CrossMark

Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of the Netherlands

Bas Kempen [a], Dick J. Brus [b],[*], Folkert de Vries [b]

[a] ISRIC — World Soil Information, P.O. Box 353, 6700AJ Wageningen, The Netherlands
[b] Alterra, Wageningen University and Research Centre, P.O. Box 47, 6700AA Wageningen, The Netherlands

ARTICLE INFO

ABSTRACT

This paper presents a pedometric approach to updating the Dutch 1:50,000 national soil map for the peatlands, and illustrates this approach for a 187,525 ha area in the northern peatlands. This is the first time that digital soil mapping replaces conventional soil mapping in a nationwide, government-funded soil survey program in the Netherlands. Soil classes were updated indirectly through mapping two quantitative diagnostic soil properties: the thickness and starting depth of the peat layer. From these, five major soil groups could be constructed. Because the point data were zero-inflated, a two-step simulation approach was implemented. First, peat presence/absence indicators were simulated from probabilities of peat occurrence that were predicted with a generalized linear model. Second, conditional peat thickness values were simulated from kriging with external drift predictions. The indicator and peat thickness simulations were combined to obtain simulations of the unconditional peat thickness. A similar approach was followed for the starting depth. From the simulated soil properties, probability distributions of soil groups were derived. These groups were refined with information on (static) soil properties derived from the 1:50,000 map to obtain soil classes according to the 1:50,000 legend. The updated raster map was then incorporated in the 1:50,000 polygon map. The prediction models were calibrated with legacy point data, that were updated for peat thickness before being used, in addition to a set of newly acquired point data. The uncertainty associated to the updated peat thickness values in the legacy dataset was quantified and accounted for by the prediction models. The peat thickness map and a map with three soil orders were validated with independent probability sample data. The overall purity of the soil order map was 66% for both subareas. For subarea 1 this was a 12% purity improvement compared to the current 1:50,000 map, for subarea 2 this was 3%. For subarea 1, the mean absolute error of the predicted peat thickness was 23.5 cm, and the R^2 is 0.50. For subarea 2 these accuracy measures were 30.9 cm and 0.65. We conclude that nationwide updating the 1:50,000 map with pedometric techniques is feasible. In order to increase the value and usability of the legacy point data as well as the large set of newly acquired field observations and the updated 1:50,000 map, we recommend installation of a soil monitoring network in the Dutch peatlands.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The national soil map at scale 1:50,000 (Steur and Heijink, 1991) is the main source of soil information in the Netherlands. This map was initially created for soil suitability analysis of various land-use systems (van Lynden et al., 1985; Sonneveld et al., 2010), but is since the 1990s increasingly used for environmental and agro-economic analyses in support of policy-making. Examples include modeling of nutrient and pollutant fluxes in the soil (van der Salm et al., 1996; Hack-ten Broeke et al., 1999; Kros et al. 2011), inventories and monitoring of carbon stocks (Schulp and Veldkamp, 2008; Reijneveld et al., 2009), modeling

soil subsidence (Hoogland et al., 2012), implementation of the EU Thematic Soil Strategy (Bouma and Droogers, 2007) and simulation studies on greenhouse gas emissions from peat soils (Nol et al., 2010; van Beek et al., 2011).

Organic soils cover 527,000 ha, or almost 16% of the land surface area of the Netherlands. The Dutch soil classification system (de Bakker and Schelling, 1966) distinguishes two main types of organic soils: peat soils (peat layer > 0.4 m thick and starting within 0.4 m from the surface) and peaty soils (peat layer 0.05–0.4 m thick and starting within 0.4 m from the surface). Intensive agricultural use and deep drainage of these soils have resulted in major changes in soil conditions since the completion of the 1:50,000 survey in 1995 (the first map sheets date from the 1960s). The peat oxidation and compaction rate is estimated between 5 and 10 mm year^{-1} (van den Akker et al., 2008; Hoogland et al., 2012). As a consequence, peat soils may have changed into peaty soils, and peaty

soils into mineral soils. A reconnaissance survey of peat soils in the eastern part of the Netherlands showed that the acreage of peat soils was reduced by about 50% (van Kekem et al., 2005; de Vries et al., 2009). This clearly illustrated the need for an updated soil map.

The Dutch national government recognizes the importance of good quality, up-to-date soil information and has acknowledged the need for a map update. The government commissioned an extensive updating program that aims to have an updated map for 365,000 ha of peatlands ready by the end of 2014. The deep peat soils (peat layer starting within 0.4 m below the surface and extending deeper than 1.2 m below surface) of the western and northern fen peat areas (162,000 ha) were not part of the updating program. The peat layer in these areas can be up to 6 m deep. Here, oxidation and compaction of peat do not directly result in a change of soil class. The national soil map was, therefore, considered to be up-to-date for these areas. Updating by means of conventional soil mapping started in 2009, but this soon turned out to be too expensive and too slow. In 2011 it was decided to continue the map update program by means of digital soil mapping (DSM) (McBratney et al., 2003). Kempen et al. (2012c) have shown that DSM can be an efficient alternative to conventional soil mapping for updating the national soil map in terms of accuracy and costs, but the use of DSM for map updating in the Netherlands has so far been experimental and applied to small case study areas only (Kempen et al., 2009, 2012b). This was the first time that DSM was going to be made operational in a government-funded, nationwide soil mapping program.

Making DSM operational for updating the national soil map for the peatlands brings two challenges. The first relates to the use of point data from different sources and different quality. The Dutch soil information system *BIS* stores spatially referenced soil profile descriptions from over 325,000 locations, which were collected during surveys and research projects since the 1950s. These data are an important resource for DSM (Carré et al., 2007; Sulaeman et al., 2013), but, given their age, may not properly represent actual field conditions for dynamic soil properties such as the thickness of the peat layer. Though this might limit their utility for updating, these data may still provide relevant information. Since we aimed to make most out of existing data, we decided to update the legacy soil point data and use these data, together with newly acquired field data, in our modeling framework. Hereby taking into account the uncertainty associated to the updated point data.

The second challenge concerns the mapping methodology. Kempen et al. (2009, 2012b) have shown that pedometric mapping of soil classes can be challenging, especially when one wants to model the spatial correlation structure. Calibrating a generalized linear geostatistical model (GLGM) is complex and computationally demanding (Diggle et al., 1998; Christensen, 2004). Alternatively, multinomial logistic regression, which is much easier to implement than the GLGM, has as disadvantage that certain structures in the data can cause numerical problems when fitting the model in the presence of categorical covariates (Hosmer and Lemeshow, 2000). Furthermore, the spatial correlation structure is not accounted for. The same holds for (boosted) classification trees and random forests that become increasingly popular for mapping categorical soil attributes (e.g. Heung et al., 2014; Odgers et al., 2014; Pahlavan Rad et al., 2014; Subburayalu et al., 2014). We, therefore, decided to take a different approach to updating soil class maps. Instead of mapping soil classes, we mapped (continuous) key diagnostic soil properties as proposed by Kempen (2011). The soil classes of the 1:50,000 legend are defined by a set of measurable soil properties. We focus on those properties required to distinguish peat soils, peaty soils and mineral soils, which are the thickness and starting depth of the peat layer. The type of peat, peaty or mineral soil class according to the 1:50,000 legend can then be reconstructed by refining the predictions with information on (static) properties obtained from the current soil map. This mapping approach also better suits the information demand by several soil data users, who have expressed interest in quantitative information about the thickness of the peat layer instead of qualitative information in the form of a peat soil class.

The aim of this paper is to describe, illustrate and validate the geostatistical mapping approach for updating the Dutch national soil map at scale 1:50,000 for the northern peatlands.

2. Materials and methods

2.1. Study area

For the map update program, the Dutch peatlands were divided into six soil-geographical subareas according to peat landscape type, each of which is modeled independently. Here the results are presented for two of these subareas: the northern till plateau (subarea 1) and the northern fen peat area (subarea 2) that jointly cover 152,925 ha. Added to the latter were the deep peat soils of the Frisian part of the northern fen peat area on request of the Province of Friesland and the Frisian water board (34,600 ha; these soils were initially not scheduled for updating). In total the study area comprises 187,525 ha and stretches between 52.35–53.49° North and 5.20–7.19° East (Fig. 1).

The northern till plateau is a Pleistocene glacial till plain dissected by an extensive brook valley system. The till deposits are covered with aeolian sands. In the early Holocene, the brook valleys were filled with sedge and reed-sedge dominated fen peat. At the same time, raised peat bogs of Sphagnum peat moss developed on the plateau positions. The raised bogs were drained, excavated and reclaimed for agriculture between the early 17th century and mid-20th century. Here dominant land use is cropland (grain, potato, sugar beet). Land use in the brook valleys is dominated by intensively managed grassland for dairy farming. Some small patches of raised bogs still remain in the study area. These are nature conservation areas. Peat soils (Folic Histosols) are found in the center of the brook valleys, while peaty soils are found along the edges (Folic Gleysols Arenic and Folic Gleyic Podzols). The cultivated peat and peaty soils of the plateau are predominantly Folic Histosols Drainic and Folic Gleyic Podzols, both soils are typically covered by a 20–30-cm thick sandy Transportic horizon that has been applied to make the reclaimed soils suitable for cropping. Fibric Histosols dominate the nature conservation areas on the plateau. Peat layers are relatively thin and do not typically exceed 1.5 m.

The northern fen peat area lies in the transition zone from the Pleistocene sand landscape of the northeastern and eastern Netherlands, to the marine clay landscape of central and northern Netherlands. Peat started to form in the mid-Holocene on top of Pleistocene cover sand deposits as a result of rising groundwater levels caused by a rising sea level. By 1000 BC, vast fen peat meadows and raised Sphagnum bogs were formed. Since that time, clay was deposited on top of the peat during a series of marine transgressions. Reclamation of the fen peat area for habitation and agriculture, including peat excavation, started in the early Middle Ages. Today, the northern fen area is dominated by intensively managed grassland for dairy farming. The area is dominated by Folic Histosols with a 20–40-cm thick clayey-peat, peaty-clay or clayey topsoil, and Folic Gleyic Podzols. The peat layer can be up to 3 m deep.

2.2. Data

2.2.1. Point data

Three sources of point data were used: i) legacy point data obtained from the Dutch soil information system *BIS*[1], ii) legacy point data from the 2002–2004 peat soil survey (van Kekem et al., 2005), and iii) newly acquired point data.

From *BIS*, 21,220 soil profile descriptions were extracted: 5758 sites in subarea 1 and 15,462 sites in subarea 2. These can be split in two sets: profiles described for small-scale (1:10,000) soil surveys (19,712 points) and profiles described for large-scale surveys and other research projects (1508 points). In addition, legacy point data from the 2004 peat
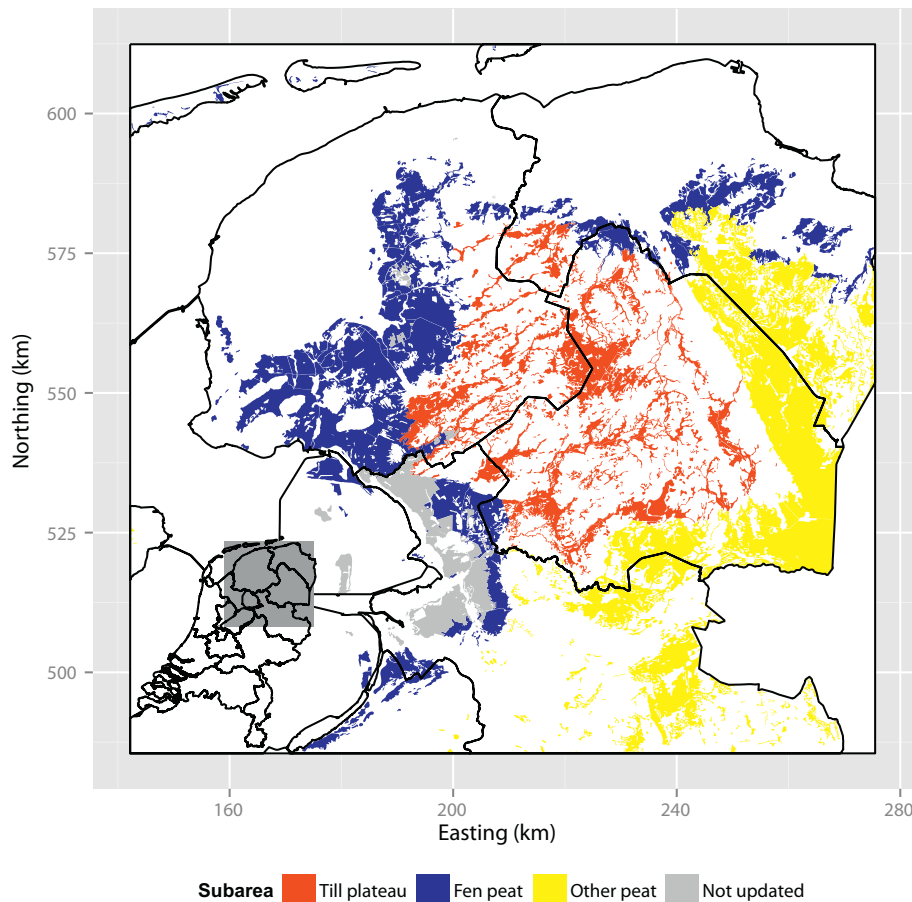
---

[1] www.bodemdata.nl.

Fig. 1. Study area of the map updating exercise. The update is carried out for the subareas 'Till plateau' and 'Fen peat' subareas. 'Other peat' denotes organic soils in other subareas and 'Not updated' denotes the deep peat soils that are not part of the update program. The inset shows the location of the study area in the Netherlands.

survey (van Kekem et al., 2005) were digitized from topographic maps. This resulted in 5296 data points: 2459 for subarea 1 and for 2837 for subarea 2. For these points only the soil classification codes could be retrieved, which were noted on the topographic maps. Soil profiles were not recorded at the time of the survey. Minimum and maximum values for peat thickness and starting depth were derived from the soil classification codes. Peat thickness as derived from these codes ranged from 5 to 215 cm, with a median of 40 cm and a mean of 50 cm. The difference between maximum and minimum peat thickness at the survey locations was on average 20 cm. In addition to legacy data, new soil profile descriptions were collected. Sampling sites were selected by spatial coverage sampling (Walvoort et al., 2010) to ensure an even geographic coverage of the study area. In subarea 1, 322 profile descriptions were collected (circa 1 per 200 ha) in 2011. In subarea 2, 2045 profile descriptions were collected (circa 1 per 60 ha) in 2012 and 2013. Fig. 2 shows the spatial distribution of the data points.

2.2.2. Environmental covariate data

A GIS database was prepared that contained 54 environmental data layers that were used as covariates in the prediction models. The covariate layers were constructed from twelve data sources:

• *The 1:50,000 national soil map* (Steur and Heijink, 1991). Eight categorical layers were derived from map unit classifications: one layer indicating peat type, two layers indicating the topsoil lithology, three layers representing peat thickness, one layer indicating the status (degraded or not degraded) of peat soils based on the results of the 2002–2004 peat survey, and one variable indicating sensitivity to oxidation according to (Finke et al., 1996).

• *Groundwater depth maps at 25-m resolution* (Berendrecht et al., 2007). The set contains two layers: one depicting the annual average maximum depth of groundwater table below surface, and one the annual average minimum depth predicted by the MIPWA groundwater model. In addition to the two continuous layers, seven categorical layers were generated. Both depth layers were reclassified into layers with five depth classes. Furthermore, the information on average maximum and minimum depth was used to construct a layer with 31 soil drainage classes according to the definitions of the national 1:50,000 groundwater table map (Steur and Heijink, 1991). This layer was further simplified to layers with six and three drainage classes, and to layers representing summer and winter drainage classes.

• *National DEM at 25-m resolution.*[2] Five relative elevation layers were derived by subtracting the average elevation at each pixel, computed for search radii of 100, 250, 500, 750, and 1000 m, from the actual elevation. These layers capture local relief at different scales. One relative elevation layer (based on the 750-m search radius) was reclassified into six layers with different combinations of two, three and four classes to account for a possible non-linear relation between soil and relative elevation.

• *Land cover maps.* This set contains five layers depicting land cover in 1900 (50-m resolution), 1940 (25 m), 1960 (25 m), 1970 (25 m), 1980 (25 m), 1990 (25 m) and 2003 (25 m) (Clement and Kooistra, 2003; Knol et al., 2003, 2004; Hazeu, 2005). The 1900 layer distinguishes ten classes and was reclassified into layers with two, three and four classes. The 2003 layer, representing current land cover, distinguishes 23 classes and was reclassified into eight layers, each with a
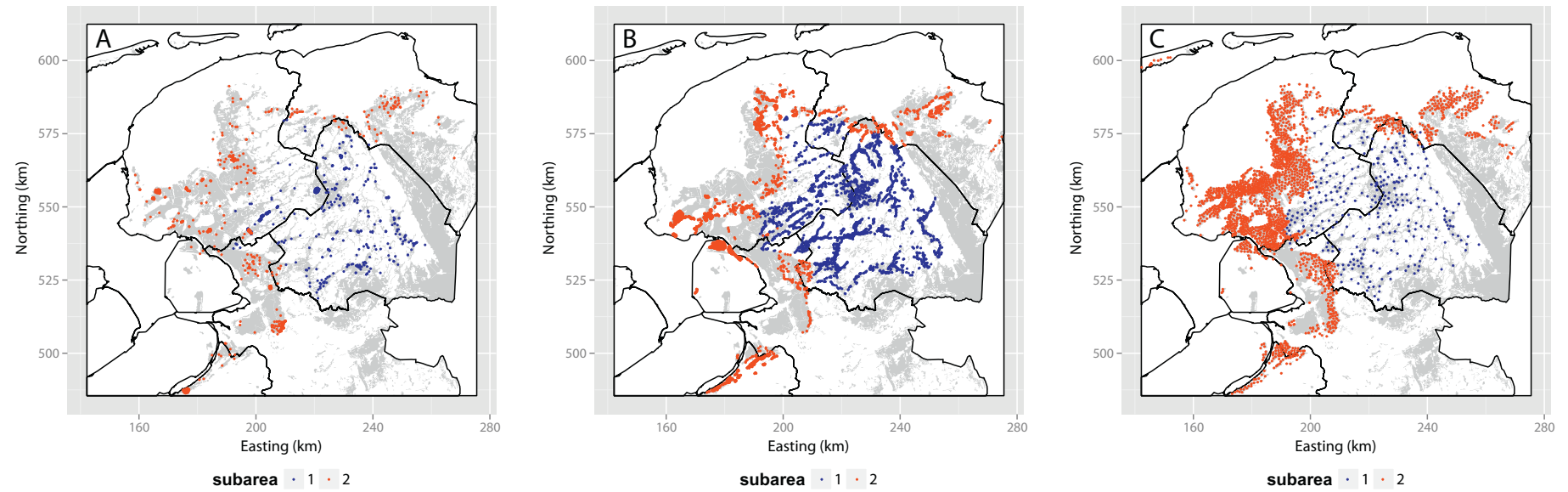
---

[2] www.ahn.nl.

Fig. 2. Distribution of the data points of the *BIS* dataset (excluding heavy-clustered points located in small-scale survey areas) (A), peat survey dataset (B), and new dataset (C).

different combination of classes. The 1940, 1960, 1970, 1980, and 1990 layers were reclassified into layers with two classes: natural land and cultivated land. Furthermore, the layers from 1900, 1940, 1960, 1970, 1980, 1990 and 2003 were combined into a map with eight reclamation period classes, i.e. the period when natural land was reclaimed for agriculture. Seven layers were derived from this map, each with a different combination of reclamation periods.

- *Geo-hydrological map at 100-m resolution* (Vernes, 2005). Map of starting depth of the Holocene deposits. Holocene deposits in the study area are marine clays and peat, and are, therefore, informative of the thickness of the peat layer.

These covariates are either drivers of processes that affect the oxidation and compaction of the peat layer, or are directly or indirectly informative of the soil class that is likely to be found currently. Mineral soils are more likely to be found in areas mapped as peaty soils on the 1:50,000 soil map than in areas in mapped as peat soils. Presence of a mineral topsoil protects the underlying peat against oxidation. Sphagnum peat is more resistant against oxidation than sedge or reed peat. Groundwater table depth (drainage) is a key driver of peat oxidation. Peat loss will be larger in areas with deeply drained soils than in areas with soils with a shallow groundwater table. Peat below the mean lowest groundwater table is protected against oxidation (though it might still be affected by compaction). Peat layers in the relatively high parts of the landscape will generally be shallower than in the lower parts of the landscape. In the former, mineral soils are more likely to be found as a result of peat oxidation than in the latter. In agricultural lands, the effect of oxidation on the thickness of the peat layer will be larger for cropland than for grassland, whereas natural lands will hardly be affected by peat oxidation.

2.3. Outline of the mapping procedure

We update the 1:50,000 soil map of the peatlands through mapping the key diagnostic soil properties *peat layer thickness* and *peat layer starting depth*. The update procedure has four stages that are outlined in brief here. The subsequent sections provide detailed descriptions of each of the stages. In this paper we only present and illustrate the mapping procedure for peat thickness. For peat starting depth, a similar procedure was followed. Fig. 4 gives a schematic overview of the mapping procedure and Table 1 gives an overview of the point data sources, and indicates their use in the various modeling stages.

In the first stage, the legacy soil point data were updated. The legacy soil profile descriptions were recorded between 1955 and 2010, with the majority between 1980 and 1990. Given their age, the recorded profiles may not represent current field conditions. Though this might limit their use for calibrating a prediction model of the current peat thickness, these data can still be informative about the current thickness. If peat was not observed in the past, then it is very likely that this is still the case today since there is no active peat growth anymore. This also means that the observed thickness in the past can be assumed the maximum thickness today. To make most out of the legacy point data, the peat thickness values were updated and the uncertainty associated to the updated thickness values was quantified. The updated point data were used as 'soft' data in our prediction models. Note that only peat thickness values were updated. The starting depth of the peat layer was assumed not to have changed since the depth was recorded, and was not updated.

In the second stage, peat thickness values were predicted at the nodes of a grid with 50-m resolution. For modeling, the (updated) legacy data were used in addition to the newly acquired data. Differences in uncertainty associated to the peat thickness values were taken into account in the model. Fig. 3 shows the frequency distribution of the observed peat thickness values for the legacy and newly acquired data. Both datasets show that a considerable proportion of the soil profiles contained no peat (peat thickness of 0 cm), i.e. the data are zero-inflated. Zero-inflated data can best be modeled by a mixture of two distributions in a two-step approach (Heilbron, 1994; Fletcher et al., 2005), which we did here. In the first step, the presence/absence of peat was modeled by a logistic regression model. In the second step, the (log-transformed) peat thickness was modeled by a spatial linear regression model, conditional on the presence of peat. To obtain a prediction of the unconditional peat thickness, the model predictions were combined. The same two-step procedure was followed for the starting depth.

In the third stage, a peat thickness class map was constructed from the simulated unconditional peat thickness values, and a map of soil groups was constructed by combining the unconditional peat thickness values with the unconditional peat starting depth values.

In the fourth stage, the updated soil group map was incorporated into the 1:50,000 polygon soil map. Updated soil classes according to the 1:50,000 map legend were obtained by refining the soil group predictions with information about (static) soil properties such as peat type and topsoil lithology that was obtained from the 1:50,000 map.
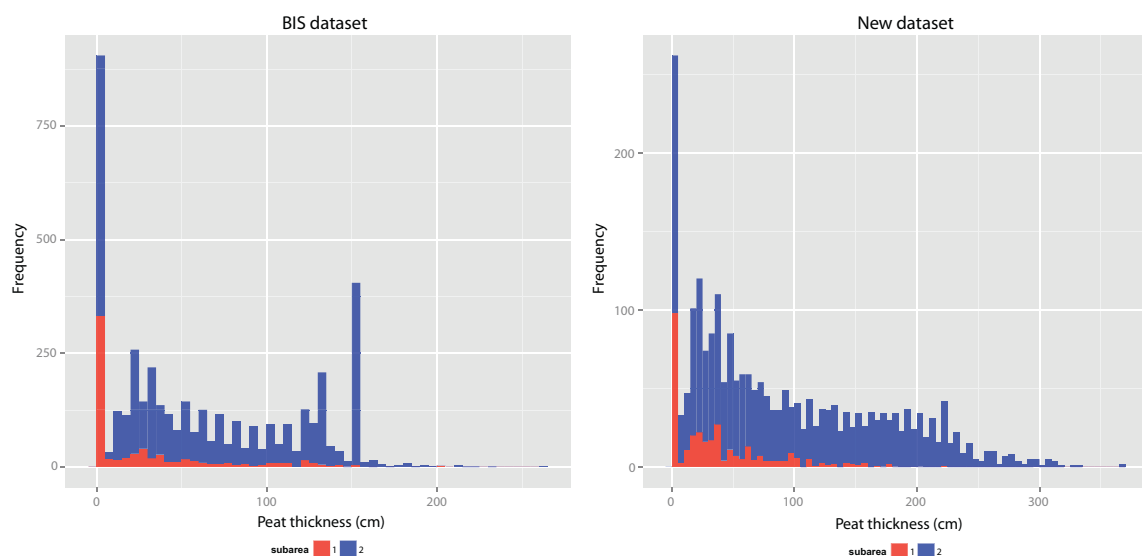

Fig. 3. Frequency distributions (for 5-cm intervals) of the observed peat thickness values for the BIS and newly acquired datasets.
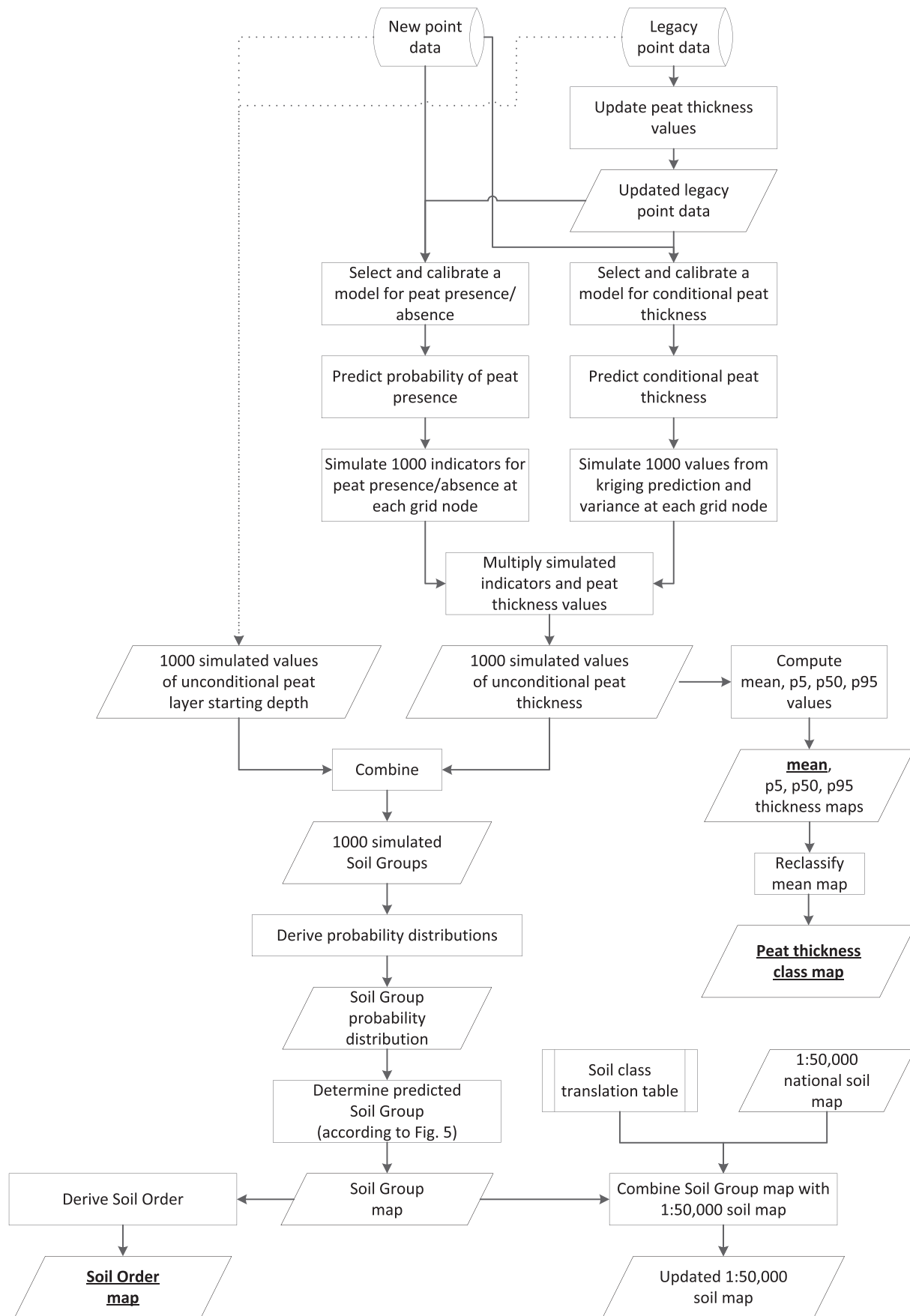
Fig. 4. Schematic overview of the mapping approach. To obtain simulated values for the peat layer starting depth, a similar approach was used as for the peat thickness values as indicated by the dashed arrow. The outputs in bold and underlined type were validated.

The outputs of the second and third stages were validated with independent probability sample data. Subareas 1 and 2 were modeled separately.

2.4. Updating legacy soil profile data

2.4.1. Legacy profile data types

We distinguish three types of legacy point data: i) hard data, ii) interval data and iii) censored data. Hard data are measurements with negligible error at the time of observation. Interval data comprise field observations for which only a thickness interval can be derived. The 2002–2004 peat survey contained many interval data points. During this survey only soil classification codes were indicated on a topographical field map. For some data points, the peat layer thickness could be deduced from these codes. For other points, only a thickness interval could be determined. For censored observations (Knotters et al., 1995) the peat layer thickness cannot be determined from the soil profile descriptions: the bottom of the peat layer exceeds the observation (augering) depth. The observed peat thickness is thus the minimum thickness at the data point.

The hard data were directly updated with the model presented below. For each interval data point, 100,000 thickness values were drawn from a uniform distribution with parameters equal to the upper and lower boundary of the interval. These simulated values are possible values of the peat thickness at the time of observation. For each censored data point, 100,000 values were drawn from a beta(a,b) distribution. Parameter a was chosen as 2, and parameter b as 5, based on expert judgment. The simulated values, that take values between 0 and 1, were multiplied by the censored thickness of the data point and then censored thickness was added. This gave us 100,000 simulated thickness values at each censored data point. The maximum (uncensored) thickness in the dataset was used as upper bound for the simulated thickness values. All simulated thickness values were updated in the same way as for hard data, which is described below.

2.4.2. Model definition

The thickness of peat layers in the soil profiles stored in *BIS* and the peat survey dataset was updated with the following model:

$$
\begin{aligned}
z_{ti} &= u_{ti} + v_{ti} \\
u_{ti} &= u_{t-1,i} \times p_i \qquad && \text{if} \quad v_{t-1,i} = 0, \quad \text{else} \quad u_{t-1,i} \\
v_{ti} &= \max\left(v_{t-1,i} - u_{t-1,i}\times(1-p_i), 0\right) && \text{if} \quad v_{t-1,i}>0, \quad \text{else} \qquad 0
\end{aligned}
\tag{1}
$$

with z_{ti} is the total peat layer thickness in year t at location i, u_{ti} is the thickness of the peat layer above the mean lowest (MLW) groundwater table in year t, v_{ti} is the thickness of the peat layer below MLW in year t, and p_i is the proportion of the thickness of the peat layer above MLW remaining after one year at location i. We assumed that i) the part of the peat layer that is below MLW is not affected by oxidation and ii) the MLW is kept at a constant level. This model is a refinement of the model described by Kempen et al. (2012a) that did not distinguish between peat above and below MLW. Application of the model of

Table 1
Point data sources and their use in the various modeling stages. The plus sign indicates used, the minus sign not used.

Modeling stage	BIS data		Peat survey data	New data
	1:10,000 survey	Other		
Updating peat thickness	+	+	+	−
Selection/calibration peat presence model	−	+	+	+
Peat presence prediction	−	+	+	+
Selection/calibration peat thickness model	−	+	−	+
Peat thickness prediction	+	+	+	+

Kempen et al. (2012a) in subarea 2 resulted in unrealistically large decreases of the peat thickness.

The model is extended by the following sub-model for p_i (Kempen et al., 2012a):

$$
\begin{aligned}
p_i &= \pi_i + \epsilon_i \\
\text{logit}(\pi_i) &= x_i^{\mathsf{T}}\beta \\
\mathrm{E}[\epsilon_i] &= 0 \\
\text{var}[\epsilon_i] &= \sigma^2\pi_i(1-\pi_i) \\
\text{cov}\left[\epsilon_i, \epsilon_j\right] &= 0 \qquad\qquad \text{for} \quad i\neq j
\end{aligned}
\tag{2}
$$

This model is a generalized linear model (GLM) (McCullagh and Nelder, 1989), fitted with a logit link function and residual variance proportional to $\pi(1 - \pi)$ (σ^2 is the dispersion parameter). The model was fitted by maximum quasi-likelihood (Wedderburn, 1974). Note that spatial independence was assumed when simulating the peat thickness at the point observation locations.

2.4.3. Model fitting and application

Kempen et al. (2012a) found no relation between $\text{logit}(\pi_i)$ and environmental covariates x, thus here π_i is constant in space. To account for uncertainty in π_i, we used the values of $\hat{\pi}$ and $\hat{\sigma}^2$ as reported by Kempen et al. (2012a) to simulate values for p_i, using a beta(a,b) distribution. The beta probability density function is only positive on [0,1], a useful property for simulating proportions. The expectation of this distribution is $a/(a + b)$, and the variance is $(ab)/[(a + b + 1)(a + b)^2]$. By choosing $\hat{\pi}_i\left(1-\hat{\sigma}^2\right)/\hat{\sigma}^2$ for a and $(1-\hat{\pi}_i)\left(1-\hat{\sigma}^2\right)/\hat{\sigma}^2$ for b, the expected value equals $\hat{p}_i$ and the variance $\hat{\sigma}^2\hat{\pi}_i(1-\hat{\pi}_i)$.

For each hard data point we simulated 100,000 values for p, whereas for each interval or censored data point one value for p was simulated for each of the 100,000 simulated original thickness values. The update model (Eq. (1)) was applied to each of the 100,000 values for p, yielding 100,000 updated peat thickness values at each legacy data point. Note that the variation in updated peat thickness for interval and censored data will be larger than for hard data, due to variation in original thickness.

The updated values were log-transformed, and the mean and variance of these log-transformed updated peat thicknesses were computed for each legacy data point. The variance of the simulated peat thicknesses reflects our uncertainty about the actual peat thickness. This uncertainty was accounted for in spatial prediction of the peat thickness (see hereafter). Note that the larger t, the larger the variance of the simulated peat thicknesses, and the smaller the weight attached to these data in spatial prediction. This means that older legacy point observations receive less weight in the prediction model than younger legacy observation points because we are more uncertain about the former than about the latter. Similarly, interval and censored data will get a smaller weight than hard data due to the larger variance of the updated peat thickness.

2.5. Modeling and mapping the thickness of the peat layer

2.5.1. Selection and calibration of a model for presence/absence of peat

The first step in the two-step approach for peat thickness mapping is to model peat presence/absence. For model selection and calibration, the legacy soil profile data with updated peat thicknesses were used in addition to the newly collected data. The strongly clustered points located in the areas where small-scale surveys have been carried out were excluded from the model selection and calibration steps to avoid bias in the estimated regression coefficients. For each data point, an indicator variable was defined, which takes the value 1 if peat is present, and 0 when peat is absent. For the legacy soil profile data, a threshold value of 1 cm for the updated peat thickness was used: if the updated peat thickness (the average of 100,000 simulated values) in a soil profile

was less than 1 cm, then the peat indicator value for this soil profile was 0.

Peat presence/absence was modeled with a GLM with a logit link function. For model selection, correlated covariates were grouped (e.g. the land cover covariates and the relative elevation covariates). From each covariate group, only one covariate was allowed in the model. Unique covariate combinations were formed with one covariate from each group. For each combination, the best model was selected using a step-wise approach on basis of Akaike's Information Criterion (AIC) (Webster and McBratney, 1989). Finally, from the set of 'best models' the overall best model was selected based on AIC. The signs of the regression coefficients were checked if these conform to our pedological knowledge. If this was not the case, then the second overall best model was selected. The deviance residuals of the model showed only very weak spatial correlation: in subarea 1 the relative nugget was very large, whereas in subarea 2 the experimental variogram fluctuated around a horizontal line. We, therefore, decided not to model the residuals.

2.5.2. Prediction of the probability of peat presence and simulation of indicators

The calibrated GLM was then used to predict the probability of peat presence at the nodes of a 50×50 m grid. Next, at each grid node 1000 peat indicator values were simulated from a Bernoulli distribution, with the predicted probability as the 'probability of observing peat'. The indicators at the grid nodes were simulated independently from each other, i.e. spatial correlation was not accounted for because for geostatistical simulation this is not needed when results are not aggregated.

2.5.3. Selection and calibration of a model for conditional peat thickness

The second step in peat thickness mapping is to model the (updated) thickness of peat conditional on the presence of peat. This means that for model selection and calibration, only those soil profiles were used where peat was observed (thickness > 0). The peat thickness data showed skew, which was removed by transformation to natural logarithms. The spatial distribution of the log-transformed peat thickness was modeled by a linear mixed model (Lark et al., 2006), i.e. the sum of a linear combination of covariates (linear trend) and a spatially correlated residual. Note that the peat survey data were not used for model selection and calibration (Table 1). The interval nature of the data, in combination with the fact that the data were updated, made the (simulated) peat thickness values highly uncertain.

Model selection was done as for the peat presence model, assuming uncorrelated residuals. To account for the uncertainty about the updated peat thickness for the legacy data points, weighted least squares (WLS) estimation was used with weights equal to the inverse of the variance of the log-transformed peat thickness simulations. The coefficients of the selected model and the variogram parameters were estimated by residual maximum likelihood (REML) (Lark et al., 2006). The uncertainty associated to the (updated) legacy data points was not accounted for in REML estimation.

2.5.4. Prediction and simulation of conditional peat thickness

The calibrated linear mixed model was used to predict the conditional peat thickness at the nodes of the prediction grid, using all data points (Table 1). The geostatistical application of the linear mixed model is referred to as kriging with an external drift. The uncertainty about the updated peat thickness values in the legacy dataset was accounted for in kriging by adding the variance of the log-transformed 100,000 simulated peat thickness values (Section 2.4) to the diagonal of the covariance matrix (Delhomme, 1978; Knotters et al., 1995). The predicted log-transformed peat thickness and the associated prediction variance were then used to simulate 1000 values per grid node, assuming a normal distribution. These simulated values were then back-transformed by exponentiation.

2.5.5. Computation of the unconditional peat thickness

For each grid node, 1000 *unconditional* peat thicknesses were obtained by element-wise multiplication of the vectors with simulated peat presence indicators and conditional peat thickness values. From these, the mean, median, 5-percentile and 95-percentile were computed.

2.6. Mapping peat thickness classes and soil groups

In addition to a continuous peat thickness map, we constructed a peat thickness class map from the simulations. Three classes were defined (<5 cm, 5–40 cm, and ≥40 cm) that correspond to peat thickness classes used in the Dutch soil classification system (de Bakker and Schelling, 1966). Each simulated value was assigned to one of the classes. Then the class frequency distribution was determined for each grid node. The thickness class with the largest frequency was used as the predicted class.

To derive the updated soil groups for a grid node, the simulated unconditional peat thickness values were combined with the simulated unconditional peat starting depth values. This resulted in 1000 simulated soil groups for each grid node. Here, the following five groups were considered: i) mineral soil, ii) mineral soil with peat layer starting deeper than 40 cm below the surface, iii) peaty soil, iv) peat soil with mineral subsoil starting within 120 cm from the surface, and v) peat soil with peat layer extending deeper than 120 cm below the surface. These groups correspond to the classification of the main soil units according to the 1:50,000 map legend. At each grid node, a soil group probability distribution was derived from the 1000 simulated soil groups. From this probability distribution, a soil group was selected and used to construct the map of updated soil groups. For this purpose we did not select the soil group with the largest probability. Instead, we used a hierarchical approach that follows the Dutch soil classification system. This approach is schematically represented in Fig. 5.

First, we determined if the soil was to be classified as a peat soil or non-peat soil (mineral or peaty) by comparing the sum of the probabilities of the two peat soil groups to a probability threshold value. If the soil was to be classified non-peat soil, then we determined if the soil was to be classified as a peaty soil or mineral soil by comparing the conditional probability of the occurrence of a mineral soil with a second threshold value (conditional on the soil being a non-peat soil). The threshold values were chosen in such a way that these minimized the sum of the absolute differences between the fractions of the three soil orders (mineral, peaty, peat) as obtained by leave-one-out cross-validation predictions at the newly acquired sampling sites and the fractions computed from the observed soil order at these sites. Using threshold values for classification allowed us to control the areal fraction of the soil orders on the updated map, and ensure that these fractions gave a realistic representation of the true areal fractions of the soil orders in the mapping area, which was important to us. For this, we assumed that the newly acquired data gave a realistic representation of the actual fractions in the mapping area, which is not unreasonable given the spatial coverage sampling design used to collect these data. Once it was determined if the mapped soil was to classified as a peat soil, peaty soil or mineral soil, the classification was continued as depicted in Fig. 5. The hierarchical classification procedure was applied to each grid node-specific soil group probability distribution, resulting in an updated map of soil groups.

For the purpose of validation, we constructed a map distinguishing three major soil orders from the map with five soil groups. The peat thickness and peat thickness class maps were validated as well.

2.7. Updating the 1:50,000 soil map

The 50-m resolution raster map of the five soil groups was converted to polygon format by constructing a smoothed contour map from the soil group boundaries. Polygons smaller than 2 ha were eliminated by
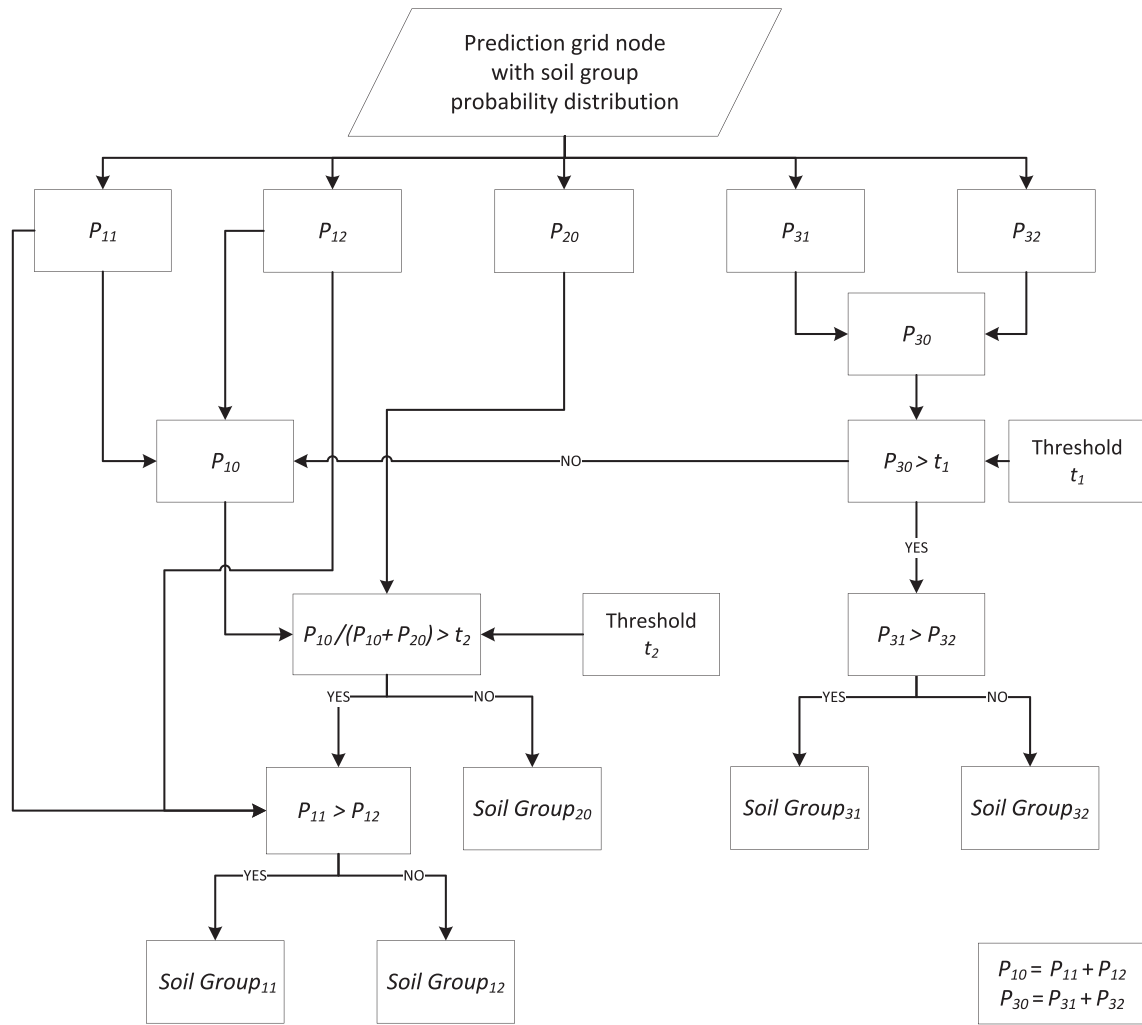
Fig. 5. Procedure for selecting the mapped soil group at a prediction node from the predicted soil group probability distribution. The 'P' indicates the probability. The numeric subscripts indicated the soil group [11) mineral soil without peat layer, 12) mineral soil with peat layer starting deeper than 40 cm below the surface, 20) peaty soil, 31) peat soil with mineral subsoil starting within 120 cm from the surface, 32) peat soil with peat layer extending deeper than 120 cm below the surface] or soil order [10) mineral soil, 20) peaty soil, 30) peat soil].

merging these with neighboring polygons with the longest shared boundary. Next, the soil group polygon map was combined with the 1:50,000 soil map to identify which of the current polygon classification codes need to be updated.

Updating of the soil codes of the 1:50,000 soil map on basis of the updated soil group map was complicated by the fact that there was no one-to-one relationship between these. For example, a former peaty soil with a mineral, podzolic subsoil that degraded to a mineral soil according to the predicted soil group, could have become a sandy podzol or a loamy podzol (whether it is sandy or loamy cannot be inferred for peaty soils from the 1:50,000 soil code). Therefore, a translation table was constructed that specifies for each combination of original soil code ($n = 51$) and updated major soil group ($n = 5$) up to eight possible updated soil codes in order of likeliness as determined by a soil survey-or. Which of the candidate soil codes is finally chosen to update the polygon, depends on the current (not-updated) polygon soil codes that are found within a 50-m radius. If this search results in a selection of two or more soil codes, then the code with the largest likelihood was selected from the translation table. If the search does not result in a match with one of the candidate codes, then the code with the largest likelihood was selected. This selection was checked by the soil surveyor. In a final generalization step polygons smaller than 3.2 ha were eliminated. These polygons were merged with neighboring polygons based on matching soil group.

2.8. Validation

The peat thickness and peat thickness class maps were validated with independent probability sample data. Validation data were collected at 89 locations in subarea 1 and at 150 locations in subarea 2. Sampling locations were selected with a stratified two-stage random sampling design. The stratification variable is the predicted peat thickness. For subarea 1 four thickness class strata were defined, in subarea 2 five. Sampling locations were allocated to the strata proportional to their areas. In the first stage, a number of 50-m grid cells were randomly selected within each stratum. In the second stage, five sampling locations are randomly selected in each selected grid cell. Sampling locations where permission was denied or proved otherwise impossible to sample were replaced with randomly selected locations from a reserve list. Fieldwork took place in 2012 and 2013. At each sampling location, the soil profile was described and classified from an auger bore observation. From these, the average peat thickness and dominant soil order (peat, peaty, mineral) were determined from the five individual measurements. If there was no dominant soil order (e.g. two observations of peat soil, two of peaty soil and one of mineral soil), then one was chosen on basis of expert judgment.

We consider three quality measures for the peat thickness class and peat soil class maps: overall purity, map unit purity (user's accuracy) and class representation (producer's accuracy) (Brus et al., 2011).

Overall purity is defined as the proportion of the mapped area in which the predicted soil type, which is the soil type as depicted on the map, equals the true soil type. In other words, it is the areal proportion correctly classified. The map unit purity is the proportion of the map unit correctly classified. The class representation for class k is the proportion of the area where in reality class k occurs that is also mapped as class k. For the peat thickness map, we computed the mean error (bias), and the mean absolute and root mean squared error (Brus et al., 2011).

3. Results

3.1. Updating legacy soil profile data

Fig. 6 shows the frequency distribution of 100,000 simulated thickness values for two locations; one sampled in 1982 (left) and one in 2007 (right). At both sites the peat thickness was 100 cm at the time of observation. The simulated values represent the peat thickness in 2012. The frequency distributions reflect the uncertainty about the yearly decrease coefficient p_i (Eq. (2)), and thus about the actual peat thickness in 2012. The shape of the frequency distributions shows that the uncertainty about the actual peat thickness at a sampling site increases with age of the soil profile description. The average of the simulations for the 1982 location is 65 cm and for the 2007 location 92 cm. Fig. 7. shows a scatter-plot of the updated versus the original (observed) peat thickness for the interval data points (left) and the hard data points (right) of subarea 1. The figure shows that the absolute decrease of the peat thickness becomes larger when the original thickness increases, and that the effect of age on the predicted actual thickness becomes smaller when the original thickness becomes smaller. This implies that the absolute annual decrease becomes smaller in time and, because a proportional model was used, the updated thickness approaches zero asymptotically. The former might be plausible since the most resistant parts of the peat will tend to accumulate. The latter is less realistic since the peat layer will eventually completely disappear, as evidenced by our field observations. The figure also shows that there can be considerable variation in updated thickness for a given original thickness, depending on the part of the peat layer that is initially below groundwater level.

3.2. Prediction models

Table 2 shows the selected regression models for the presence/absence of peat and the peat thickness for both subareas. The most important covariate was the peat thickness in three classes as derived from the existing (not updated) soil map. All regression models include a predictor associated to the groundwater level, historic land use, reclamation age, and relative elevation. In addition, topsoil lithology, current land use, oxidation risk, peat type and the thickness of the Holocene deposits were selected in some of the models. The linear regression models explained 34% of the variance of the conditional peat thickness in subarea 1, and 59% in subarea 2. The McFadden-R^2 (Menard, 2000) of the GLM for peat presence/absence was 31% for subarea 1 and 19% for subarea 2. The spatial dependence coefficient [partial sill/sill-ratio] (Lark and Cullis, 2004) for subarea 1 was 0.50, and for subarea 2 0.54, both indicating a moderate degree of spatial correlation in peat thickness. The estimated range parameter was 445 m for subarea 1 and 4767 m for subarea 2. The large difference in the range parameter might be explained by the fragmented occurrence of peat in subarea 1, whereas in subarea 2 peat occurs in larger continuous areas.

3.3. Updated soil maps

Fig. 8 shows maps constructed from the P5, P50 (median) and P95 quantiles of the 1000 simulated peat thickness values, illustrating the uncertainty associated to the predicted peat layer thickness. The average predicted peat thickness for the study area is 65 cm, as computed from the means of the 1000 simulated values. The P5 and P95 maps show that the uncertainty is relatively large. The average of the P5 map is 16.4 cm, that of the P95 map 146.3 cm.

A map depicting the three soil orders is shown in Fig. 9. According to this map, mineral soils now cover 26.9% of the area, peaty soils 27.7%, and peat soils 45.3%. Compared to the 1:50,000 national soil map, there is a considerable shift from peat soils to peaty soils and peaty soils to mineral soils. According to this map, peaty soils covered 37.4% of the study area and peat soils 62.6%. For 41% of the mapping area the updated and originally mapped soil orders differ.

In subarea 1, the thickest peat soils are found in the centers of the brook valleys. The patches of peat on the till plateau have now largely disappeared. For 45% of the area the updated soil order differs from the soil order depicted on the 1:50,000 map. Approximately 55% of the area originally mapped as peaty soils is now mapped as mineral soil, and 28% of the area originally mapped as peat soils is now mapped as peaty soil. Within the peat soil order there is a dramatic shift from the deep peat soil group (peat layer extends below 120 cm) to the shallow peat soil group (peat layer ends within 120 cm below the soil surface). Roughly 95% of the original deep peat soils are now mapped as shallow
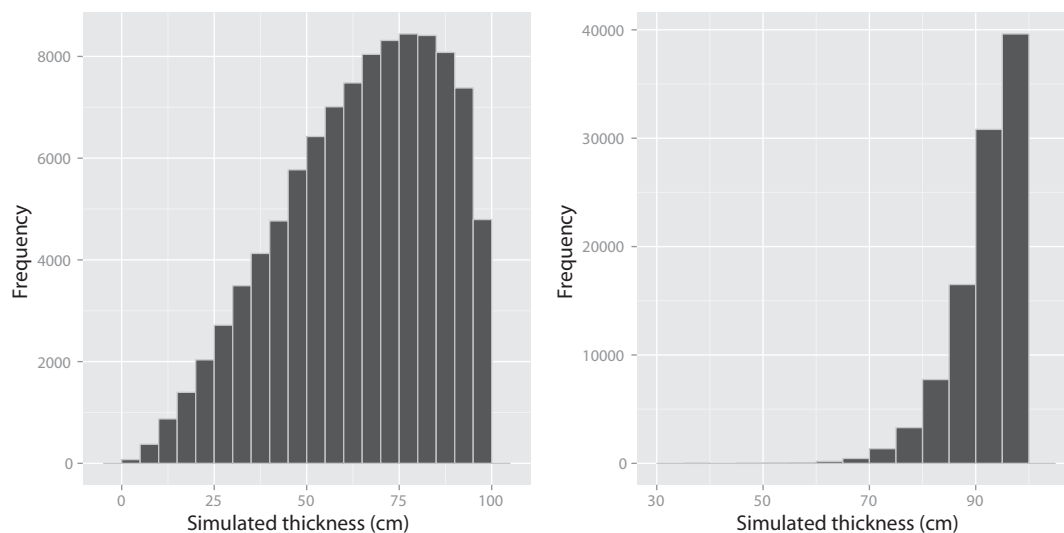

Fig. 6. Two examples of the frequency distribution of 100,000 simulated peat thickness at two sites in 2012. The sites were sampled in 1982 (left) and 2007 (right). The initial peat thickness was 100 cm at both sites. The mean lowest groundwater table was deeper than 100 cm at both sites.
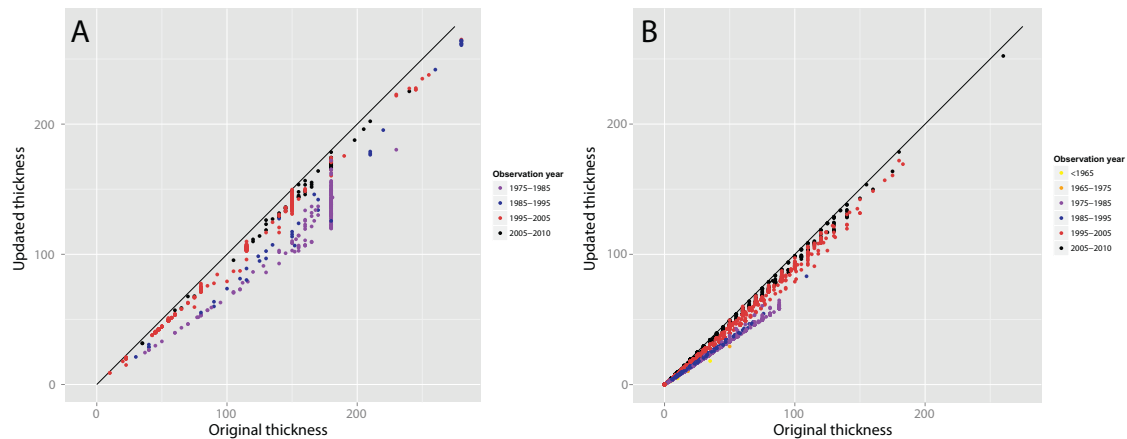
Fig. 7. Original versus updated peat thickness for uncertain (interval) (A) and hard (B) data on the original peat thickness, grouped by year of observation.

peat soils. Peat soils are predicted to cover 30% of the mapping area, peaty soils 37% and mineral soils 33%.

In subarea 2, predicted changes are somewhat less extensive. For 38% of the mapping area the updated and originally mapped soil order differ. About 63% of the area originally mapped as peaty soils is now mapped as mineral soil, and 19% of the area originally mapped as peat soils is now mapped as peaty soil. Because of these shifts, the area with mapped peaty soils is almost equal for the original and updated map. About 10% of the area originally mapped as peat soil is now mapped as mineral soil. A peat layer is predicted to be present for the majority of these mineral soils (68%). However, because this layer begins at a depth greater than 40 cm below the soil surface, the soil classifies as a mineral soil. Peat soils are predicted to cover 54% of the mapping area, peaty soils 23%, and mineral soils 23%. Peat soils originally covered 75%. Subarea 2 showed a less dramatic shift from deep peat soils to shallow peat soils than subarea 1. For 78% of the area originally mapped as deep peat soils, deep peat soils are predicted. The difference with subarea 1 can be explained by the fact that the peat layers in subarea 2 are in general thicker than in subarea 1. In subarea 2, with its fen peat plains, the peat layer can easily reach a thickness of more than 3 m, while in subarea 1, where the deep peat soils are mainly restricted to the brook valley centers, the peat layer is often less than 2 m thick. Here, degradation of the peat layer will have a stronger effect on soil classification than in subarea 2.

Fig. 10 shows two details of the updated 1:50,000 soil map and the corresponding details of the original 1:50,000 map. Soil codes containing a 'V' are peat soils, 'W' peaty soils, and 'Z' or 'H' mineral soils. The top two maps show not only shifts from peat soils to peaty soils (e.g. aVz to vWz) and from peat soils to mineral soils (e.g. zWp to cHn23), but also within the peat soil class. The 'aVc' class (peat layer extending deeper than 120 cm below surface) is remapped as 'aVz' (mineral subsoil starts within 120 cm below surface). The bottom two maps show more extensive changes within the peat soil class (xXc to xXz) and larger shifts from peaty soils to mineral soils. These maps also show the soil

surveyor expert knowledge used to determine the type of mineral soil: a peaty soil with a sandy subsoil without podzol profile (vWz) becomes a Mollic Gleysol (pZg23), whereas a peaty soil with a sandy subsoil with podzol profile becomes a Haplic Podzol (Hn21) or a Haplic Podzol Plaggic (cHn23), depending on the type of podzol found in the surrounding polygons. The figure also shows that the map unit boundaries between mineral and organic soils were kept intact, i.e. the map update took place within the original organic soil map units. We did not remap the extent of organic soils.

3.4. Validation

Table 3 presents the results of the validation with independent probability sample data. For both subareas, the models gave unbiased predictions of the peat thickness. The ME values did not differ significantly from zero at the 95% level ($p = 0.350$ for subarea 1 and $p = 0.150$ for subarea 2). For subarea 1, the MAE was 23 cm, and for subarea 2, 31 cm. The RMSE was 29 cm for subarea 1 and 48 cm for subarea 2. For both subareas, the error distributions were highly skewed, which inflates the means. We, therefore, also report the median of the squared errors, which is a more robust statistic of the 'average' error in case of highly skewed error distributions. The medians were remarkably smaller than the means. The RMedianSE was 12.6 cm for subarea 1, and 18.6 cm for subarea 2. To put the mean and median errors in perspective, the mean of the predicted peat thickness values for subarea 1 is 34.3 cm, and for subarea 2, 82.4 cm. For subarea 1, the correlation coefficient was 0.71, and for subarea 2, 0.80. This corresponds to an R^2-value of 0.50 for subarea 1, and 0.65 for subarea 2.

For subarea 1, the overall purity of the peat thickness class map was 64.2%, and that of the soil order map 66.3%. For subarea 2, the overall purity of the peat thickness map was 72.2%. When incorporating information about the starting depth of the peat layer to derive the soil orders, the purity dropped to 66.3%. Table 4 shows the error matrices with the map unit purities (MUPs) and class representations (CRs)

Table 2

Overview of the selected predictors for each model in subarea 1 and 2.

Model	Covariates[a]
Subarea 1	
Peat presence/absence	PEATTHK3 + GTS3 + LC1990 + RECLAM3 + REL1000 + LITH2 + COV2 + OXRISK
Conditional peat thickness	PEATTHK3 + GT3 + LC1970 + RECLAM3 + REL1000
Subarea 2	
Peat presence/absence	PEATTHK3 + MHW5 + LC1970 + RECLAM3 + REL100 + PEATTYP
Conditional peat thickness	PEATTHK3 + MHW5 + LC1990 + RECLAM3 + REL750 + LITH4 + PEATTYPE + OXRISK + THICKHOL

[a] PEATTHK3: peat thickness (3 classes); GTS3: summer groundwater table level (3 classes); GT3: groundwater table level (3 classes); MHW5: mean highest groundwater table (5 classes); RECLAM3: land reclamation age (3 classes); LC1970/LC1990: historic land cover in 1970/1990 (2 classes); COV2: current land cover (2 classes); LITH2/4: topsoil lithology (2/4 classes), REL100/750/1000: relative elevation within search radii of 100/750/1000 m; PEATTYP: peat type (2 classes); OXRISK: peat oxidation risk (2 classes); THICKHOL: thickness of the Holocene deposits.
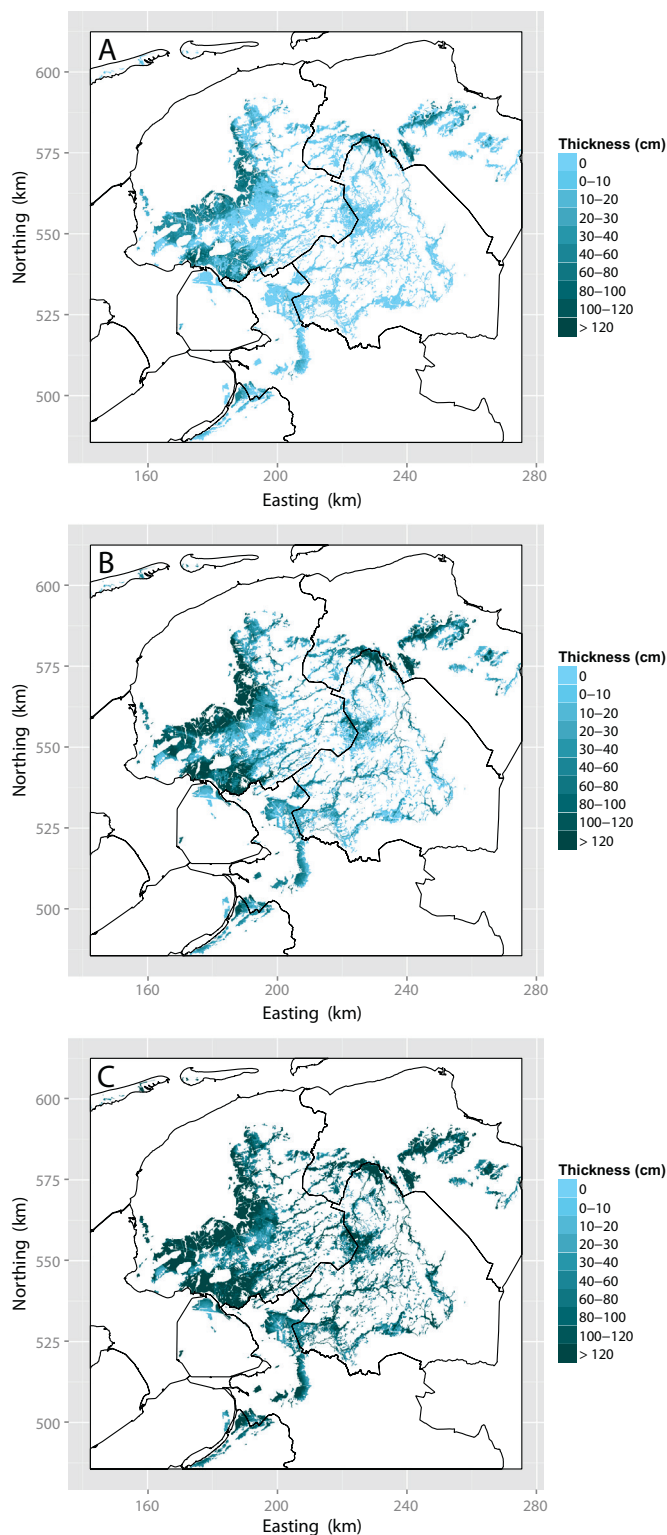
Fig. 8. Maps depicting the 5% (A), 50% (B) and 95% (C) quantiles of the 1000 simulated peat thickness values.

that are computed from the marginals. (Note that we can derive the MUPs and CRs directly from the matrices because sampling locations were allocated to the sampling strata proportional the strata areas. If this would not have been the case, then more sophisticated estimators should be used, see for example Kempen et al. (2009).) The statistics in this table shows that for subarea 1 the MUPs and CRs were mostly between 60 and 70%. There are no peat thickness classes and soil orders

that are either very well or very poorly predicted or represented. Subarea 2 showed larger differences between the peat thickness classes and between the soil orders. The peat soil order has the largest MUP and CR. The mineral soils are mainly confused with peaty soils and peaty soils mainly with peat soils. The 0–5-cm peat thickness class is poorly mapped with a MUP of 33% (though it should be noted that this statistic is based on only 6 validation points). Furthermore, the 0–5 cm class is not well represented by the updated map. For only 11% of the area where in reality peat layers less than 5 cm thick are found, these were mapped.

The validation sample also provided an estimate of the true areal fractions of the peat thickness and soil orders. In subarea 1 there now is an almost equal distribution of the three soil orders, which is well represented by the updated soil order map (Table 4). Peaty and peat soils cover an almost similar extent. For subarea 2, the validation sample showed that organic soils still cover 81% of the area, with peat soils covering 54% of the total area. Also for this area there is a fair agreement between the estimated true areal fractions of the three soil orders and the mapped areal fractions. The extent of mineral soils is slightly too large and the extent of peat soils slightly too small, although we do not think that these differences are significant at the 0.05-level. The observed peat thickness class data shows that for 87% of the area a peat layer of at least 5 cm is present, indicating the presence of peat in some mineral soil profiles.

4. Discussion

4.1. Methodology

In this paper we presented a pedometric mapping approach that was developed to update the national soil map of the Netherlands at scale 1:50,000 for the peatlands. Though quantitative methods for soil inventories were introduced in the Netherlands in the mid-1980s and 1990s, operational use of DSM has so far been limited to regional studies and for mapping quantitative properties (e.g. Brus et al., 2002; Knotters et al., 2007). This was the first time in the Netherlands that digital soil mapping was made operational in a government-funded (nationwide) soil survey program to replace conventional survey. (Although it should be mentioned that similar techniques were used to map groundwater dynamics for 1.8 million ha of sandy soils in the eastern and southern part of the Netherlands (Finke et al., 2004).) The mapping methodology was developed and applied for 187,000 ha of peatlands in the north of the Netherlands, and is currently being applied to update an additional 213,000 ha. The map update is expected to be completed early 2015.

There is an increasing interest recently in the use of pedometric methods for generating, disaggregating or updating soil class maps (e.g. Yang et al., 2011; Häring et al., 2012; Subburayalu and Slater, 2013; Subburayalu et al., 2014; Odgers et al., 2014; Nauman and Thompson, 2014; Adhikari et al., 2014; van Zijl et al., 2014; Pahlavan Rad et al., 2014). Random forests and (boosted) classification trees are currently the most popular methods to predict the spatial distribution of soil classes from environmental covariates and soil profile data. There are three methodological characteristics that set our study apart from what has recently been done. First, instead of direct prediction of soil classes, we predicted soil class indirectly through mapping of quantitative, diagnostic soil properties. Mapping quantitative properties is more straightforward than categorical properties, especially when one wants to take the spatial correlation structure into account. Such approach is possible here since i) only two properties need to be considered, and ii) we do not aim to increase the detail of the original map, as many of the recent studies aim for, so that this map can be used to derive the detailed soil class from the soil groups predicted by the model by using information on (static) properties obtained from the map. Second, contrary to most studies, an updated polygon map was one of the outputs, which was requested by the client. This implied that the raster maps had to be converted to polygon format and integrated in the
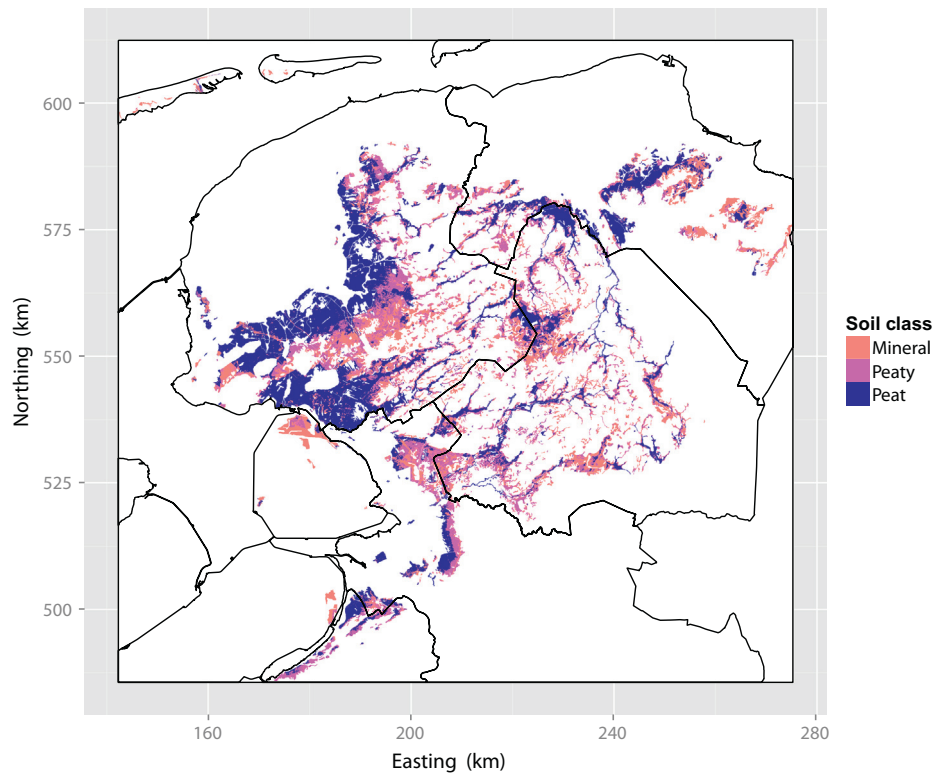
Fig. 9. Updated soil order map derived from peat thickness and starting depth simulations.

1:50,000 map, which appeared not to be a trivial task. An automated GIS workflow was designed for this purpose. The workflow assigns 1:50,000 legend codes to the updated polygons based on soil surveyor expert knowledge and spatial context information. Third, legacy point data from different sources and quality are combined with newly acquired data, and that the differences in uncertainty associated to these data sources are accounted for by the prediction model. The latter is more easily accomplished when modeling and mapping quantitative properties.

The two-step simulation approach presented here is a straightforward way to deal with zero-inflated data and allows for a full quantification of uncertainty. Gastaldi et al. (2012) used a similar two-step approach to map the occurrence and thickness of soil horizons within profiles. However, unlike these authors we took a simulation approach in order to fully quantify the uncertainty associated to the predicted peat thickness values and the updated soil groups derived from these. A two step approach has been used in the past to model occurrence and intensity of rainfall in a GLM framework. Coe and Stern (1982) model rainfall occurrence with a GLM with a logit link-function and the conditional intensity as a Gamma-distributed variable with log-link as an alternative to the Gaussian linear model we used here. Though Coe and Stern (1982) do not model rainfall occurrence and intensity spatially, their approach is worthy of further research. If a Gamma distribution is used, then attention should be paid to the distribution and variance of the model residuals if kriging interpolation of these is considered.

4.2. Legacy soil point data

Integration of legacy soil point data in the prediction framework was one of the main challenges that we faced when developing the methodology. First, there were different sorts of data: hard, interval and censored data. For the latter two, the exact peat thickness values at the time of observation are unknown. For both data types, we choose a simulation approach to generate possible thickness values. Second, the

legacy point data are subject to change and do not represent current field conditions. Hence we chose to update the legacy point data before using these in our prediction model. Calibration of the update model (Eq. (1)) proved to be difficult. Data from 95 sampling locations, that were revisited in 2007, were available for this purpose. After screening, only 44 data points remained (Kempen et al., 2012a). With so few points we were unable to establish relationships with factors that we would expect to affect the oxidation rate such as land cover and land use history. Furthermore, the fact that many sampling locations were strongly disturbed by human activities during the period between the original and new observation, and the inaccurate georeferencing of the original sampling location added to the difficulties we had with calibrating the updating model.

We went to considerable length in making the legacy soil point data useful for this mapping effort. A relevant issue that we did not address but is worth further exploring is the added value of using legacy soil point data for mapping the peat thickness: does use of these data in addition to new data improve mapping accuracy? This will likely depend on the number of newly acquired observations. In subarea 1 only 300 new observations were collected. Here we would expect a positive effect of using legacy point data on map accuracy. In subarea 2 over 2000 new observations were collected with an even geographic coverage. In this case the effect of adding extra points (with a relative large uncertainty about peat layer thickness) on map accuracy is likely to be very limited.

4.3. Updated soil maps

The overall purity of the soil order map was about 66% for both areas (Table 3). For subarea 1, this is a considerable improvement compared to the overall of the original 1:50,000 map, which was 54.0% on the soil order level. For subarea 2, however, the improvement is marginal (3.3%) and not statistically significant. The overall purity of the original 1:50,000 map for this area was 62.9% on the soil order level. There are several reasons we can think of to explain this.
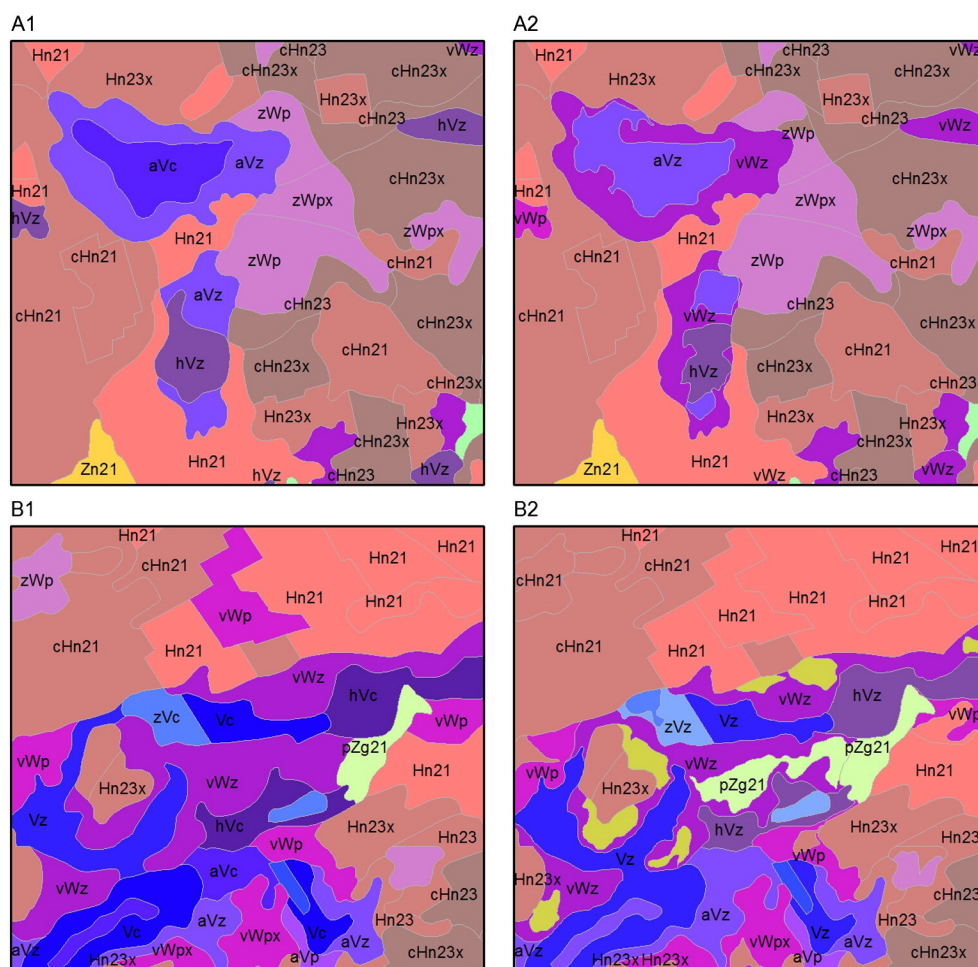
Fig. 10. Fragments of the original 1:50,000 soil map (A1, B1) and corresponding fragments of the updated 1:50,000 soil map (A2, B2).

First, the peat layers in subarea 2 are relatively thick, especially when compared to subarea 1. Here, even a substantial decrease in peat layer thickness does not always result in a change in soil class. This might explain the relatively large purity of the 1:50,000 map. In subarea 1, peat layers rarely exceed a thickness of 120 cm so here changes in thickness will have a more pronounced effect on soil classification. In other areas that still require updating, such as the cultivated peatlands of the Northeast, changes in soil class are known to be more extensive (van Kekem et al., 2005; Kempen et al., 2012b). So in these areas we expect that a map update results in a larger gain in accuracy then for subareas 1 and 2.

Second, the prediction model had difficulties in predicting the occurrence of mineral and peaty soils as evidenced by their class representations and map unit purities, which were about 43%, though the predicted areal fractions are in the correct order of magnitude when compared to the validation sample data (Table 4). A peaty soil is predicted at 43% of the locations where a mineral soil was observed. The occurrence of mineral soils did not show strong correlation with the covariates, as evidenced by the McFadden-R^2 of the GLM, despite using a large set of good-quality point data and covariates. This can be partly explained by the strong (and sometimes erratic) anthropogenic effects on the peat thickness that we are unable to capture with our covariates. The Dutch peatlands are intensively used and managed that causes a strong short-distance variation in the thickness of the peat layer. This was evidenced by the validation sample data, which showed large variations in peat thickness within the 50×50 m validation blocks. At 31% of the locations where a peaty soil was found a peat soil is predicted. These results point to an over-prediction of peat thickness at locations where mineral and peaty soils are found, despite that on average the predicted peat thickness is unbiased. At locations were a mineral soil was found the predicted peat thickness is on average 20 cm thicker than the observed thickness. For peaty soils this is 18 cm. Despite soil order-specific bias in peat thickness, the validation results of the peat thickness maps are fairly good, and are similar to the results found by Kempen et al. (2012c). There is a good correspondence between observed and predicted thicknesses as shown by the correlation coefficients.

Table 3
Validation results for subareas 1 and 2. ME is the mean error, MAE the mean absolute error, MedAE the median absolute error, RMSE the root mean square error, RMedSE the root median square error. Values between brackets are the estimated standard errors.

Error measure	Subarea 1	Subarea 2
Peat thickness (cm)		
ME	− 1.09 (2.82)	3.96 (3.96)
MAE	19.6 (1.95)	30.9 (2.95)
RMSE	28.8	47.8
RMedSE	12.6	18.6
Correlation coefficient/R^2	0.71/0.50	0.81/0.65
Peat thickness class		
Overall purity (%)	64.2 (5.2)	72.2 (3.3)
Soil order		
Overall purity (%)	66.3 (5.1)	66.3 (3.6)

Table 4

Sample error matrices showing the counts of predicted versus observed peat thickness classes and soil orders in the validation sample. The rows represent the mapped classes and orders, the columns the observed. 'P' indicates peat soils, 'Py' peaty soils, 'M' mineral soils, 'MUP' map unit purity, 'CR' class representation, $\hat{f}$ the estimated true areal fraction, and f_{map} is the predicted area fraction.

Peat thickness class

Subarea 1

	0–5 cm	5–40 cm	>40 cm	Total	MUP
0–5 cm	14	8	3	25	0.56
5–40 cm	6	20	9	35	0.57
>40 cm	2	4	23	28	0.79
Total	22	32	35	89	
CR	0.64	0.63	0.66		
$\hat{f}$	0.31	0.34	0.35		
f_{map}	0.39	0.35	0.27		

Subarea 2

	0–5 cm	5–40 cm	>40 cm	Total	MUP
0–5 cm	2	3	1	6	0.33
5–40 cm	16	21	8	45	0.47
>40 cm	1	13	85	99	0.86
Total	19	37	94	150	
CR	0.11	0.57	0.90		
$\hat{f}$	0.13	0.24	0.63		
f_{map}	0.06	0.33	0.62		

Soil order

Subarea 1

	M	Py	P	Total	MUP
M	13	6	2	27	0.62
Py	8	21	8	34	0.57
P	2	4	25	28	0.81
Total	23	31	35	89	
CR	0.57	0.68	0.71		
$\hat{f}$	0.33	0.32	0.35		
f_{map}	0.33	0.37	0.30		

Subarea 2

	M	Py	P	Total	MUP
M	12	9	6	27	0.44
Py	12	15	9	36	0.42
P	4	11	72	87	0.83
Total	28	35	87	150	
CR	0.43	0.43	0.83		
$\hat{f}$	0.19	0.23	0.59		
f_{map}	0.23	0.23	0.54		

4.4. Implications

In contrary to conventional mapping, multiple outputs are generated here (Fig. 4), of which the updated 1:50,000 soil class map is only one. With this collection of maps, we hope that the variety of wishes of the soil data user community is better served than with a single soil map. In addition to the updated 1:50,000 map and the map with the soil orders (that was validated), there are maps with the soil order and soil group probabilities that give an indication of the uncertainty associated predictions. Perhaps more valuable for soil data users than the soil class map, is the map of predicted peat thickness values and their associated uncertainties (Fig. 8). This map might help, for example, to improve estimates of soil carbon stocks that have to be reported on yearly basis to the United Nations Framework Convention on Climate Change (de Groot et al., 2005), to better identify areas where protective measures must be taken to decrease the rate of soil subsidence as a result of peat oxidation (Jansen et al., 2010), or to better quantify soil-related greenhouse gas emissions (de Vries et al., 2009). Until now, the 1:50,000 soil map is used for these purposes.

In the course of the map update program, the need for a national network for monitoring the peat thickness became more and more apparent. At this moment there is very little (spatially explicit) data available on temporal change of the thickness of the peat layers in the Netherlands. This makes it very challenging to estimate peat oxidation rates and model the annual decrease of peat layer thickness. Now we have used a simple model (Eq. (1)) for this purpose that only takes the effect of groundwater table into account. We think monitoring data will allow a better calibration of this model. It is expected that the peat layers will continue to degrade in the (near) future in large parts of the country. Peat oxidation rates are estimated to be up to 1 cm year^{-1} (van den Akker et al., 2008; Hoogland et al., 2012). So within a decade or so, the maps will again be outdated. The update program invests much effort in field data collection. It is expected that by the end of the program over 5000 new soil profile descriptions are collected that will provide a valuable insight in the current status of the soils of the Dutch peatlands. The value and usability of these data would be greatly increased if these could be updated at regular intervals (e.g. every 5 years) with a reliable model calibrated with the data from the soil monitoring network. The updated field data can then be used to generate updated peat thickness and soil class maps with the methodology that is now in place.

Operationalizing DSM for soil survey in the Netherlands does not only influence the way soil maps are generated. It also has implications for the Dutch soil information system *BIS* (Kempen, 2011). *BIS* is now a system that is designed to store a limited number of maps, currently the 1:50,000 and 1:250,000 national soil class maps and the soil point data. Its functionality should be expanded to system that not only stores a multitude of maps in both polygon and raster format at different scales and resolutions, but also DSM toolboxes with prediction models that can be run to generate maps on-demand according to user-defined specifications (Heuvelink et al., 2010).

5. Conclusions

Digital soil mapping was made operational for nationwide updating of the Dutch 1:50,000 soil map for the peatlands. Soil classes were updated through updating of diagnostic, quantitative soil properties that are subject to change. The updated map was integrated in the 1:50,000 soil map. The DSM methodology allows the combined use of both legacy and newly acquired data while accounting for differences in uncertainty, and can deal with zero-inflated data through a two-step simulation approach. The method allows full quantification of the uncertainty associated to the predicted peat thickness values and soil groups. Validation results were promising and suggest that the method can be applied to update the map for the remaining parts of the peatlands.

In order to increase the value and usability of the point data and soil maps generated in the mapping program, installation of a soil monitoring network in the peatlands is recommended. The functionality of the Dutch soil information system *BIS* should be extended to accommodate not only a new-generation of soil maps generated with DSM, but also new possibilities that DSM offers such as on-demand generation of soil maps according to user-specific needs. The added value of using legacy point data in addition to newly acquired data was not assessed and merits further exploration.

References

Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H., 2014. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. Geoderma 214–215, 101–113.

Berendrecht, W.L., Snepvangers, J.J.J.C., Minnema, B., Vermeulen, P.T.M., 2007. MIPWA: a methodology for interactive planning for water management. In: Oxley, L., Kulasiri, D. (Eds.), MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp. 330–334 (December).

Bouma, J., Droogers, P., 2007. Translating soil science into environmental policy: a case study on implementing the EU soil protection strategy in The Netherlands. Environ. Sci. Pol. 10 (5), 454–463.

Brus, D.J., de Gruijter, J.J., Walvoort, D.J.J., de Vries, F., Bronswijk, J.J.B., Romkens, P.F.A.M., de Vries, W., 2002. Mapping the probability of exceeding critical thresholds for cadmium concentrations in soils in the Netherlands. J. Environ. Qual. 31 (6), 1875–1884.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62 (3), 394–407.

Carré, F., McBratney, A.B., Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma 141 (1–2), 1–14.

Christensen, O.F., 2004. Monte Carlo maximum likelihood in model-based geostatistics. J. Comput. Graph. Stat. 13 (3), 702–718.

Clement, J., Kooistra, L., 2003. Eerste bosstatistiek digitaal; opbouw van een historisch basisbestand. Technical Report 744. Alterra.

Coe, R., Stern, R., 1982. Fitting models to daily rainfall data. J. Appl. Meteorol. 21 (7), 1024–1031.

de Bakker, H., Schelling, J., 1966. A System of Soil Classification for The Netherlands: The Higher Levels. Pudoc, Wageningen.

de Groot, W.J.M., Visschers, R., Kiestra, E., Kuikman, P.J., Nabuurs, G.J., 2005. Nationaal systeem voor de rapportage van voorraad en veranderingen in bodem-C in relatie tot landgebruik en landgebruikveranderingen in Nederland aan de UNFCCC. Technical Report 1035–3. Alterra.

de Vries, F., Lesschen, J.P., van den Akker, J.J.H., Petrescu, A.M.R., van Huissteden, J., van den Wyngaert, I., 2009. Bodemgerelateerde emissie van broeikasgassen in Drenthe; de huidige situatie. Technical Report 1859. Alterra.

Delhomme, J., 1978. Kriging in the hydrosciences. Adv. Water Resour. 1 (5), 251–266.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C Appl. Stat. 47 (3), 299–325.

Finke, P.A., Groot Obbink, D.J., Rosing, H., de Vries, F., 1996. Actualisatie Gt-kaarten 1: 50 000 Drents deel kaartbladen 16 Oost (Steenwijk) en 17 West (Emmen). Technical Report 439. DLO-Staring Centrum.

Finke, P.A., Brus, D.J., Bierkens, M.F.P., Hoogland, T., Knotters, M., de Vries, F., 2004. Mapping groundwater dynamics using multiple sources of exhaustive high resolution data. Geoderma 123 (1–2), 23–39.

Fletcher, D., MacKenzie, D., Villouta, E., 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environ. Ecol. Stat. 12 (1), 45–54.

Gastaldi, G., Minasny, B., McBratney, A., 2012. Mapping the occurrence and thickness of soil horizons within soil profiles. , pp. 145–148.

Hack-ten Broeke, M.J.D., Schut, A.G.T., Bouma, J., 1999. Effects on nitrate leaching and yield potential of implementing newly developed sustainable land use systems for dairy farming on sandy soils in the Netherlands. Geoderma 91 (3–4), 217–235.

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: a decision-tree based approach in Bavarian forest soils. Geoderma 185–186, 37–47.

Hazeu, G.W., 2005. Landelijk Grondgebruiksbestand Nederland (LGN5). Vervaardiging, nauwkeurigheid en gebruik. Technical Report 1213. Alterra.

Heilbron, D.C., 1994. Zero-altered and other regression models for count data with added zeros. Biom. J. 36 (5), 531–547.

Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: a Random Forest approach. Geoderma 214–215, 141–154.

Heuvelink, G.B.M., Brus, D.J., de Vries, F., Vašát, R., Walvoort, D.J.J., Kempen, B., Knotters, M., 2010. Implications of digital soil mapping for soil information systems. In: Napoli, R., Francaviglia, R. (Eds.), Proceedings of the 4th Global Workshop on Digital Soil Mapping, p. 67 (Rome, 24–26 May). Available at: http://edepot.wur.nl/160764.

Hoogland, T., van den Akker, J.J.H., Brus, D.J., 2012. Modeling the subsidence of peat soils in the Dutch coastal area. Geoderma 171–172, 92–97.

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression. 2nd edition. John Wiley & Sons, New York.

Jansen, P., Hendriks, R., Kwakernaak, C., 2010. Behoud van veenbodems door ander peilbeheer. Maatregelen voor een robuuste inrichting van het westelijk veenweidegebied. Technical Report 2009. Alterra.

Kempen, B., 2011. Updating Soil Information With Digital Soil Mapping. (PhD thesis). Wageningen University.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151 (3–4), 311–326.

Kempen, B., Brus, D.J., De Vries, F., Engel, B., 2012a. Updating legacy soil data for digital soil mapping. In: Minasny, B., Malone, B., McBratney, A. (Eds.), Digital Soil Assessments and Beyond, pp. 91–96.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., 2012b. Soil type mapping using the generalised linear geostatistical model: a case study in a Dutch cultivated peatland. Geoderma 189–190, 540–553.

Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M., de Vries, F., 2012c. Efficiency comparison of conventional and digital soil mapping for updating soil maps. Soil Sci. Soc. Am. J. 76 (6), 2097–2115.

Knol, W., Kramer, H., Dorland, G., Gijsbertse, H., 2003. Historisch Grondgebruik Nederland: tijdreeksen grondgebruik Noord-Holland van 1950 tot 1980. Technical Report 751. Alterra.

Knol, W., Kramer, H., Gijsbertse, H., 2004. Historisch Grondgebruik Nederland: een landelijke reconstructie van het grondgebruik rond 1900. Technical Report 573. Alterra.

Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67 (3–4), 227–246.

Knotters, M., Brus, D., Heidema, A., 2007. Heavy metal concentrations in the top soils of 'Achterhoek-Kreis Borken'; Mapping across the Dutch–German border. Technical Report, Wageningen, Alterra-Report 1455.

Kros, J., Frumau, K., Hensen, A., de Vries, W., 2011. Integrated analysis of the effects of agricultural management on nitrogen fluxes at landscape scale. Environ. Pollut. 159 (11), 3171–3182.

Lark, R.M., Cullis, B.R., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. Eur. J. Soil Sci. 55 (4), 799–813.

Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor E-BLUP with REML. Eur. J. Soil Sci. 57 (6), 787–799.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.

McCullagh, P., Nelder, J., 1989. Generalized Linear Models. Chapman & Hall/CRC.

Menard, S., 2000. Coefficients of determination for multiple logistic regression analysis. Am. Stat. 54 (1), 17–24.

Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213, 385–399.

Nol, L., Heuvelink, G.B.M., Veldkamp, A., de Vries, W., Kros, J., 2010. Uncertainty propagation analysis of an N₂O emission model at the plot and landscape scale. Geoderma 159 (1–2), 9–23.

Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma 214–215, 91–100.

Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., Bogaert, P., 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. Geoderma 232–234, 97–106.

Reijneveld, A., van Wensem, J., Oenema, O., 2009. Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. Geoderma 152 (3–4), 231–238.

Schulp, C.J.E., Veldkamp, A., 2008. Long-term landscape–land use interactions as explaining factor for soil organic matter variability in Dutch agricultural landscapes. Geoderma 146 (3–4), 457–465.

Sonneveld, M.P.W., Hack-ten Broeke, M.J.D., van Diepen, C.A., Boogaard, H.L., 2010. Thirty years of systematic land evaluation in the Netherlands. Geoderma 156 (3–4), 84–92.

Steur, G.G.L., Heijink, W., 1991. Bodemkaart van Nederland, schaal 1:50.000; Algemene begrippen en indelingen (4e editie). Staring Centrum, Wageningen.

Subburayalu, S., Slater, B., 2013. Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Sci. Soc. Am. J. 77 (4), 1254–1268.

Subburayalu, S., Jenhani, I., Slater, B., 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. Geoderma 213, 334–345.

Sulaeman, Y., Minasny, B., McBratney, A.B., Sarwani, M., Sutandi, A., 2013. Harmonizing legacy soil data for digital soil mapping in Indonesia. Geoderma 192, 77–85.

van Beek, C.L., Pleijter, M., Kuikman, P.J., 2011. Nitrous oxide emissions from fertilized and unfertilized grasslands on peat soil. Nutr. Cycl. Agroecosyst. 89 (3), 453–461.

van den Akker, J., Kuikman, P., de Vries, F., Hoving, I., Pleijter, M., Hendriks, R., Wolleswinkel, R., Simões, R., Kwakernaak, C., 2008. Emission of CO2 from agricultural peat soils in the Netherlands and ways to limit this emission. In: Farrell, C., Feehan, J. (Eds.), After Wise Use—The Future of Peatlands, Proceedings of the 13th International Peat Congress, Vol. 1 — Oral Presentations, Tullamore, Ireland, 8–13 June 2008. International Peat Society, Jyväskylä, Finland, pp. 645–648.

van der Salm, C., de Vries, W., Kros, J., 1996. Modelling trends in soil solution concentrations under five forest–soil combinations in the Netherlands. Ecol. Model. 88 (1–3), 19–37.

van Kekem, A.J., Hoogland, T., van der Horst, J.B.F., 2005. Uitspoelingsgevoelige gronden op de kaart; werkwijze en resultaten. Technical Report 1080. Alterra.

van Lynden, K.R., van Soesbergen, G.A., van Wallenburg, C., Westerveld, G.J.W., 1985. De interpretatie van de bodemkaarten in Nederland. Cultuurtechnisch Tijdschr. 25 (2), 58–68.

van Zijl, G., Bouwer, D., van Tol, J., le Roux, P., 2014. Functional digital soil mapping: a case study from Namarroi, Mozambique. Geoderma 219–220, 155–161.

Vernes, R.W., 2005. Van gidslaag naar een hydrogeologische eenheid. toelichting bij de totstandkoming van de dataset regis ii. Technical Report Rapport 05-038-B. Nederlands Instituut voor Toegepaste Geowetenschappen TNO.

Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36 (10), 1261–1267.

Webster, R., McBratney, A.B., 1989. On the Akaike Information Criterion for choosing models for variograms of soil properties. J. Soil Sci. 40 (3), 493–496.

Wedderburn, R.W.M., 1974. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. Biometrika 61, 439–447.

Yang, L., Jiao, Y., Fahmy, S., Zhu, A.-X., Hann, S., Burt, J.E., Qi, F., 2011. Updating conventional soil maps through digital soil mapping. Soil Sci. Soc. Am. J. 75 (3), 1044–1053.