

## Field and crop specific manure application on a dairy farm based on historical data and machine learning



Herman Mollenhorst<sup>a,b,\*</sup>, Michel H.A. de Haan<sup>b</sup>, Jouke Oenema<sup>c</sup>, Claudia Kamphuis<sup>a</sup>

<sup>a</sup> Wageningen University and Research, Animal Breeding and Genomics, PO Box 338, 6700 AH Wageningen, the Netherlands

<sup>b</sup> Wageningen University and Research, Livestock and Environment, PO Box 338, 6700 AH Wageningen, the Netherlands

<sup>c</sup> Wageningen University and Research, Agrosystems Research, PO Box 16, 6700 AA Wageningen, the Netherlands

### ARTICLE INFO

#### Keywords:

Crop yield  
Dairy farm  
Grass  
Machine learning  
Manure application  
Nutrient management

### ABSTRACT

An important factor in a circular agricultural system is the efficient use of animal manure. Until now, the applied quantity of manure is regulated by law at farm level, based on fixed phosphorus (P) application norms. However, a first step towards more efficient manure application is to better balance P input and output at field level by predicting future P yields. Machine learning techniques can be useful in this respect, because they can be trained with many variables without prior knowledge regarding their interrelationship. This study's objective, therefore, was to predict P yields based on detailed records of on-farm data as recorded on an experimental farm combined with open source weather data. The dataset contained 657 records of annual crop yields per field between 1993 and 2016, and the boosted regression model was used for model development. Validation on the final five years of the dataset resulted in an RMSE of 7.3 kg P per ha per year, an R-squared of 0.46 and a correlation between observed and predicted values of 0.68, outperforming legal norms. We conclude that with the limited but detailed data available, prediction of P yield, and therewith, defining flexible P application norms before first manure application, is already feasible. This conclusion, together with the expected increasing availability of data through proximal and remote sensing technologies, opens the way to further improve nutrient management and move towards circular agriculture in the future.

### 1. Introduction

Nutrients, like nitrogen (N), potassium (K) and phosphorus (P), are essential for plant growth. Phosphate rock, the main source of P, is a non-renewable finite resource. Proper management of the P cycle, therefore, is necessary to guarantee future availability of P (Cordell et al., 2009; Van Kernebeek et al., 2018) and governmental policies have focused on P application norms for decades already (e.g., EC, 1991, 2000; MinEZ, 1986; MinLNV, 2005). Furthermore, agricultural and societal thinking and policy making is moving towards a circular agricultural system in which residuals of agricultural production and food processing are kept within the food system as renewable resources (MinLNV, 2018; WUR, 2018). An important factor in a circular agricultural system is the efficient use of animal manure for growing crops, which means balancing input and output of nutrients. Until now, however, manure application is regulated by law at farm level, based on fixed P application norms, with only three levels based on the P status of the soil of the field and two types of land use, grass and arable land. A

first step towards more efficient manure application could be to better balance P input and output at field level by predicting future P yields based on field specific, historical data. Models predicting future P yield may use information on fields, crops, soil parameters and open source weather data. Machine learning techniques can be useful in this respect, because they can be trained with many variables, without prior knowledge regarding their interrelationship or distribution function and with noisy data (Witten and Frank, 2005). In a review on the use of big data in precision dairy farming, Lokhorst et al. (2019) noticed that the vast majority of papers were on the animal sublevel and none regarding manure management. Recently, Nikoloski et al. (2019) showed the feasibility of a machine learning approach to predict grass yields under Irish circumstances, but did not include predictions into the future.

This study's objective is to predict future P yields based on detailed records of an experimental farm and open source weather data and to provide an indication for the suitability of machine learning models for P yield predictions.

\* Corresponding author.

E-mail address: [erwin.mollenhorst@wur.nl](mailto:erwin.mollenhorst@wur.nl) (H. Mollenhorst).

## 2. Materials and methods

### 2.1. Data sets

Data from a twenty-four year period (between 1993 and 2016) were available from experimental dairy farm 'De Marke', located on a light sandy soil in the eastern part of the Netherlands. The soils at 'De Marke' are characterized by a 25–30 cm anthropogenic upper layer, with an average organic matter content of 4.8%, overlying a layer of yellow sand, very low in organic matter and hardly penetrable by roots (Aarts et al., 2000). The P status of the soils was characterized with the P-AL value (Egnér et al., 1960; NEN-ISO 15923-1, 2013; NEN 5793, 2010), which indicates readily available P during the growing season. Over all fields and years the P-AL value was  $62.4 \pm 19.1$  mg P<sub>2</sub>O<sub>5</sub> per 100 g dry soil (mean  $\pm$  standard deviation). Fields with P-AL values > 50 are currently categorized as 'high' in Dutch manure policy and, therefore, receive more strict application norms. These high levels are due to historical high manure application rates (Reijneveld et al., 2010). At the establishment of the experimental farm in 1989 the average P-AL value was 75 and decreased by 25% in the subsequent 17 years (Verloop et al., 2010). Within the period 1996–2016 P-AL values decreased with 0.47 per year ( $P < 0.001$ , based on least square regression on the data used in the current study).

Every year in the dataset consisted of 26 to 28 fields, with 24 fields being present in all years. The average field size was 2 ha (range 0.5–4.2 ha), resulting in a total farm plan per year of 55 ha (range 51.9–56.5 ha). Six fields were permanent grassland, while the majority of the fields were used in a rotational crop scheme. This scheme, in general, consisted of three years grassland, two years of maize and one year a cereal crop, although the crop rotation varied sometimes depending on the farm planning and cereals were left out from 2013 onwards. During the first year maize after grass, in general, no manure or artificial fertilizer was applied. Italian ryegrass was used as a catch crop on maize fields and ploughed-in some months before sowing the next crop. On grassland, fertilizer was applied from March 1st till the end of August, whereas on second year maize, fertilizer was applied shortly after sowing in early May (Verloop et al., 2006).

The dataset contained 657 records of annual crop yields per field with information on N and P input and output, irrigation and soil status, all at field scale, and local weather data. Precipitation and temperature were measured at an on-farm weather station, whereas data on wind, sunshine, irradiation and evaporation were from the nearest weather station, located 21 km east of the farm (KNMI, 2018).

As most fields were present in the dataset for all, or at least several, years, the history of each field could be derived from the dataset. Furthermore, the dataset contained a variable stating the number of consecutive years the specific field was already grassland or arable land. Field inputs were reported as N and P inputs from artificial fertilizer, manure application, animal excreta during grazing, deposition and N-fixation by legumes. Data from yearly soil analyses being performed in autumn, reporting N, P, K and organic matter content per field, were used as input variables for the next year. Total dry matter and P yields per field were split into grazing and mowing for grassland, and, if applicable, in main and by-products for arable land. Yields of all products were measured, or estimated in the case of grazing, and composition, e.g. P content, was based on routine laboratory analyses. For the data analysis, the total P yield from all products of a field, expressed as kg P per ha per year, was used as the predicted variable. Predictions were made for a whole year before the first manure application, without knowledge of future weather circumstances.

### 2.2. Data pre-processing

Initially, besides grass, seven crops were reported in the dataset, namely maize, triticale, barley, barley combined with peas, fodder beets, lucerne and hemp. Maize, harvested as whole plant silage or as

**Table 1**

Overview of aggregation methods for weather data (Y = yes, N = no).

Variable per day	Monthly aggregate	Cumulative sum	Rolling average
Average temperature	Mean	Y	Y
Minimum temperature	Minimum	N	N
Minimum temperature at 10 cm <sup>1</sup>	Minimum	N	Y
Maximum temperature	Maximum	N	N
Precipitation	Sum	Y	N
Days dry	Count	N	N
Evaporation <sup>2</sup>	Sum	Y	N
Natural precipitation surplus <sup>3</sup>	Sum	Y	Y
Artificial precipitation surplus <sup>4</sup>	Sum	Y	Y
Average wind speed	Mean	N	Y
Hours of sunshine	Sum	Y	Y
Percentage of sunshine <sup>5</sup>	Mean	N	Y
Irradiation	Sum	Y	Y
Humidity	Mean	N	Y

<sup>1</sup> Measured at 10 cm above field level. All other temperature variables are measured at 1,5 m above field level.

<sup>2</sup> Reference crop evapotranspiration (modified Makkink equation, De Bruin and Lablans, 1998).

<sup>3</sup> Natural precipitation surplus = Precipitation – Evaporation.

<sup>4</sup> Artificial precipitation surplus = Precipitation + Irrigation – Evaporation; this is the only weather variable that is field specific.

<sup>5</sup> Percentage of maximum time possible.

maize ear silage and straw, was grouped together as maize. Furthermore, triticale, barley, and barley combined with peas were grouped together as grains, irrespective whether the barley was harvested as silage or as grain and straw. Because there were only 2 records of hemp, only 4 records of lucerne and only 7 records of fodder beets these records were deleted together with 4 records that lacked P yield figures, totalling to 640 records eligible for further analyses.

Field history was characterized by two lag-variables, stating the crop in the previous year and the crop two years back. Both lag-variables were used as input variables. Field N and P balance were calculated as total inputs, thus including N fixation and deposition, minus total outputs from main and by-products. Field balances in the past year and averaged over the past two years were used as input variables. Furthermore, the average P yield of grass, maize or grains in the past seven years on a specific field was used as input variable. Field history variables were set to missing when the required information from previous years was not available, for example in the first years of a field.

Weather data were available per day, but were aggregated in three different ways. Table 1 provides an overview of which weather variables were used as input variables and in what way. All weather variables were aggregated to month by reporting their means, minimum, maximum, sum or count. Cumulative sums were calculated starting January 1st of every year, which is common for temperature (Jagtenberg, 1970), and also applied to other weather variables in this study. Furthermore, for each day the average of the preceding five days (rolling average) was calculated for some weather variables. For cumulative sums and rolling averages of weather data, only values on the 5th, 10th, 15th, 20th, 25th, 30th for each month were used as input variables. Furthermore, data from November and December were discarded as those were outside and after the growing and harvesting season. In total, 1131 input variables were used during model development of which 1084 were weather related.

### 2.3. Model development

One of the basic and probably most studied machine learning techniques is decision tree induction (Witten and Frank, 2005). Decision trees repetitively split the data at each node into segments of

similar data points. Selecting the variable to split the data at each node is done based on which variable reduces the squared error the most. Advantages of decision trees are insensitiveness to the distributions of input variables, ability to handle missing values and their robustness against irrelevant or correlated input variables (Friedman, 2001). The main disadvantage of decisions trees, however, is their inaccuracy in prediction. To alleviate this problem, multiple models can be combined using ensemble methods, like bagging, boosting and stacking, of which boosting is considered to be the most powerful (Witten and Frank, 2005). Boosting is a stage-wise iterative method, which means that the method obtains the predictive results of one tree, computes the error and uses these errors to build the next tree, and thereby improves on the mistakes made by earlier trees. In the case of a continuous outcome variable, each iteration uses the residuals of the previous model (Witten and Frank, 2005). In this study, the Gradient Boosting Machine (GBM; h2o.gbm function (h2o version 3.20.0.2)) was used. Some model parameters were adapted to speed up the analysis. Model parameters used were 1000 trees (ntrees), i.e. number of iterations during model building, with a maximum of 3 splits per tree (max\_depth) and learning rate of 0.01 (learn\_rate). All analyses were performed in RStudio (version 1.1.423 running R version 3.5.0).

2.4. Model validation

To validate the model on an independent validation set, we used data from the years 1993–2011 as training data and validated on data from 2012. Subsequently, we used 1993–2012 as training data and validated on data from 2013, and so on for 2014, 2015 and 2016 as validation data. Next to this independent validation on the next year, we also split the records of the training set randomly in 70% training and 30% test set to test the performance of the model in the same time period as well (Fig. 1). For model testing and validation, weather data of the growing season, March to October, were set to missing, as they are not known at the moment of first fertilizer application.

2.5. Model performance criteria

Model performance was evaluated by plotting observed versus predicted values (Piñeiro et al., 2008), including a  $y = x$  line and linear regression line, and computing the root-mean-square error (RMSE), coefficient of determination ( $R^2$ ) and correlation coefficient. Performance metrics were calculated per dataset, which means separately for the training, test and validation dataset per iteration (Fig. 1). The RMSE was calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{pred} - y_{obs})^2}{n}} \tag{1}$$

with  $y_{pred}$  being the predicted value and  $y_{obs}$  the corresponding observed value (RMSE metric from h2o.performance function (h2o version 3.20.0.2)). The  $R^2$ , calculated relative to the  $y = x$  line was calculated as

$$R^2 = 1 - \frac{\sum (y_{obs} - \bar{y}_{obs})^2}{\sum (y_{obs} - y_{pred})^2} \tag{2}$$

with  $\bar{y}_{obs}$  being the mean of all  $y_{obs}$  corresponding with the predicted values and remainder as defined above (r2 metric from h2o.performance function (h2o version 3.20.0.2)). By calculating the  $R^2$  relative to the  $y = x$  line, the performance of the model is evaluated by comparing the predicted values with the observed ones and tells what fraction of the variation in observed values is covered by the predicted values. When  $R^2$  is 1 the model is able to predict the observed values perfectly. When  $R^2$  is 0, the model is just as good a predictor as the mean of the observed values. And when  $R^2$  is negative, the model is a worse predictor than the mean of the observed values. The fit of the regression line is represented by the Pearson correlation coefficient ( $r$ ), which was calculated using the cor function from R Stats package. This performance parameter, therefore, shows how well the trend in observed values is predicted by the model. Furthermore, observed versus predicted values of all five years of validation data were plotted together and the RMSE,  $R^2$  and correlation coefficient were calculated. These predictions were compared to legal manure application norms for fields classified as ‘high’ (> 50 mg  $P_2O_5$  per 100 g dry soil) with regard to available P level, as the reference for current practice. Statistical significance of differences in performance were tested using a simple randomization test with 1000 iterations for MSE, i.e., the squared RMSE (Van der Voet, 1994) or using a  $t$ -test (tsum.test function in R BSDA package, version 1.2.0) on Fisher’s-z transformed correlation coefficients (or square root of  $R^2$  values) (Cox, 2008).

3. Results

3.1. Descriptive statistics dataset

Yearly P yields ranged from 16 to 64 kg P per ha ( $n = 393$ ;  $36.4 \pm 7.9$ ; mean  $\pm$  SD) for grassland, 11–36 kg P per ha ( $n = 202$ ;  $22.3 \pm 4.7$ ; mean  $\pm$  SD) for maize and 18 to 43 kg P per ha ( $n = 45$ ;  $28.3 \pm 6.2$ ; mean  $\pm$  SD) for cereal crops for all years in the dataset (Table A.1). These data indicate that P yields with grass are higher ( $P < 0.001$ ) than with maize. P yields from cereal crops are intermediate and different ( $P < 0.001$ ) from P yields from both grass and maize. It has to be notified, however, that cereal yields include yields from grass as successive crop. Specified per validation year, figures are shown in Table 2. These figures indicate that P yields were high in 2014 for maize ( $P < 0.001$ ) and to a lesser extent also for grass ( $P = 0.06$ ) and relatively low in 2013 for grass ( $P < 0.001$ ) and to a lesser extent also for maize ( $P = 0.09$ ). Variation in grass P yields was high in 2014 and 2015.

Other descriptive statistics concerning management (Table A.1) and weather (Table A.2) related variables are shown in the Appendix.

3.2. Model performance

Model performance metrics are reported in Table 3. The correlation

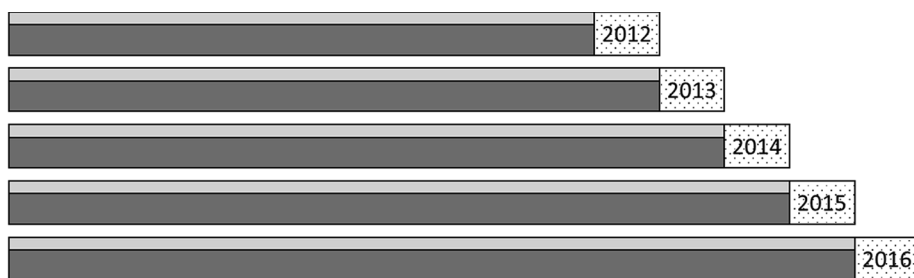


Fig. 1. Model validation strategy. Five iterations of model building on training data (dark grey; 70% randomly selected records) and testing on test data (light grey; remaining 30%) from the same period, and validation on an independent dataset (dotted) from a new year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**  
Number of records (n), mean and standard deviation (SD) of P yields per validation year and in the remainder of the dataset (1993–2011).

Year (s)	Grass			Maize			Cereals		
	n	Mean	SD	n	Mean	SD	n	mean	SD
< 2012	304	36.5	7.5	161	22.0	4.6	42	28.4	6.4
2012	18	34.3	5.4	7	22.6	7.1	3	26.7	1.3
2013	18	29.7	4.9	10	20.8	2.6			
2014	18	42.1	12.5	9	28.1	3.2			
2015	17	38.5	11.0	9	23.2	3.2			
2016	18	36.0	6.5	6	22.2	5.1			

**Table 3**  
Coefficient of determination ( $R^2$ ), root-mean-square error and Pearson correlation coefficient (r, all with  $P < 0.001$ ) of the models for 2012 to 2016 on train (70% of previous years), test (remaining 30% of previous years) and validation (listed year) datasets.

Year	Train			Test			Validation		
	$R^2$	RMSE	r	$R^2$	RMSE	R	$R^2$	RMSE	r
2012	0.96	1.95	0.98	0.71	4.56	0.85	0.63	4.52	0.80
2013	0.96	1.86	0.98	0.70	5.33	0.84	-0.64	7.61	0.80
2014	0.95	1.93	0.98	0.70	5.18	0.85	0.26	10.35	0.80
2015	0.95	2.11	0.98	0.63	5.53	0.80	0.58	7.43	0.77
2016	0.96	1.89	0.98	0.66	5.74	0.82	0.63	5.09	0.82

coefficients (r) on the validation sets, all around 0.8, indicate that the models were able to predict the trend from low to high P yield. The  $R^2$  and RMSE on the validation sets, however, indicate that the models had difficulty with predicting the exact levels of P yield in the years 2013 and 2014, while for the other 3 years,  $R^2$  was around 0.6 and RMSE was below 8 kg P per ha per year. The differences between observed and predicted values, or in other words the distance of the points to the  $y = x$  line, fully determine the RMSE (formula 1). The  $R^2$ , on the other hand, is dependent on the differences between observed and predicted values (numerator in formula 2) and on the variation in observed values (denominator of formula 2). As variation in observed P yields was low in 2013 (Table 2),  $R^2$  was very low (-0.64), even negative, with a reasonable RMSE (7.61), while in 2014 the variation in observed P yields was high (Table 2), resulting in a moderate  $R^2$  (0.26), while the RMSE was high (10.35).

Model performance on the train sets, that is the 70% of the data that were used for model development, was very good and decreased somewhat when the models were applied on the test sets, the 30% of that data that was used to test the model on data from the same time period (Table 3). Also over the years, model performance on the test sets decreased.

When the data from all five years were plotted together (Fig. 2), differences between years are visualised. For 2013, in general, the model overestimated the P yields, while for 2014, the model predominantly underestimated the P yield. Model performance criteria were also calculated for the combined results (bottom right in figure).

Furthermore, model performance criteria were calculated for the observed values plotted against the legal P application norms (Fig. 3). Comparison of these two plots showed that the GBM model performed better, both with respect to predicting the trend in observed P yields ( $r = 0.68$  for the model compared to 0.59 for the application norms,  $P < 0.001$ ) as with respect to predicting the exact levels ( $R^2 = 0.46$  vs. 0.32 ( $P < 0.001$ ) and RMSE = 7.33 vs. 8.22 ( $P < 0.05$ )).

#### 4. Discussion

This study has shown, as a proof of principle, that it is possible to predict P yields of individual fields before the first manure application

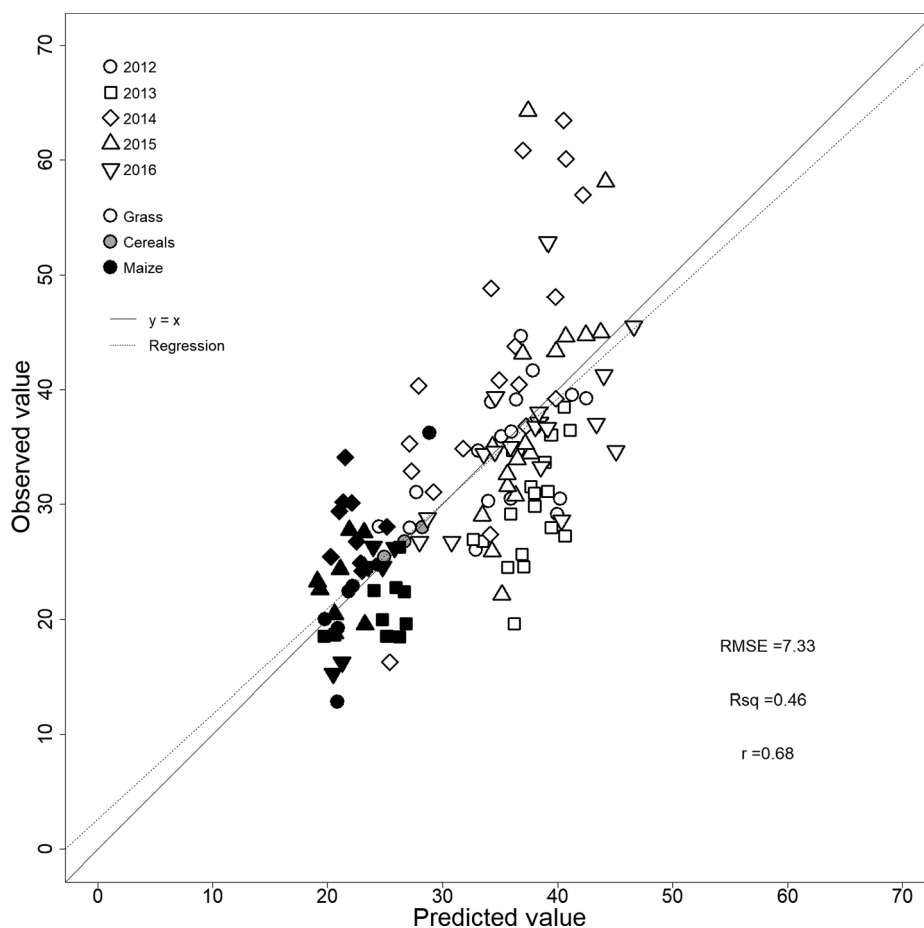
reasonably well with a data driven method like GBM. These predicted P yields could be used to support decisions on manure application which are closer to the realized yields than the current legal manure application norms. A closer balance between P application and P yield helps to decrease losses and leads to a more circular agricultural system.

This proof of principle, however, was based on a limited number of records (640) from only one farm, which means, in general, one management and one soil type. The developed prediction model, therefore, is not directly applicable on other farms and soil types. Furthermore, P levels in the soil were relatively high, especially in the early years of the dataset, which was quite common in the Netherlands. The somewhat limited range of soil P status could have diminished the influence of this parameter in the model. In this study, we have chosen for this particular farm, because of the availability of detailed and accurate data at field scale for a long time period. To develop and validate a model that is more widely applicable, data from many more farms on different soil types are needed, but these farms don't keep such detailed records yet. Including data from farms with less detailed records could result in a model that is applicable in a wider range of circumstances, while, on the other hand, it is not sure whether this would result in an improvement of the predictive performance of the final model. Considerable improvements, however, could be expected when more detailed information from many more farms will become available for model development and validation as machine learning techniques need lots of data to discover meaningful patterns. Data from other farms can be used by the algorithm to fill gaps in the data from a specific farm, for example situations with specific weather conditions or new crops. When a similar approach of data collection as applied in this study would be applied on commercial farms, this would result in considerable costs for additional recordings by the farmer, contractors and researchers, and costs for laboratory analyses. But in addition to these financial costs, the feasibility of this approach heavily relies on the willingness of farmers to spend their time on collecting the data. With the advent of precision technologies, like in-line manure or crop analysis sensors (Büscher et al., 2014), and proximal (Dennis et al., 2015) and remote sensing techniques, based on satellites (Ali et al., 2016; Cicore et al., 2016; Punalekar et al., 2018) or drones (Larsen et al., 2018), for yield monitoring, however, new opportunities arise. Currently, these technologies are not applied yet at a large scale on commercial farms.

In contrast to one prediction before the first manure application, as done in the current study, fertilizing management could be adapted during the season for grassland, since manure is applied on grasslands several times a year. Prediction could then be based on actual weather and field conditions as well as weather forecasts. In this way, weather data of March till the moment of last fertilizer application, which were ignored in the current study, could be utilized. In this study, however, all fields in all years, thus including use as arable land, were considered together and, therefore, yields were aggregated to whole year level. Results, therefore, need to be considered as a whole year fertilizing advice at the start of the growing season. Considering all fields in all years yielded the advantage that also information on field performance with different land use could be considered by the model. When a larger dataset is available it could be tested whether developing models per crop is more appropriate and probably a within season model for grassland could be a next step in data driven fertilizing management.

Capturing the year to year variation in P yields was difficult, as shown by some lower  $R^2$  and higher RMSE values, especially for the years 2013 and 2014. This was most probably caused by differences in weather. The GBM model overestimated P yields for 2013 and underestimated P yields for 2014 (Fig. 2), years that were recognized by some of the co-authors and by a colleague of the experimental farm as 'bad' and 'good' years, respectively, for growing grass and crops. Weather data were used during model development to correct historic P yields, but are unknown at the start of the growing season, and, therefore, set at missing in the test and validation datasets. Deviations in predictions





**Fig. 2.** Observed versus predicted plot of P yields of all five years of validation data including linear regression line (dotted line). Models were trained per year (different symbols), as described in Section 2.4 Model validation. Different fill colours represent different crops.

for certain years, therefore, were expected. But also when weather data were included in the validation dataset, the picture was very much alike (data not shown). This could have been caused by the difficulty of capturing the weather influence, especially when yields have to be predicted for years with weather conditions that deviate from those present in the training dataset. However, predictions based just on linearly extrapolated historic crop yields combined with field specific performance (data not shown) were not adequate and performed worse than the GBM model ( $P < 0.05$  for RMSE and  $P < 0.001$  for  $R^2$  and  $r$ ). To approach the influence of the weather on the yield more closely, other variables could be searched for, like ones indicating soil water content (Chen et al., 2014) or modelling water availability more precisely (Kroes and Van Dam, 2003).

Including the deviant years 2013 and 2014 in the train and test sets could have caused the lower performance of the models on the test sets in the years 2015 and 2016. Deviant data enrich the training set, but could also add noise and as such distract from the normal patterns. Furthermore, some input variables are defined as looking back just one or two years, which could hamper model performance when two deviating years appear in a row.

Another difficulty in predicting P yields from grassland is the fact that some fields are grazed and that grazing intensity differs greatly among fields. In this study, grazing was captured by one input variable stating whether a field was grazed or not during the season. Including variables that cover grazing intensity and frequency, e.g., number and type of animals, duration of grazing periods or estimated percentage of yield by grazing or mowing, should be considered when developing models on larger datasets, because different grazing systems will be

applied when more farms are included. Furthermore, the accuracy of the observed P yield values, was lower for grazing compared to mowing. Yield prediction during grazing is less precise, as it was based on calculations and assumptions instead of actual measurements, and it is more difficult to obtain a representative sample for P content determination due to regrowth during long grazing periods. With, additionally, a higher P content of younger grass (Wilman, 1975; Wilson and McCarrick, 1967) compared to older grass, errors will be enlarged. Additionally, the manure deposited during grazing is less well spread than when collected in the barn and spread by a machine, which could lead to less efficient usage of nutrients and more losses to the environment (Van Middelkoop et al., 2016).

Although we applied a data driven approach, when reflecting on the process of this study, it became clear that domain knowledge was also very important, next to data analytical knowledge and skills. Variables in the input dataset have to be defined in such a way that they make sense, but also in the interpretation of the results it became clear that almost all outliers could tell a story. Some appeared to be from a field with adverse conditions or a pest, information not captured by the used input variables. Others could be explained by grazing management and focussed attention to the fact that also the measured P yield in this study was not always really measured; e.g., uptake by grazing was based on a calculation of cows' requirements instead of real yield measurements. Even though we did not exclude these outliers, the model still performed equal or better than the currently used application norms in aiming for a closer balance between input and output of nutrients at field level and, therewith, to a more circular agricultural system.

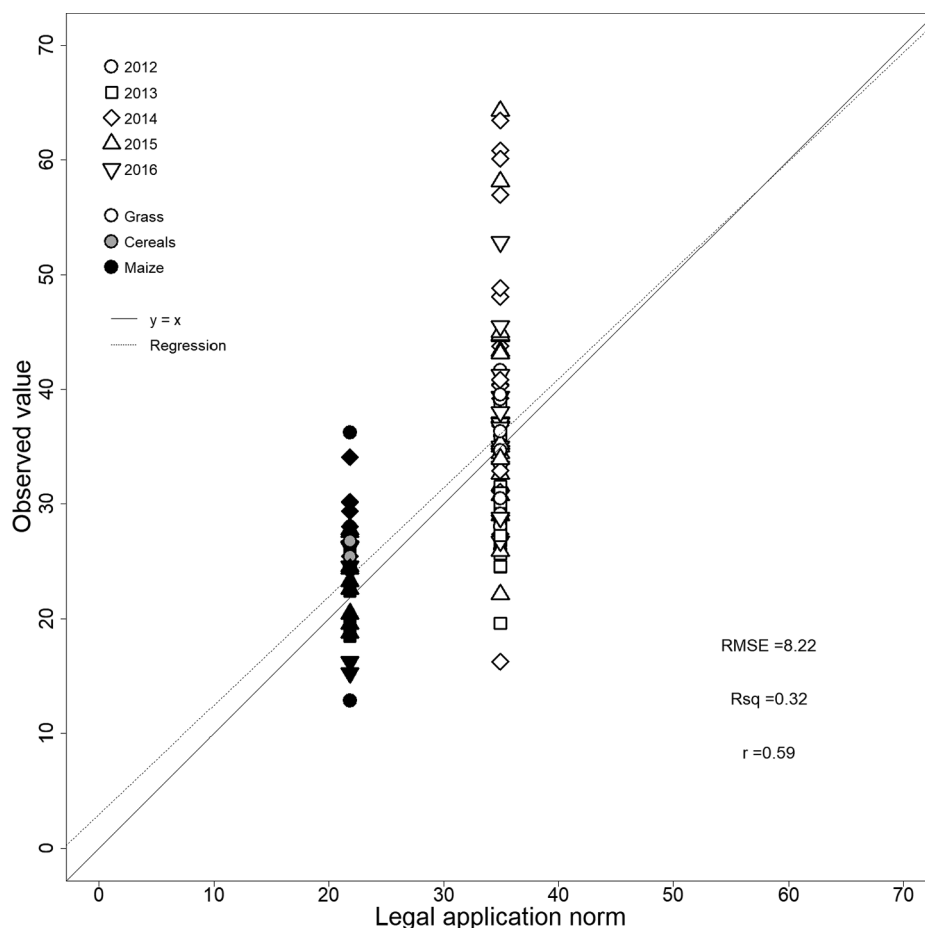


Fig. 3. Observed P yields plotted against the legal P application norms of all five years of validation data including linear regression line (dotted line). Models were trained per year (different symbols), as described in Section 2.4 Model validation. Different fill colours represent different crops.

## 5. Conclusions

We conclude that with the limited but detailed data available, prediction of P yield, and therewith, defining flexible P application norms before first manure application, is better than current fixed application norms. This conclusion, together with the expected increasing availability of data through proximal and remote sensing technologies, opens the way to further improve nutrient management and move towards circular agriculture in the future.

### CRedit authorship contribution statement

**Herman Mollenhorst:** Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft. **Michel H.A. Haan:** Investigation, Writing - review & editing. **Jouke Oenema:** Investigation, Data curation, Writing - review & editing. **Claudia Kamphuis:** Conceptualization, Methodology, Writing - review & editing, Supervision.

## Appendix

Appendix to COMPAG\_2019\_250 (Mollenhorst et al., Field and crop specific manure application on a dairy farm based on historical data and machine learning)

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors thank Knowledge Transfer Centre “De Marke”, Hengelo, Gelderland, the Netherlands, and especially Gerjan Hilhorst, for supplying the data and sharing valuable information and comments on the results. Rita Hoving is thanked for her supportive work during the project and Roel Veerkamp for his contribution in the initiation of this project.

This study was part of KB-27-001-001 ‘Big data for healthy resources utilisation’, funded by the Dutch Ministry of Economic Affairs.

**Table A.1**

Mean, standard deviation (SD), minimum and maximum of management related variables for the years 1993–2016.

Variable	Grass				Maize				Cereals			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
N input (kg per ha)	432.8	88.3	245.6	687.7	113.9	62.3	31.0	376.4	255.4	57.9	177.6	380.5
N output (kg per ha)	273.6	63.1	107.5	528.8	138.1	25.9	80.0	205.5	151.3	40.1	78.3	248.7
N surplus (kg per ha)	159.2	69.5	-33.0	331.8	-24.2	73.0	-165.0	223.6	104.2	48.0	31.4	234.9
P input (kg per ha)	39.9	9.1	19.4	68.4	12.6	9.8	1.0	50.9	26.7	8.5	13.2	49.9
P output (kg per ha)	36.4	7.9	16.3	64.3	22.3	4.7	11.5	36.2	28.3	6.2	18.1	43.1
P surplus (kg per ha)	3.5	10.8	-22.6	37.9	-9.7	11.7	-35.2	27.7	-1.5	8.9	-17.4	31.7
Total number of cuts	5.6	1.2	3.0	10.0	2.1	0.6	1.0	3.0	3.1	0.5	2.0	5.0
Number of mowing cuts	3.4	1.2	1.0	7.0					2.1	0.3	1.0	3.0
Average yield of mowing cut (kg DM per ha)	2205.9	497.1	952.0	4736.0					1553.6	514.1	786.5	3817.0
Number of grazing days	20.9	25.9	0.0	238.0	5.3	4.1	2.0	16.5	0.1	0.3	0.0	2.0

**Table A.2**

Monthly weather data for the years 1993–2016. The mean, minimum and maximum of the daily mean temperatures, and the lowest and highest temperature per month (in degrees Celsius). And monthly mean, minimum and maximum precipitation (in mm).

Month	Temperature				Precipitation			
	Mean	Lowest mean	Highest mean	Min.	Max	Mean	Min.	Max.
Jan	2.7	-2.2	6.3	-22.1	15.1	66.2	2.8	117.2
Feb	3.3	-1.6	7.3	-19.5	17.4	51.1	12.0	91.0
Mrt	5.8	1.7	8.6	-14.9	23.2	51.9	9.7	107.8
Apr	9.5	5.4	13.7	-8.3	28.0	40.5	0.2	81.8
Mei	13.3	9.7	15.4	-3.2	33.2	67.4	12.8	148.4
Jun	15.9	13.7	17.7	1.4	34.6	63.8	7.2	189.4
Jul	18.0	15.1	22.0	3.6	37.3	80.6	15.2	164.5
Aug	17.4	14.6	20.1	0.4	36.3	82.2	24.2	215.8
Sep	14.3	10.2	17.7	-1.4	32.4	65.3	4.6	210.0
Okt	10.2	5.8	13.6	-8.1	25.4	64.7	7.2	154.0
Nov	6.1	1.7	9.7	-10.4	19.3	61.3	8.8	128.0
Dec	3.2	-2.4	9.6	-15.5	16.1	66.7	16.0	175.0

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2020.105599>.

## References

- Aarts, H.F.M., Habekotté, B., Van Keulen, H., 2000. Ground water recharge through optimized intensive dairy farms. *J. Environ. Qual.* 29, 738–743. <https://doi.org/10.2134/jeq2000.00472425002900030008x>.
- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. *J. Plant Ecol.* 9, 649–671. <https://doi.org/10.1093/jpe/rtw005>.
- Büscher, W., Twickler, P., Maack, C., Marquering, J., Müller, M., 2014. NIRS Sensor controlled total-mixed-ration for nutrient optimized feeding of dairy cattle. *Proceedings of the 12th International Conference on Precision Agriculture. International Society of Precision Agriculture, Sacramento, CA.*
- Chen, T., De Jeu, R.A.M., Liu, Y.Y., Van der Werf, G.R., Dolman, A.J., 2014. Using satellite based soil moisture to quantify the water driven variability in NDVI: A case study over mainland Australia. *Remote Sens. Environ.* 140, 330–338. <https://doi.org/10.1016/j.rse.2013.08.022>.
- Cicore, P., Serrano, J., Shahidian, S., Sousa, A., Costa, J.L., Da Silva, J.R., 2016. Assessment of the spatial variability in tall wheatgrass forage using LANDSAT 8 satellite imagery to delineate potential management zones. *Environ. Monit. Assess.* 188. <https://doi.org/10.1007/s10661-016-5512-z>.
- Cordell, D., Drangert, J.-O., White, S., 2009. The story of phosphorus: Global food security and food for thought. *Global Environ. Change* 19, 292–305. <https://doi.org/10.1016/j.gloenvcha.2008.10.009>.
- Cox, N.J., 2008. Speaking Stata: Correlation with confidence, or Fisher's z revisited. *Stata J.* 8, 413–439. <http://www.stata-journal.com/article.html?article=pr0041>.
- De Bruin, H.A.R., Lablans, W.N., 1998. Reference crop evapotranspiration determined with a modified Makkink equation. *Hydrol. Process.* 12, 1053–1062. [https://doi.org/10.1002/\(SICI\)1099-1085\(199806\)12:7<1053::AID-HYP639>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-1085(199806)12:7<1053::AID-HYP639>3.0.CO;2-E).
- Dennis, S.J., Taylor, A.L., O'Neill, K., Clarke-Hill, W., Dynes, R.A., Cox, N., Van Koten, C., Jowett, T.W.D., 2015. Pasture yield mapping: Why & how. *J. New Zealand Grasslands* 77, 41–46.
- EC, 1991. Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources. *Official J. L* 3751–3758.
- EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official J. L* 327, 1–73.
- Egnér, H., Reihm, H., Dominga, W.R., 1960. Untersuchungen über die chemische Bodemanalyse als Grundlage für die Beurteilung des Nährstoffzustandes der Böden II. Chemische extraktionsmethoden zur Phosphor- und Kaliumbestimmung (in German; Soil tests as basis of establishing soil nutrient status. II. Chemical extraction methods for phosphorus and potassium). *Kungliga Landbruks Högskole Annaler* 26, 119–215.
- Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29, 1189–1232. <http://www.jstor.org/stable/2699986>.
- Jagtenberg, W.D., 1970. Predicting the best time to apply nitrogen to grassland in spring. *Grass Forage Sci.* 25, 266–271. <https://doi.org/10.1111/j.1365-2494.1970.tb01202.x>.
- KNMI, 2018. Daggegevens van het weer in Nederland (Daily weather data in the Netherlands). The Royal Netherlands Meteorological Institute (KNMI).
- Kroes, J.G., Van Dam, J.C., 2003. Reference manual SWAP version 3.0.3, Alterra-report 773. Alterra, Green World Research, Wageningen, the Netherlands.
- Larsen, D., Skovsen, S., Steen, A.K., Grooters, K., Eriksen, J., Green, O., Jørgensen, N.R., 2018. Autonomous mapping of grass-clover ratio based on unmanned aerial vehicles and convolutional neural networks. *Proceedings of the 14th International Conference on Precision Agriculture. International Society of Precision Agriculture, Montreal, Quebec, Canada.*
- Lokhorst, C., De Mol, R.M., Kamphuis, C., 2019. Invited review: Big Data in precision dairy farming. *Animal* 13, 1519–1528. <https://doi.org/10.1017/S1751731118003439>.
- MinEZ, 1986. Wet van 27 november 1986, houdende regelen inzake het verhandelen van meststoffen en de afvoer van mestoverschotten (in Dutch). Ministry of Economic Affairs, The Hague, the Netherlands.
- MinLNV, 2005. Besluit van 9 november 2005, houdende regels ter uitvoering van de Meststoffenwet (Uitvoeringsbesluit Meststoffenwet) (in Dutch). Ministry of Agriculture, Nature and Food Quality, The Hague, the Netherlands.

- MinLNV, 2018. Landbouw, natuur en voedsel: waardevol en verbonden - Nederland als koploper in kringlooplandbouw (in Dutch). Ministry of Agriculture, Nature and Food Quality, The Hague, the Netherlands.
- NEN-ISO 15923-1, 2013. Water quality - Determination of selected parameters by discrete analysis systems - Part 1: Ammonium, nitrate, nitrite, chloride, orthophosphate, sulfate and silicate with photometric detection. Nederlands Normalisatie Instituut (Dutch Standardisation Institute), Delft, the Netherlands.
- NEN 5793, 2010. Soil - Determination of phosphate in soil extractable with ammoniumlactate-acetic acid buffer (P-AL). Nederlands Normalisatie Instituut (Dutch Standardisation Institute), Delft, the Netherlands.
- Nikoloski, S., Murphy, P., Kocev, D., Dzeroski, S., Wall, D.P., 2019. Using machine learning to estimate herbage production and nutrient uptake on Irish dairy farms. *J. Dairy Sci.* <https://doi.org/10.3168/jds.2019-16575>.
- Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecol. Model.* 216, 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>.
- Punalekar, S.M., Verhoef, A., Quaife, T.L., Humphries, D., Birmingham, L., Reynolds, C.K., 2018. Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sens. Environ.* 218, 207–220. <https://doi.org/10.1016/j.rse.2018.09.028>.
- Reijneveld, J.A., Ehlert, P.A.I., Termorshuizen, A.J., Oenema, O., 2010. Changes in the soil phosphorus status of agricultural land in the Netherlands during the 20th century. *Soil Use Manag.* 26, 399–411. <https://doi.org/10.1111/j.1475-2743.2010.00290.x>.
- Van der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomization test. *Chemometrics Intellig. Lab. Syst.* 25, 313–323. [https://doi.org/10.1016/0169-7439\(94\)85050-X](https://doi.org/10.1016/0169-7439(94)85050-X).
- Van Kernebeek, H.R.J., Oosting, S.J., Van Ittersum, M.K., Ripoll-Bosch, R., De Boer, I.J.M., 2018. Closing the phosphorus cycle in a food system: Insights from a modeling exercise. *Animal* 12, 1755–1765. <https://doi.org/10.1017/S1751731118001039>.
- Van Middelkoop, J.C., Van der Salm, C., Ehlert, P.A.I., De Boer, I.J.M., Oenema, O., 2016. Does balanced phosphorus fertilisation sustain high herbage yields and phosphorus contents in alternately grazed and mown pastures? *Nutr. Cycl. Agroecosyst.* 106, 93–111. <https://doi.org/10.1007/s10705-016-9791-0>.
- Verloop, J., Boumans, L.J.M., Van Keulen, H., Oenema, J., Hilhorst, G.J., Aarts, H.F.M., Sebek, L.B.J., 2006. Reducing nitrate leaching to groundwater in an intensive dairy farming system. *Nutr. Cycl. Agroecosyst.* 74, 59–74. <https://doi.org/10.1007/s10705-005-6241-9>.
- Verloop, J., Oenema, J., Burgers, S.L.G., Aarts, H.F.M., Van Keulen, H., 2010. P-equilibrium fertilization in an intensive dairy farming system: Effects on soil-P status, crop yield and P leaching. *Nutr. Cycl. Agroecosyst.* 87, 369–382. <https://doi.org/10.1007/s10705-010-9344-x>.
- Wilman, D., 1975. Nitrogen and Italian ryegrass 2. Growth up to 14 weeks: Nitrogen, Phosphorus, Potassium content and yield. *J. British Grassland Soc.* 30, 243–249. <https://doi.org/10.1111/j.1365-2494.1975.tb01384.x>.
- Wilson, R.K., McCarrick, R.B., 1967. A Nutritional Study of Grass Swards at Progressive Stages of Maturity: 1. The Digestibility, Intake, Yield and Chemical Composition of Dried Grass Harvested from Swards of Irish Perennial Ryegrass, Timothy and a Mixed Sward at Nine Progressive Stages of Growth. *Irish J. Agric. Res.* 6, 267–279 <http://www.jstor.org/stable/25555453>.
- Witten, I.H., Frank, E., 2005. *Data mining: Practical machine learning tools and techniques*, Second ed. Elsevier/Morgan Kaufmann, San Fransisco, CA.
- WUR, 2018. *Circular agriculture: A new perspective for Dutch agriculture*. Wageningen UR, Wageningen, the Netherlands.