil P balances varied between -27 and +55 kg P ha^{-1} y^{-1} (Figure A.2.16). At one coarser scale of analysis, the crop level, soil N balances varied between -50 and +160 kg N ha^{-1} y^{-1} for the five dairy farms (Figure 2.5), whereas soil P balances varied between -21 and +8 kg P ha^{-1} y^{-1} (Figure 2.5). These data demonstrated that field-level nutrient balances showed a broader range of nutrient surpluses and deficits than crop-level nutrient balances, suggesting that field-level nutrient balances are more informative than nutrient balances at the crop level or farm level. This implies that field-level nutrient balances are more relevant than crop- or farm-level nutrient balances for decision-making by farmers, to reduce nutrient losses to the environment.

Besides the spatial variation of soil nutrient balances that was partly related to soil variation, calculated nitrogen leaching may be soil-dependent if the used leaching factors are soil-dependent. In this thesis I showed that this was the case for the Annual farm Nutrient Cycle Assessment (ANCA, Schröder et al., 2016b), applied in Chapter 2. The values of the soil-specific leaching factors were determined using soil series on the 1:50,000 Dutch soil map, as the values of the leaching factors depend on the combination of texture and groundwater table depths (Fraters et al., 2015). Hence, this assessment showed that the level of required detail in the calculations matters: in the case of high soil variation and the presence of different leaching factors across fields, upscaling of the soil-specific leaching factors to crop or farm level can lead to a bias in calculated N leaching compared to calculations at field level (Figure 2.8 and 2.9). Moreover, N leaching calculations at crop or farm level are likely underestimating N leaching, because N deficits (negative soil N balances) are reducing the average soil N balance at crop or farm level, hence a lower N leaching is estimated (Figure 2.8). Especially for farms with the highest variation in soil types and/or management, these

findings also support the conclusion that nutrient balances at field level are more relevant for decision-making by farmers to reduce nutrient losses to the environment (Chapter 2; Van Beek et al., 2003), strengthened by the fact that N leaching calculations at field level are more accurate than calculations at crop or farm level (Chapter 2).

A challenge that remains, however, is the availability of soil maps and the data needed to calculate nutrient balances at field level. Soil maps, for example, are often not available at a detailed spatial scale needed for nutrient management. World soil maps are developed (e.g. SoilGrids at 250 m resolution; Hengl et al., 2017), but those soil maps do not yet have the spatial detail required for nutrient management at field level. In that case, conventional soil survey techniques, such as soil profile description and classification (FAO, 2006; IUSS Working Group WRB, 2014), could be used to determine soil series at field level. Also, nutrient balances are not always available at field level. In that case, proxies for nutrient balances could be used to identify the sustainability of the management practices and the soil quality. The appropriate nutrient balance proxies, however, depend on the intensity of the agricultural system (Smaling et al., 1997). In general, in intensive agricultural systems located in temperate regions, nutrient surpluses are found (Smaling et al., 1997). In that case, the environmental impact of agricultural land is mainly determined by the amount of added nutrients, which is a management factor. Nutrient inputs, therefore, could be taken as proxy for the nutrient surpluses and nitrogen leaching. To illustrate this, for the five dairy farms studied in Chapter 2, field-level balances showed that soil N inputs are positively correlated with soil N surpluses (R^2=0.50), and P inputs are positively
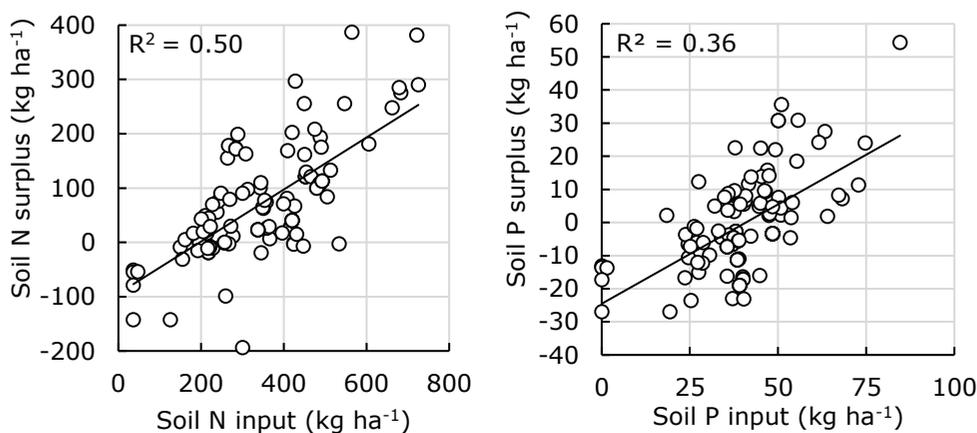
Figure 5.2. Correlation between soil N and P inputs, and soil N and P surpluses at field level, for grasslands of the five dairy farms studied in Chapter 2. Inputs: fertilizers, manure, N fixation, N deposition, and N mineralization of soil organic matter (in case of grass-maize rotation); outputs: mowed and grazed grass, N_2O emission, and N accumulation in soil organic matter (in case of grass-maize rotation).

correlated with soil P surpluses (R²=0.36; Figure 5.2). For the N balances on dairy farms in the Netherlands, it is reported that soil N surpluses are positively correlated with N leaching (Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2012). This indicates that for soil N and P balances, soil N and P inputs can be taken as a proxy for nutrient balances. Also for total soil N and P inputs, however, it is a challenge to estimate the nutrient inputs from manure during grazing. The field-level data showed that instead of total soil N inputs, the applied nutrients (chemical fertilizer and organic fertilizer) could be taken as an indicator for N surplus (R²=0.34, data not shown). For P inputs, the correlation between the applied nutrients (chemical fertilizer and organic fertilizer) and P surplus was poor (R²=0.11, data not shown). These results suggest that for intensive agricultural systems the best proxy for the field-level N balances could be the manually applied N fertilizers, and the best proxy for the field-level P balances could be the total field-level P inputs. In contrast to intensive agricultural systems, in extensive agricultural systems located in e.g. Africa, nutrients are often depleted from the soil (Smaling et al., 1997). Nevertheless, soils can be productive due to high inherent soil fertility (Smaling et al., 1997). Long-term nutrient depletion, however, is not sustainable for soil, and therefore measurements of soil nutrient stocks are essential to indicate sustainability of agricultural practices. Like the availability of nutrient balances, soil nutrient stocks are not easily obtainable. Proximal sensors, however, could be an opportunity to measure soil nutrient stocks. Proximal sensors use the electromagnetic spectrum to estimate soil chemical or physical properties (Rossel et al., 2011). Those estimates are maybe less accurate than laboratory analyses, but proximal sensing allows to take larger amounts of soil analyses because it is fast and cost-effective, once a device is purchased (Rossel et al., 2011). To conclude, the preferred spatial scale of nutrient balances is the field level. If soil maps are not available, visually described soil profiles could be used instead. If nutrient balances are not available at field level, for intensive agricultural systems manually applied N inputs (chemical and organic fertilizer) at field level could be used as a proxy for field-level N balances, and total P inputs at field level could be used as a proxy for field-level P balances. For extensive agricultural systems, the best proxy for nutrient balances and the sustainability of crop production is the combination of nutrient inputs and nutrient stocks, and the development of proximal sensors is an opportunity to estimate nutrient stocks.

5.1.2   To what extent are quantitative visual soil observations reproducible, and do they correlate with standard field or laboratory measurements?

Chapter 3 evaluated the reproducibility of quantitative visual observations (sensitivity to different users), and the correlations between quantitative visual observations and standard field or laboratory measurements. The research findings are summarized in Figure 5.3. The reproducibility of a set of quantitative visual observations was studied

as an effect of observers' background (farmer or soil scientist) and as an effect of individual observers' choices. We did not find significant subjectivity as a result of the observers' background, for six out of seven quantitative visual observations. Only for the depth of soil compaction subjectivity due to the observers' background was significant: on average farmers estimated the depth of soil compaction 8 cm deeper than soil scientists did (the overall mean observed compaction depth was 25 cm; Table 3.2, Figure 3.3S). The subjectivity by individual observers was present in seven out of eight quantitative observations, and could be categorized as systematic subjectivity and random subjectivity. As discussed in Chapter 3, systematic subjectivity leads to systematic errors, and therefore the observations may be still useable as relative differences between sites can be assessed. This was the case for the fraction of largest soil structural elements, earthworm count, and the depth of soil compaction (Table 3.2). In contrast, random subjectivity was found for grass cover, biopore count and root count (relatively high residuals, Table 3.2). For grass cover and biopore count, the group mean observed values revealed significant differences across the sites (Table 3.2). This means that estimates of grass cover and biopore count are reproducible when estimates are based on an average of several observations (e.g. more observations at a site, or the average observed value from a group of observers, Ball et al., 2007). For root count, however, no significant differences could be detected across the sites, meaning that root count was not reproducible. In contrast, results indicated that the observations of gley mottles was the most reproducible quantitative visual observation, as the residuals were relatively low; there was no significant difference found between farmers or soil scientists; there was no subjectivity by the individual observers found; and significant differences between sites were detected (Table 3.2).

This study showed that for grass cover and biopore count reproducibility is the highest when an average observed value is taken from a group of observers, or from several observations at the same site (Ball et al., 2007). Except for root count, the other quantitative visual observations (fraction of largest soil structural elements, earthworm count, gley mottles and the depth of soil compaction) are reproducible when observed by one person. It was shown that for these quantitative visual observations relative differences between sites can be detected.

The reproducibility study (Chapter 3) was conducted on five relatively contrasting soil types, which resulted in a broad range of observed values (Figure 3.3). If the same study would be repeated on five sites that are more similar to each other, it is likely that the reproducibility decreases because the differences between the sites is smaller and thus the random errors will be relatively larger. In that case, reproducibility improves if an average is taken of a larger group of observers (Ball et al., 2007); when observers regularly cross-check their findings with each other (Ball et al., 2015; Guimarães et al., 2011); and it likely improves when observers are better trained.

(1) Grass cover: valid, moderately reproducible[F&S, AVG]

(5) Root count: moderately valid[1], not reproducible

(4) Earthworm count: not valid, reproducible[F&S, SYS]

(4) Biopore count: not valid, reproducible[F&S, AVG]

Soil colour: moderately valid[1], *reproducibility NA*

Soil structure shape: not valid, *reproducible NA*

(2) Soil structure fraction largest elements: moderately valid[1], reproducible[F&S, SYS]

(3) Gley mottles: not valid, reproducible

Soil pit:

(5) Depth soil compaction: not valid, moderately reproducible [SYS]

Degree soil compaction: moderately valid[1], *reproducibility NA*

Maximum rooting depth: moderately valid[1], *reproducibility NA*

Figure 5.3. Summary of the evaluated reproducibility of quantitative visual observations, and the correlation with standard field or laboratory measurements (denoted as 'valid'). The darker green the colour of a box is, the higher the reproducibility of the quantitative visual observation and the stronger the correlation with a standard measurement. For the uncoloured boxes reproducibility was not studied. Note superscripts: '1' indicates that validation is soil type dependent; 'F&S' indicates that there is agreement among farmers and soil scientists; 'SYS' refers to systematic errors made; and 'AVG' indicates that inconsistent observations are made, hence the average of a group of observations may be the most reproducible option.

Next to the evaluation of the reproducibility of quantitative visual observations, the correlation between quantitative visual observations and standard field or laboratory measurements was assessed (Chapter 3). The strongest correlations between quantitative visual observations and standard measurements were found for grass cover, root count, soil colour value, fraction of soil structural elements, degree of soil compaction and maximum rooting depth (Figure 3.5). Correlations were affected by soil type, which can be seen from the opposite sign for the correlation between root count and root dry matter on peat and clay soils (Figure 3.5C), and the significant correlation between Munsell soil colour value and soil organic matter content for clay soils (Figure 3.5E). For sandy soils, however, all soils corresponded with the darkest Munsell soil colour chip possible irrespective of soil organic matter contents (Figure 3.8B). Furthermore, correlations were affected by the dry soil conditions during field work (this especially negatively affected the visual assessment of soil structure and compaction) and by possibly less suitable chosen field or laboratory measurements to validate with the visual observations. Several combinations of visual observation and validators did not represent the same soil properties, for example, number of biopores (count) and bulk density (g cm^{-3}); shape of soil structural elements (angular, sub-angular, or granular) and mean weight diameter (index for the size proportion of soil structural elements); and number of earthworms (count) and mean weight diameter. Accuracy of the validation study would likely improve if the soil moisture conditions are closer to field capacity at the time of VSE deployment (Guimarães et al., 2017; Shepherd, 2009), and if validators are chosen that have similar soil properties and measurement units as the visual observations.

Based on the reproducibility study and the validation study (Chapter 3), it is concluded that the most reliable quantitative visual observations investigated in Chapter 3 are grass cover and the soil structure fraction of largest elements, followed by gley mottles (Figure 5.3). Chapter 3 showed that most quantitative visual soil observations include uncertainty, given the fact that individual observers make systematic and random errors, and given the fact that more than half of the quantitative visual soil observations could not be validated.

5.1.3 Can quantitative visual soil observations be used to assess soil functioning?

Chapter 4 investigated whether visual soil evaluation correlated with several soil functions that determine soil quality. On five Dutch dairy farms 25 sites were selected (5 sites per farm) that were located on sandy and clayey soils. Simple linear regression showed that quantitative visual soil observations correlated with indicators for crop production (crop yield gaps, oxygen and water stress, and plant available water) and the storage, filtering and transformation of water and nutrients (water storage

capacity, nitrate and phosphate concentrations in drain or groundwater, Table 4.3). It was found that including the interaction with clay content improved the correlation between several quantitative visual observations and soil function indicators (Table 4.4). Furthermore, a stepwise linear regression approach was used to derive multiple linear regression models to predict a soil function indicator based on a set of quantitative visual observations. Except for drought stress in a dry year, given the high degree of explained variation for the regression models (adjusted R^2 between 0.66 and 0.93, Table 4.5), it was shown that visual assessment of soil physical properties can be used to assess various soil functions at the same time. For drought stress in a dry year (2003) the best obtained regression model (adjusted $R^2=0.17$, Table 4.4, Figure 4.3G) was based on the maximum rooting depth and the interaction with clay content. Surprisingly, no other quantitative visual observations could be used to predict drought stress in a dry year, while drought stress was mainly affected by soil physical properties, such as plant available water, the groundwater table and the (absence of) capillary rise of water to the root zone (Kroes, 2018).

Chapter 4 showed that soil functions can be assessed using a set of quantitative visual observations. As VSE mainly assesses soil physical properties, soil functions that depend on soil physical properties can be accurately assessed (such as yield gaps, that are affected by a combination of plant oxygen or drought stress). In contrast, if chemical soil properties are mainly determining soil functions (such a as yield gaps that are the result of nutrient shortage), VSE may be not the best indicator for soil functioning. In that case, soil chemical analysis could provide more insight in soil functioning. Besides, it was shown that clay content affected the correlation between soil function indicators and quantitative visual observations, which shows the importance of developing site-specific VSEs.

5.2 General discussion and conclusions
Currently many soil sensing techniques are developed that vary in complexity and ease of use, to assess and monitor soil quality (Bünemann et al., 2018). Basic techniques that use easily obtainable soil information, however, can be used to assess soil functioning (Chapter 3 and 4). In this thesis I showed that detailed soil maps (1:50,000) provide soil information that can be used to obtain soil-specific leaching factors at field level. If within-farm soil variation leads to variation in leaching factors, N leaching calculations at field level are more accurate than calculations at a coarser spatial scale (Chapter 2). The Annual farm Nutrient Cycle Assessment was used to demonstrate this consideration, but these findings may also apply to other environmental impact assessments that aggregate soil- or site-specific information to a certain spatial scale, such as life cycle assessments (Garrigues et al., 2012; Oberholzer

et al., 2012). I also showed that, apart from within-farm soil variation, the level of detail in a nutrient balance matters. At a more detailed spatial scale (field level) extremes or 'hotspots' in nutrient balances can be identified (such as soil nitrogen or phosphorus surpluses, and nitrate leaching), whereas at crop or farm level those hotspots may be less visible because of averaging (Chapter 2). The fact that soil maps could not explain all the variation in nutrient balances within a farm (Chapter 2), was explained by the fact that management factors (such as fertilizer and irrigation applications, and manure inputs) and manageable soil quality properties also influence the nutrient balances, and that these factors can vary considerably within soil series (Droogers and Bouma, 1997). Visual soil evaluation (VSE) may be used to complement soil map information for the assessment of soil functions, as VSE assesses visual soil quality properties that can be changed by management (Chapter 4).

An additional advantage of VSE is that by literally digging in the soil, it raises the awareness of the importance of soil and the understanding of soil quality, to many sorts of stakeholders (Ball et al., 2018; Bünemann et al., 2018). The reliability of a broad set of visual soil observation in VSE was tested, which was based on the Visual Soil Assessment of Shepherd (2009). Chapter 3 showed that across sandy, peaty and clayey soils on dairy farms, most quantitative visual soil observations were reproducible among farmers and soil scientists, and several visual observations corresponded with standard field or laboratory measurements. Chapter 3 showed that for 25 Dutch dairy farmed sites located on sandy and clayey soils, soil function indicators (crop yield gaps as function of oxygen stress and drought stress; oxygen stress in a wet year, water storage capacity, plant available water, nitrate and phosphate concentrations in drain or groundwater) correlated with a set of quantitative visual soil observations. Also, clay content significantly affected various correlations between a single visual observation and a soil function. This suggests that future VSEs improve when interaction with soil texture is included in the evaluation of soil functions.

The main conclusion of this thesis is that easily obtainable soil information can be used to assess soil functioning. More specifically, the use of soil series can improve environmental assessments such as nutrient balances and life cycle assessment. Furthermore, visual soil evaluation can be used to assess soil functions that are driven by soil physical properties. This can contribute to a better understanding of soil quality and the environmental performance of agricultural land. For VSE, the found regression models presented in Chapter 4 hold for dairy farmed soils in the Netherlands, but the methodology presented can be repeated to other regions, bearing in mind that soil functions may be driven by other factors than in the Netherlands.

5.3 Implications for using easily obtainable soil information to assess soil functioning

5.3.1 *Easily obtainable soil information: when to use which method?*

This thesis showed that easily obtainable soil information can be used to assess soil functioning, in order to improve the environmental performance of agricultural land. Instead of focussing on a single soil function, soil quality assessments ideally assess various soil functions at the same time, because of synergies and trade-offs between soil functions (Schröder et al., 2016a). For each soil function the question remains what the preferred assessment method is: soil series or VSE? Table 5.1 and 5.2 provide an overview of the overall challenges and opportunities of using soil series and VSE to assess soil functions, followed by an overview of the use of soil series and VSE for assessing each soil function.

Chapter 2 addressed the use of soil series to provide soil-specific information in nutrient balances, and the required spatial scale in nutrient balances, given any soil variation. Other soil functions that likely can be assessed with soil series are the soil functions 'serving as a carbon pool' and 'serving as archive for geological and cultural heritage' (Table 5.1, 5.2). To identify areas that have most capacity to store carbon, soil

Table 5.1. Overview of strengths, challenges and opportunities of soil series and visual soil evaluation, in relation to the assessment of soil functions (Table 5.2).

	Soil series	Visual soil evaluation (VSE)
Strengths	Insight in spatial variation of a mapped area. Identification of inherent soil properties that determine the soils' potential to function (Karlen et al., 2001).	Insight in spatial variation of soils when deploying various VSEs. Identification of dynamic soil properties that reflect management impacts (Karlen et al., 2001). Raising awareness of soil quality (Ball et al., 2018).
Challenges	Availability of detailed soil maps.	Availability of VSE and soil function data to develop region-specific scoring functions (Section 5.3.4).
Opportunities	Combination of soil series and measured soil physical and hydrological properties (e.g. BOFEK and HYPRES: Wösten et al., 2013, 1999).	VSE can be easily extended with soil profile descriptions and classification, if soil maps are not available.
	Combination of soil maps and VSEs to assess soil functions.	

Table 5.2. Overview of the potential suitability of using soil series and visual soil evaluation to assess soil functions.

Soil function	Soil series	Visual soil evaluation (VSE)
Crop production	**Maybe**: indicator for inherent fertility of soils (Smaling et al., 1997), but needs to be accomplished with management data and soil physical data (Chapter 4).	**Yes**: when crop production is mainly affected by soil physical properties (Chapter 4).
Storing, filtering, and transforming nutrients and water	For nutrients: **yes**; can provide site-specific soil information for nutrient balances (Chapter 2). For water: **yes;** in combination with soil physical and hydrological properties (e.g. BOFEK and HYPRES).	For nutrients: **no**; VSE is an indirect indicator (Chapter 4). For water: **yes**; water storage is based on soil physical properties (Chapter 4).
Maintaining biodiversity	**No:** biodiversity is affected by complex soil interactions that cannot be deduced from soil maps.	**Maybe**: biomass and functional group diversity of earthworms as visual indicator (Pulleman et al., 2012; Shepherd, 2009).
Serving as carbon pool	**Yes**: the soils' potential to store carbon can be assessed using soil maps in combination with land use and climate data (e.g. S-World: Stoorvogel et al., 2017).	**Yes**: Munsell soil colour value in combination with soil texture (Wills et al., 2007), in combination with land use and climate data (Stoorvogel et al., 2017). Might be not suitable for various sandy soils (Van Leeuwen et al., 2018).
Serving as archive for geological and cultural heritage	**Yes**: soil series can provide information about the build-up, drainage conditions, and the origin of soil (Van Beek, Pers. Comm. 11-03-2019).	**Yes**: indicators could be added to VSE, such as the number and description of found artefacts in a soil block, the amount of soil disturbance, and soil profile descriptions (e.g. FAO, 2006).

series need to be supplemented with climate data, and land use data (e.g. Stoorvogel et al., 2017). For the soil function 'serving as archive for geological and cultural heritage' it is questionable how to quantify this function. Indicators derived from the soil map could be the layering of soil, drainage conditions (e.g. preservation of peat bodies requires anaerobe conditions) and the origin of soil (e.g. the manmade 'enk' earth soils, or 'plaggic anthrosols', are literally an archive of cultural heritage). These indicators provide details about the 'protection potential' of the soil, to serve as geological archive and to preserve archaeological artefacts (Van Beek, Pers. Comm. 11-03-2019). Soil series can furthermore be an indirect indicator for crop growth, depending on the climate region and intensity of farming systems, because soil series can provide information about inherent soil fertility (Smaling et al., 1997). As discussed in Chapter 4, however, crop production is preferably assessed based on a combination of soil physical and chemical properties, and management data. Furthermore, soil series can be related to soil hydraulic parameters by implementing pedotransfer functions (e.g. BOFEK for the Netherlands Wösten et al., 2013, and HYPRES for Europe Wösten et al., 1999), which may improve the assessment of soil functions that are driven by soil hydrological properties. Overall, soil series may show most potential in the assessment of the soil functions 'storing, filtering, and transforming nutrients and water', 'serving as a carbon pool', and 'serving as archive for geological and cultural heritage' (Table 5.1, 5.2).

VSE is a more direct indicator than the use of soil series, to assess soil functioning. Next to assessing crop production and storing, filtering, and transforming nutrients and water using VSE (Chapter 4), VSE has potential to assess the soil functions 'maintaining biodiversity', 'serving as carbon pool', and 'serving as archive for geological and cultural heritage' (Table 1). A visual indicator for biodiversity is the biomass and functional group diversity of earthworms (Pulleman et al., 2012; Shepherd, 2009). More challenging, however, is to assess the soils' potential to maintain biodiversity, because many soil processes interact with biodiversity. For the soil function 'serving as archive for geological and cultural heritage', currently VSE does not include indicators. Indicators that may be added to VSE are the number and description of found archaeological artefacts in a soil block (e.g. pieces of plastic, bones, charcoal, brick and clay pipes); and the degree of soil disturbance (e.g. mixing of soil layers and drainage of peat). Also, soil profile descriptions (e.g. FAO, 2006) could be used to assess the presence of soil layers that protect deeper soil layers containing archaeological artefacts (Van Beek, Pers. Comm. 11-03-2019). Next to the assessment of 'crop production' and 'storing water', VSE shows potential to assess the soil functions 'serving as carbon pool' and 'serving as archive for geological and cultural heritage'.

Soil series and VSE could be used separately to assess soil functions, but the combination of soil series and VSE is an opportunity for assessing soil functioning. First of all, soil series provide insight into the spatial variation of soil, which can be used to determine the VSE sampling locations (as done in Chapter 4, and e.g. Sonneveld et al., 2014). Secondly, soil maps provide the soil inherent characteristics that can be used to define the soils' potential to function, and VSE reflect dynamic soil properties that can easily change after management impacts (e.g. Karlen et al., 2001). The combination of soil series and VSE, therefore, provide the soil information that is needed to assess and evaluate soil functions in relation to the soils' potential to function (Karlen et al., 2001). If soil maps are not available to complement VSE, visual soil profile descriptions and classifications can be performed using e.g. FAO (2006). To conclude: soil series and VSE can be used to assess various soil functions, but the combination of the two methods is ideal.

In the following sections the main focus is on VSE, as there are still several challenges for the development of scoring functions to evaluate soil functions.

5.3.2   Dealing with spatial soil variation in soil quality assessment

The aim of this thesis was to evaluate how easily obtainable soil information can be used to assess soil functioning, in order to improve the environmental performance of agricultural land. VSEs can be used for different purposes, such as to identify potential limiting soil physical properties at point level (Guimarães et al., 2017), and to assess soil functioning within a farm. Especially when assessing soil functioning within a farm, it is important to realize that soil quality can vary within (Chapter 3) and between fields (Chapter 2), and that the VSE locations selected will have an effect on the quantified soil function.

To my knowledge, there is a lack of VSE studies that indicate which visual properties are most variable within a field, or at a small spatial scale. Results of the reproducibility study (Chapter 3) suggest that root count was the most variable quantitative visual observation within a field. The residuals term for root count was relatively large (Table 3.2), which was the combined result of small-scale spatial variation and random errors of the observers. Those two factors could not be further separated by the linear mixed-effect model, but these results could suggest that for root count the highest small-scale spatial variation was found. Other VSE studies reported high small-scale spatial variation of visually evaluated soil structure, earthworm abundance and soil moisture contents. Johannes et al. (2017) found poor correlations between visual evaluation of soil structure and standard soil physical measurements, which were attributed to small-scale spatial variation of soil moisture content. Spatial variation of soil moisture content could be partly explained by the spatial variation of earthworm abundance, as earthworm burrows increase infiltration rates (Blouin et al., 2013). Other studies

136

reported also a positive correlation between earthworm abundance and soil moisture condition, although causality was not further explained (Auerswald et al., 1996). To account for possible small-scale spatial variation of soil properties in the quantification of soil functions within a farm, Sonneveld et al. (2014) recommended to take at least two VSE samples in an area with a similar crop and similar soils.

Next to small-scale spatial variation of soil properties, soil properties can vary between fields. Chapter 2 showed that there can be a considerable variation of N and P balances between fields, which was discussed to be the combined effect of soil series, management and possible other soil quality characteristics. If the aim of VSE is to evaluate soil functioning within a farm, it is important to choose VSE sampling locations that represent the farm. To this end, farmland can be divided into strata that have similar properties, and in each stratum a VSE can be deployed. Strata can be defined, for example, based on crop type, crop performance, management characteristics and soil series. Subsequently, VSE locations can be randomly assigned within each stratum (De Gruijter et al., 2006). The minimum number of VSEs within a stratum is two (Sonneveld et al., 2014), to account for possible spatial variation of soil properties within a stratum. Quantified soil functions in each stratum could be averaged to farm level (using weights based on surface area of the stratum) to assess overall farm performance (Sonneveld et al., 2014) and to be able to compare farms with each other. To optimize soil functions and management decisions, however, quantification of soil functions at the stratum level is most appropriate, as averaging information to a coarser spatial scale implies that information is lost (as was illustrated in Chapter 2).

5.3.3   Dealing with temporal soil variation in soil quality assessment

Next to spatial variation of visual soil properties, visual soil properties can vary during the growing season due to varying soil moisture conditions, and due to management impacts on soil (Ball and Munkholm, 2015; Karlen et al., 2001). The timing of VSE deployment, therefore, can affect quantification of soil functions. This is not a problem when assessing the impact of an individual management activity on soil quality. In that case VSE can be deployed before and after the management activity (Ball et al., 2017). For monitoring long-term soil quality changes, however, timing of VSE is more important. Long-term (>10 year) soil quality monitoring likely results in trends that include annual variations, as a result of weather conditions or management activities (Karlen et al., 2001; Reijneveld et al., 2009). Monitoring a long term enables to evaluate soil quality changes in relation to management practices, and to exclude the weather effects, when the monitoring time is long enough. Weather conditions may in particularly affect soil moisture conditions, which can affect soil structure (Ball et al., 2017), crop root and earthworm abundance (Mueller et al., 2009).

For monitoring soil functioning using VSE, it is important to deploy VSEs in the defined strata (Section 5.3.2), and to deploy VSEs when soils are close to field capacity in order to exclude weather effects on soil properties (Guimarães et al., 2017; Shepherd, 2009).

In this thesis, the VSE data that were used to construct the multiple linear regression models to predict soil function indicators (Chapter 4), were collected in the late growing season (September 12 to October 5, 2016). Nevertheless, results indicated that VSE can be used to assess soil functions (Chapter 4). If future soil function evaluations are based on regression models that include more data, including data that are collected under varying soil moisture conditions, it is likely that the evaluations are more robust, and that soil function results are less influenced by weather conditions.

5.3.4 *Region-specific drivers for soil functioning and implications for VSEs*
Next to spatial and temporal variation of visual properties and soil functions, the drivers for soil functions can be region-specific (Karlen et al., 2001). This means that the relevance of visual soil properties in relation to soil functions may be region-specific (Wienhold et al., 2009). For example, crop yield gaps alone can be driven by various factors, such as water surpluses, water limitations, nutrient limitations, or toxicity, which can be the result of topography and the amount of run-off, inherent soil fertility and investments of farmers to manage their land (Affholder et al., 2013). In the case of water surplus or limitation for crops, such as the modelled situation in Chapter 4, visual assessment of soil physical properties were appropriate indicators for yield gaps, as soil physical properties affect water availability and water drainage (Håkansson and Lipiec, 2000; Kroes, 2018). In the case of nutrient-limited yield, soil nutrient balances in combination with the soil nutrient stocks are the best indicators for crop yield and sustainable management of soils (Smaling et al., 1997), as soils with nutrient depletion (higher outputs than inputs) can produce high crop yields because of high inherent soil quality. Those soils have a good quality considering crop production, but management is not sustainable as it depletes the soil on the long term. As discussed in Section 5.1.2, in that case nutrient stocks are a better indicator for crop production than soil physical properties. These examples show that drivers for soil functions are likely region-specific, and that VSE evaluations should be region-specific.

5.3.5 *Development of universal visual soil evaluations*
As discussed in Section 5.3.4, the drivers for soil functions can be region-specific, which implies that the appropriate soil quality indicators likely vary between regions (Karlen et al., 2001). The stepwise linear regression approach (Chapter 4) can be used to design evaluation functions, or scoring functions, specific for different regions. For developing the evaluation functions many data is required, which is a major challenge

(Wienhold et al., 2009). Also, when collecting data, it is important that data are collected and analysed using the same methods, to allow comparison of data (Van Beek et al., 2010). Currently many VSEs have been collected worldwide (e.g. see table 4.1 in Ball and Munkholm, 2015), and combining these data in open access databases could create more insight into soil variation between regions, sites and farming systems (as mentioned on the ISTRO workshop in 2017, Ireland). In line with this, all VSE data collected for this thesis are archived by Van Leeuwen et al. (2019b and c).

An opportunity for data collection could be the involvement of the general public to collect data. For example, farmers, land managers, and students who carry out VSE, could upload their data to online databases. Soil apps that already make use of this so called 'citizen science' concept, are the 'SoilInfo App' for excessing and uploading soil profile data (https://www.isric.org/explore/soilinfo), and 'LandPKS' for excessing soil quality information and uploading basic soil property data such as soil colour, texture and land cover (https://landpotential.org/landpks/). Even simpler forms of soil monitoring can be performed, such as only uploading the time and place where signs of erosion have been observed (Prager et al., 2014), or when and where signs of soil trampling were observed. Citizen science is becoming more popular and is successful for data collection (Irwin, 2018). Nevertheless, in building online databases, the privacy of land owners should be safeguarded.

Once the VSE and soil function data are available, regions can be defined based on crop type or land use, soil series and climate, in order to develop region-specific evaluation functions. Chapter 4 showed that soil texture affects the correlation between visual observations and soil functions across sandy and clayey soils. In fact, the sandy and clayey soils could be considered as different regions. What needs to be studied further, is where to put the boundaries between regions. Some grouping of soil series would reduce the complexity for developing scoring functions. Wösten et al. (1985) determined soil moister retention characteristics for nine typical sandy soils in the Netherlands, and showed that the nine soil series could be grouped into five groups based on similar soil moisture retention characteristics. If soil functions are driven by soil physical properties, like in Chapter 4, then grouping soil series based on soil moisture retention characteristics could be a good starting point to group soils. In this way, scoring functions can be developed that take into account 'the soils' capacity to function', which is needed to make a fair soil evaluation (Karlen et al., 2001).

5.4 How to increase the use of visual soil evaluation and to improve soil management?

This thesis showed that easily obtainable soil information can be used to assess soil functioning (Chapter 4). It is, however, questionable whether farmers and land managers use e.g. VSE methods, and whether they change management to improve soil quality and the sustainability of agricultural land management. Bünemann et al. (2018) sent a questionnaire about the use of VSEs to 17 international scientists who were involved in developing VSEs. It turned out that the main developers of VSEs were scientists, and that the main users were government agencies and agricultural advisors, followed by farmers and land managers. Soil quality in agricultural land is mainly affected by the land management practices of farmers and land managers. It is, therefore, important that farmers and land managers use VSEs more often, in order to identify limiting soil properties and to improve management.

The most direct way to raise awareness about soil quality among farmers and land managers, is by literally looking at soil (Ball et al., 2018). VSE serves as a good tool to raise awareness about the importance of soil, as it requires direct soil observations. Advisors and trainers, furthermore, can place the observed soil quality in the context of various soil functions, and they can give site-specific management recommendations that can improve the understanding of soils among farmers and land managers (Ball et al., 2018; Bünemann et al., 2018). Even after VSE demonstration in the field, further guidance along the way of improving soil quality is preferred. To illustrate this, in 2018, on about 100 Dutch agricultural sites soil was visually assessed and specific advice was provided by advisors, about how to improve soil quality. Approximately 20-40% of the farmers took action to improve soil quality. More guidance could have increased this number (Van Essen, Pers. Comm. 25-01-2019). To improve management practices, VSE helps raising awareness about the importance of soil quality and it identifies limiting soil properties. Further training and guidance is expected to improve management with regards to soil functions.

Not only advisors, but also smartphone or web applications could guide VSE users in soil quality assessments. For example, the Visual Evaluation of Soil Structure is available through a smartphone application (Polla, 2018). For each assessment step the application tells the user what to do. A picture of the soil can be uploaded, as well as the soil data. Such an application could be extended with training videos and comparison of the users' observations with an online database. Smartphone or web applications could also provide management recommendation. This is also one of the most challenging parts of an application, because management recommendations are often site-specific due to specific soils, weather patterns and management options (Schröder et al., 2016a). Site-specific management recommendations through

smartphone or web application, however, could guide farmers and land managers in improving soil quality and the environmental performance of agricultural crop production.

5.5 Challenges and future directions

While this thesis contributed to a better understanding of using soil series and visual soil evaluation (VSE) in relation to the assessment of soil functions, there are still some challenges, in particular for VSE.

- Availability of detailed soil maps and field-level nutrient balances. Soil maps can provide site-specific information in nutrient balances.
- Availability of VSE data, soil physical measurements or quantified soil functions. VSE and soil function data are needed to further develop region-specific VSEs for various soil functions.
- Identification of region-specific soil processes and socio-economic factors, that may affect the relationship between VSE and soil functions. A stepwise linear regression approach can be used to select appropriate soil properties for the scoring functions, but additional soil chemical data or soil management data (such as nutrient inputs) can give further insight, or can increase the accuracy in the estimation of soil function performance.
- A major challenge is how to improve management after assessing soil functions. Many general advices exist, for example, in how to avoid soil compaction. Soils and farms are site-specific, hence the best management practices are site-specific.
- Instead of focussing much on negative management impacts on soils, success stories are needed that show how soil quality improved after improving management. This can be a challenge, as soil quality improvement usually takes years. Success stories, however, can further promote the use of VSE.
- Developing online open-access soil databases has many advantages for further soil research, but we should guarantee farmers' privacy.
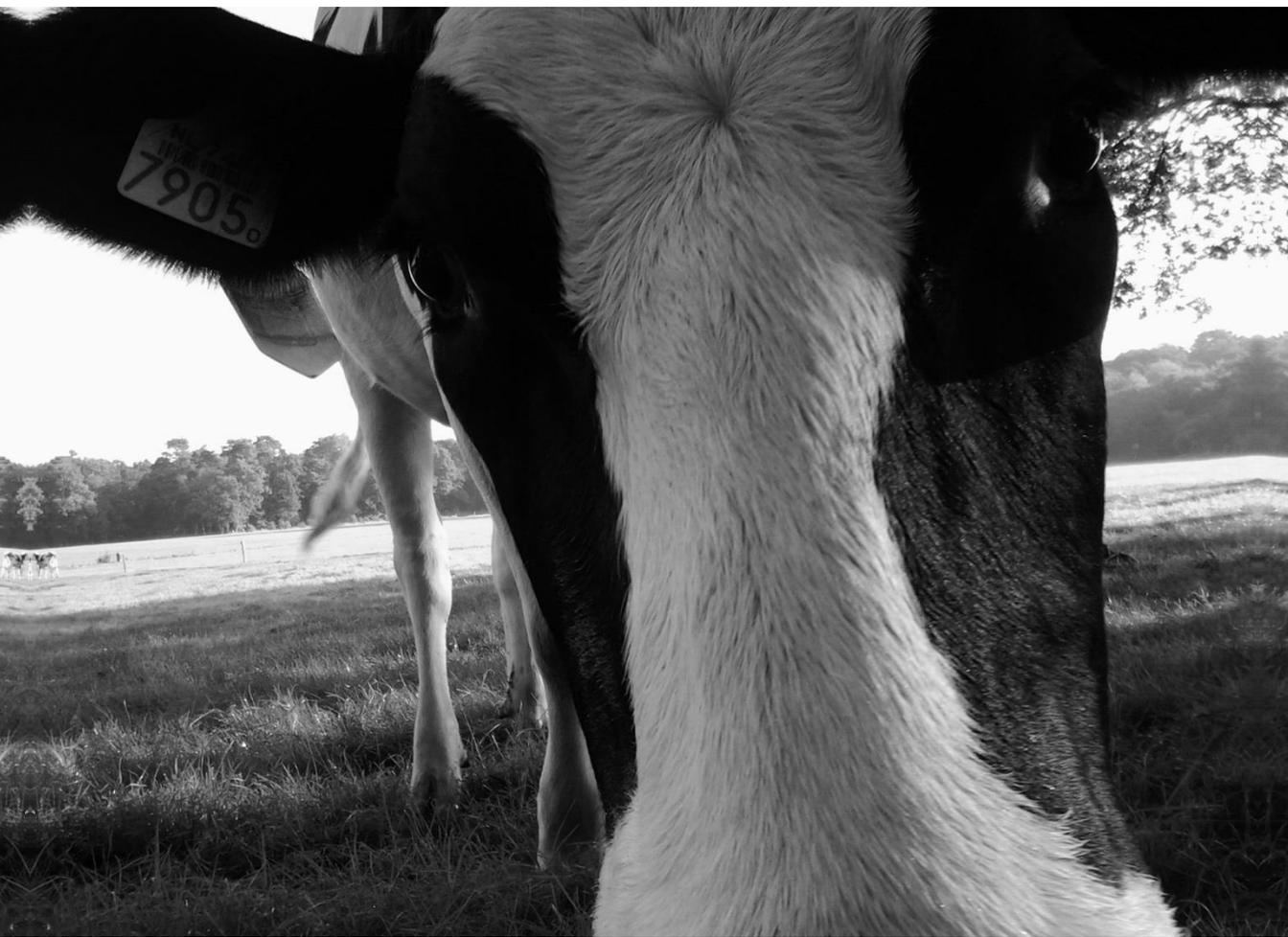
References

(CEC) Commission of the European Communities, 2006. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Thematic strategy for soil protection. COM231 final. Brussels.

Aarts, H.F.M., De Haan, M.H.A., Schröder, J.J., Holster, H.C., De Boer, J.A., Reijs, J.W., Oenema, J., Hilhorst, G.J., Sebek, L.B., Verhoeven, F.P.M., Meerkerk, B., 2015. Quantifying the environmental performance of individual dairy farms – the annual nutrient cycling assessment. Grassl. Sci. Eur. 20, 377–380.

Aarts, H.F.M., Habekotté, B., Hilhorst, G.J., Koskamp, G.J., Van der Schans, F.C., De Vries, C.K., 1999. Efficient resource management in dairy farming on sandy soil. Netherlands J. Agric. Sci. 47, 153–167.

Affholder, F., Poeydebat, C., Corbeels, M., Scopel, E., Tittonell, P., 2013. The yield gap of major food crops in family agriculture in the tropics: assessment and analysis through field surveys and modelling. F. Crop. Res. 143, 106–118.

Alterra, 2014. Standaardwerkvoorschrift. Bepaling van de Aggregaat fracties in lucht-droge grond en de bepaling an de instabiele en stabiele Aggregaten. Versie 2.1.

Alterra, 2006. Bodemkaart van Nederland, schaal 1 : 50 000, met veenkartering, versie 2006. Wageningen UR. Wageningen.

Alterra, 2005. Project "Koeien en kansen & Telen met de toekomst". Schaal 10000.

Auerswald, K., Weigand, S., Kainz, M., Philipp, C., 1996. Influence of soil properties on the population and activity of geophagous earthworms after five years of bare fallow. Biol. Fertil. Soils 23, 382–387.

Ball, B.C., Batey, T., Munkholm, L.J., 2007. Field assessment of soil structural quality – a development of the Peerlkamp test. Soil Use Manag. 23, 329–337.

Ball, B.C., Batey, T., Munkholm, L.J., Guimarães, R.M.L., Boizard, H., McKenzie, D.C., Peigné, J., Tormena, C.A., Hargreaves, P., 2015. The numeric visual evaluation of subsoil structure (SubVESS) under agricultural production. Soil Tillage Res. 148, 85–96.

Ball, B.C., Guimarães, R.M.L., Cloy, J.M., Hargreaves, P.R., Shepherd, T.G., Mckenzie, B.M., 2017. Visual soil evaluation: a summary of some applications and potential developments for agriculture. Soil Tillage Res. 173, 114–124.

Ball, B.C., Guimarães, R.M.L., Cloy, J.M., Hargreaves, P.R., Shepherd, T.G., McKenzie, B.M., 2017. Visual soil evaluation: A summary of some applications and potential developments for agriculture. Soil Tillage Res. 173, 114–124.

Ball, B.C., Hargreaves, P.R., Watson, C.A., 2018. A framework of connections between soil and people can help improve sustainability of the food system and soil functions. Ambio 47, 269–283.

Ball, B.C., Munkholm, L.J. (Eds.), 2015. Visual soil evaluation: realizing potential crop production with minimum environmental impact. CABI.

Ball, D.F., 1964. Loss-on-ignition as an estimate of organic matter and organic carbon in non-

calcareous soils. J. Soil Sci. 15, 84–92.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48.

Batey, T., 2009. Soil compaction and soil management – a review. Soil Use Manag. 25, 335–345.

Batey, T., McKenzie, D.C., 2006. Soil compaction: identification directly in the field. Soil Use Manag. 22, 123–131.

Batjes, N.H., 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma 269, 61–68.

Bengough, A.G., Campbell, D.J., O'Sullivan, M.F., 2001. Penetrometer techniques in relation to soil compaction and root growth, in: Smith, K.A., Mullins, C.E. (Eds.), Soil and Environmental Analysis: Physical Methods, Revised, and Expanded. Marcel Dekker Inc., New York, Basel, pp. 377–397.

Bindraban, P., Stoorvogel, J., Jansen, D., Vlaming, J., Groot, J.J., 2000. Land quality indicators for sustainable land management: proposed method for yield gap and soil nutrient balance. Agric. Ecosyst. Environ. 81, 103–112.

Blair, J., Lacy, M.G., 2000. Statistics of ordinal data. Sociol. Methods Res. 28, 251–280.

Blouin, M., Hodson, M.E., Delgado, E.A., Baker, G., Brussaard, L., Butt, K.R., Dai, J., Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D., Brun, J.J., 2013. A review of earthworm impact on soil function and ecosystem services. Eur. J. Soil Sci. 64, 161–182.

Bouma, J., 1994. Sustainable land use as a future focus for pedology? (A guest editorial). Soil Sci. Soc. Am. J. 58, 645–646.

Bramley, R.G.V., 2009. Lessons from nearly 20 years of Precision Agriculture research, development, and adoption as a guide to its appropriate application. Crop Pasture Sci. 60, 197–217.

Buckley, C., Wall, D.P., Moran, B., Murphy, P.N.C., 2015. Developing the EU Farm Accountancy Data Network to derive indicators around the sustainable use of nitrogen and phosphorus at farm level. Nutr. Cycl. Agroecosystems 102, 319–333.

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., De Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., Van Groenigen, J.W., Brussaard, L., 2018. Soil quality – a critical review. Soil Biol. Biochem. 120, 105–125.

CBS, 2017. Meeste grond op melkveebedrijven [WWW Document]. URL https://www.cbs.nl/nl-nl/achtergrond/2017/30/meeste-grond-op-melkveebedrijven (accessed 3.8.19).

Cherry, K.A., Shepherd, M., Withers, P.J.A., Mooney, S.J., 2008. Assessing the effectiveness of actions to mitigate nutrient loss from agriculture: a review of methods. Sci. Total Environ. 406, 1–23.

Costantini, A., 1996. Relationships between cone penetration resistance, bulk density, and moisture content in uncultivated, repacked, and cultivated hardsetting and non-hardsetting soils from the coastal lowlands of south-east Queensland. New Zeal. J. For. Sci. 26, 395–412.

Curry, J.P., 2004. Factors affecting the abundance of earthworms in soils, in: Edwards, C.A. (Ed.), Earthworm Ecology. CRC Press, pp. 91–108.

Cuttle, S.P., Bourne, P.C., 1993. Uptake and leaching of nitrogen from artificial urine applied to grassland on different dates during the growing season. Plant Soil 150, 77–86.

da Silva, A.P., Ball, B.C., Tormena, C.A., Giarola, N.F.B., Guimarães, R.M.L., 2014. Soil structure and greenhouse gas production differences between row and interrow positions under no-tillage. Sci. Agric. 71, 157–162.

Da Silva, A.P., Kay, B.D., Perfect, E., 1997. Management versus inherent soil properties effects on bulk density and relative compaction. Soil Tillage Res. 44, 81–93.

Daatselaar, C., Reijs, J., Oenema, J., Doornewaard, G., Aarts, H., 2015. Variation in nitrogen use efficiencies on Dutch dairy farms. J. Sci. Food Agric. 3055–3058.

De Bakker, H., Schelling, J., 1989. Systeem voor bodemclassificatie voor Nederland: de hogere niveaus. English summary. Centrum voor Landbouwpublikaties en Landbouwdocumentatie, Wageningen, pp. 169–197.

De Gruijter, J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for natural resource monitoring. Springer, Berlin.

De Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., Van Kraalingen, D., Supit, I., Van der Wijngaart, R., Van Diepen, K., 2019. 25 years of the WOFOST cropping systems model. Agric. Syst. 168, 154–167.

Douglas, J.T., 1997. Soil compaction effects on second-harvest yields of perennial ryegrass for silage. Grass Forage Sci. 52, 129–133.

Droogers, P., Bouma, J., 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61, 1704–1710.

Emmet-Booth, J.P., Forristal, P.D., Fenton, O., Ball, B.C., Holden, N.M., 2016. A review of visual soil evaluation techniques for soil structure. Soil Use Manag. 32, 623–634.

European Commission, 1991. Official Journal of the European Communities.

European Environment Agency., 2017. NEC Directive reporting status 2017 - The need to reduce air pollution in Europe. Copenhagen, Denmark.

FAO, 2006. Guidelines for soil description, 4th ed. Food and Agriculture Organization of the United Nations, Rome.

Fraters, D., Van Leeuwen, T., Boumans, L., Reijs, J., 2015. Use of long-term monitoring data to derive a relationship between nitrogen surplus and nitrate leaching for grassland and arable land on well- drained sandy soils in the Netherlands. Acta Agric. Scand. Sect. B — Soil Plant Sci. 65, 144–154.

Freese, D., Van der Zee, S.E.A.T.M., Van Riemsdijk, W.H., 1992. Comparison of different models for phosphate sorption as a function of the iron and aluminium oxides of soils. J. Soil Sci. 43, 729–738.

Gaiser, T., Perkons, U., Martin, P., Kautz, T., Uteau-puschmann, D., Ewert, F., Enders, A., Krauss, G., 2013. Modeling biopore effects on root growth and biomass production on soils with pronounced sub-soil clay accumulation. Ecol. Modell. 256, 6–15.

Garrigues, E., Corson, M.S., Angers, D.A., Van der Werf, H.M.G., Walter, C., 2012. Soil quality in life cycle assessment: towards development of an indicator. Ecol. Indic. 18, 434–442.

Giarola, N.F.B., Da Silva, Á.P., Tormena, C.A., Guimarães, R.M.L., Ball, B.C., 2013. On the Visual Evaluation of Soil Structure: the Brazilian experience in Oxisols under no-tillage. Soil Tillage Res. 127, 60–64.

Głab, T., Kulig, B., 2008. Effect of mulch and tillage system on soil porosity under wheat (Triticum aestivum). Soil Tillage Res. 99, 169–178.

Gourley, C.J.P., Aarons, S.R., Hannah, M.C., Awty, I.M., Dougherty, W.J., Burkitt, L.L., 2015. Soil phosphorus, potassium and sulphur excesses, regularities and heterogeneity in grazing-based dairy farms. Agric. Ecosyst. Environ. 201, 70–82.

Gourley, C.J.P., Dougherty, W.J., Weaver, D.M., Aarons, S.R., Awty, I.M., Gibson, D.M., Hannah, M.C., Smith, A.P., Peverill, K.I., 2012. Farm-scale nitrogen, phosphorus, potassium and sulfur balances and use efficiencies on Australian dairy farms. Anim. Prod. Sci. 52, 929–944.

Gourley, C.J.P., Powell, J.M., Dougherty, W.J., Weaver, D.M., 2007. Nutrient budgeting as an approach to improving nutrient management on Australian dairy farms. Aust. J. Exp. Agric. 47, 1064–1074.

Gregorich, E.G., Carter, M.R., Angers, D.A., Monreal, C.M., Ellert, B.H., 1994. Towards a minimum data set to assess soil organic matter quality in agricultural soils. Can. J. Soil Sci. 74, 367–385.

Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Van Es, H.M., Wolfe, D.W., Moebius-Clune, B.N., Thies, J.E., Abawi, G.S., 2009. Cornell Soil Health Assessment Training Manual, Edition 2.0. Geneva, New York.

Guimarães, R.M.L., Ball, B.C., Tormena, C.A., 2011. Improvements in the visual evaluation of soil structure. Soil Use Manag. 27, 395–403.

Guimarães, R.M.L., Ball, B.C., Tormena, C.A., Giarola, N.F.B., da Silva, Á.P., 2013. Relating visual evaluation of soil structure to other physical properties in soils of contrasting texture and management. Soil Tillage Res. 127, 92–99.

Guimarães, R.M.L., Lamandé, M., Munkholm, L.J., Ball, B.C., Keller, T., 2017. Opportunities and future directions for visual soil evaluation methods in soil structure research. Soil Tillage Res. 173, 104–113.

Hack-ten Broeke, M.J.D., Kroes, J.G., Bartholomeus, R.P., Van Dam, J.C., De Wit, A.J., Supit, I., Walvoort, D.J.J., Van Bakel, P., Ruijtenberg, R., 2016. Quantification of the impact of hydrology on agricultural production as a result of too dry, too wet or too saline conditions. SOIL 2, 391–402.

Håkansson, I., Lipiec, J., 2000. A review of the usefulness of relative bulk density values in studies of soil structure and compaction. Soil Tillage Res. 53, 71–85.

Harrel Jr, F.E., Dupont, C., and many others, 2015. Hmisc: Harrell Miscellaneous. [WWW Document]. URL http://cran.r-project.org/package=Hmisc

Harrell Jr, F.E., Dupont, C., Others., M., 2018. Hmisc: Harrell Miscellaneous. R package version 4.1-1. [WWW Document].

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Gonzales, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, 1–40.

Henkens, P.L.C.M., Van Keulen, H., 2001. Mineral policy in the Netherlands and nitrate policy within the European Community. Netherlands J. Agric. Sci. 49, 117–134.

Hoogland, T., Knotters, M., Pleijter, M., Walvoort, D.J.J., 2014. Actualisatie van de grondwatertrappenkaart van holoceen Nederland; Resultaten van het veldonderzoek. Wageningen.

IPCC, 2006. IPCC Guidelines for National Greenhouse Gas Inventories. Chapter 11: N2O emissions from managed soils, and CO2 emissions from lime and urea application 11.1-11.54.

Irwin, A., 2018. Citizen science comes of age. Nature 562, 480–482.

IUSS Working Group WRB, 2014. World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106.

IUSS Working Group WRB, 2007. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome.

Jackson, R.B., Canadell, J., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.D., 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108, 389–411.

Johannes, A., Weisskopf, P., Schulin, R., Boivin, P., 2017. To what extent do physical measurements match with visual evaluation of soil structure? Soil Tillage Res. 173, 24–32.

Karlen, D.L., Andrews, S.S., Doran, J.W., 2001. Soil quality: current concepts and applications. Adv. Agron. 74, 1–40.

Karlen, D.L., Mausbach, M.J., Cline, R.G., Schuman, G.E., Doran, J.W., Harris, R.F., 1997. Soil quality: a concept, definition, and framework for evaluation (a guest editorial). Soil Sci. Soc. Am. J. 61, 4–10.

Katuwal, S., Norgaard, T., Moldrup, P., Lamandé, M., Wildenschild, D., De Jonge, L.W., 2015. Linking air and water transport in intact soils to macropore characteristics inferred from X-ray computed tomography. Geoderma 237–238, 9–20.

Kemp, D.R., Culvenor, R.A., 1994. Improving the grazing and drought tolerance of temperate perennial grasses. New Zeal. J. Agric. Res. 37, 365–378.

Killourhy, C.C., Crane, D., Stehman, S.V., 2016. Precision and accuracy of visual estimates of aquatic habitat. Freshw. Sci. 35, 1062–1072.

Klimeš, L., 2003. Scale-•dependent variation in visual estimates of grassland plant cover. J. Veg. Sci. 14, 815–821.

Knowledge Transfer Centre De Marke [WWW Document], 2017. URL https://www.wur.nl/nl/Expertises-Dienstverlening/Onderzoeksinstituten/livestock-research/Innovatiecentra-en-faciliteiten/de-marke/Over-De-Marke/Bedrijf.htm (accessed 12.13.17).

Koeien & Kansen, 2017. Verantwoorde Veehouderij [WWW Document]. URL https://www.verantwoordeveehouderij.nl/nl/koeien-kansen-1.htm (accessed 12.13.17).

Kroes, J.G., 2018. Soil hydrological modelling and sustainable agricultural crop production at multiple scales. PhD thesis. Wageningen, the Netherlands.

Kroes, J.G., Van Dam, J.C., Bartholomeus, R.P., Groenendijk, P., Heinen, M., Hendriks, R.F.A., Mulder, H.M., Supit, I., Van Walsum, P.E.V., 2017. SWAP version 4; theory description and user manual. Wageningen.

Lanyon, L.E., 1994. Dairy manure and plant nutrient management issues affecting water quality and the dairy industry. J. Dairy Sci. 77, 1999–2007.

Larsbo, M., Jarvis, N., 2003. MACRO 5.0: a model of water flow and solute transport in macroporous soil. Swedish University of Agricultural Science.

Leao, T.P., Da Silva, A.P., Perfect, E., Tormena, C.A., 2005. An algorithm for calculating the least

limiting water range of soils. Agron. J. 97, 1210–1215.

Leip, A., Billen, G., Garnier, J., Grizzetti, B., Lassaletta, L., Reis, S., Simpson, D., Sutton, M.A., De Vries, W., Weiss, F., Westhoek, H., 2015. Impacts of European livestock production: nitrogen, sulphur, phosphorus and greenhouse gas emissions, land-use, water eutrophication and biodiversity. Environ. Res. Lett. 10, 1–13.

Lipiec, J., Stępniewski, W., 1995. Effects of soil compaction and tillage systems on uptake and losses of nutrients. Soil Tillage Res. 35, 37–52.

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213.

McGarry, D., 2004. A methodology of a Visual Soil - Field Assessment Tool - to support, enhance and contribute to the LADA program. FAO, Rome.

McKenzie, D.C., 2013. Visual soil examination techniques as part of a soil appraisal framework for farm evaluation in Australia. Soil Tillage Res. 127, 26–33.

McKenzie, D.C., 2001. Rapid assessment of soil compaction damage I. The SOILpak score, a semi-quantitative measure of soil structural form. Aust. J. Soil Res. 39, 117–125.

McKenzie, D.C., 2001. Rapid assessment of soil compaction damage II. Relationships between the SOILpak score, strength and aeration measurements, clod shrinkage parameters, and image analysis data on a Vertisol. Aust. J. Soil Res. 39, 127–141.

McKenzie, D.C., 1998. SOILpak for cotton growers, third edition.

Mihailescu, E., Murphy, P.N.C., Ryan, W., Casey, I.A., Humphreys, J., 2014. Nitrogen balance and use efficiency on twenty-one intensive grass-based dairy farms in the South of Ireland. J. Agric. Sci. 152, 843–859.

Mosier, A., Kroeze, C., Nevison, C., Oenema, O., Seitzinger, S., Cleemput, O. van, 1998. Closing the global N2O budget: nitrous oxide emissions through the agricultural nitrogen cycle inventory methodology. Nutr. Cycl. Agroecosystems 52, 225–248.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12, 513–522.

Mueller, L., Kay, B.D., Hu, C., Li, Y., Schindler, U., Behrendt, A., Shepherd, T.G., Ball, B.C., 2009. Visual assessment of soil structure: evaluation of methodologies on sites in Canada, China and Germany. Part I: comparing visual methods and linking them with soil physical data and grain yield of cereals. Soil Tillage Res. 103, 178–187.

Mueller, L., Shepherd, G., Schindler, U., Ball, B.C., Munkholm, L.J., Hennings, V., Smolentseva, E., Rukhovic, O., Lukin, S., Hu, C., 2013. Evaluation of soil structure in the framework of an overall soil quality rating. Soil Tillage Res. 127, 74–84.

Munkholm, L.J., Heck, R.J., Deen, B., 2013. Long-term rotation and tillage effects on soil structure and crop yield. Soil Tillage Res. 127, 85–91.

Munsell Color, 1975. Munsell Soil Color Chart. Munsell Color, Baltimore, MD, USA.

Nakamoto, T., 1997. The distribution of maize roots as influenced by artificial vertical macropores. Japanese J. Crop Sci. 66, 331–332.

Newell-Price, J.P., Whittingham, M.J., Chambers, B.J., Peel, S., 2013. Visual soil evaluation in relation to measured soil physical properties in a survey of grassland soil compaction in England and Wales. Soil Tillage Res. 127, 65–73.

Noardlike Fryske Wâlden, 2016. Noardlike Fryske Wâlden - Home_nl- [WWW Document]. URL

http://www.noardlikefryskewalden.nl/ (accessed 4.1.16).

O'Sullivan, L., Creamer, R.E., Fealy, R., Lanigan, G., Simo, I., Fenton, O., Carfrae, J., Schulte, R.P.O., 2015. Functional land management for managing soil functions: a case-study of the trade-off between primary productivity and carbon storage in response to the intervention of drainage systems in Ireland. Land use policy 47, 42–54.

Oberholzer, H.R., Freiermuth Knuchel, R., Weisskopf, P., Gaillard, G., 2012. A novel method for soil quality in life cycle assessment using several soil indicators. Agron. Sustain. Dev. 32, 639–649.

Öborn, I., Edwards, A.C., Witter, E., Oenema C, O., Ivarsson, K., Withers, P.J.A., Nilsson, S.I., Richert Stinzing, A., 2003. Element balances as a tool for sustainable nutrient management: A critical appraisal of their merits and limitations within an agronomic and environmental context. Eur. J. Agron. 20, 211–225.

Oenema, J., Burgers, S., Van Keulen, H., Van Ittersum, M. 2015. Stochastic uncertainty and sensitivities of nitrogen flows on dairy farms in The Netherlands. Agric. Syst. 137, 126–138.

Oenema, J., Burgers, S., Verloop, K., 2010. Multiscale effects of management, environmental conditions, and land Usu. J. Environ. Qual. 39, 2016–2028.

Oenema, J., Koskamp, G.J., Galama, P.J., 2001. Guiding commercial pilot farms to bridge the gap between experimental and commercial dairy farms; the project 'Cows & Opportunities.' Netherlands J. Agric. Sci. 49, 277–296.

Oenema, J., Šebek, L.B., Schröder, J.J., Verloop, J., De Haan, M.H.A., Hilhorst, G.J., 2017. Toetsing van de KringloopWijzer; Gemeten en voorspelde stikstof- en fosfaatproducties van mest en gewas.

Oenema, J., Van Keulen, H., Schils, R.L.M., Aarts, H.F.M., 2011. Participatory farm management adaptations to reduce environmental impact on commercial pilot dairy farms in the Netherlands. NJAS - Wageningen J. Life Sci. 58, 39–48.

Oenema, O., Kros, H., De Vries, W., 2003. Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies. Eur. J. Agron. 20, 3–16.

Oenema, O., Van Liere, L., Plette, S., Prins, T., Van Zeijts, H., Schoumans, O., 2004. Environmental effects of manure policy options in The Netherlands. Water Sci. Technol. 49, 101–108.

PBL, 2017. Evaluatie Meststoffenwet 2016: Syntheserapport. Den Haag.

Peerlkamp, P.K., 1959. A visual method of soil structure evaluation. Instituut voor Bodemvruchtbaarheid, Groningen.

Peigné, J., Vian, J.-F., Cannavacciuolo, M., Lefevre, V., Gautronneau, Y., Boizard, H., 2013. Assessment of soil structure in the transition layer between topsoil and subsoil using the profil cultural method. Soil Tillage Res. 127, 13–25.

Pinheiro, J.C., Bates, D.M., 2000. Mixed-effects models in S and S-PLUS. Springer, New York.

Polla, M., 2018. VESS - Visual Evaluation of Soil Structure (beta).

Powell, J.M., Gourley, C.J.P., Rotz, C.A., Weaver, D.M., 2010. Nitrogen use efficiency: a potential performance indicator and policy tool for dairy farms. Environ. Sci. Policy 13, 217–228.

Prager, K., Baggaley, N., Donnelly, D., Mckee, A., Lilly, A., Cooksley, S., 2014. How can we employ citizen science to determine the extent of soil erosion in Scotland? Edinburgh.

References

149

Pulido Moncada, M., Helwig Penning, L., Timm, L.C., Gabriels, D., Cornelis, W.M., 2014. Visual examinations and soil physical and hydraulic properties for assessing soil structural quality of soils with contrasting textures and land uses. Soil Tillage Res. 140, 20–28.

Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Pérès, G., Rutgers, M., 2012. Soil biodiversity, biological indicators and soil ecosystem services — an overview of European approaches. Curr. Opin. Environ. Sustain. 4, 529–538.

R Core Team, 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ramsey, J.B., 1969. Tests for specification errors in classical linear least-squares regression analysis. J. R. Stat. Soc. 31, 350–371.

Reijneveld, A., Wensem, J. Van, Oenema, O., 2009. Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. Geoderma 152, 231–238.

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2016. Landelijk Meetnet effecten Mestbeleid [WWW Document]. URL https://www.rivm.nl/landelijk-meetnet-effecten-mestbeleid

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2012. De uitspoeling van het stikstofoverschot naar grond- en oppervlaktewater op landbouwbedrijven. Herberekening van uitspoelfracties.

Rijksinstituut voor Volksgezondheid en Milieu (RIVM), 2007. De uitspoeling van het stikstofoverschot naar grond- en oppervlaktewater op landbouwbedrijven.

Roger-estrade, J., Boizard, H., Peigné, J., Sasal, M.C., Guimaraes, R., Piron, D., Tomis, V., Vian, J., Cadoux, S., Ralisch, R., 2016. Improvements of the Profil Cultural Method for a better low-tech field assessment of soil structure under no-till. EGU General Assembly 2016, p. 7724.

Rossel, R.A.V., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal soil sensing: an effective approach for soil measurements in space and time, Advances in Agronomy. Elsevier Inc.

Royal Netherlands Meteorological Institute, 2018. Monthly and yearly mean daily temperatures and amount of precipitation [WWW Document]. De Bilt.

Royal Netherlands Meteorological Institute, 2017. Monthly and yearly mean daily temperatures and amount of precipitation, measured in Leeuwarden. Data retrieved 05-09-2017. [WWW Document]. URL http://www.knmi.nl/nederland-nu/klimatologie/maandgegevens

RVO, 2017. Mestbelijd 2014-2017 Tabellen. Tabel 2 Fosfaatgebruiksnormen [WWW Document]. URL https://www.rvo.nl/sites/default/files/2015/04/Tabel 2 Fosfaatgebruiksnormen 2014-2017%282%29.pdf (accessed 1.8.18).

Sassenrath, G.F., Schneider, J.M., Gaj, R., Grzebisz, W., Halloran, J.M., 2013. Nitrogen balance as an indicator of environmental impact: Toward sustainable agricultural production. Renew. Agric. Food Syst. 28, 276–289.

Schoumans, O.F., Van der Salm, C., Groenendijk, P., 2013. PLEASE: a simple model to determine P losses by leaching. Soil Use Manag. 29, 138–146.

Schröder, J.J., Neeteson, J.J., 2008. Nutrient management regulations in The Netherlands. Geoderma 144, 418–425.

Schröder, J.J., Schulte, R.P.O., Creamer, R.E., Delgado, A., Van Leeuwen, J., Lehtinen, T.,

Rutgers, M., Spiegel, H., Staes, J., Tóth, G., Wall, D.P., 2016a. The elusive role of soil quality in nutrient cycling : a review. Soil Use Manag. 32, 476–486.

Schröder, J.J., Šebek, L.B., Reijs, J.W., Oenema, J., Goselink, R.M.A., Conijn, J.G., De Boer, J., 2016b. Rekenregels van de KringloopWijzer; Achtergronden van BEX, BEA, BEN, BEP en BEC: actualisatie van de 4 maart 2014 versie., PRI-rapport 640.

Schulte, R.P.O., Creamer, R.E., Donnellan, T., Farrelly, N., Fealy, R., O'Donoghue, C., O'hUallachain, D., 2014. Functional land management: a framework for managing soil-based ecosystem services for the sustainable intensification of agriculture. Environ. Sci. Policy 38, 45–58.

Shepherd, T.G., 2009. Visual Soil Assessment. Volume 1. Field guide for pastoral grazing and cropping on flat to rolling country, 2nd ed. Horizons Regional Council, Palmerston North, 119 pp. Rome.

Shipitalo, M.J., Dick, W.A., Edwards, W.M., 2000. Conservation tillage and macropore factors that effect water movement andthe fate of chemicals. Soil Tilllage Res. 53, 167–183.

Šimůnek, J., Šejna, M., Van Genuchten, M.T., n.d. HYDRUS-1D, version 4.16.0110. Code for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated porous media.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Tillage Res. 79, 7–31.

Smaling, E.M.A., Nandwa, S.M., Janssen, B.H., 1997. Soil fertility in Africa is at stake, in: Replenishing Soil Fertility in Africa. American Society of Agronomy and Soil Science Society of America, Madison, USA, pp. 47–61.

Sonneveld, M.P.W., Bouma, J., Veldkamp, A., 2002. Refining soil survey information for a Dutch soil series using land use history. Soil Use Manag. 18, 157–163.

Sonneveld, M.P.W., Heuvelink, G.B.M., Moolenaar, S.W., 2014. Application of a visual soil examination and evaluation technique at site and farm level. Soil Use Manag. 30, 263–271.

Sonneveld, M.P.W., Schoorl, J.M., Veldkamp, A., 2006. Mapping hydrological pathways of phosphorus transfer in apparently homogeneous landscapes using a high-resolution DEM. Geoderma 133, 32–42.

Stolte, J., 1997. Determination of the saturated hydraulic conductivity using the constant head method. In: Manual for soil physical measurements: version 3. Wageningen.

Stoorvogel, J.J., Bakkenes, M., Temme, A.J.A.M., Batjes, N.H., Brink, B.J.E., 2017. S-World: a global soil map for environmental modelling. L. Degrad. Dev. 33, 22–33.

Sykes, J.M., Horrill, A.D., Mountford, M.D., 1983. Use of visual cover assessment as quantitative estimators of some British Woodland taxa. J. Ecol. 71, 437–450.

Thomas, S.M., Ledgard, S.F., Francis, G.S., 2005. Improving estimates of nitrate leaching for quantifying New Zealand's indirect nitrous oxide emissions. Nutr. Cycl. Agroecosystems 73, 213–226.

TNO Geologische Dienst Nederland, 2018. DINOloket. Data en Informatie van de Nederlandse Ondergrond [WWW Document].

United Nations, 2019. Sustainable Development Goals [WWW Document]. URL https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed 1.21.19).

United Nations, 2017. The Sustainable Development Goals Report. New York.

van Bavel, C.H.M., 1949. Mean weight diameter of soil aggregates as a statistical index of aggregation. Soil Sci. Soc. Am. J. 14, 20–23.

Van Beek, C.L., Brouwer, L., Oenema, O., 2003. The use of farmgate balances and soil surface balances as estimator for nitrogen leaching to surface water. Nutr. Cycl. Agroecosystems 67, 233–244.

Van Beek, C.L., Tóth, T., Hagyó, A., Recatalá Boix, L., Añó Vidal, C., Malet, J.P., Maquaire, O., Van den Akker, J.J.H., Van der Zee, S.E.A.T.M. Verzandvoort, S., Simota, C., Kuikman, P.J., Oenema, O., 2010. The need for harmonizing methodologies for assessing soil threats in Europe. Soil Use Manag. 26, 299–309.

Van Beek, C.L., Van der Salm, C., Plette, A.C.C., Van de Weerd, H., 2009. Nutrient loss pathways from grazed grasslands and the effects of decreasing inputs: experimental results for three soil types. Nutr. Cycl. Agroecosystems 83, 99–110.

Van den Berg, M.W., Den Otter, C., 1993. Toelichtingen bij de Geologische Kaart van Nederland 1:50.000, blad Almelo Oost / Denekamp (28O/29). Rijks Geologische Dienst, Haarlem.

Van den Bosch, H., Gitari, J.N., Ogaro, V.N., Maobe, S., Vlaming, J., 1998. Monitoring nutrient flows and economic performance in African farming systems (NUTMON). III. Monitoring nutrient flows and balances in three districts in Kenya. Agric. Ecosyst. Environ. 71, 63–80.

Van der Sluijs, P., De Gruijter, J.J., 1985. Water table classes: a method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. Agric. Water Manag. 10, 109–125.

Van Es, H.M., Sogbedji, J.M., Schindelbeck, R.R., 2006. Effect of manure application timing, crop, and soil type on nitrate leaching. J. Environ. Qual. 35, 670–679.

Van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44, 892–898.

Van Groenigen, J.W., Lubbers, I.M., Vos, H.M.J., Brown, G.G., De Deyn, G.B., Van Groenigen, K.J., 2014. Earthworms increase plant production: a meta-analysis. Sci. Reprots 4, 1–7.

Van Groenigen, J.W., Velthof, G.L., Oenema, O., Van Groenigen, K.J., Van Kessel, C., 2010. Towards an agronomic assessment of N2O emissions: a case study for arable crops. Eur. J. Soil Sci. 61, 903–913.

Van Leeuwen, M.M.W.J., Heuvelink, G.B.M., Wallinga, J., De Boer, I.J.M., Van Dam, J.C., Van Essen, E.A., Moolenaar, S.W., Verhoeven, F.P.M., Stoorvogel, J.J., Stoof, C.R., 2018. Visual soil evaluation: reproducibility and correlation with standard measurements. Soil Tillage Res. 178, 167–178.

Van Leeuwen, M.M.W.J., Van Middelaar, C.E., Oenema, J., Van Dam, J.C., Stoorvogel, J.J., Stoof, C.R., De Boer, I.J.M., 2019. The relevance of spatial scales in nutrient balances on dairy farms. Agric. Ecosyst. Environ. 269, 125–139.

Vellinga, T. V, André, G., 1999. Sixty years of Dutch nitrogen fertiliser experiments, an overview of the effects of soil type, fertiliser input, management and of developments in time. Netherlands J. Agric. Sci. 47, 215–241.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Fourth Edition. Springer, New York.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., Van Genuchten, M.T., 2010. Using pedotransfer functions to estimate the van Genuchten – Mualem soil

hydraulic properties: a review. Vadose Zo. J. 9, 795–820.

Wageningen Environmental Research, 2017. Grondwatertrappenkaart 2017 - Geactualiseerde grondwatertrappen (Gt) kaart 2017.

Wageningen Environmental Research (Alterra), 2015. Bodemkaart50000_versie2014 - Groepslayer Bodemkaart 50000 met update voor de veengebieden (versie 2014).

Watson, C.A., Atkinson, D., 1999. Using nitrogen budgets to indicate nitrogen use efficiency and losses from whole farm systems: a comparison of three methodological approaches. Nutr. Cycl. Agroecosystems 53, 259–267.

Wendroth, O., Ehlers, W., Hopmans, J.W., Kage, H., Halbertsma, J., Wösten, J.H.M., 1993. Reevaluation of the evaporation method for determining hydraulic functions in unsaturated soils. Soil Sci. Soc. Am. J. 57, 1436–1443.

WEnR, 2016. Landelijk Grondgebruik Nederland versie 7 met Basis Registratie Percelen 2015.

Wienhold, B.J., Karlen, D.L., Andrews, S.S., Stott, D.E., 2009. Protocol for indicator scoring in the soil management assessment framework (SMAF). Renew. Agric. Food Syst. 24, 260–266.

Wills, S.A., Burras, C.L., Sandor, J.A., 2007. Prediction of soil organic carbon content using field and laboratory measurements of soil color. Soil Sci. Soc. Am. J. 71, 380–388.

Wind, G.P., 1968. Capillary conductivity data estimated by a simple method. In: P.E. Rijtema and H. Wassink, Water in the unsaturated zone: Proceedings of the Wageningen Symposium. 19–23 June 1966. Vol. 1. Int. Assoc. Sci. Hydrol. Publ., Gentbrugge, the Netherlands. 181–19.

Wooldridge, J.M., 2009. Introductory Econometrics, Fourth Edition, 4th ed. South-Western Cengage Learning, Mason, USA.

Wösten, H., De Vries, F., Hoogland, T., Massop, H.T.L., Veldhuizen, A.A., Vroon, H., Wesseling, J., Heijkers, J., Bolman, A., 2013. BOFEK2012, de nieuwe, bodemfysische schematisatie van Nederland [BOFEK2012; the new soil physical schematization of the Netherlands, in Dutch]. Alterra Rep. 92.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185.

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251, 123–150.

Zeileis, A., Hothorn, T., 2002. Diagnostic Checking in Regression Relationships. R News 2, 7–10.

Datasets

Van Leeuwen, M.M.W.J., Bakker, E.J., Stoof, C.R., De Boer, I.J.M., Van Dam, J.C., De Haan, M.H.A., Oenema, J. 2019b. Quantitative visual soil examination to evaluate soil functions on dairy farms. 4TU. https://doi.org/10.4121/uuid:b3b1ccce-0d53-43b0-9802-ae99f72154ee

Van Leeuwen, M.M.W.J., Schols, A.M., Quik, C. 2019c. Quantitative visual soil observations for visual soil evaluation on dairy farms. 4TU. https://doi.org/10.4121/uuid:237f0826-d04b-43e0-82d3-d2f4a6bf4ce4

References

Summary

Soils have many functions, such as producing crops, filtering water by binding nutrients and solutes, serving as a carbon pool and providing a habitat for all kinds of soil fauna. Soil quality can be assessed using easily obtainable soil information, such as soil maps and visual soil evaluation (VSE). Soil series, obtainable from soil maps, can be used in nutrient balances to provide site-specific soil information. This is relevant, as it was unknown whether nutrient balances at farm level represent nutrient balances at field level, given the potential within-farm diversity of soil series. Regarding VSE, in this thesis quantitative visual observations were used instead of using qualitative soil quality scores. For quantitative visual observations, the reproducibility was unknown, as well as the correlation with standard field or laboratory measurements. Furthermore, it was unknown how VSE correlates with several soil functions, such as crop production and storing, filtering, and transforming nutrients and water. Also, it was unknown how soil texture affects the correlation between VSE and these soil functions. The aim of this thesis is, therefore, to evaluate whether easily obtainable soil information (soil maps and visual soil evaluation) can be used to assess soil functions. Fieldwork, laboratory analyses and environmental modelling were performed to study the relationships between easily obtainable soil information and other indicators for soil functioning. This thesis thereby contributes to the assessment of soil quality in agricultural land, which can contribute to environmentally sustainable crop production. To reach the thesis aim and to bridge the knowledge gaps, three research questions were formulated: 1) to what extent does the required spatial scale in nutrient balances depend on the level of soil variation, 2) to what extent are quantitative visual soil observations reproducible, and do they correlate with standard field or laboratory measurements, and 3) can quantitative visual soil observations be used to assess soil functioning?

After the general introduction (Chapter 1), Chapter 2 discusses the role of spatial scales in nutrient balances on dairy farms. On five Dutch dairy farms, field level nitrogen (N) and phosphorus (P) balances were associated with soil series that were obtained from a 1:50,000 soil map. The analyses shows that field nutrient balance variation is not only explained by soil series, but also by management factors and soil conditions. Nevertheless, nutrient balances at field level are more informative than nutrient balances at the crop level or farm level, because at field level extremes in N and P deficits or surpluses are visible, whereas at crop or farm level these extremes are not visible because they disappear by averaging the field balances. This analysis therefore

shows that field-level balances are more relevant than farm-level balances, for decision making by farmers to reduce nutrient losses to the environment.

Chapter 2 furthermore discusses the effect of the used spatial scale of soil-dependent leaching factors on estimated nitrate leaching at farm level. Leaching factors are parameters that are multiplied with the soil nitrogen balance, to calculate the nitrate leaching. For this, the Dutch nutrient balance model 'Annual farm Nutrient Cycle Assessment' was used. Results indicate that if soil variation within a farm is high, estimated nitrate leaching at crop or farm level does not represent nitrate leaching at field level, because leaching factors vary across fields due to contrasting soil types. In contrast, if soil variation within a farm is low, estimated nitrate leaching at crop or farm level is at a sufficiently detailed spatial scale, because leaching factors across fields are similar due to similar soil types. Also, nitrate leaching calculations at crop or farm level underestimate nitrate leaching if fields have nitrogen deficits, because fields with a nitrogen deficit will lower the averaged nitrogen surplus at crop or farm level. To conclude, nutrient balances at field level are more informative than nutrient balances at the crop or farm level. This was concluded from the fact that extremes in N and P deficits or surpluses at field level are not visible at crop or farm level, and the fact that nitrate leaching calculations at field level are more accurate than calculations at crop or farm level. This research therefore shows that field-level nutrient balances are relevant for decision making by farmers who aim to reduce nutrient losses to the environment.

Chapter 3 uses quantitative visual observations in VSE. The two aims of this Chapter were to assess the reproducibility of quantitative visual observations, and to evaluate the correlation of quantitative visual observations with standard field or laboratory measurements. First, the reproducibility of quantitative visual observations was assessed using a linear mixed-effect modelling approach. Five Dutch dairy farmed sites were selected that were located on clayey (n=2), peaty (n=1) and sandy (n=2) soils. Tested potential subjectivity in the visual observations is the effect of the observers' background (farmer, n=8, or soil scientist, n=11) and the result of systematic or random observations by the individual observers. The subjectivity due to the observers' background is only significantly present in the observations of soil compaction depth: on average farmers observe the depth of soil compaction 8 cm deeper than what soil scientists observe (the overall mean observed compaction depth was 25 cm). The subjectivity due to individual observers' assessment is present in seven out of eight of the quantitative visual observations. Systematic subjectivity is found in the observations of the fraction of largest soil structural elements, earthworm count, gley mottles and the depth of soil compaction. For these visual observations relative soil quality differences between the sites can be detected if one person assesses the various sites. Next to systematic subjectivity, random subjectivity is found in the observations

of grass cover and biopore count. For these observations the reproducibility is the highest when the average observed value is taken from a group of observers, or from several observations at the same site. The visual observation of root count is the least reproducible observation, and the observation of gley mottles is the most reproducible observation.

Chapter 3 does not only asses the reproducibility of quantitative visual observations, but also the correlation between quantitative visual observations and standard field or laboratory measurements. For this study, 26 sites were sampled that were located in the North of the Netherlands, on sandy (n=11), peaty (n=7) and clayey (n=8) soils. Quantitative visual observations that strongly correlate with standard field or laboratory measurements are grass cover, root count, Munsell soil colour value, fraction of soil structural elements, degree of soil compaction and maximum rooting depth. The correlations between each of these quantitative visual observations and standard field or laboratory measurements are affected by soil texture. The remaining quantitative visual soil observations could not be validated, likely because of the influence of the dry soil conditions (which negatively affected the visual assessment of soil compaction), and validators were chosen that did not represent similar soil properties and measurement units as the visual observations. The reproducibility study and the validation study show that quantitative visual soil observations are moderately reliable, given the high probability that systematic errors are made by observers, and given the fact that more than half of the quantitative visual soil observations could not be validated. Reproducibility, however, can be improved with more training. It is expected that a repetition of the validation study can result in more accurate visual observations if the soil moisture conditions are closer to field capacity at the time of VSE deployment, and if the right validators are chosen.

Chapter 4 investigates the correlation between quantitative visual soil observations and indicators for crop production (crop yield gaps as result of water surplus or water limitation, oxygen and water stress, and plant available water) and the storage, filtering and transformation of water and nutrients (water storage capacity, nitrate and phosphate concentrations in drain or groundwater). For this study, 25 sites were selected that were located on five Dutch dairy farms on sandy and clayey soils. On each site quantitative visual soil observations were performed. On the same sites, soil samples were collected to measure the soil function indicators plant available water and water storage capacity. Ecohydrological model SWAP (Soil-Water-Atmosphere-Plant) was used to determine the following soil function indicators: yield gap, oxygen and drought stress for a wet year (2001), a dry year (2003), and a 'normal' year (2016). Furthermore, measured nitrate and phosphate concentrations in drain and groundwater were used as soil function indicators. A stepwise linear regression approach showed that, except for crop drought stress in a dry year, the soil function

157

indicators significantly correlate with a set of quantitative visual soil observations. As it was found that the interaction with clay content improved the correlation between several quantitative visual observations and soil function indicators, it may be ideal to include soil texture effects in future VSEs. This Chapter shows that soil functions can be assessed using a set of quantitative visual observations.

The synthesis (Chapter 5) presents the overall conclusions of the previous chapters, which is that easily obtainable soil information can be used to assess soil functioning. The use of soil series can improve environmental assessments such as nutrient balances and life cycle assessment. Besides, soil series can be used in the assessment of the soil functions 'serving as a carbon pool', and 'serving as archive for geological and cultural heritage'. Next to the use of soil series to assess soil functions, VSE can be used to assess soil functions that are driven by soil physical properties, such as crop production and storing, filtering and transforming nutrients and water. Also, VSE shows potential to assess the soil functions 'serving as carbon pool' and 'serving as archive for geological and cultural heritage'. The combination of soil series and VSE is an opportunity for assessing soil functioning. Soil maps provide the soil inherent characteristics that can be used to define the soils' potential to function, and VSE assesses dynamic soil properties that can easily change after management actions. The combination of soil series and VSE, therefore, provides the soil information that is needed to assess and evaluate soil functions in relation to the soils' potential to function. An opportunity is the implementation of pedotransfer functions, which relate soil series to soil hydraulic parameters that can be used to derive several crop production indicators and an indicator for storing, filtering, and transforming nutrients and water (Chapter 4). The combination of soil series and VSE contributes to a better understanding of soil quality and the environmental performance of agricultural land.

Chapter 5 also discusses the availability of soil maps and field-level nutrient balances. Visual soil profile descriptions could be used to identify the soil series if detailed soil maps are not available. If field-level nutrient balances are not available, they could be estimated using known parameters. For intensive agricultural systems, for example, manually applied N inputs (chemical and organic fertilizer) at field level could be used as a proxy for field-level N balances. Likewise, total P inputs at field level could be used as a proxy for field-level P balances. For extensive agricultural systems, the best proxy for nutrient balances is the combination of nutrient inputs and nutrient stocks. The development of proximal sensors is an opportunity to estimate nutrient stocks.

Chapter 5 furthermore discusses the quantitative visual observations that can be used to evaluate soil quality. For each soil function indicator the number and type of selected visual observations is different. Also, the selected set of quantitative visual

observations likely varies between regions: the context of climate, management and soil affects the drivers behind soil functioning and the importance of each visual observation in relation to each soil function. This thesis, however, shows that VSE can be used to assess soil functions, and that the methodology shows potential to be implemented in other regions. VSE thereby contributes to better understanding of soil quality on dairy farms. Lastly, VSEs should be complemented with site-specific management options in order to be able to improve the environmental sustainability of agricultural land management after VSE deployment.

Summary

Samenvatting

Bodems hebben veel functies, zoals het produceren van gewassen, het filteren van water door het binden van nutriënten en opgeloste stoffen, het dienen als opslag voor koolstof en het bieden van een habitat voor allerlei soorten bodemfauna. De bodemkwaliteit kan worden beoordeeld met behulp van gemakkelijk verkrijgbare bodeminformatie, zoals bodemkaarteenheden (aangeduid op bodemkaarten) en visuele bodemevaluatie. Bodemkaarteenheden kunnen worden gebruikt in nutriëntenbalansen om locatie-specifieke bodeminformatie te leveren. Dit is van belang, omdat het niet bekend was of nutriëntenbalansen op bedrijfsniveau representatief zijn voor nutriëntenbalansen op veldniveau, gegeven de potentiële diversiteit van bodems binnen een bedrijf. Met betrekking tot de visuele bodemevaluatie zijn in dit proefschrift kwantitatieve visuele waarnemingen gebruikt, in plaats van de meer gebruikelijke kwalitatieve scores voor bodemkwaliteit. Voor kwantitatieve visuele waarnemingen was de reproduceerbaarheid onbekend, evenals de correlatie met standaard veld- of laboratoriummetingen. Verder was het onbekend hoe visuele bodemevaluaties correleren met verschillende bodemfuncties, zoals gewasproductie en opslag, filteren en transformatie van nutriënten en water. Het is ook onbekend hoe de bodemtextuur de correlatie tussen visuele bodemevaluatie en deze bodemfuncties beïnvloedt. Het hoofddoel van dit proefschrift is daarom om te evalueren of gemakkelijk verkrijgbare bodeminformatie (bodemkaarten en visuele bodemevaluatie) kan worden gebruikt om bodemfuncties te beoordelen. Veldwerk, laboratoriumanalyses en modelleerwerk werden uitgevoerd om de relaties te bestuderen tussen gemakkelijk verkrijgbare bodeminformatie en andere bodemfunctie-indicatoren. Dit proefschrift draagt daarmee bij aan de beoordeling van de bodemkwaliteit in landbouwgrond, wat kan bijdragen aan duurzame gewasproductie. Om het hoofddoel te bereiken werden drie onderzoeksvragen geformuleerd: 1) in hoeverre is de vereiste ruimtelijke schaal in nutriëntenbalansen afhankelijk van de mate van bodemvariatie, 2) in hoeverre zijn kwantitatieve visuele bodemwaarnemingen reproduceerbaar, en correleren ze met standaard veld- of laboratoriummetingen, en 3) kunnen kwantitatieve visuele bodemwaarnemingen worden gebruikt om het functioneren van de bodem te beoordelen?

Na de algemene inleiding (Hoofdstuk 1), bespreekt Hoofdstuk 2 de rol van ruimtelijke schalen in nutriëntenbalansen op melkveebedrijven. Op vijf Nederlandse melkveebedrijven werden op veldniveau stikstof (N) en fosfor (P) in verband gebracht met bodemtypes (bodemkaarteenheden) die werden verkregen uit een 1:50.000 bodemkaart. Uit de analyses blijkt dat de variatie in nutriëntenbalansen op veldniveau

niet alleen wordt verklaard door bodemdiversiteit, maar ook door managementfactoren en bodemgesteldheid. Desondanks zijn nutriëntenbalansen op veldniveau informatiever dan nutriëntenbalansen op gewasniveau of op bedrijfsniveau, omdat op veldniveau extremen in N en P tekorten of overschotten zichtbaar zijn, terwijl op gewas- of bedrijfsniveau deze extremen niet zichtbaar zijn omdat ze verdwijnen door het uitmiddelen van de veldbalansen. Deze analyse toont daarom aan dat veldbalansen relevanter zijn voor besluitvorming door boeren dan bedrijfsbalansen, om de nutriëntenverliezen naar het milieu te verminderen.

Hoofdstuk 2 gaat dieper in op het effect van de gebruikte ruimtelijke schaal van bodemafhankelijke uitspoelfracties, op de geschatte hoeveelheid nitraatuitspoeling op bedrijfsniveau. Een uitspoelfractie is een constante die wordt vermenigvuldigd met het bodem stikstofoverschot, om de hoeveelheid nitraatuitspoeling te berekenen. Voor deze vraag werd de Nederlandse nutriëntenbalans 'KringloopWijzer' gebruikt. Uit de resultaten blijkt dat als de bodemvariatie binnen een melkveebedrijf hoog is, de geschatte nitraatuitspoeling op gewas- of bedrijfsniveau niet de nitraatuitspoeling op veldniveau vertegenwoordigt, omdat uitspoelfracties variëren tussen de velden als gevolg van contrasterende bodemtypes. In tegenstelling, als de bodemvariatie binnen een melkveebedrijf laag is, is de geschatte nitraatuitspoeling op gewas- of bedrijfsniveau op een voldoende gedetailleerde ruimtelijke schaal, omdat uitspoelfracties tussen velden vergelijkbaar zijn vanwege de vergelijkbare bodemtypes. Ook onderschatten de berekeningen van nitraatuitspoeling op gewas- of bedrijfsniveau de totale nitraatuitspoeling als er velden zijn met een stikstoftekort, omdat tekorten op veldniveau (negatieve bodem stikstofbalansen) de gemiddelde stikstofbalans op gewas- of bedrijfsniveau verlagen. Concluderend, nutriëntenbalansen op veldniveau zijn informatiever dan nutriëntenbalansen op gewas- of bedrijfsniveau. Dit werd geconcludeerd uit het feit dat extremen in N en P tekorten of overschotten op veldniveau niet zichtbaar zijn op gewas- of bedrijfsniveau, en het feit dat berekeningen van nitraatuitspoeling op veldniveau nauwkeuriger zijn dan berekeningen op gewas- of bedrijfsniveau. Dit onderzoek toont daarom aan dat nutriëntenbalansen op veldniveau relevant zijn voor de besluitvorming door boeren die ernaar streven de nutriëntenverliezen naar het milieu te verminderen.

Hoofdstuk 3 gebruikt kwantitatieve visuele waarnemingen in visuele bodemevaluatie. De twee doelstellingen van dit hoofdstuk waren om de reproduceerbaarheid van kwantitatieve visuele waarnemingen te beoordelen en de correlatie van kwantitatieve visuele waarnemingen met standaard veld- of laboratoriummetingen te evalueren. Eerst werd de reproduceerbaarheid van kwantitatieve visuele waarnemingen beoordeeld met behulp van lineair gemengde ('*linear mixed-effect*') modellen. Er werden vijf locaties geselecteerd die onderdeel waren van een melkveebedrijf, en die zich bevonden op zandige (n=2), venige (n=1) en kleiige (n=2) bodems in Nederland.

De geteste potentiële subjectiviteit in de visuele waarnemingen is het effect van de achtergrond van de waarnemer (boer, n=8, of bodemwetenschapper, n=11) en het resultaat van systematische of willekeurige waarnemingen door de individuele waarnemers. De subjectiviteit vanwege de achtergrond van de waarnemer is alleen significant aanwezig bij de waarnemingen van de diepte waar bodemcompactie begint: gemiddeld observeren boeren de diepte van bodemcompactie 8 cm dieper dan wat de bodemwetenschappers observeren (de totale gemiddelde waargenomen diepte van bodemcompactie was 25 cm). De subjectiviteit als gevolg van de waarnemingen door individuele waarnemers is aanwezig in zeven van de acht kwantitatieve visuele waarnemingen. Systematische subjectiviteit is aanwezig in de waarnemingen van de fractie van de grootste bodemstructuurelementen, het aantal wormen, het aantal roestvlekken en de diepte van de bodemcompactie. Voor deze waarnemingen geldt dat als één persoon verschillende locaties beoordeelt, de relatieve verschillen in bodemkwaliteit tussen de locaties kunnen worden gedetecteerd. Naast systematische subjectiviteit is willekeurige subjectiviteit gevonden in de waarnemingen van grasbedekking en het aantal bioporiën. Voor deze waarnemingen is de reproduceerbaarheid het hoogst wanneer de gemiddelde waargenomen waarde wordt genomen van een groep waarnemers of van een aantal waarnemingen op dezelfde plaats. De minst reproduceerbare waarneming is het aantal wortels, en de meest reproduceerbare waarneming is het aantal roestvlekken.

In hoofdstuk 3 wordt naast de reproduceerbaarheid van kwantitatieve visuele waarnemingen ook de correlatie tussen kwantitatieve visuele waarnemingen en standaard veld- of laboratoriummetingen beoordeeld. Voor deze studie werden 26 locaties bemonsterd die zich in het noorden van Nederland bevonden, op zandige (n=11), venige (n=7) en kleiige (n=8) bodems. Kwantitatieve visuele waarnemingen die sterk correleren met standaard veld- of laboratoriummetingen zijn grasbedekking, het aantal wortels, Munsell bodemkleur '*value*', fractie van bodemstructuurelementen, mate van bodemcompactie en de maximale worteldiepte. De correlatie tussen elk van deze visuele waarnemingen en standaard veld- of laboratoriummetingen wordt beïnvloed door bodemtextuur. De andere kwantitatieve visuele bodemwaarnemingen konden niet worden gevalideerd, waarschijnlijk vanwege de droge bodems (waardoor de visuele beoordeling van bodemcompactie negatief werd beïnvloed), en doordat er validatiemetingen werden gekozen die geen vergelijkbare bodemeigenschappen of meeteenheden hadden als de visuele waarnemingen. De studies van reproduceerbaarheid en de validatie tonen aan dat kwantitatieve visuele bodemwaarnemingen matig betrouwbaar zijn, gezien de grote waarschijnlijkheid dat systematische fouten worden gemaakt door waarnemers, en doordat meer dan de helft van de kwantitatieve visuele bodemwaarnemingen niet gevalideerd kon worden. De reproduceerbaarheid kan echter worden verbeterd met meer training. Een herhaling

van de validatiestudie zou kunnen resulteren in meer accurate visuele waarnemingen, als op het moment van waarnemen de bodem dichtbij veldcapaciteit is, en als de juiste validatiemetingen worden gekozen.

Hoofdstuk 4 onderzoekt de correlatie tussen kwantitatieve visuele bodemwaarnemingen en indicatoren voor gewasproductie (verschil tussen de potentiële en de actuele gewasopbrengst als gevolg van wateroverschot of waterlimitatie, zuurstof- en waterstress en het beschikbaar water voor de plant) en de opslag, filtering en transformatie van water en nutriënten (bergingsvermogen van water, nitraat- en fosfaatconcentraties in drainage- of grondwater). Voor dit onderzoek werden 25 locaties geselecteerd die zich bevonden op vijf Nederlandse melkveebedrijven op zandige en kleiige bodems. Op deze locaties werden bodemmonsters genomen om de volgende bodemfunctie-indicatoren te meten: voor planten beschikbaar water en het bergingsvermogen. Ecohydrologisch model SWAP (bodem-water-atmosfeer-plant) werd gebruikt om de volgende indicatoren van de bodemfuncties te bepalen: verschil tussen de potentiële en de actuele gewasopbrengst, zuurstof- en droogtestress voor een nat jaar (2001), een droog jaar (2003) en een 'normaal' jaar (2016). Verder werden de gemeten nitraat- en fosfaatconcentraties in drainage- en grondwater gebruikt als bodemfunctie-indicatoren. Een stapsgewijze lineaire regressiebenadering toonde aan dat, afgezien van droogtestress in een droog jaar, bodemfunctie-indicatoren significant correleren met een set van kwantitatieve visuele bodemwaarnemingen. Omdat bleek dat de interactie met kleigehalte de correlatie tussen verschillende kwantitatieve visuele waarnemingen en bodemfunctie-indicatoren verbeterde, zou het raadzaam zijn om bodemtextuureffecten mee te nemen in toekomstige visuele bodemevaluaties. Dit hoofdstuk laat zien dat bodemfuncties kunnen worden beoordeeld op basis van een set kwantitatieve visuele waarnemingen.

De synthese (hoofdstuk 5) presenteert de algemene conclusies van de voorgaande hoofdstukken, namelijk dat gemakkelijk verkrijgbare bodeminformatie kan worden gebruikt om het functioneren van de bodem te beoordelen. Het gebruik van bodemkaarteenheden kan omgevingsstudies zoals nutriëntenbalansen en levenscyclusanalyses verbeteren. Daarnaast kunnen bodemkaarteenheden worden gebruikt bij de beoordeling van de bodemfuncties 'dienen als opslag voor koolstof' en 'dienen als archief voor geologisch en cultureel erfgoed'. Naast het gebruik van bodemkaarteenheden om bodemfuncties te bepalen, kan visuele bodemevaluatie worden gebruikt voor het beoordelen van bodemfuncties die worden aangestuurd door bodemfysische eigenschappen, zoals gewasproductie en de opslag, filteren en transformeren van nutriënten en water. Ook toont visuele bodemevaluatie potentie om de bodemfuncties 'dienen als opslag voor koolstof' en 'dienen als archief voor geologisch en cultureel erfgoed' te beoordelen. De combinatie van

bodemkaarteenheden en visuele bodemevaluatie is een kans om het functioneren van de bodem te beoordelen. Bodemkaarten bevatten informatie over de inherente bodemeigenschappen, die kunnen worden gebruikt om de potentie van de bodem om te functioneren te bepalen. Een visuele bodemevaluatie bepaalt de dynamische bodemeigenschappen die gemakkelijk kunnen veranderen als gevolg van beheer. De combinatie van bodemkaarteenheden en visuele bodemevaluatie biedt daarom de informatie die nodig is voor het beoordelen en evalueren van bodemfuncties in relatie tot de potentie van de bodem om te functioneren. Een kans is de implementatie van pedotransferfuncties, die bodemhydrologische parameters kunnen relateren aan bodemkaarteenheden, die vervolgens gebruikt kunnen worden om verschillende gewasproductie-indicatoren af te leiden en een indicator voor het opslaan, filteren en transformeren van nutriënten en water (hoofdstuk 4). De combinatie van bodemkaarteenheden en visuele bodemevaluatie draagt bij aan een beter begrip van de bodemkwaliteit en milieuprestaties van landbouwgrond.

Hoofdstuk 5 bespreekt ook de beschikbaarheid van bodemkaarten en nutriëntenbalansen op veldniveau. Als er geen gedetailleerde bodemkaarten beschikbaar zijn kunnen visuele bodemprofielbeschrijvingen worden gebruikt om het bodemtype te bepalen. Als er geen nutriëntenbalansen op veldniveau beschikbaar zijn kunnen die worden geschat met behulp van andere parameters die wel bekend zijn. Voor intensieve landbouwsystemen kan bijvoorbeeld aangewende N uit chemische en organische mest (op veldniveau) worden gebruikt als een schatting voor de N balans op veldniveau. De totale P input op veldniveau kan worden gebruikt als een schatting voor de P balans op veldniveau. Voor extensieve landbouwsystemen is de combinatie van nutriënteninputs en nutriëntenvoorraden in de bodem de beste indicator voor nutriëntenbalansen. De ontwikkeling van sensoren is een mogelijkheid om bodem nutriëntenvoorraden te schatten.

Hoofdstuk 5 bespreekt verder de kwantitatieve visuele waarnemingen die nodig zijn om de bodemkwaliteit te evalueren. Voor iedere bodemfunctie is het aantal en type geselecteerde visuele waarnemingen verschillend. De geselecteerde set van kwantitatieve visuele waarnemingen varieert waarschijnlijk ook tussen regio's: de context van klimaat, beheer en bodem beïnvloedt de mechanismes die het functioneren van de bodem bepalen, maar ook de belangrijkheid van iedere visuele waarneming in relatie tot iedere bodemfunctie. Dit proefschrift laat echter zien dat visuele bodemevaluatie kan worden gebruikt om bodemfuncties te beoordelen en dat de methodologie potentie heeft om in andere regio's te worden toegepast. Visuele bodemevaluatie draagt daarmee bij aan een beter begrip van de bodemkwaliteit op melkveebedrijven. Ten slotte moeten visuele bodemevaluaties worden aangevuld met locatie-specifieke suggesties om de duurzaamheid van het beheer van landbouwgrond na de inzet van visuele bodemevaluatie te kunnen verbeteren.

165

About the author

Maricke van Leeuwen was born at 27 October 1990, in Haarlem, the Netherlands. She grew up in a warm family with two brothers and two sisters. Maricke was always fascinated by nature, she enjoyed being outside or going for holidays to the Swiss Alps. It is not surprising, therefore, that Maricke studied Soil, Water, Atmosphere (BSc, 2008-2012) and Earth and Environment (MSc, 2011-2013) at Wageningen University. During her MSc she was member of the Navigators student association in Wageningen. This period was valuable for Maricke, as she learnt about prayer lead by the Holy Spirit, and about God's loving father heart.

After finishing her studies, Maricke continued with a PhD in the Soil Geography and Landscape group at Wageningen University (2014-2019). She studied whether visual soil evaluation could be used to assess soil quality on dairy farms. Although the PhD was often difficult, there were many moments to be thankful for. The most exciting moments to be thankful for were the fieldwork campaigns to examine soils at dairy farms; soil-physical laboratory work; visiting the Soil and Water Lab at Cornell University; going to conferences; following PhD courses; assisting in soil-related field practicals for students; crying because of laughing very hard in the office; Cake Fridays with colleagues; learning to overcome/dealing with negative thoughts and fear of failure; and after all the struggles with scientific writing: to get the first papers published. Furthermore, another exciting period was that she married with Arco Modderkolk, who is a fantastic husband (20 April 2018). Loving her husband and having a passion for soils, it is all connected now in her new family name (Modderkolk) that has something to do with mud.

Right after submitting the PhD, Maricke and Arco went to Spain to work in a hostel (Albergue Oasis Trails) where pilgrims stay who walk the Camino de Santiago. Maricke knows that not only pilgrims make a walk, but everyone has their story and walks a personal journey. In her walk, Maricke wants to trust the true Shepherd and follows him, as she believes that he leads you on right paths to green pastures of peace, and he protects you when going through deep valleys. After coming back from Spain, Maricke is excited to continue working on soil topics in the consultancy company Aequator Groen & Ruimte.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)
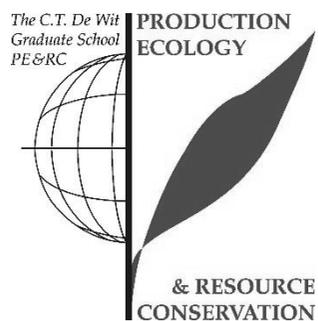
Review of literature (4.5 ECTS)
- Review of visual soil assessment techniques

Writing of project proposal (1 ECTS)
- Visual soil assessment for improved farm and ecosystem management

Post-graduate courses (3.6 ECTS)
- Land dynamics: getting to the bottom of Mount Kenya; PE&RC, WASS, SENSE (2015)
- Mixed linear models; PE&RC (2017)

Laboratory training and working visits (3 ECTS)
- Assessment methods for soil health; Cornell University (2015)

Deficiency, Refresh, Brush-up courses (2.1 ECTS)
- Sustainable development of animal systems; Animal Production Systems group (2014)
- Basic statistics; PE&RC (2014)

Competence strengthening / skills courses (4.5 ECTS)
- Personal leadership and self-direction while performing; HRM/ESG (2014)
- Stress identification & management; WGS (2014)
- Competence assessment; WGS (2014)
- Techniques for writing and presenting a scientific paper; WGS (2015)
- PhD Peer consultation; PE&RC (2015, 2016)
- Career orientation; WGS (2017)
- ESG Career coaching; ESG (2017, 2018)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.7 ECTS)
- PE&RC First year's weekend (2014)

- PE&RC Middle year's weekend (2015)
- PE&RC Last year's weekend (2017)
- WGS PhD Workshop carousel (2018)

Discussion groups / local seminars / other scientific meetings (4.7 ECTS)
- Project meeting: 'Goud van oud' (2014)
- BodemConditieScore workshop for farmers in Heythuysen (2014)
- NBV Theme day: 'Het hoogveenlandschap en turfwinning' (2014)
- BodemConditieScore for farmers in Groningen (2014)
- BodemConditieScore workshop for farmers in study group 'Bodemmanagement'; Groningen (2014)
- BodemConditieScore workshop for farmers and policy makers of Ministry of Economic Affairs; Bodegraven (2015)
- Seminar at Soil and Water Lab; Cornell University: 'Quantitative approach on visual soil assessment: validation and reproducibility' (2015)
- Plant soil interactions discussion group meeting (2016)
- Soil health seminar in Wageningen (2016)
- Alterra symposium: 'Bodemverdichting maakt veel los' (2016)
- 1st Soil quality in life cycle management workshop (2015)
- LCA Food conference; Dublin
- Sustainable intensification of agricultural systems discussion group meeting (2017)
- NBV Theme day: 'Bodemdaling in het veenweidegebied' (2017)
- NWO symposium: 'Co-creation? Naturally!' (2017)
- Meeting advisory board project Koeien&Kansen (2017)
- Arts meets science (2018)

International symposia, workshops and conferences (10.8 ECTS)
- Wageningen soil conference; poster presentations (2015,2017)
- European geosciences union, annual meeting; PICO and poster presentation (2016), oral presentation (2017)
- Grassland and tillage visual soil evaluation workshop organized by the International Soil Tillage Research Organization; oral presentation; Ireland (2017)

Lecturing / Supervision of practicals / tutorials (7.95 ECTS)
- Integrated natural resource management in organic agriculture (2014)
- Integration course soil, water, atmosphere (2014, 2015, 2016)
- Landscape geography (2015, 2016, 2017)
- Soil I (2016, 2017)
- Capita selecta (2018)