Agricultural Systems 156 (2017) 25–33

Contents lists available at ScienceDirect

Agricultural Systems

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

Benchmarking nutrient use efficiency of dairy farms: The effect of epistemic uncertainty

W. Mu[a,*,1], E.A. Groen[a,1], C.E. van Middelaar[a], E.A.M. Bokkers[a], S. Hennart[b], D. Stilmant[b], I.J.M. de Boer[a]

[a] Animal Production Systems group, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands
[b] Farming Systems, Territory and Information Technologies Unit, Walloon Agricultural Research Centre, 100 rue du Serpont, 6800 Libramont, Belgium

ARTICLE INFO

Keywords:
Nutrient use efficiency
Clustering farms
Farm comparison
Monte Carlo simulation
Environmental performance
Global sensitivity analysis

ABSTRACT

The nutrient use efficiency (NUE) of a system, generally computed as the amount of nutrients in valuable outputs over the amount of nutrients in all inputs, is commonly used to benchmark the environmental performance of dairy farms. Benchmarking the NUE of farms, however, may lead to biased conclusions because of differences in major decisive characteristics between farms, such as soil type and production intensity, and because of epistemic uncertainty of input parameters caused by errors in measurement devices or observations. This study aimed to benchmark the nitrogen use efficiency (NUE_N; calculated as N output per unit of N input) of farm clusters with similar characteristics while including epistemic uncertainty, using Monte Carlo simulation. Subsequently, the uncertainty of the parameters explaining most of the output variance was reduced to examine if this would improve benchmarking results. Farms in cluster 1 (n = 15) were located on sandy soils and farms in cluster 2 (n = 17) on loamy soils. Cluster 1 farms were more intensive in terms of milk production per hectare and per cow, had less grazing hours, and fed more concentrates compared to farms in cluster 2. The mean NUE_N of farm in cluster 1 was 43%, while in cluster 2 it was 26%. Input parameters that explained most of the output variance differed between clusters. For cluster 1, input of feed and output of roughage were most important, whereas for cluster 2, the input of mineral fertilizer (or fixation) was most important. For both clusters, the output of milk was relatively important. Including the epistemic uncertainty of input parameters showed that only 37% of the farms in cluster 1 (out of 105 mutual comparisons) differed significantly in terms of their NUE_N, whereas in cluster 2 this was 82% (out of 120 comparisons). Therefore, benchmarking NUE_N of farms in cluster 1 was no longer possible, whereas farms in cluster 2 could still be ranked when uncertainty was included. After reducing the uncertainties of the most important parameters, 72% of the farms in cluster 1 differed significantly in terms of their NUE_N, and in cluster 2 this was 87%. Results indicate that reducing epistemic uncertainty of input parameters can significantly improve benchmarking results. The method presented in this study, therefore, can be used to draw more reliable conclusions regarding benchmarking the NUE of farms, and to identify the parameters that require more precision to do so.

1. Introduction

Nitrogen (N) is an essential nutrient for milk production. The input of N into European milk production systems has increased in the past decades, mainly via purchase of fertilizer and feed, but also via atmospheric deposition and biological fixation (Powell et al., 2010). These increased N inputs have also increased N losses to the environment, via leaching of nitrate (NO_3^-) and emissions of N-gases, such as nitrous oxide (N_2O) and ammonia (NH_3). These N losses contribute to environmental problems, such as eutrophication, acidification and global warming (Whitehead, 1995; Smith et al., 1999). To tackle this problem, the European Union introduced legislation, such as the Nitrates Directive (EU, 2006), which set limits on N application per hectare to reduce NO_3^- leaching.

There have been on-going studies and discussions on how to reduce N losses of dairy farms in Europe (e.g. Aarts et al., 1992; Schröder et al., 2003; Nevens et al., 2006; Phuong et al., 2013; Mihailescu et al., 2015). Calculating the nutrient balance at farm level is the most commonly used approach to evaluate these losses. In the Netherlands, for example, dairy farms are obliged to quantify their annual nitrogen and

* Corresponding author.
E-mail address: wenjuan.mu@wur.nl (W. Mu).
[1] Combined first authorship.

http://dx.doi.org/10.1016/j.agsy.2017.04.001
Received 25 January 2016; Received in revised form 27 March 2017; Accepted 2 April 2017
Available online 02 June 2017

phosphorus balance from 2016 onwards (Veeteelt, 2015). A nutrient balance reflects the difference in nutrients entering and leaving a system, and allows computation of environmental indicators, such as the nutrient use efficiency (NUE) or the nutrient surplus per ha of a farming system (Spears et al., 2003). NUE generally is computed as the amount of nutrients in valuable outputs of a system over the amount of nutrients in all inputs of that system (Nevens et al., 2006).

Due to the simplicity of the method and relatively low data requirement, the nutrient balance has been used as a tool to benchmark the environmental performance of farms (Oenema et al., 2003; Schröder et al., 2003). Benchmarking is defined by Camp (1989) as "the search for those best practices that will lead to the superior performance" and, in this study, relates to the comparison of farms based on their environmental performance in order to identify differences and potentially, improvement options. Benchmarking farms based on, for example, their NUE, however, may lead to biased conclusions because of two reasons. First, as pointed out by Schröder et al. (2003), comparing the NUE of farms is justified only if they have similar major decisive characteristics. These characteristics can be based on: (unmanageable) physical factors, such as soil type and climatic conditions (Roberts, 2008; Powell et al., 2010); and long term strategic decisions, such as the degree of self-sufficiency (e.g. grass-based versus concentrate-based), production intensity, or manure management system (Nevens et al., 2006). Other characteristics that have an influence of the NUE of a farm include short term tactical decisions, such as choice of the feed crop, or grazing regime; operational decisions (i.e., day to day decisions); and other management skills of the farmer, such as the capacity to reduce losses (e.g. losses of feed, nutrients, milk or cows (culling)) (Nevens et al., 2006). Benchmarking NUE of farms should be based on differences in short term strategic and tactical decision-making, rather than differences in physical factors and long term decisions. Second, comparing NUE of farms may be affected by epistemic uncertainty of input data, caused by errors in measurement devices or errors around observations. Epistemic uncertainty can arise from e.g. errors in practically determining the N fixation by clover, measurement errors around the feed intake of the cows or estimations around the N-content of the animals (Oenema et al., 2015). Increasing knowledge or better measurements can reduce epistemic uncertainty (Walker et al., 2003; Groen et al., 2016).

Previous studies focused on examining the epistemic uncertainties of nutrient flows by looking into e.g. quantity of nutrient inputs (Mulier et al., 2003; Gourley et al., 2012; Oenema et al., 2015). However, they did not examine the impact of epistemic uncertainties on benchmarking results, nor did they benchmark farms with similar decisive farm characteristics.

The objectives of this study were to benchmark the nutrient losses by comparing nitrogen use efficiency (NUE_N) of farms with similar decisive characteristics while including epistemic uncertainty, and to examine which input parameters explain most uncertainty of NUE_N results. In addition, the epistemic uncertainties of input parameters that explain most of the output variance were reduced, to illustrate how this will improve benchmarking results.

2. Materials and methods

2.1. Case study: European specialized dairy farms

We used data of specialized dairy farms from Dairyman. Dairyman was a project directed at improving regional prosperity through better resource utilization on 113 dairy farms in different European countries (Dairyman, 2010). From the 113 farms, 32 specialized dairy farms were selected. Specialized dairy farms were defined as farms that have < 5% non-dairy purpose animals, and < 10% of their agricultural area in use for non-dairy purpose activities. These 32 dairy farms were located in different countries and regions (i.e. Netherlands (7), Ireland (13), Belgium (Flanders 8, Wallonia 2), Germany (1) and Luxembourg (1)).

Table 1
Characteristics of the 32 European specialized dairy farms used in this study.

Characteristics	Unit	Mean	Minimum	Maximum
Agricultural area	ha	65	25	270
Herd size	number of dairy cows	90	37	384
Milk production	kg milk cow^{-1} $year^{-1}$	7689	5700	9853
Milk production	kg milk ha^{-1} $year^{-1}$	12,598	3448	26,300
Grazing hours	h $year^{-1}$	2857	0	5146
Concentrate usage	kg cow^{-1} $year^{-1}$	1215	317	2459

Selected dairy farms differed in soil types (i.e. sandy soil, loam soil), milk production (i.e. milk production per cow and per ha), grazing hours per year, and feed import (i.e. kg concentrate usage per cow per year; Table 1). Whereas data on soil type, milk production and feed import were based on measured farm data, data on grazing hours per year were based on estimations by the farmers. Farm data from the year 2010 were used as baseline values to determine all N-flows.

2.2. Defining homogenous farm clusters

To enable benchmarking of NUE_N of farms with similar characteristics, farms were sorted into homogenous groups (i.e. typologies) based on their characteristics (Table 1). For this purpose, we used a two-step cluster analysis, because it allows using both continuous and categorical variables as clustering criteria (Chiu et al., 2001). To perform a cluster analysis with n criteria, a sample size of 2^n farms is required (Formann, 1984). Since our sample size included 32 farms, we selected 5 criteria for the cluster analysis, namely grazing hours, soil type, concentrate per cow per year, milk production per cow per year and milk production per ha (De Vries et al., 2015; Daatselaar et al., 2015). The analysis was performed in the statistical software package IBM SPSS statistics 22 (SPSS, 2015).

2.3. System boundary and model assumptions of calculating NUE_N

The NUE_N was quantified at farm level, implying that only on-farm flows and losses were considered. The N-flows through a dairy farm included in this study are visualized in Fig. 1. Inputs of N include N in mineral fertilizers, manure, animals, concentrates, roughages, biological N fixation and atmospheric N deposition. Outputs of N include N in animals, milk, manure and roughage. Stock changes (defined as final stock minus initial stock) of the mineral fertilizers, manure, animals, concentrates and roughages were taken into consideration during the computation processes. Manure output was subtracted from the total fertilizer input (i.e. through mineral fertilizer and manure). If the total manure output of the farm exceeded its total fertilizer input, excessive manure was treated as a loss. The internal N-flow from crop production to feed storage was based on the energy requirements of the herd, minus feed input and stock changes of feed. The calculation rules are specified in the Supplementary material. Losses of N from manure storage were based on storage type (i.e. slurry, solid) and the baseline values of manure N in all calculations (EEA, 2013).

2.4. Matrix based calculation for on-farm NUE_N

We used the matrix-based approach developed by Suh and Yee (2011) to quantify the N-efficiency of the 32 dairy farms. This approach was used to describe the herd and crop balance (Fig. 1) in one equation, which facilitates the global sensitivity analysis to examine epistemic uncertainty. A matrix-based approach allows for the presence of loops and parallel components, as is often the case on dairy farms (e.g. manure is used for the production of feed crops, which are consequently fed to the animals, producing manure). This approach requires a detailed insight into the nutrient flows within the farm.

The difference between the matrix-based approach to assess the

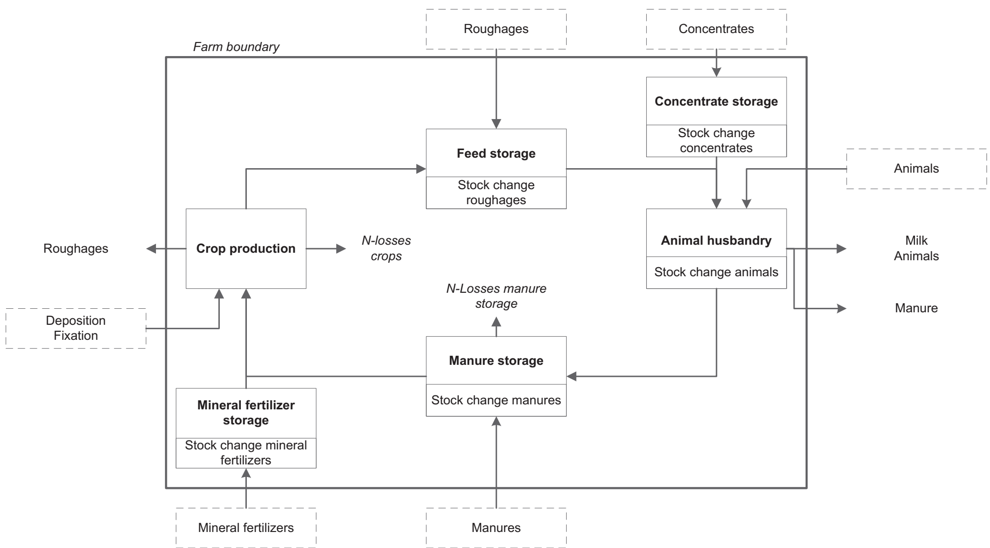

Fig. 1. N-flows on a dairy farm to assess nutrient use efficiency; the production processes are given by the solid boxes, the N-flows are given by the arrows. A detailed description of the input parameters can be found in Table 2.

Table 2
Description of the parameters and their epistemic uncertainty given by the relative uncertainty (CV), which was taken from Oenema et al. (2015).

Process	Type	Parameter		CV (%)	Remark
Crop production	Resource input	N-fixation (kg N)	Grassland area (ha)	5.0	
			Legume yield (kg/ha)	10	
			N-fixation (kg N/kg legume)	30	
		Deposition (kg N)	Farm area (ha)	5.0	
			N-deposition (kg N/ha)	17	
		Mineral fertilizer (kg N)	Mineral fertilizer (kg)	2.5	
			N-content mineral fertilizer (kg N/kg)	2.5	
		Stock change mineral fertilizer (kg N)	Stock change mineral fertilizer (kg)	7.5	
			N-content stock change mineral fertilizer (kg N/kg)	2.5	
	Export	Roughage (kg N)	Roughage (kg)	7.5	
			N-content roughage (kg N/kg)	7.5	
	Losses	N-losses crops (kg N)	n.a.	n.a	Function
Feed storage	Resource input	Roughage (kg N)	Roughage (kg)	7.5	
			N-content roughage (kg N/kg)	7.5	
		Stock change roughage (kg N)	n.a.	17	GEP[a]
Fertilizer storage	Resource input	Manure (kg N)	Manure (kg)	5.0	
			N-content manure (kg N/kg)	7.5	
		Stock change manure (kg N)	n.a.	22	GEP
	Losses	N emissions from manure storage (kg N)	n.a.		Fixed
Milk and animal production	Resource inputs	Animals (kg N)	Number of animals (−)	2.0	
			Life-weight per animal (kg)		n.a.
			N-content per animal (kg N/kg)	5.0	
		Stock change animals (kg N)	n.a.	5.6[8]	GEP
		Concentrates (kg N)	Concentrates (kg)	2.5	
			N-content concentrates (kg N/kg)	2.5	
		Stock change concentrates (kg N)	n.a.	11	GEP
	Final use	Milk (kg N)	Milk (kg)	1.0	
			N-content milk (kg N/kg)	2.0	
		Animals (kg N)	Number of animals (−)	2.0	
			Life-weight per animal (kg)		n.a.
			N-content animal (kg N/kg)	5.0	
	Export	Manure (kg N)	Manure (kg)	5.0	
			N-content manure (kg N/kg)	7.5	

[a] GEP: Gaussian error propagation is used to determine the CV of parameters when there is a lack of information to separate the N-content from the items in the stock change and therefore only the kg N of stock change is available (e.g., roughage can include different items with different N contents). Details on the method can be found in (Heijungs and Lenzen, 2014).

farm N-balance and the common nutrient balance approach is that in case of the matrix-based approach the internal flows are considered (e.g. the flows between manure storage and crop production, or crop production and feed storage), just as in a substance flow analysis. In the common nutrient balance, the farm is considered as a black box (e.g. Oenema et al., 2015 and Mu et al., 2016). For more details, see the Supplementary material.

In the matrix-based approach, the internal N-flows in Fig. 1 are described by the **V** and **U**-matrix, where the **V**-matrix describes how much kg N is supplied to each production process. The **U**-matrix describes how much kg N is used by each production process (Suh and Yee, 2011). The N-flows are corrected for the stock changes (**s**) on the farms. Combined, they are quantified in a matrix **A** for each (intermediate) process. T refers to the transpose. The vector (**b**) gives the amount of nutrients extracted (**r**) to produce 1 unit of final product, which, in this case, is determined by the valuable outputs of the farm:

$$\mathbf{b} = \mathbf{r}(\mathbf{V}^T - \mathbf{U} + \hat{\mathbf{s}})^{-1} = \mathbf{r}\mathbf{A}^{-1} \quad (1)$$

Table 3
Results of the cluster analysis, showing the farm characteristics for 15 farms in cluster 1 and 17 farms in cluster 2, given by the mean (standard deviation) of each characteristic or a categorical characteristic per cluster.

Characteristics[a]	Unit	Cluster 1	Cluster 2
Soil type	n.a.	Sandy	Loam
Milk production	kg milk cow^{-1} year^{-1}	8519 (854)	6956 (878)
Milk production	kg milk cow^{-1} ha^{-1}	15,970 (5108)	9623 (3792)
Grazing hours	h cow^{-1} year^{-1}	1115 (1099)	4393 (1175)
Concentrate use	kg cow^{-1} year^{-1}	1719 (499)	770 (207)
NUE_N	%	43 (10)	26 (12)

[a] Characteristics of these two clusters are significantly different ($p < 0.05$). The order of importance of the characteristics in determining the final clusters are: grazing hours > concentrate use > milk production per cow > soil type > milk production per ha.

In our case, the four elements in **b** represent the production processes of Fig. 1 (animal husbandry, manure storage, crop production, feed storage). The nitrogen use efficiency (NUE_N) for the production process of the animal husbandry is quantified by:

$$NUE_N = 1/b_{husbandry} \quad (2)$$

A detailed example of this procedure can be found in the Supplementary material of Suh and Yee (2011).

2.5. Quantifying the effect of epistemic uncertainty on benchmarking

To quantify the effect of epistemic uncertainties of the input parameters on the benchmarking of farms based on their NUE_N, the distribution functions of the parameters need to be defined first. Subsequently, the input uncertainties are propagated through the NUE_N model.

2.5.1. Defining distribution functions

Each parameter in the NUE_N model was considered as an uncertain parameter, only the N-flow from crop production to feed storage and the N losses during manure storage were fixed. The N-flow from crop production was fixed, because it was based on the energy requirements of the herd. The N losses during manure storage were fixed, because they were based on storage specific emission factors. All input parameters are assumed to be normally distributed. Fixation was assumed to be truncated normally distributed to avoid drawing negative numbers. The coefficient of variation ($CV = \sigma/\mu$) described the epistemic uncertainty of the parameters and was based on Oenema et al. (2015) (Table 2). Based on the equation for the CV, the standard deviation was calculated per farm, because each farm had a different (i.e. farm specific) mean.

2.5.2. Quantifying the effect of epistemic uncertainty on benchmarking

The propagation of the uncertainties of the input parameters through the NUE_N model (Eq. (1)) was done using Monte Carlo simulation and was performed for all farms in each cluster. The code for performing the uncertainty and global sensitivity analysis is available at: http://evelynegroen.github.io. From each distribution function (Table 1) a random value was drawn, and used to calculate the NUE_N. The output uncertainty was given by the variance:

$$var(NUE_N) = \frac{1}{n-1}\sum_{i=1}^{n}(NUE_{Ni} - \overline{NUE_N})^2$$

where the mean is given by: $\overline{NUE_N} = \frac{1}{n}\sum_i NUE_{Ni}$, for a sample size of $n = 5000$. We performed a discernibility analysis (Heijungs and Kleijn, 2001) to determine if the input uncertainties had an effect on benchmarking. To determine if there was a significant difference between farms the farms within a cluster were pairwise compared for the results for each Monte Carlo run. This means that we counted how many times the NUE_N of one farm was better than another farm, expressed as a frequency. A significance level of 5% was chosen (Heijungs and Kleijn, 2001; Henriksson et al., 2015). This means, for example, that if farm A has a lower NUE_N than farm B in 630 out of 1000 runs, difference in NUE_N of the two farms was considered as not significant (63% > 2.5%). But, if farm A had a lower NUE_N than farm C in 24 out of 1000 runs, than farm C was considered as significantly better than farm A (2.4% < 2.5%).

2.6. Explaining output uncertainty for different farm typologies

To identify which input parameter contributed most to the output uncertainty within a specific farm cluster, a global sensitivity analysis was performed by calculating the squared standardized regression coefficients (S_j) as a measure for the sensitivity index (Saltelli et al., 2008; Groen et al., 2016):

$$S_j = \frac{var(p_j)}{var(NUE_N)}(b_j)^2$$

where var(p_j) gives the variance of each input parameter (p_j) based on Table 2 and b_j is equal to the regression coefficient.

3. Results

3.1. Farm clusters

Two homogeneous groups of farms, i.e. farm clusters, were derived from the cluster analysis. Farms in the first group, further referred to as farms in cluster 1, are located on sandy soils and relatively intensive in terms of milk production per cow and per hectare (Table 3). The number of grazing hours is low, whereas the amount of purchased concentrates per cow per year is high relative to the farms in the other cluster. Farms in cluster 2 are located on loam soils, and are less intensive when compared to farms in cluster 1. The number of grazing hours is higher, whereas the amount of concentrates per cow per year is lower than on farms in cluster 1. The average NUE_N of farms in cluster 1 is 43%, and for farms in cluster 2 this is 26%. The difference in NUE_N between the two clusters result from a combination of all 5 characteristics that specify the group of farms in each cluster (Table 3).

3.2. The effect of epistemic uncertainties on benchmarking

For each farm, the input uncertainties of Table 2 were propagated through the NUE_N model (Eqs. (1) and (2)). For each farm in both clusters, a median and a variance were derived (Fig. 2, cluster 1; Fig. 3, cluster 2). Results show that each cluster has one outlier: farm 1 in cluster 1 and farm 2 in cluster 2. For farm 1 in cluster 1, the output of manure exceeds the input of fertilizer. Because we subtracted manure output from fertilizer inputs, the input of fertilizer was set to 0. This leads to the high NUE_N of this farm. Farm 2 in cluster 2 is an organic farm with only grassland and no cropland. The imported feed inputs are low, and there is no input of synthetic fertilizer. Due to the low N inputs and high N outputs of the farm, it has a high NUE_N.

The results of the discernibility analysis for cluster 1 can be found in Table 4. For example, farm 5 had a lower NUE_N than farm 1, and a higher NUE_N than the other farms, except when compared to farm 8 and farm 14. In case of farm 6, only 52% of the Monte Carlo runs show a higher NUE_N than farm 3, meaning their performance is almost indistinguishable taking the epistemic uncertainties of the input parameters into account.

For farm 1, approximately 4% of the Monte Carlo runs resulted in a negative value for N losses of crop production. This is explained by the importance of deposition as an N input on this farm, and the large uncertainty of this parameter (CV = 17%; Table 2). The negative values, therefore, are more likely related to the uncertainty of deposition,

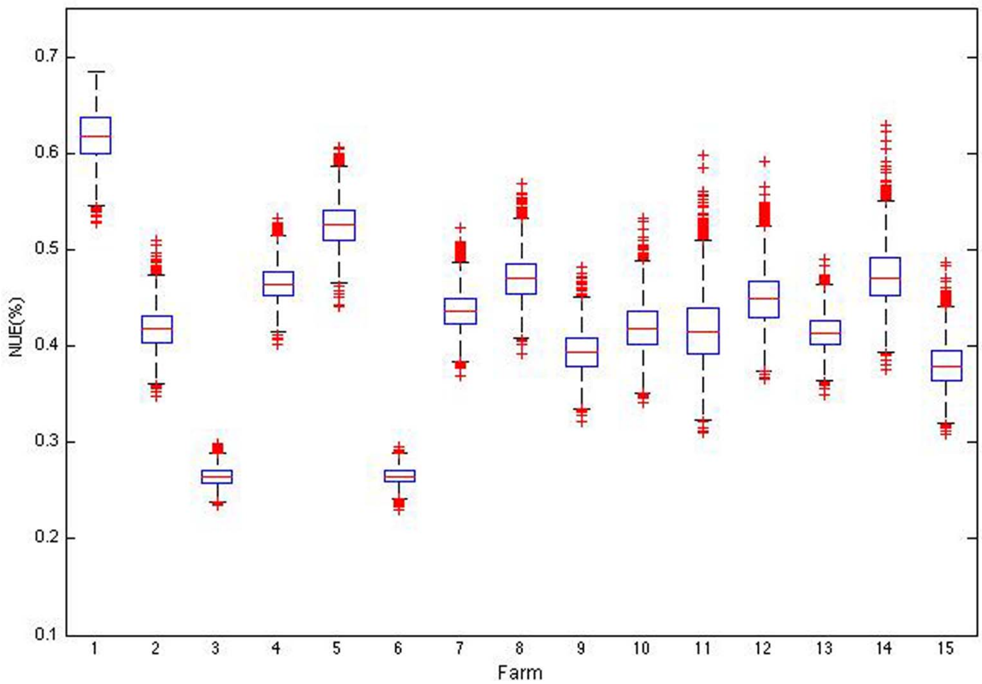

Fig. 2. Box plot of NUE_N for the 15 farms in cluster 1. The horizontal line in each box gives the median, the box gives the 25–75% interval, and the plusses are realizations that appear outside the 10–90% interval.

than to display a realistic model outcome. The drawings from the Monte Carlo simulation that included a negative value for N losses of crop production, therefore, were removed from the analysis.

Applying the 5% significance level, results show that farm 1 is most efficient when taking the epistemic uncertainty of the input parameters into account, followed by farm 5, which is only not significantly better than farm 8 and 14. The two least efficient farms are farm 3 and 6. The NUE_N of the other farms turned out to be very similar (Table 4).

The results of the discernibility analysis for cluster 2 are found in Table 5. For farm 2, approximately 46% of the Monte Carlo runs resulted in a negative value for N losses of crop production. This is explained by the importance of N fixation on this farm, in combination with a relatively large uncertainty of this parameter (CV = 30%). Because we analysed quite intensive and productive farms, these outcomes are more likely to result from the high CV than to display a realistic situation. Similar to farm 1 in cluster 1, negative values were assumed to display an unrealistic model outcome. Because of the high percentage of unrealistic model outcomes, it was decided to remove farm 2 from further analysis. The large number of unrealistic model outcomes illustrates the need to reduce CVs by improving measurements on farms.

Applying the 5% significance level, results show that farm 1 is most efficient (only not significantly higher than farm 10 and 17). Of the 120 farm comparisons, 17% is significantly different (Table 5). Contrary to the first cluster, including the epistemic uncertainties still allowed for some kind of ranking, although most farms overlapped with at least two other farms.

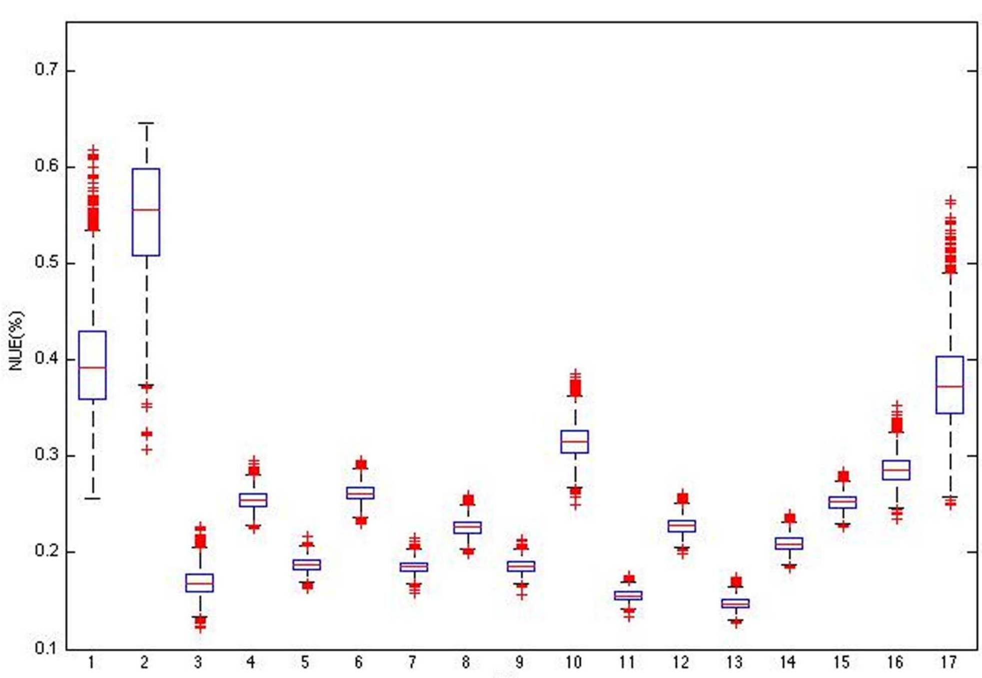

Fig. 3. Box plot of NUE_N for the 17 farms in cluster 2. The horizontal line in each box gives the median, the box gives the 25–75% interval, and the plusses are realizations that appear outside the 10–90% interval.

Table 4
Results of discernibility analysis for cluster 1 based on pairwise comparing Monte Carlo runs between farms. The column and row numbers 1 to 15 represent the 15 farms. The percentages show how often a farm (row) has a higher NUE_N than another farm (column). When α-value of 0.05 is applied, values between 2.5% and 97.5% indicate that the NUE_N of the farms are no longer considered as significantly different. The significant different farms are given by the bold-printed percentages.

%	1[a]	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1[a]		**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
2	**0**		**100**	5	**0**	**100**	25	5	78	47	52	19	54	7	89
3	**0**	**0**		**0**	**0**	48	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
4	**0**	95	**100**		**2**	**100**	85	43	**99**	92	88	70	97	42	**100**
5	**0**	**100**	**100**	**98**		**100**	**100**	95	**100**	**100**	**99**	**98**	**100**	91	**100**
6	**0**	**0**	52	**0**	**0**		**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
7	**0**	75	**100**	15	**0**	**100**		13	93	69	71	37	79	16	97
8	**0**	95	**100**	57	5	**100**	87		**99**	93	90	73	97	49	**100**
9	**0**	22	**100**	**1**	**0**	**100**	7	**1**		22	31	6	24	**1**	67
10	**0**	53	**100**	8	**0**	**100**	31	7	78		55	23	57	9	88
11	**0**	48	**100**	12	**1**	**100**	29	10	69	45		23	50	11	81
12	**0**	81	**100**	30	**2**	**100**	63	27	94	77	77		84	28	**97**
13	**0**	46	**100**	3	**0**	**100**	21	3	76	43	50	16		5	88
14	**0**	93	**100**	58	9	**100**	84	51	**99**	91	89	72	95		**99**
15	**0**	11	**100**	**0**	**0**	**100**	3	**0**	33	12	19	3	12	**1**	

[a] Approximately 4% of the Monte Carlo runs were excluded from the analysis due to unrealistic model outcomes.

3.3. Explaining the output variance

The global sensitivity analyses shows how much of the output variance can be explained by the variance of the individual input parameters. The results of the global sensitivity analysis can be found in Fig. 4 (cluster 1) and Fig. 5 (cluster 2).

Results show that in case of cluster 1, the input of concentrates, roughage, mineral fertilizer, and deposition, and the output of milk, roughage, and manure explain most of the output variance. Input of animals and manure, stock change of each of the inputs, and output of animals did not show up as important explanatory parameters in any of the farms, except for stock change of mineral fertilizer for farm 12. Further analysis showed that both the quantity as well as the N content of each parameter is approximately equally important in terms of their contribution to the output variance.

Fig. 5 shows that in case of cluster 2, for most of the farms the input of mineral fertilizer and the output of milk and animals explain most of the output variance. For a few farms, the most important parameter in terms of contribution to the output variance is fixation (input); while for one farm, this is the output of roughage. Input of concentrates, roughage and manure and stock change of animals did not show up as important explanatory parameters in any of the farms in cluster 2.

3.4. Effect of decreasing uncertainty on benchmarking

To analyse if decreasing epistemic uncertainty can improve benchmarking, we reduced the uncertainty of the most important input parameters and reran the discernibility analysis. For cluster 1, the input uncertainty was reduced to 1% for: input of concentrates, roughage, mineral fertilizer, deposition and the output of milk, roughage, and manure (Fig. 4). For cluster 2, the input uncertainty was reduced to 1% for: input of mineral fertilizer, and the output of milk and animals (Fig. 5). Table 6 shows how many pairwise comparisons were made in both cluster, and how many were significantly different, before and after reducing input uncertainty. Results show that reducing the uncertainty of the most important input parameters based on the global sensitivity analysis, improved the ability to find significant differences between the NUE_N of the farms in both clusters. Benchmarking, therefore, can be improved when input uncertainties are reduced, especially for the farms in the first cluster.

4. Discussion

This study builds on, and extends the principles regarding epistemic uncertainty of nitrogen flows on dairy farms presented by Oenema et al. (2015). Although we used the same coefficients of variations of input

Table 5
Results of discernibility analysis for cluster 2 based on pairwise comparing Monte Carlo runs between farms. The column and row numbers 1 to 17 represent the 16 farms[a]. The percentages show how often a farm (row) has a higher NUE_N than another farm (column). When α-value of 0.05 is applied, values between 2.5% and 97.5% indicate that the NUE_N of the farms are no longer considered as significantly different. The significant different farms are given by the bold-printed percentages.

%	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1		**100**	**100**	**100**	**100**	**100**	**100**	**100**	95	**100**	**100**	**100**	**100**	**100**	**99**	62
3	**0**		**0**	12	**0**	14	**0**	14	**0**	82	**0**	92	**1**	**0**	**0**	**0**
4	**0**	**100**		**100**	31	**100**	**99**	**100**	**0**	**100**	**98**	**100**	**100**	57	**4**	**0**
5	**0**	88	**0**		**0**	59	**0**	59	**0**	**100**	**0**	**100**	**2**	**0**	**0**	**0**
6	**0**	**100**	69	**100**		**100**	**100**	**100**	**0**	**100**	**100**	**100**	**100**	76	9	**0**
7	**0**	86	**0**	41	**0**		**0**	49	**0**	**100**	**0**	**100**	**1**	**0**	**0**	**0**
8	**0**	**100**	**1**	**100**	**0**	**100**		**100**	**0**	**100**	45	**100**	93	**2**	**0**	**0**
9	**0**	86	**0**	41	**0**	51	**0**		**0**	**100**	**0**	**100**	**1**	**0**	**0**	**0**
10	5	**100**	**100**	**100**	**100**	**100**	**100**	**100**		**100**	**100**	**100**	**100**	**100**	90	8
11	**0**	18	**0**	**0**	**0**	**0**	**0**	**0**	**0**		**0**	83	**0**	**0**	**0**	**0**
12	**0**	**100**	**2**	**100**	**0**	**100**	55	**100**	**0**	**100**		**100**	94	**2**	**0**	**0**
13	**0**	8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	17	**0**		**0**	**0**	**0**	**0**
14	**0**	**99**	**0**	**98**	**0**	**99**	7	**99**	**0**	**100**	6	**100**		**0**	**0**	**0**
15	**0**	**100**	43	**100**	24	**100**	**99**	**100**	**0**	**100**	**98**	**100**	**100**		**2**	**0**
16	**1**	**100**	**96**	**100**	91	**100**	**100**	**100**	11	**100**	**100**	**100**	**100**	**98**		**1**
17	37	**100**	**100**	**100**	**100**	**100**	**100**	**100**	92	**100**	**100**	**100**	**100**	**100**	**99**	

[a] Approximately 46% of the Monte Carlo runs were excluded from the analysis due to unrealistic model outcomes of farm 2, therefore, this farm was excluded from further analysis.

Parameter (%)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Input animals		0		0				0				0			0
Input concentrates	30	1	29	8	4	39	30	15	16	10	4	3	9	16	8
Input roughage	0		17	0	3		0	49	25	51	20	2	28	22	30
Input min. fertilizer	10	8	14	11	7	15	16		3	1	0	20	7	2	7
Input manure		2								1	0			1	
Deposition	21	4	4	7	3	6	7	15	7	9	13	16	30	9	7
Fixation									0		52			1	
SC animal	0	0	0	0	0	0	0		0	0	0	0	0	0	0
SC concentrates				0	0		0		0	0	0	0	0	0	0
SC roughage									0	1	2	1	1	4	2
SC min. fertilizer									0	1		31		0	
SC. manure									2	2	1		3	3	2
Output animal	1	0	5	1	0	4	1	1	2	1	1	1	2	1	1
Output milk	13	3	29	8	5	35	20	16	13	11	6	5	16	10	11
Output roughage	4	81	0	63	77							19	3		
Output manure	35	0	4				26		29	12	2		2	30	33
Explained variance	n/a	100	100	99	99	100	100	97	99	100	100	100	100	100	100

Fig. 4. Sensitivity indices (S_j) for each input parameter, explaining how much each parameter contributes to the output variance for each farm in cluster 1: the darker a cell, the higher the contribution. SC: stock change. An empty cell means that these parameters were zero for that farm; 0% means that this parameter contributed 0% to the output variance; n/a not applicable, approximately 4% of the Monte Carlo runs were excluded from the analysis due to unrealistic model outcomes, therefore the partial variances were not considered independent and could not be added.

parameters, results of our study and Oenema et al. (2015) show important differences. Based on our analysis, input of concentrates and roughage, and output of milk and roughage explain most of the output variance in cluster 1. Input of mineral fertilizer and fixation, and output of animals and milk explain most of the output variance in cluster 2. Oenema et al. (2015), however, concluded that N fixation, atmospheric deposition and stock changes of roughage and manure explain most of the output variance when determining the N surplus of dairy farms. Differences between our study and Oenema et al. (2015) can be explained by two reasons. First, the characteristics of the farms were different. In general, Oenema et al. (2015) included farms with a lower input of feed, but a higher stock change of roughage, and a higher N input through fixation compared to the farms in our study. Uncertainties related to stock changes of feed are higher than uncertainties related to input of feed, whereas uncertainties related to N fixation is highest among all N flows. Second, Oenema et al. (2015) used a different approach to determine N intake during grazing. In our study, N intake from grazing and on-farm roughage production was fixed based on feed requirements and the baseline values of input of purchased feed (see Supplementary materials). Oenema et al. (2015) changed the N-intake from grazing with a change in roughage and concentrate intake, which consequently influenced the importance of feed parameters. The contribution of the input of feed to the output variance was therefore found to be lower in Oenema et al. (2015) than in our study.

Table 6
Effect of decreasing the input uncertainties of the most important parameters to 1%, for both clusters.

	Cluster 1	Cluster 2
Total number of pairwise comparisons	105	120
Significantly different farms before reducing input uncertainty	39 (37%)	99 (83%)
Significantly different farms after reducing input uncertainty	76 (72%)	104 (87%)

Dairy farms in Europe show different decisive characteristics. For example, most farms in the Netherlands have a high stocking density because land resources are limited. The main N inputs on these farms are through purchased concentrates and roughages. In Ireland, however, most farms are grass-based extensive farms. The main N inputs on

Parameter (%)	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Input animals		0										12				0
Input concentrates	0	0	0	2	4	2	2	2	0	3	2	0	1	13	0	0
Input roughage		3	0	8	0	0	0	0	1	1		1	0	2		6
Input min. fertilizer		1	70	45	54	44	66	61	12	55	57	45	60	38	59	
Input manure	3	8														
Deposition	9	3	2	1	5	2	1	1	1	1	3	1	3	5	1	14
Fixation	82	78														68
SC animal	0	0	0	2	0	1	0	0	0	0	1	1	0	0	0	0
SC concentrates																0
SC roughage																1
SC min. fertilizer						13			10	0			0		15	
SC. manure																
Output animal	1	2	10	21	12	13	6	11	1	15	14	25	23	13	2	1
Output milk	2	5	17	21	25	27	25	24	2	22	21	15	14	28	8	1
Output roughage									73						11	8
Explained variance	97	99	99	100	100	100	100	98	100	99	97	100	100	99	97	99

Fig. 5. Sensitivity indices (S_j) for each input parameter, explaining how much each parameter contributes to the output variance for each farm in cluster 2: the darker a cell, the higher the contribution. SC: stock change. An empty cell means that these parameters were to zero for that farm; 0% means that this parameter contributed 0% to the output variance. Output of manure is not included because none of the farms in cluster 2 exported manure. Farm 2 was excluded from the global sensitivity analysis.

these farms are through purchased mineral fertilizer and N fixation. Comparing NUE of Dutch and Irish farms can lead to biased conclusions because of inherent differences between systems. Clustering of farms into groups with similar decisive characteristics, therefore, is a prerequisite for benchmarking the NUE of farms and facilitates the identification of major parameters. When comparing results of the global sensitivity analysis between the two farm clusters, for example, input of feed and output of roughage show up to be most important in case of cluster 1, whereas the input of mineral fertilizer (or fixation) is most important in case of cluster 2. Results show that the importance of parameters can vary between farm types (clusters). Methods to improve benchmarking of farms, therefore, should account for differences in decisive characteristics. The method presented in this study, can contribute to more solid conclusions regarding the performance of farms in terms of their NUE.

In this study, we used a matrix-based approach to assess NUE. The advantage of this approach is that it facilitates the uncertainty and sensitivity analysis. All input parameters are sampled at the same time, and are subsequently used to calculate the internal flows for each Monte Carlo run. Another advantage of the matrix-based approach is that it is easy to extend the system boundary beyond the farm: production of crops can be easily incorporated as additional production flows.

Several methodological limitations could have affected the results of this study. The first limitation is the choice of the parameter distribution function and, in link to our choice of Gaussian distribution, of CV values. A Gaussian or normal distribution represents a symmetrical uncertainty range which seems correct in case of (most) measurement errors. Future studies, however, could use parameter specific distribution functions to improve the impact assessment. The CVs we used were based on Oenema et al. (2015), focussing on Dutch pilot commercial farms only. Farms in our study are from different countries in Europe. Results of the uncertainty and global sensitivity analysis might have been different if country specific coefficients of variation were applied, but such information was not available. In addition, we used farm data from the year 2010, which might not hold for any year. However, since a similar measurement error over years can be expected, we do not expect a big change in the CVs, but mainly in the mean values of the N-flows on the farms. In general, it takes quite a big change in the CV to influence the result of the sensitivity analysis as seen in Figs. 4 and 5. In addition, only drastically decreasing the CVs of the most important parameters (e.g. from 30% to 1%) influenced the number of significantly different farms (mainly in the first cluster; Table 6). Differences in the CVs because of yearly variations, therefore, are not expected to influence the results. Nevertheless, the methodological procedure that was presented in this study can be used to assess the impact of epistemic uncertainty on different farms and based on different CVs. Results show that to benchmark the NUE of farms, epistemic uncertainty of input parameters has to be reduced.

Secondly, changes in soil N-stock were not considered in this study due to data limitations. Assessing changes in soil N stock at the farm level is difficult but can significantly improve interpretation of nutrient balance results (Godinot et al., 2014).

Thirdly, uncertainty related to on-farm crop and grass production was not included in the model, because this was estimated based on the energy requirements of the dairy herd and the energy in purchased feed and stock changes of feed. Incorporating uncertainty of crop production would increase the uncertainty of the model output. This would mainly affect the farms relying more heavily on on-farm produced roughage, such as the farms in cluster 2. Considering that input of mineral fertilizer or fixation and output of N via milk an animals are the main explanatory factors of the uncertainty around the NUE of these farms, we do not expect that the additional uncertainty of on-farm crop (grass) production would influence the results much.

Fourthly, to prevent purchase-resale bias (Godinot et al., 2014), the output of roughage was subtracted from the input of roughage, and the output of manure was subtracted from the input of fertilizers. As a result, exported manure is valued for its fertilizer capacity similar to (synthetic) fertilizer inputs. The disadvantage of the approach is that the output of manure results in an artificial reduction of fertilizer input, while an actual reduction should form the basis for ecological intensification and an improved NUE. The importance of manure output and the impact of these methodological choices should be addressed when benchmarking the NUE of dairy farms.

Fifthly, clustering of farms was based on 5 characteristics reflecting physical and long term strategic decisions. In practice, farming systems are much more complex than we considered in our study. Including other (unmanageable) factors that affect NUE could influence the clustering of farms and hence, the benchmarking of those farms. Nevertheless, this study is a first step towards improving benchmarking farms based on their NUE. Results emphasize the need to benchmark NUE by comparing farms with similar decisive characteristics, and that the importance of parameters that contribute to the uncertainty of the NUE results differ among farm types.

Sixthly, this study focused on NUE_N at farm level. Nitrogen losses related to the production of purchased feed and fertilizers were not considered. It should be kept in mind that, as a result, the NEU of a farm increases with a decrease in self-sufficiency. This approach, therefore, can contribute to biased conclusions and problem swapping, when on-farm nutrient losses related to feed production are reduced at the expense of off-farm losses. Furthermore, NUE provides insight into the efficiency of production rather than into the environmental impact related to nutrient losses. To gain insight into the impact of losses, information on nutrient losses per hectare should be combined with site specific knowledge of local eco-systems. In addition to this, it should be noted that the results of this study are limited to benchmarking the NUE_N of specialized dairy farms in Europe. For another indicator or another set of farms, the impact of uncertainty on benchmarking the environmental performance of dairy production could be different.

Reducing epistemic uncertainty and benchmarking NUE of farms with similar decisive characteristics can contribute to the identification of improvement options. Based on the variability between farms within a cluster, farm specific management options can be identified. Evaluating the (causes of) variability between farms within a farm cluster, therefore, can be a next step for further improving the NUE of farms.

5. Conclusion

Benchmarking the NUE of dairy farms requires an approach that accounts for differences in major decisive characteristics among farms, and for the impact of epistemic uncertainties of input parameters. The parameters that are most important in terms of epistemic uncertainty (i.e., explain most of the output variance), however, can vary among farm types. Clustering farms based on their main characteristics and understanding and reducing the impact of epistemic uncertainty of major parameters can significantly improve benchmarking results. The method presented in this study, therefore, can contribute to more solid conclusions regarding the performance of farms in terms of their nutrient use efficiency.

Acknowledgements

The Seventh Framework Programme (FP7) EU provided funding of E.A. Groen, grant agreement no: 286141. This research is also a part of the Wageningen UR strategic programme ‘Mapping for sustainable intensification’, 2012–2016, funded by the strategic funds of Wageningen UR, and the WASS and WIAS graduate schools of Wageningen University. We gratefully acknowledge Jouke Oenema for his helpful explanation on his paper referred to in this study, Reinout Heijungs for his suggestions on the statistical analysis and Michaël Mathot for his comments. We thank Jouke Oenema (Nl, Wageningen), Andy Boland

(Ir, TEAGASC), Lies Debruyne (Be, ILVO), Aurélie Grignard (Be, CRA-W), Thomas Jilg (Ge, LAZBW) and Jeff Boonen (Lu, LTA), funded by the INTERREG IV program for North-West-Europe, in the context of DAIRYMAN project, for data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.agsy.2017.04.001.

References

Aarts, H.F.M., Biewinga, E.E., Van Keulen, H., 1992. Dairy farming systems based on efficient nutrient management. Neth. J. Agric. Sci. 40, 285–299.

Camp, R.C., 1989. Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance. ASQC Quality Press, Milwaukee, WI.

Chiu, T., Fang, D., Chen, J., Wang, Y., Jeris, C., 2001. A robust and scalable clustering algorithm for mixed type attributes in large database environment. In: Proceedings of the 7th ACM SIGKDD International Conference in Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, CA, pp. 263–268.

Daatselaar, C.H., Reijs, J.R., Oenema, J., Doornewaard, G.J., Aarts, H.F., 2015. Variation in nutrient use efficiencies on Dutch dairy farms. J. Sci. Food Agric. 95, 15.

Dairyman, 2010. http://www.interregdairyman.eu/en/dairyman.htm Last access on 14-01-2016.

European Environment Agency (EEA), 2013. EMEP/EEA Air Pollutant Emission Inventory Guidebook. http://www.eea.europa.eu/publications/emep-eea-guidebook-2013.

European Union (EU), 2006. (Good Agricultural Practice for Protection of Waters) Regulations. S.I. No. 378 of 2006. Government Publications Office, Stationery Office.

Formann, A.K., 1984. Die Latent-Class-Analyse: Einführung in die Theorie und Anwendung. Weinheim, Beltz.

Godinot, O., Carof, M., Vertes, F., Leterme, P., 2014. SyNE: an improved indicator to assess nitrogen efficiency of farming systems. Agric. Syst. 127, 41–52.

Gourley, C.J.P., Aarons, S.R., Powell, J.M., 2012. Nitrogen use efficiency and manure management practices in contrasting dairy production systems. Agric. Ecosyst. Environ. 147, 73–81.

Groen, E.A., Bokkers, E.A.M., Heijung, R., de Boer, I.J.M., 2016. Methods for global sensitivity analysis in life cycle assessment. Int. J. Life Cycle Assess. http://dx.doi.org/10.1007/s11367-016-1217-3.

Heijungs, R., Kleijn, R., 2001. Numerical approach towards life cycle interpretation. Int. J. Life Cycle Assess. 6, 141–148.

Heijungs, R., Lenzen, M., 2014. Error propagation methods for LCA—a comparison. Int. J. Life Cycle Assess. 19, 1445–1461.

Henriksson, P.J.G., Heijungs, R., Dao, H.M., Phan, L.T., de Snoo, G.R., Guinee, J.B., 2015. Product carbon footprints and their uncertainties in comparative decision contexts. PLoS One 1–11.

Mihailescu, E., Ryan, W., Murphy, P.N.C., Casey, I.A., Humphreys, J., 2015. Economic impacts of nitrogen and phosphorus efficiency on nineteen intensive grass-based dairy farms in the South of Ireland. Agric. Syst. 132, 121–132.

Mu, W., van Middelaar, C.E., Bloemhof, J.M., Oenema, J., de Boer, I.J.M., 2016. Nutrient balance at chain level: a valuable approach to benchmark nutrient losses of milk production systems. J. Clean. Prod. 112, 2419–2428.

Mulier, A., Hofman, G., Baecke, E., Carlier, L., De Brabander, D., De Groote, G., 2003. A methodology for the calculation of farm level nitrogen and phosphorus balances in Flemish agriculture. Eur. J. Agron. 20, 45–51.

Nevens, F., Verbruggen, I., Reheul, D., Hofman, G., 2006. Farm gate nitrogen surpluses and nitrogen use efficiency of specialized dairy farms in Flanders: evolution and future goals. Agric. Syst. 88, 142–155.

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

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.

Phuong, H.N., Friggens, N.C., de Boer, I.J.M., Schmidely, P., 2013. Factors affecting energy and nitrogen efficiency of dairy cows: a meta-analysis. J. Dairy Sci. 96, 7245–7259.

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. Pol. 13, 217–228.

Roberts, T.L., 2008. Improving nutrient use efficiency. Turk. J. Agric. For. 32, 177–182.

Saltelli, A., Ratto, M., Andres, T., Corporation, E., 2008. Global Sensitivity Analysis: The Primer. Wiley Online Library.

Schröder, J.J., Aarts, H.F.M., Ten Berge, H.F.M., Van Keulen, H., Neeteson, J.J., 2003. An evaluation of whole-farm nitrogen balances and related indices for efficient nitrogen use. Eur. J. Agron. 20, 33–44.

Smith, V.H., Tilman, G.D., Nekola, J.C., 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environ. Pollut. 100, 179–196.

Spears, R.A., Kohn, R.A., Young, A.J., 2003. Whole-farm nitrogen balance on western dairy farms. J. Dairy Sci. 86, 4178–4186.

SPSS, 2015. Statistical Package for Social Sciences. SPSS Software Package 22.0 for Microsoft Windows, Chicago, Illinois, USA.

Suh, S., Yee, S., 2011. Phosphorus use-efficiency of agriculture and food system in the US. Chemosphere 84, 806–813.

Veeteelt, 2015. http://veeteelt.nl/nieuws/kringloopwijzer-2016-verplicht-voor-alle melkveebedrijven. Last accessed on 02-11-2015.

de Vries, W., Kros, J., Dolman, M.A., Vellinga, Th.V., de Boer, H.C., de Boer, I.J.M., Sonneveld, M.P.W., Bouma, J., 2015. Environmental impacts of innovative dairy farming systems aiming at improved internal nutrient cycling: a multi-scale assessment. Sci. Total Environ. 536, 432–442.

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P., Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr. Assess. 4, 5–17.

Whitehead, D.C., 1995. Volatilization of ammonia. In: Whitehead, D.C. (Ed.), Grassland Nitrogen. CAB International, Wallingford, pp. 152–179.