

# Prediction of soil pH patterns in nature areas on a national scale

G. W. Wieger Wamelink  | Dennis J. J. Walvoort | Marlies E. Sanders |  
Henk A. M. Meeuwse | Ruut M. A. Wegman | Rogier Pouwels | Martin Knotters

Wageningen University & Research,  
Wageningen, The Netherlands

## Correspondence

G. W. Wieger Wamelink, Wageningen  
University & Research, Wageningen, The  
Netherlands.

Email: [Wieger.wamelink@wur.nl](mailto:Wieger.wamelink@wur.nl)

## Funding information

The Dutch Ministry of Agriculture, Nature  
and Food Quality through the Statutory  
Research Tasks Unit for Nature & the  
Environment.

Co-ordinating Editor: Kris Verheyen

## Abstract

**Question:** To assess the acidification process, nationwide information about soil pH on a site level is called for. Measurements of soil pH may be used, however there are not sufficient measurements available to map soil pH nationwide on site level. Instead we developed a soil pH map based on vegetation data.

**Location:** Natural terrestrial areas in The Netherlands.

**Methods:** 271,693 vegetation plots were used to estimate average soil pH per plot with indicator values, based on field measurements, of plant species. By spatial interpolation average pH values between the plots, with the soil type, groundwater table and vegetation management type as ancillary explanatory variables we created a soil pH map. The map covers all terrestrial nature areas (all areas that are not built up areas, agricultural areas and infrastructural areas) in the Netherlands with a map resolution of  $25 \times 25 \text{ m}^2$  raster cells.

**Results:** The predicted pH of the map varied between 3.0 and 8.6 with standard errors between 0.13 and 0.93. Most of the standard errors range from 0.4 to 0.55, with an average just below 0.5 pH unit. Cross-validation shows that for 33% the difference between observed and predicted is between  $-0.1$  and  $0.1$  pH-unit and for 83% the difference is between  $-0.5$  and  $0.5$  pH-unit. Validation shows that the pH map is unbiased (mean error is almost zero), accurate (root mean squared error is 0.64) and nicely captures spatial patterns ( $r = 0.77$ ). We applied the pH map to assess the impact of acidification on the abiotic quality of nature areas in the Netherlands.

**Conclusions:** The model fit in the predicted soil pH is in good resulting in a low standard error and a high correlation. The measures taken to prevent acidic deposition causing further acidifying of nature areas can be considered as successful.

## KEYWORDS

acidity, kriging, pH, relevé, soil type, vegetation

## 1 | INTRODUCTION

Since the beginning of the industrial revolution acid deposition increased with a major peak around 1970 in Europe and later in North America forming a major problem to human health, cultural heritage and biodiversity. Countermeasures were taken to decrease the sulphur emissions causing the acidification of soils and thus a lower soil pH (Buijsman et al., 2010; Driscoll et al., 2001; Schöpp, Posch, Mylona, & Johansson, 2003). Acidification was substantially reduced between 1990 and 2010 in sensitive nature areas that were subjected to acid deposition of excess sulphur and nitrogen compounds (Kros et al., 2016; Schöpp et al., 2003; Stevens et al., 2010). At the same time a restoration programme was started to finance research and measures to restore soil pH, original low nutrient levels etcetera (Bootsma, Van Den Broek, Barendregt, & Beltman, 2002; De Graaf, Verbeek, Bobbink, & Roelofs, 1998; Lamers et al., 2015). This programme was necessary to restore and create environmental conditions that make the sustainable conservation of biodiversity possible. To evaluate the effects of the measures taken or still planned to be taken as a result of government policies, a map of the status of the environmental conditions contains essential information.

Soil parameters such as nutrient availability, soil pH and water availability affect the growth of plants (Bobbink, Hornung, & Roelofs, 1998; Roem & Berendse, 2000). Soil pH determines nutrient availability, as well as the availability of heavy metals (Deng, Ye, & Wong, 2004; Patra & Sharma, 2000; Schwertmann, 1991). High soil pH values, in combination with calcium, may reduce the phosphorous availability, whereas low soil pH values may cause metals, like the toxic aluminium, to become available and harm plant growth. Soil pH also influences bacterial activity thus influencing the form of available nitrogen: nitrate versus ammonium (Marschner, Crowley, & Yang, 2004; Stevens et al., 2011). These direct and indirect effects of soil pH determine plant species occurrence (Falkengren-Grerup, 1986; Roem & Berendse, 2000; Silvertown et al., 2006; Wamelink, Goedhart, van Dobben, & Berendse, 2005). As a consequence, plant species occurrence can be used as indicators for soil pH (see e.g., Wamelink et al., 2005; Wasof et al., 2013).

Soil pH in natural areas is not only, indirectly, influenced by industrial activities, but since long by other anthropogenic activities, such as land use, manuring, nitrogen deposition and nowadays even human induced climate change (Bobbink et al., 1998; Falkengren-Grerup, 1986; Roem & Berendse, 2000; Van Den Berg et al., 2005). Soil pH is also influenced by natural processes such as mineralisation, that causes a decrease in pH, weathering, soil displacement (for example by digging of animals), erosion and sedimentation, that may cause an increase in soil pH (Van Breemen, Mulder, & Driscoll, 1983). Furthermore, soil pH depends on the soil type, the supply of carbonate rich water, and plant litter.

For the evaluation of the effects of measures taken or planned in the field (environmental scenario's) as a result of government policies to mitigate acidifying effects or nature restoration insight

in the soil pH is necessary. Therefore, we developed a soil pH map for terrestrial nature areas in the Netherlands. Because of the lack of sufficient costly field measurements, we used the Dutch vegetation database and the responses of plant species established by Wamelink et al. (2005, 2012) to estimate soil pH. In combination with information about groundwater table, soil type and vegetation type we predicted a soil pH map for terrestrial nature areas in the Netherlands. We define terrestrial nature areas here as all terrestrial areas that are not built up areas, agricultural areas and infrastructural areas.

## 2 | MATERIAL & METHODS

### 2.1 | Overview of the method

The basis for the soil pH map is National Dutch vegetation (NDV) database (Hennekens & Schaminée, 2001; Schaminée, Hennekens, & Ozinga, 2012) and the indicator values for soil pH developed by Wamelink et al. (2005). The NDV contains vegetation relevés from different vegetation types and are spread over all nature areas in the Netherlands. The soil pH of these relevés is unknown, the soil pH is instead estimated. For this we use the indicator values for soil pH. These indicator values are based on field measurements of the soil pH (see further Appendix S1 and Wamelink et al., 2005). The estimated soil pH, as the average of the indicator values of the species present in the relevé, where the basis for the estimation of the pH map. To go from point data to the grid cell level we applied spatial interpolation, to interpolate between data points. This model does not only include the estimated pH but also groundwater table, soil type and vegetation type. The standard error of the estimated soil pH with the indicator values was used to weigh the estimated soil pH. With the regression model a soil pH and standard error was estimated for each terrestrial nature area grid cell. Subsequently a cross validation and a validation on independent data were carried out. The map was applied to evaluate the current state of the Dutch nature areas by comparing the pH requirements of the vegetation types with the estimated soil pH on our map.

The applied method to estimate the soil pH map is in short:

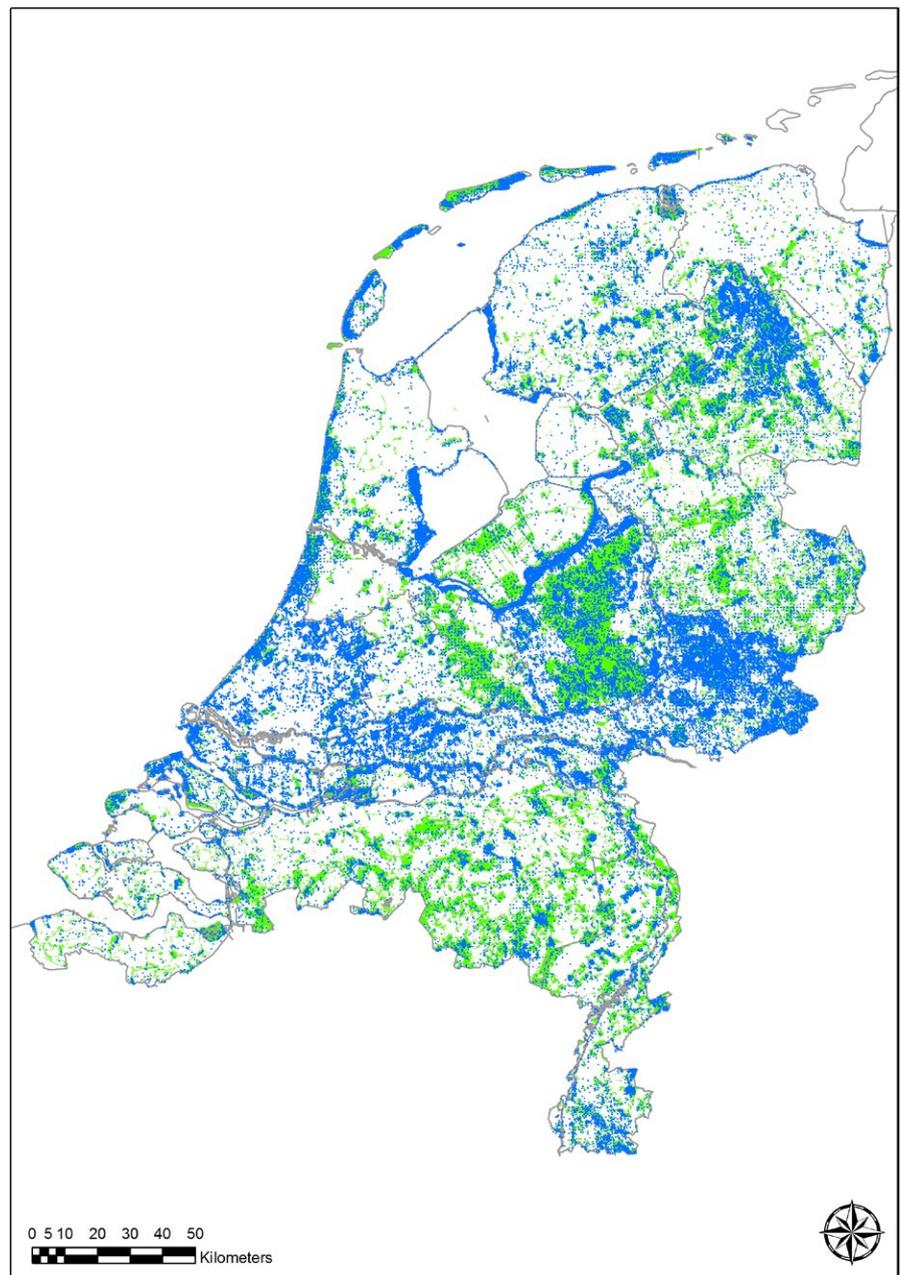
1. Our data are vegetation derived mean pH values at almost 300,000 locations including an associated standard error. The model to obtain these pH values and standard errors is given in Wamelink et al. (2005).
2. To interpolate these values to a dense  $25 \times 25 \text{ m}^2$  raster, we fitted a random forest model by correlating these mean pH values with the following covariates: soil type, groundwater class and vegetation type. Contrary to the mean pH values, these covariates were available for all  $25 \times 25 \text{ m}^2$  raster cells. By applying the random forest model, predictions of pH were obtained for all  $25 \times 25 \text{ m}^2$  raster cells.
3. The random forest predictions were subsequently used as trend component in a kriging model. In this step, we also took the

standard errors of step 1 into account (Delhomme, 1978). Hence, more weight is given to locations with smaller standard errors.

## 2.2 | Database and spatial prediction of soil pH

The vegetation plots (relevés) used for the estimation of soil pH were retrieved from the National Dutch vegetation database (Hennekens & Schaminée, 2001; Schaminée et al., 2012). The database has an entry in the "Global index of vegetation-plot databases (GIVD)" under number EU-NL-001 (<https://www.givd.info/ID/EU-NL-001>). This database contains relevés where at least all higher plants were identified, a part of them also include bryophytes and lichens. They were made by many authors for many different purposes. All relevés were made following Braun-Blanquet (1964) method or similar methods. We only used georeferenced relevés made after 1 January 1990.

Based on the indicator values per species estimated by Wamelink et al. (2005, 2012) the soil pH  $H_2O$  was estimated as the arithmetic mean for each relevé (including bryophytes and lichens, if available). Relevés without species with an indicator value were omitted. The indicator values per species are based on field data. Per species a response function was estimated for measured soil pH measured in water extract. Only soil samples of the upper soil layer (20 cm) were included. The indicator value of a species is the optimum value from the response curve (a spline function; see further Wamelink et al., 2005 for the method and Wamelink et al., 2012 for the description of the database; this database has also an entry in the GIVD under number EU-00-006; <https://www.givd.info/faces/databases.xhtml>). The final selection included 271,693 relevés, spread all over the country (Figure 1). For each relevé we calculated the standard error of the mean as well. Both the arithmetic mean and the standard error



**FIGURE 1** Location of the relevés used to predict the soil pH map (blue dots). In the background (in green) the terrestrial nature areas in the Netherlands

were used in the statistical analyses for mapping soil pH. The standard error was used to give different weights to the mean soil pH. A pH map was made for terrestrial and semi-terrestrial vegetation management types only (including e.g., swamps, fens and so on). For the spatial interpolation of soil pH we used a soil type, groundwater table and vegetation type as ancillary information. A more extensive description of the data and the used method can be found in Appendix S1.

### 2.3 | Soil type and groundwater table

Soil type and soil pH are related and can therefore be used to improve prediction of soil pH. We used the national Dutch soil map scale 1:50,000 (De Vries, 1999; De Vries, de Groot, Hoogland, & Denneboom, 2003) generalized to 43 soil mapping units. We also used the groundwater table class of the soil map that distinguishes eight different classes of depths at which groundwater tables seasonally fluctuate.

### 2.4 | Vegetation management type

The vegetation management map consists of 37 terrestrial vegetation management types (Appendix S2). This map was primarily set up for the benefit of nature managers and one of the major goals was to appoint subsidies to the nature areas for vegetation management such as mowing and grazing and the improvement of the nature (abiotic) quality. The map was set up by field experts and is independent from the other variables used. The vegetation management types are rather broad and not defined at the more precise phytosociological classes (association) level. For instance, pine, oak and beech forest have comparable management in the Netherlands so they ended up in one vegetation management type 'pine, oak and beech forest' (N15.02). Most of these types are more or less semi-natural, since all are managed, even the raised bogs (removal of trees, hydrological management). Some nature areas on the map are defined at a landscape level, and have their own section in the typology. Because these landscape types are too broad for our purposes, we added additional information to split these types up to the level of the semi-natural vegetation management types. This was done for Dune and Salt Marsh landscapes (N01.02), River and Marsh landscapes including flood plains (N01.03) and Sand and Chalk Landscape (N01.04). This downscaling was carried out based on soil type, groundwater table and satellite data.

### 2.5 | Statistical analyses

Spatial interpolation is often used to predict values at grid points given values at neighbouring points. The result is a raster map. To apply spatial interpolation the polygon and point maps we used, were all transformed into raster maps with  $25 \times 25 \text{ m}^2$  cells. A well-known spatial interpolation technique is kriging. We applied this geostatistical technique to predict pH-H<sub>2</sub>O on a regular grid for the Netherlands. Our general model is:

$$y(\mathbf{s}) = m(x_1, x_2, \dots, x_p; \mathbf{s}) + r(\mathbf{s})$$

where:

$y(\mathbf{s})$ : the pH at location  $\mathbf{s}$ , with  $\mathbf{s}$  a vector of spatial co-ordinates

$m(\cdot; \mathbf{s})$ : a model that gives the spatial trend of the pH at location  $\mathbf{s}$   
 $x_1 \dots x_p$ : explanatory variables that are related to  $y(\mathbf{s})$

$r(\mathbf{s})$ : a spatially correlated residual at location  $\mathbf{s}$ , i.e., the spatial variation not accounted for by model  $m(\cdot; \mathbf{s})$ .

Our geostatistical model is often referred to as kriging with an external drift (Goovaerts, 1997). It is well known that kriging predictions generally benefit from taking auxiliary explanatory variables  $x_1, \dots, x_p$  into account.

The deterministic trend model  $m(x_1, x_2, \dots, x_p; \mathbf{s})$  is usually written as a weighted linear combination of  $p$  explanatory variables  $x_i$ ,  $i = 1 \dots p$ :

$$m(x_1, x_2, \dots, x_p; \mathbf{s}) = b_0 + b_1 x_1(\mathbf{s}) + b_2 x_2(\mathbf{s}) + \dots + b_p x_p(\mathbf{s})$$

In our approach we will take a different view by writing this expression as

$$m(x_1, x_2, \dots, x_p; \mathbf{s}) = b_0 + b_1 f(x_1, x_2, \dots, x_p; \mathbf{s})$$

where  $f(\cdot; \mathbf{s})$  is a random forest model (Breiman, 2001). A random forest model is a machine learning model that uses bootstrap samples from the original data to build an entire forest of regression trees models. The explanatory variables,  $x_i$ ,  $i = 1 \dots p$ , are run down each tree to get predictions of pH. The predictions of all trees are combined to an ensemble prediction of pH.

A random forest model is able to capture nonlinear structures in the data and is relatively easy to fit. On top of that, it provides out-of-bag validation statistics which can be used to judge the fit. On the downside, a random forest prediction may be biased, but this potential bias will be corrected for by parameters  $b_0$  and  $b_1$  of our geostatistical model.

We took soil types, vegetation management types and groundwater table classes as explanatory variables in our random forest model  $f(\cdot; \mathbf{s})$  to predict the spatial trend of pH for each grid cell. We used the random Forest package (Liaw & Wiener, 2002) in R (R Core Team, 2016) for fitting the random forest model. The gstat-package (Pebesma, 2004) in R is subsequently used for semivariogram estimation/modelling and kriging. We used the functions randomForest and krige for, random forest and kriging respectively. The uncertainty in the "observations"  $y(\mathbf{s})$ , which are in fact estimates from indicator values of pH for species, is explicitly taken into account by incorporating the (square of the) standard errors of the pH-estimates in the kriging equations following the procedure of Delhomme (1978), referred to as "kriging with uncertain data."

### 2.6 | Uncertainty

Solving our geostatistical model not only leads to predictions of pH, but also to the associated variance of the prediction error.

Smaller variances indicate more accurate predictions. The variances of the prediction error are usually zero at the observations points. In our situation, this is generally not the case, as we took the standard error of each pH estimate explicitly into account in the kriging system.

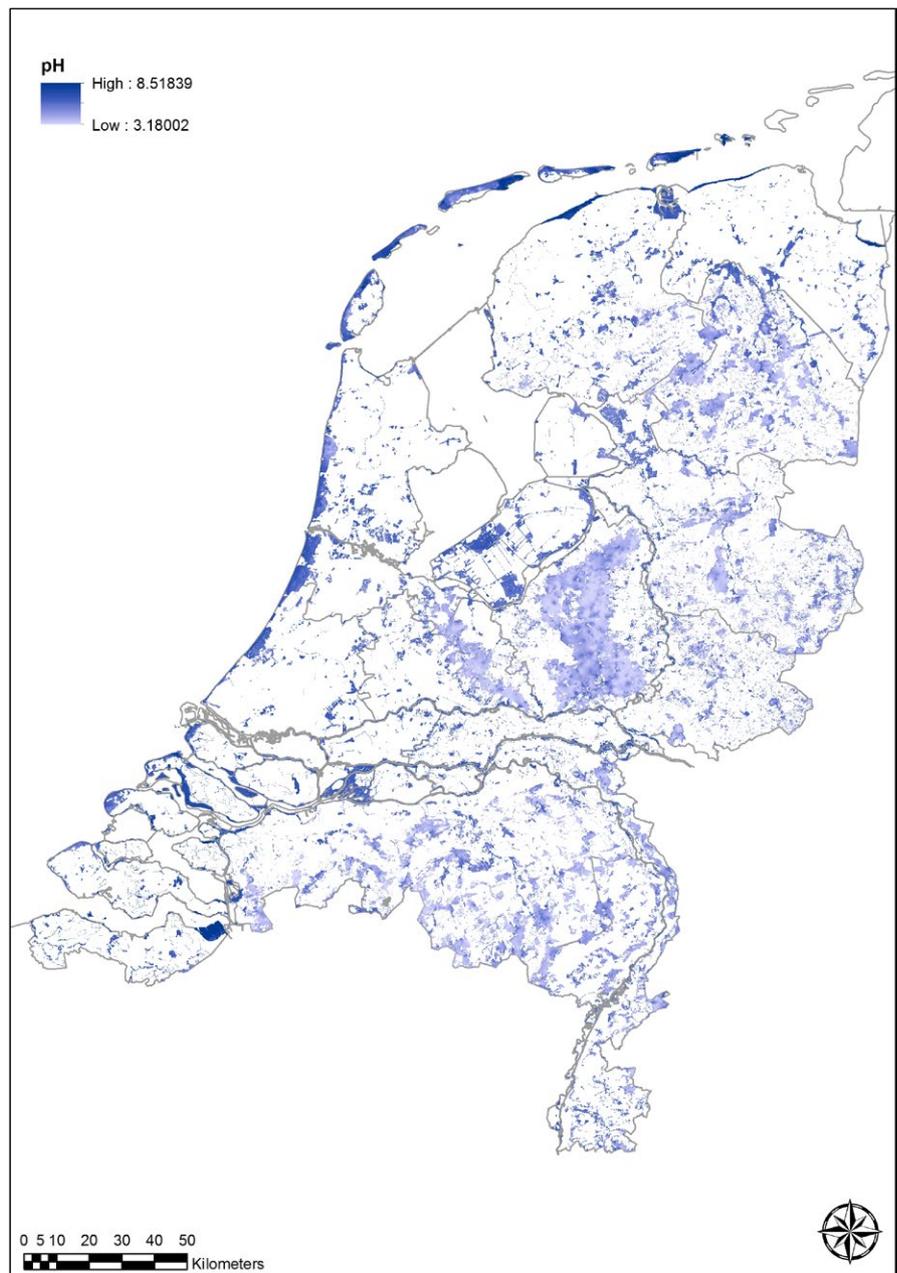
## 2.7 | Cross-validation

To validate the predictions of the pH-map, we used leave-one-out cross-validation (Efron & Gong, 1983; Efron & Tibshirani, 1994; Wehrens, Putter, & Buydens, 2000). This is a well-established validation method where each observation is successively removed from the data and its value is predicted given the remaining data. In this way, an error can be computed for each data location. The errors are

then summarised by a statistic like the mean error, as a measure of bias, or the root mean squared error, as a measure of accuracy.

## 2.8 | Validation

The predictions of the pH-map were also validated against an independent dataset. The BIS dataset (<http://maps.bodemdata.nl/bodemdata.nl/index.jsp>, last visited 17-6-2018) contains field measurements of soil pH that were not used for the estimation of the map. The BIS database contains only a limited amount of measurements in nature areas, \* in total, since it mostly focusses on agricultural areas. The database contains data of soil pH measured in H<sub>2</sub>O, CaCl<sub>2</sub> en KCl extract, all three were used for the validation. Field data were linked to the pH map via the coordinates of the field data.



**FIGURE 2** Predicted soil pH map for the Netherlands based on vegetation relevés, soil map, groundwater table and vegetation type for nature areas

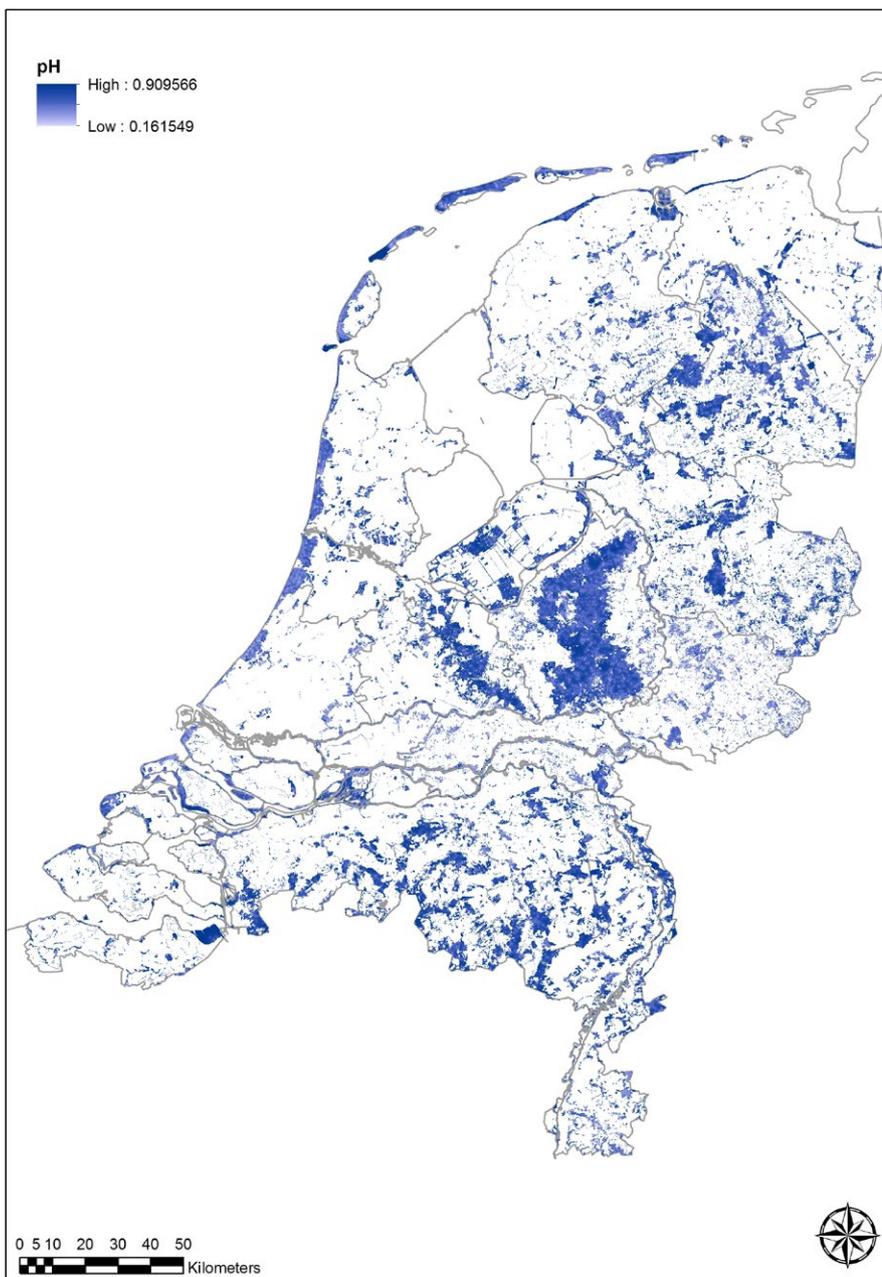
National assessment of the soil pH, linked to soil quality. The predicted soil pH map was used to assess the abiotic quality of designated vegetation management types for all terrestrial nature areas in the Netherlands. We used threshold values per vegetation type to assess the quality. The threshold values were derived from vegetation relevés that are linked to and are exemplary for the vegetation types. A response curve was estimated per vegetation type and the threshold values were defined as the 5 and 25 percentiles of the response curve (spline; Wamelink et al., 2011). The defined threshold values of the optimal conditions for soil pH per vegetation management type at a certain raster cell were compared with the predicted soil pH of the same raster cell. This was only done for the vegetation management types that have defined threshold values of the optimal conditions for soil pH (Appendix S2). If the predicted soil pH is above the maximum threshold value for pH for the vegetation management type, we

assume that the abiotic quality for soil pH is considered good. If the predicted soil pH lies between the threshold values for maximum and minimum pH for the vegetation management type, the abiotic quality for soil pH is considered moderate. If the predicted value is below the minimum threshold value for the vegetation type the abiotic quality is considered bad; The threshold values are based on field measurements similar to the estimation of the response functions for the plant species (Wamelink et al., 2005, 2011).

### 3 | RESULTS

#### 3.1 | Patterns in predicted soil pH

The soil pH, based on the indicator values (Wamelink et al., 2005), was predicted in a raster map for all natural areas by means of



**FIGURE 3** Standard error for the soil pH map for the Netherlands

kriging and varies between 3.0 and 8.6 (Figure 2). As was expected the sandy soils in the East show on average lower pH values than the clayish areas in the West and the North, and the more calcareous soils in the South and in the coastal dune area. The map also indicates that soil pH may vary largely within a few kilometres, also linked to differences in vegetation management type, soil type and groundwater table. In the dune area (Appendix S3) the highest pH values are found closest to the sea, in the youngest yellow dunes. The older dunes in the eastern part have lower values, which can be expected from older grey dunes. Here natural processes lead to succession to heathers and forest with leaching of calcium as a result of natural acidification due to decomposition and build-up of organic matter (Kreutzer 1995; Xu, Tang, & Chen, 2006). In the East of the Netherlands, where the nature areas are scattered in an agricultural landscape, soil pH also seems locally highly variable within the individual areas (Appendix S4). This indicates that the method in even such a fragmented landscape is able to create gradients. The Dutch nature areas have two distinct large peaks in their soil pH, one around pH 4 and one around pH 6, with a minor peak around pH 7.7 (Appendix S5). This agrees with the many more acid sandy soils that occur in The Netherlands than light acidic clayey soils under nature. The pH peak of 7.7 is caused by calcareous soils in the South-East and the salty calcareous soils (dunes and salt marshes) in the coastal zone. Hence, the predicted soil pH patterns meet all our expectations based on expert knowledge.

### 3.2 | Uncertainty

For each area we also estimated the uncertainty in the predicted soil pH (Figure 3 and Appendix S6). Standard errors of prediction errors vary from 0.16 till 0.91. The highest standard error seem to be located on the sandy soils in a zone in the middle of the Netherlands. Here are, in general, also lower pH values predicted. Lack of vegetation relevés may partly be the cause of this, since especially in the South-East of the East Netherlands detail map (Appendix S6) the standard error is in general low, combined with a high density of relevés. The detail map of the Dune area (Appendix S6) shows that especially the inland areas have a higher uncertainty, where in some areas relevés are totally absent. Most of the standard errors are in the range of 0.4–0.55. On average the standard error is just below 0.5 pH unit, which is comparable with the results of Wamelink et al. (2005) when they validated predicted soil pH with measured soil pH on a European scale.

### 3.3 | Cross-validation

The difference between predicted minus observed pH ranges from –2.82 till 2.36 pH-units, with an RMSE of 0.42 pH-units. One third of the predictions has a difference between –0.1 and 0.1 pH-units and 83% has a difference between –0.5 and 0.5 units (see also Appendix S7 for the distribution of the difference and Appendix S5). The difference between predicted and observed shows a

normal distribution with zero mean, indicating that there is no systematic bias present. Major prediction errors seem to be present all over the country and not directly related to one of the underlying maps or the pH value itself (Appendix S8). Given the fact that the error in a pH measurement is about 0.1 pH-units the predicted soil pH values do resemble the observed values rather good, though at some sites the difference between predicted and field values may be large.

### 3.4 | Validation

The correlation between field data and predicted pH are quite reasonable till good and ranges between 0.66 till 0.85 (Table 1; see also Appendix S9). The mean difference between predicted and measured soil pH is for pH H<sub>2</sub>O almost zero indicating that the predicted pH is almost the same as the measured pH. The mean error is negative for KCl and CaCl<sub>2</sub>. This means that the predicted pH based on vegetation is on average higher than pH CaCl<sub>2</sub> and KCl of the soil. That is expected as both K and Ca will remove protons (H) from the adsorption complex by cation exchange.

The period of the validation data 2000–2009 is a subset of the period of the pH data that has been used to create the map (1990 and later). This may influence the outcome of the validation. The validation are legacy data and have not been collected with validation of the pH map in mind. Therefore, the data are not evenly distributed over the Netherlands. Sandy areas are for instance over-represented. This may bias the validation results to these areas. The validity of the results are therefore limited to the validation areas.

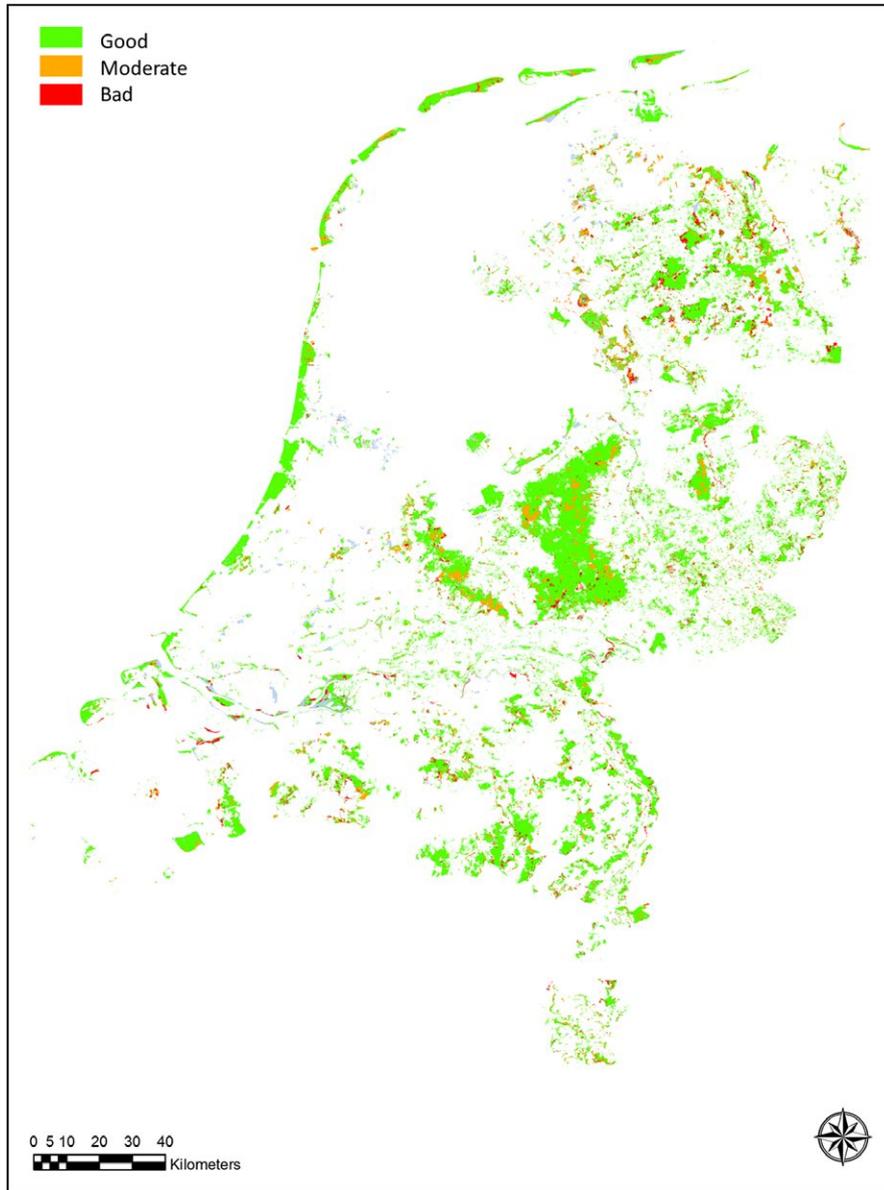
### 3.5 | National assessment of the soil pH, linked to soil quality

The quality of soil pH is in most cases considered good for the present or planned vegetation management types (Figure 4). There is no pattern visible in soil pH that is considered bad; small areas can be found scattered all over the country, both in areas with a natural high soil pH as well as areas with a low soil pH. When looking at the vegetation management types separately, especially the semi natural grasslands still seem to suffer from soil pH values that are too low as well as part of the swamps (Figure 5). For health and for most of the surface no calculations are available since some of the types do not have a threshold value for soil pH.

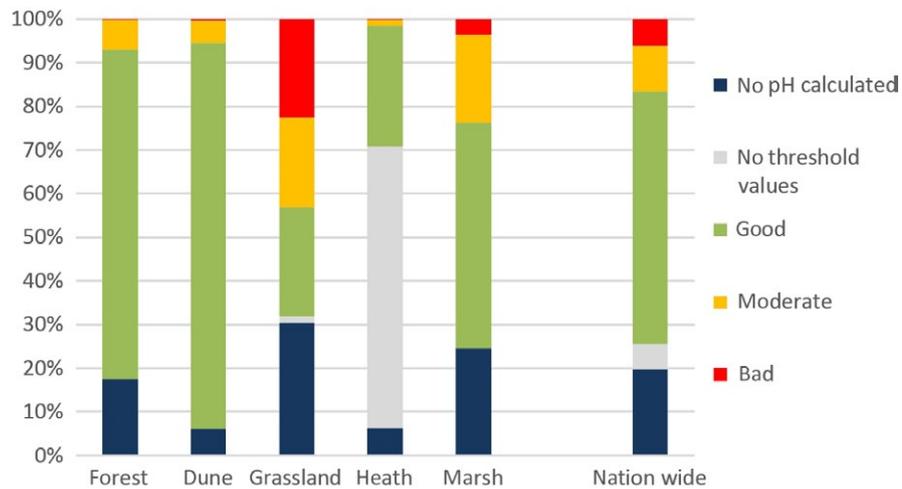
**TABLE 1** Validation statistics, with n: number of samples, r: Pearson's correlation coefficient,

Extraction	n	r	ME	RMSE
CaCl <sub>2</sub>	22	0.85	–0.51	1.10
H <sub>2</sub> O	55	0.77	–0.03	0.64
KCl	212	0.68	–0.50	1.14

ME: mean error (error = pH soil – pH from the predicted map); RMSE: root mean squared error.



**FIGURE 4** National map of the soil pH quality per vegetation type



**FIGURE 5** Relative quality (surface based) of the soil pH for (lumped) vegetation management types

## 4 | DISCUSSION

Acidification is considered one of the main pressures causing biodiversity loss. To reduce this pressure the Dutch government took measures to decrease acid emission and restore soil pH in nature areas since the 1980s of the last century. The predicted soil pH conditions, based on indicator values (Wamelink et al., 2005), are considered good in most ecosystems, and hence, we assume that acidification of the soil is not a major problem on a national scale anymore in the Netherlands. Though, in some areas the soil pH is still too low. Especially, for semi natural grasslands and swamps the pH is too low. In these case extra measures, such as liming (Kreutzer, 1995), may be necessary. Our findings that acidification is not a major problem anymore agrees with earlier research inside and outside the Netherlands (Bobbink et al., 1998; Likens, Driscoll, & Buso, 1996; Wamelink et al., 2013). However, this is not the case everywhere; especially in Asian countries acid deposition is still high and expected to raise in the future (Larssen et al., 2006). We think the predicted soil pH map is a useful tool for evaluations of both government policies as well as nature conservation management on a regional till national scale.

A major question is if soil pH is still the indicator we should be looking at regarding the effects of acidification of the past. Due to the acidification, especially on the nutrient poor sandy soils, base cations (Ca and Mg) may have been leached to the deeper groundwater, making the cations unavailable for the plants. This lack in base cations may cause growth problems for many plant species even though soil pH is still optimal. Earlier research also indicates that there may be a Ca and Mg shortage in many sandy soils in the Netherlands and abroad due to acid deposition (Graveland, van der Wal, van Balen, & van Noordwijk, 1994; Högberg, Fan, Quist, Binkley, & Tamm, 2006). Replenishment of the base cations only takes place by weathering of the soil material; a rather slow process. An alternative way to improve the availability of these cations is by liming. Doing so would also influence soil pH, making it a good indicator of the effect of calcium and or magnesium addition.

The number of relevés with an estimated soil pH is sufficient, given the validation results. On average we have 0.19 relevé per hectare available for the nature area in the Netherlands. However, the plots are not evenly distributed. They are the result of many research interests and the result of many different surveys, made by many different people. In an ideal situation the data points should have been more evenly distributed with additional points close to each other to be able to estimate short distance variation and then visited all in the same year. Such a set is not available and will not likely be available in the future either. The here applied procedure partly overcomes this 'omission' by taking data density into account. Despite an on average high data density, there are still nature areas without any available data. Here the predicted soil pH is based on the soil type, groundwater table and vegetation map and interpolation of data of nearby nature areas with relevés. As might be expected the uncertainty in the predicted soil pH is larger in these areas.

We choose 1990 as our starting year to include relevés, hence we used relevés made over a 25 year period. The period used

is considered in relation to what we thought was an optimal data density. Given the data density it also should be possible to make a pH map based on a less optimal data density for example data collected every 5 or 10 years. If there is sufficient data available in the Netherlands to do this from 1950 this may lead to a series of maps that could show a change in soil pH over time as a result of acidification. Another problem related to this is the so called time-lag of plant species as indicators of soil circumstances. Since some species are long-lived they may have germinated and grown up under circumstances that are different than the present ones. Consequently their presence may be an echo of the past giving a false indication, the less species present in a relevé the bigger this chance is. A shorter interval in years, as suggested 5 or 10 years may overcome this problem partly.

To create a soil pH map, if possible, direct measurements of soil pH should be used. Currently, there are not sufficient measurements available to map soil pH nationwide on a site level. However, it is possible to improve our method by using field data as extra information. The data can then be used with a higher weight than the estimated pH values of the relevés.

Cross-validation shows that the predicted soil pH are accurate in relation to estimated pH and the pH patterns meet our expectations. Also the validation on a limited number of independent pH measurements shows that the soil pH is predicted quite well. Thus we assume that the results are robust enough to make an accurate soil pH map on a national level. An earlier validation of predicted soil pH compared with field measurements on a European scale also showed a good relation between predicted and measure soil pH (Wamelink et al., 2005). This opens the way of the evaluations of both government policies as well as direct field management on a regional till national scale. Our results show clearly that the measures taken to mitigate acidification has had its positive effect and that in many areas acidification does not lower the soil pH anymore. Acidification may still have an impact on base cation availability and thus affecting both plant and animal species occurrence.

## ACKNOWLEDGEMENTS

We like to thank all people (way over 100) that were willing to share their valuable field measurements for this research. This paper was commissioned and funded by the Dutch Ministry of Agriculture, Nature and Food Quality through the Statutory Research Tasks Unit for Nature & the Environment. The comments of two reviewers helped to improve and focus the ms.

## AUTHOR CONTRIBUTION

GWWW and DJJW wrote the article, DJJW, GWWW and MK designed and performed the statistical method, GWWW and DJJW designed and performed the validation, MES and GWWW did the National assessment of the soil pH, linked to soil quality, HAMM, RP and GWWW created the vegetation map and RMAW made the figures.

## ORCID

G. W. Wiegner Wamelink  <https://orcid.org/0000-0002-9795-2248>

## REFERENCES

- Bobbink, R., Hornung, M., & Roelofs, J. G. M. (1998). The effects of airborne nitrogen pollutants on species diversity in natural and semi-natural European vegetation. *Journal of Ecology*, *86*, 717–738. <https://doi.org/10.1046/j.1365-2745.1998.8650717.x>
- Bootsma, M. C., Van Den Broek, T., Barendregt, A., & Beltman, B. (2002). Rehabilitation of acidified floating fens by addition of buffered surface water. *Restoration Ecology*, *10*, 112–121. <https://doi.org/10.1046/j.1526-100X.2002.10112.x>
- Braun-Blanquet, J. (1964). *Pflansensoziologie*, 3rd ed. Vienna, Austria: Springer.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buijsman, E., Aben, J. J. M., Hettelingh, J. P., van Hinsberg, A., Koelemeijer, R. B. A., & Maas, R. J. M. (2010). *Zure regen, Een analyse van 30 jaar verzuringsproblematiek in Nederland*. PBL-rapport no. 500093007, ISBN 978-90-78645-47-4.
- De Graaf, M. C. C., Verbeek, P. J. M., Bobbink, R., & Roelofs, J. G. M. (1998). Restoration of species-rich dry heaths: The importance of appropriate soil conditions. *Acta Botanica Neerlandica*, *47*, 89–111.
- De Vries, F. (1999). *Karakterisering van Nederlandse gronden naar fysisch-chemische kenmerken*. Wageningen, Staring Centrum. Rapport 654.
- De Vries, F., de Groot, W., Hoogland, T., & Denneboom, J. (2003). *De Bodemkaart van Nederland digitaal; Toelichting bij inhoud, actualiteit en methodiek en korte beschrijving van additionele informatie*. Rapport 811, Alterra, Research Instituut voor de Groene Ruimte., Wageningen.
- Delhomme, J. P. (1978). Kriging in the hydrosocieties. *Advances in Water Resources*, *1*, 251–266. [https://doi.org/10.1016/0309-1708\(78\)90039-8](https://doi.org/10.1016/0309-1708(78)90039-8)
- Deng, H., Ye, Z. H., & Wong, M. H. (2004). Accumulation of lead, zinc, copper and cadmium by 12 wetland plant species thriving in metal-contaminated sites in China. *Environmental Pollution*, *132*, 29–40. <https://doi.org/10.1016/j.envpol.2004.03.030>
- Driscoll, C. T., Lawrence, G. B., Bulger, A. J., Butler, T. J., Cronan, C. S., Eagar, C., ... Weathers, K. C. (2001). Acidic deposition in the north-eastern United States: Sources and inputs, ecosystem effects, and management strategies. *BioScience*, *51*, 180–198. [https://doi.org/10.1641/0006-3568\(2001\)051\[0180:ADITNU\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0180:ADITNU]2.0.CO;2)
- Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jack-knife, and cross-validation. *The American Statistician*, *37*, 36–48.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press, 436 pp.
- Falkengren-Grerup, U. (1986). Soil acidification and vegetation changes in deciduous forest in southern Sweden. *Oecologia*, *70*, 339–347. <https://doi.org/10.1007/BF00379494>
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*, New York, NY: Oxford University Press, 483 pp.
- Graveland, J., van der Wal, R., van Balen, J. H., & van Noordwijk, A. J. (1994). Poor reproduction in forest passerines from decline of snail abundance on acidified soils. *Nature*, *368*, 446–448. <https://doi.org/10.1038/368446a0>
- Hennekens, S. M., & Schaminée, J. H. J. (2001). Turboveg, a comprehensive database management system for vegetation data. *Journal of Vegetation Science*, *12*, 589–591. <https://doi.org/10.2307/3237010>
- Högberg, P., Fan, H., Quist, M., Binkley, D., & Tamm, C. O. (2006). Tree growth and soil acidification in response to 30 years of experimental nitrogen loading on boreal forest. *Global Change Biology*, *12*, 489–499. <https://doi.org/10.1111/j.1365-2486.2006.01102.x>
- Kreutzer, K. (1995). Effects of forest liming on soil processes. *Plant and Soil*, *168–169*, 447–470. <https://doi.org/10.1007/BF00029358>
- Kros, J., Mol-Dijkstra, J. P., Wamelink, G. W. W., Reinds, G. J., van Hinsberg, A., & de Vries, W. (2016). Modelling impacts of acid deposition and groundwater level on habitat quality and plant species diversity. *Ecological Processes*, *5*, 22. <https://doi.org/10.1186/s13717-016-0066-0>
- Lamers, L. P., Vile, M. A., Grootjans, A. P., Acreman, M. C., Van Diggelen, R., Evans, M. G., ... Smolders, A. J. (2015). Ecological restoration of rich fens in Europe and North America: From trial and error to an evidence-based approach. *Biological reviews of the Cambridge Philosophical Society*, *90*, 182–203. <https://doi.org/10.1111/brv.12102>
- Larssen, T., Lydersen, E., Tang, D., He, Y., Gao, J., Liu, H., ... Luo, J. (2006). Acid rain in China. *Environmental Science and Technology*, *40*, 418–425. <https://doi.org/10.1021/es0626133>
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*, 18–22.
- Likens, G. E., Driscoll, C. T., & Buso, D. C. (1996). Long-term effects of acid rain: Response and recovery of a forest ecosystem. *Science*, *272*, 244–246. <https://doi.org/10.1126/science.272.5259.244>
- Marschner, P., Crowley, D., & Yang, C. H. (2004). Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. *Plant and Soil*, *261*, 199–208. <https://doi.org/10.1023/B:PLSO.0000035569.80747.c5>
- Patra, M., & Sharma, A. (2000). Mercury toxicity in plants. *Botanical Review*, *66*, 379–422. <https://doi.org/10.1007/BF02868923>
- Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*, 683–691. <https://doi.org/10.1016/j.cageo.2004.03.012>
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roem, W. J., & Berendse, F. (2000). Soil acidity and nutrient supply ratio as possible factors determining changes in plant species diversity in grassland and heathland communities. *Biological Conservation*, *92*, 151–161. [https://doi.org/10.1016/S0006-3207\(99\)00049-X](https://doi.org/10.1016/S0006-3207(99)00049-X)
- Schaminée, J. H. J., Hennekens, S. M., & Ozinga, W. A. (2012). The Dutch national vegetation database. *Biodiversity & Ecology*, *4*, 201–209. <https://doi.org/10.7809/b-e>
- Schöpp, W., Posch, M., Mylona, S., & Johansson, M. (2003). Long-term development of acid deposition (1880–2030) in sensitive freshwater regions in Europe. *Hydrology and Earth System Sciences*, *7*, 436–446. <https://doi.org/10.5194/hess-7-436-2003>
- Schwertmann, U. (1991). Solubility and dissolution of iron oxides. *Plant and Soil*, *130*, 1–25. <https://doi.org/10.1007/BF00011851>
- Silvertown, J., Poulton, P., Johnston, E., Edwards, G., Heard, M., & Biss, P. M. (2006). The Park Grass Experiment 1856–2006: Its contribution to ecology. *Journal of Ecology*, *94*, 801–814. <https://doi.org/10.1111/j.1365-2745.2006.01145.x>
- Stevens, C. J., Dupre, C., Dorland, E., Gaudnik, C., Gowing, D. J. G., Bleeker, A., ... Dise, N. B. (2010). Nitrogen deposition threatens species richness of grasslands across Europe. *Environmental Pollution*, *158*, 2940–2945. <https://doi.org/10.1016/j.envpol.2010.06.006>
- Stevens, C. J., Manning, P., van den Berg, L. J. L., de Graaf, M. C. C., Wamelink, G. W. W., Boxman, A. W., ... Dorland, E. (2011). Ecosystem responses to reduced and oxidised nitrogen inputs in European terrestrial habitats. *Environmental Pollution*, *159*, 665–676. <https://doi.org/10.1016/j.envpol.2010.12.008>
- Van Breemen, N., Mulder, J., & Driscoll, C. T. (1983). Acidification and alkalization of soils. *Plant and Soil*, *75*, 283–308. <https://doi.org/10.1007/BF02369968>

- Van Den Berg, L. J. L., Dorland, E., Vergeer, P., Hart, M. A. C., Bobbink, R., & Roelofs, J. G. M. (2005). Decline of acid-sensitive plant species in heathland can be attributed to ammonium toxicity in combination with low pH. *New Phytologist*, *166*, 551–564. <https://doi.org/10.1111/j.1469-8137.2005.01338.x>
- Wamelink, G. W. W., de Knecht, B., Pouwels, R., Schuiling, C., Wegman, R. M. A., Schmidt, A. M., ... Sanders, M. E. (2013). Considerable environmental bottlenecks for the Habitats and Birds directives Species in the Netherlands. *Biological Conservation*, *165*, 43–53. <https://doi.org/10.1016/j.biocon.2013.05.012>
- Wamelink, G. W. W., Goedhart, P. W., Malinowska, A. H., Frissel, J. Y., Wegman, R. M. A., Slim, P. A., & van Dobben, H. F. (2011). Ecological ranges for the pH and NO<sub>3</sub> of syntaxa: A new basis for the estimation of critical loads for acid and nitrogen deposition. *Journal of Vegetation Science*, *22*, 741–749. <https://doi.org/10.1111/j.1654-1103.2011.01286.x>
- Wamelink, G. W. W., Goedhart, P. W., van Dobben, H. F., & Berendse, F. (2005). Plant species as predictors of soil pH: Replacing expert judgement by measurements. *Journal of Vegetation Science*, *16*, 461–470. <https://doi.org/10.1111/j.1654-1103.2005.tb02386.x>
- Wamelink, G. W. W., van Adrichem, M. H. C., van Dobben, H. F., Frissel, J. Y., den Held, M., Joosten, V., ... Wegman, R. M. A. (2012). Vegetation relevés and soil measurements in the Netherlands; a database. *Biodiversity and Ecology*, *4*, 125–132. <https://doi.org/10.7809/b-e>
- Wasof, S., Lenoir, J., Gallet-Moron, E., Jamoneau, A., Brunet, J., Cousins, S. A. O., ... Decocq, G. (2013). Ecological niche shifts of understorey plants along a latitudinal gradient of temperate forests in north-western Europe. *Global Ecology and Biogeography*, *22*, 1130–1140. <https://doi.org/10.1111/geb.12073>
- Wehrens, R. H., Putter, L., & Buydens, M. C. (2000). The bootstrap: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, *54*, 35–52. [https://doi.org/10.1016/S0169-7439\(00\)00102-7](https://doi.org/10.1016/S0169-7439(00)00102-7)
- Xu, J. M., Tang, C., & Chen, Z. L. (2006). The role of plant residues in pH change of acid soils differing in initial pH. *Soil Biology and Biochemistry*, *38*, 709–719. <https://doi.org/10.1016/j.soilbio.2005.06.022>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**APPENDIX S1.** Description of the used databases and method to estimate the soil pH per vegetation type.

**APPENDIX S2.** Description of the vegetation management types and vegetation type map.

**APPENDIX S3.** Detail of the predicted soil pH map for the Dune area in the North-West part of the Netherlands.

**APPENDIX S4.** Detail of the predicted soil pH map for the East of the Netherlands along the German border.

**APPENDIX S5.** Distribution of the soil pH of Dutch terrestrial nature.

**APPENDIX S6.** Detail of the standard error of the soil pH map for the Dune area in the North-West part of the Netherlands and the East along the German border.

**APPENDIX S7.** The number of location per difference class between predicted and observed pH.

**APPENDIX S8.** The predicted minus observed soil pH for nature areas in the Netherlands.

**APPENDIX S9.** Validation results per pH version.

**How to cite this article:** Wamelink Wiegier GW, Walvoort DJJ, Sanders ME, et al. Prediction of soil pH patterns in nature areas on a national scale. *Appl Veg Sci*. 2019;00:1–11. <https://doi.org/10.1111/avsc.12423>