An evaluation of remote sensing derived soil pH and average spring groundwater table for ecological assessments

Roelofsen, H.D.; Bodegom, P.M. van; Kooistra, L.; Amerongen, J.J. van; Witte, J-P.M.

International Journal of Applied Earth Observation and Geoinformation 43 (2015) 149–159

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

An evaluation of remote sensing derived soil pH and average spring groundwater table for ecological assessments

Hans D. Roelofsen[a,*], Peter M. van Bodegom[b], Lammert Kooistra[c], Jorg J. van Amerongen[c], Jan-Philip M. Witte[a,d]

[a] KWR Watercycle Research Institute, Post Box 1072 3430 BB, Nieuwegein, The Netherlands
[b] Department Conservation Biology, Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333CC Leiden, The Netherlands
[c] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB, Wageningen, The Netherlands
[d] Department of Ecological Sciences, Sub Department Systems Ecology, VU University, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 30 September 2014
Accepted 11 May 2015
Available online 19 June 2015

Keywords:
Site factors
Soil pH
Ground water
Indicator values
Airborne hyperspectral
Ecosystem management

ABSTRACT

Ecological assessments such as species distribution modelling and benchmarking site quality towards regulations often rely on full spatial coverage information of site factors such as soil acidity, moisture regime or nutrient availability. To determine if remote sensing (RS) is a viable alternative to traditional data sources of site factor estimates, we analysed the accuracy (using ground truth validation measurements) of traditional and RS sources of pH and mean spring groundwater level (MSL, in m) estimates. Traditional sources were a soil map and hydrological model. RS estimates were obtained using vegetation indicator values (IVs) from a Dutch national system as an intermediate between site factors and spectral response. IVs relate to those site factors that dictate vegetation occurrence, whilst also providing a robust link to canopy spectra. For pH, the soil map and the RS estimate were nearly as accurate. For MSL, the RS estimates were much closer to the observed groundwater levels than the hydrological model, but the error margin of the estimates still exceeded the tolerance range of moisture sensitive vegetation. The relatively high accuracy of the RS estimates was made possible by the availability of local calibration points and large environmental gradients in the study site. In addition, the error composition of the RS estimates could be analysed step-by-step, whereas the traditional sources had to be accepted 'as-is'. Also considering that RS offers high spatial and temporal resolution at low costs, RS offered advantages over traditional sources. This will likely hold true for any other situation where prerequisites of accurate RS estimates have been met.

1. Introduction

Natural area management at local scale requires full spatial coverage knowledge on site factors such as groundwater tables, nutrient levels and soil acidity (pH). For example, plant species and plant community distributions can be derived from site factors, on the premise that these have specific demands on e.g. groundwater regime (Bartholomeus et al., 2012; Runhaar et al., 1997; Tüxen, 1954), soil acidity (Cirkel et al., 2012; Damgaard et al., 2014) and Nitrogen availability (Douma et al., 2012), or a combination thereof (Roelofsen et al., 2014). Monitoring of natural areas is also possible from site factors, for example, when assessing the influence of management measures, succession or the impact of climate change and N-deposition on natural vegetation (Cousins and Lindborg, 2004; Ellenberg, 1992; Hannerz and Hånell, 1997; Olde Venterink and Wassen, 1997). Likewise, the European Union (EU) habitat directive requires its member states to report every six years on the conservation status of habitats and species. A possible approach is to benchmark observed site factors against known requirements (e.g. (Runhaar et al., 2009) for The Netherlands) of the target habitat types to reveal hotspots of environmental deficits.

While full coverage estimates of site factors are a versatile management aid, sampling to obtain the spatial and temporal variability of site factors is often constrained by time and resources. Therefore, researchers and managers resort to traditional sources to provide the required data, such as paper maps or previous model output on an area. However, these sources can be flawed by e.g. high hetero-

* Corresponding author. Tel: 0031 30 60 69 647; Fax: +0031 30 60 61 165.
E-mail addresses: hd.roelofsen@gmail.com (H.D. Roelofsen), p.m.van.bodegom@cml.leidenuniv.nl (P.M. van Bodegom), lammert.kooistra@wur.nl (L. Kooistra), jorgva@gmail.com (J.J. van Amerongen), flip.witte@kwrwater.nl (J.-P.M. Witte).

http://dx.doi.org/10.1016/j.jag.2015.05.005
0303-2434/

geneity within polygons and outdated content for maps, or coarse spatial resolution for models. Whatever the source, it is a risk to use them for something else than their original purpose. For example, a soil map may have been compiled with regional agricultural suitability assessment in mind, so that its sampling density and minimum mapping area preclude reliable statements about small parcels. Moreover, information from traditional sources does not necessarily coincide with those site factors that are relevant for nature management.

With the increasing spatial, spectral and temporal resolution of earth observation sensors, remote sensing (RS) is an alternative source of site factor estimates. Examples include soil moisture content (Bastiaanssen et al., 2005; Panciera et al., 2009; Wigneron et al., 2007), soil iron content (Bartholomeus et al., 2007) and soil salinity (Bell et al., 2001). These examples relate RS signals directly to the site factor of interest. In the context of nature management, an interesting alternative is to derive site factors through their impact on the natural vegetation. Plant species and communities have a narrow tolerance towards one or more site factors, making their presence indicative for the sites conditions. If one or more site factors have a harmonised and consistent influence on the vegetation physiology and phenology, this influence can be observable from RS data. For example, regional groundwater tables can be derived indirectly from thermal RS data (Kaiser et al., 2012). The compounded impact of groundwater tables on plant physiology and evapotranspiration was summarised into 'ecological moisture values' (Kaiser et al., 2012) that were statistically related to the RS canopy temperature data.

Site factors filtering on species employing certain morphological and structural properties that accommodate survival in that situation have been observed and quantified in many situations (Bartholomeus et al., 2008; Ordoñez et al., 2009; Reich et al., 2007; Soudzilovskaia et al., 2013). At the same time, plant properties such as leaf water content (Roelofsen et al., 2013), leaf area index and leaf orientation (Asner, 1998) and chlorophyll (Doughty et al., 2010) modulate the absorption and reflection intensity of sunlight in different wavelengths and together shape the canopy spectrum. Therefore, we hypothesise that plant properties can connect RS data with site factors. This holds particularly true when the site factor is the main driver of vegetation patterns inside the spatial extent of a RS image, thereby exhibiting a clear gradient in intensity that relates to the spectral signal. Indeed, the relation between RS and site factors can deteriorate by overruling factors such as planted vegetation and active management. Likewise, plant species that tolerate a wide range of site conditions are less indicative for site conditions.

Nonetheless, the advantage of using vegetation properties to derive site factors is that these relate as closely as possible to the site factors as experienced by the vegetation. Vegetation responds slowly to site factor dynamics and a single measurement in time and space may not consistently explain species occurrence. From an ecological management perspective, it is sensible to pursue long term site factor dynamics, rather than their daily and seasonal fluctuations. Although this approach is being explored (Feilhauer et al., 2011; Kaiser et al., 2012; Roelofsen et al., 2014; Schmidtlein et al., 2011), the accuracy of the site factor estimates is not always established (Kaiser et al., 2012). Furthermore, the merits of this approach compared to legacy site factor information sources are unexplored.

Therefore, the objective of this study is to evaluate the RS and traditional data methods to estimate site factors in a natural area. For this, we employ these two approaches to derive two site factor estimates for a natural area in The Netherlands. The accuracy of these estimates is assessed from an independent dataset of site factor measurements. While the data source with lowest deviation from the validation set should be considered most accurate, other aspects of the data sources such as the spatial and temporal resolution will also be considered.

We focus on two site factors that are known drivers of vegetation and for which different data sources were easily available. Firstly, soil pH, which modulates abundance of toxic substances such as aluminium and manganese, as well as essential nutrients such as phosphorus, calcium, magnesium and potassium. Because plant have to employ complex adaptations to mitigate the effects of low or high pH (Cirkel et al., 2012), soil pH is a strong driver of species distribution. Secondly, we derive mean spring groundwater level (MSL), a characteristic groundwater table that is frequently used in The Netherlands (Runhaar et al., 1997). Groundwater table is often assessed during spring because water table is still high from past winter and plant evapotranspiration has not yet begun (Runhaar et al., 1997). As such MSL determines oxygen and water availability to plant roots during this critical stage for species survival (Knapp et al., 2008). MSL is defined as the average groundwater-surface distance in m. during March–April over at least 30 years. In this paper, below surface level is expressed as a positive and flooding as a negative value.

To derive pH and MSL from traditional sources, we employ a soil map and a nationwide hydrological model. For the RS approach, plant indicator values (IVs) are estimated first and then translated to pH and MSL using earlier established relations. IVs are continuous numerical values that express the realised optimum of a plant species with respect to site factors (Diekmann, 2002). IVs are not physical plant properties that can be measured, but are rather derived from expert judgement (Ellenberg, 1992) or vegetation plot data analysis (Witte et al., 2007). IVs can be considered an umbrella descriptor that synthesises plant properties and physiology into a single number. As a consequence, IVs relate strongly to site factors that influence vegetation properties (Bartholomeus et al., 2008; Cirkel et al., 2012), as well as to the spectral (Roelofsen et al., 2014; Schmidtlein, 2005) and topographic variables (Ecker et al., 2010). The latter relation is never causal but indirect, because IVs merely relate to plant properties that modulate leaf reflectance and absorbance.

Since differences in site related plant features can best be detected by focusing on one vegetation structure, we limited our study to short vegetation (grasses, herbs, short shrubs), all the more since trees in The Netherlands species are nearly always planted and not very informative about their site conditions. Within this short vegetation class, we expect that the availability of water and nutrients is expressed by canopy height, leaf density, and the 'greenness' of the vegetation. These approaches are detailed further in the next section, following the description of the study site.

2. Materials & methods

2.1. Study site

'Kampina' is a Natura2000 protected nature reserve natural area in the south of The Netherlands (5°16′00″E 51°34′00″N), located on Pleistocene cover sand. Within it, a 15 km^2 study site was delineated that excluded forests, artificial structures, agriculture, tall shrubs and open water (this is motivated in Section 2.3.1) but leaves natural vegetation of short shrubs, grasses and forbs. Notable topographic features of the study site are a small river flowing north-east bound through the southern half of the study site and several elevated sand dune ridges in the north. Rainwater fed fens are found in between these ridges. Elevation differences up to 12 m are observed. These features create a gradient in pH and MSL along with other abiotic factors such as nutrient availability, which in turn drive considerable variation in vegetation. Dry and moist heather species (*Caluna vulgaris*, *Erica tetralix* and *Molinia caerulea*) dominate the oligotrophic northern part of the study site. Along and in between the fens, moist, acidic and nutrient poor species such as *Drosera rotundifolia* and peat mossess are found. Grasslands along the small river in the south are moist to wet and mesotrophic to eutrophic and host species such *Cirsium dissectum* and *Succisa pratensis*. Small patches of wet, alkaline oligotrophic grasslands with rare threatened species such as *Carex pulicaris* are present as well. Along walking routes and parking places eutrophic vegetation is found containing among other species *Urtica dioica* and *Phragmitus australis*.

Several parcels on the south and south eastern fringes of the study site were converted from agricultural to natural areas in 2000–2005 by excavating the nutrient rich topsoil and raising the water table. Local water management aims to improve quality and quantity of the ground and surface water. This is done by clearing pine trees and filling up ditches. A lock in the small river prevents rapid drainage of the upriver area and part of rivers vicinity is diked to retain water in case of flooding.

The main motivation to select this study site was the large pH and MSL gradient which would aid in relating IVs to the spectral data (see Section 2.4). In addition, the site was well accessible and within driving range of laboratory facilities. Several traditional data sources covering this site were readily available. Finally, the site was conveniently close to the operating base of the APEX hyperspectral scanner 2012 flight campaign (see Section 2.4.1), which increased the chances of acquiring a RS image of the site.

2.2. Traditional sources

2.2.1. pH data from a soil map

Since 1964, soil maps of The Netherlands have been published, with the last regions being mapped in the 1990's (Steur and Heijink, 1991). As such, a soil map of the study site was readily available (Damoiseaux and Teunissen van Manen, 1984). The study site was mapped in the period 1964–1977 at scale 1:50,000. Twelve different soil types were mapped as polygons in the study site, with a minimum mapping area of 0.25 km^2. Approximately, 0.55 km^2 was mapped as 'open water' or as 'disturbed soil' and was not assigned a soil type. A hypothetical soil sample within a polygon on the soil map will also, to a certain extent, contain different soil types than suggested by the soil map (Buringh et al., 1962). This is due to interpretations during the field survey, mapping scale and soil development, variation over small distances, the fuzziness of soil type extent borders and the need to ignore too small variations. Approximately 2 km^2 within the study site was labelled as having >30% probability of a mismatch between the reported soil type and a random sample (Steur and Heijink, 1991), but the accuracy of the soil map has not been specified beyond this. Land use changes (e.g. agriculture to nature conversion) since the mapping period will lead to additional mismatches. Finally, water bodies on the soil map and in the RS image deviate to some extent.

An earlier study determined physical and chemical characteristics for all soil types in The Netherlands, specifically aiming to provide quantitative soil properties for use in modelling studies (de Vries, 1999). This was achieved by linking measurement data from point soil samples to the soil unit in which the soil sample was taken. This way, each soil type in the study site was attributed a pH KCl value and a spatially explicit pH map was created, save for the polygons classified as open water or disturbed soil. Soil map derived pH estimates are referred to as pH_{trad}.

2.2.2. MSL from a hydrological model

We employed the output from a Dutch national ground and surface water model, called the 'National Hydrological Instrument' (NHI) (de Lange et al., 2014), as an estimate of MSL in the study site. NHI is freely available and the only (except for local measured data) spatially explicit source of groundwater level data. NHI couples four components that each model a specific part of the Dutch water system: the saturated and unsaturated groundwater zone (modelled with MODFLOW (Harbaugh et al., 2000) and MetaSWAP (van Walsum and Veldhuizen, 2011), respectively), regional surface water (modelled with MOZART (Vermulst et al., 1998)) and national surface water (modelled with DM-SOBEK) (de Lange et al., 2014). NHI output is daily ground water levels in 250 m grid cells for the entire Netherlands during 1966–1995. This time frame was chosen so that long term groundwater tables are represented and influence of extraordinary years is minimised. From its output, we computed MSL as the average of the simulated groundwater levels at 14 March, 28 March and 14 April. NHI derived MSL estimates are referred to as MSL_{trad}.

2.3. Remote sensing estimates of site factors through indicator values

2.3.1. Estimating indicator values

We use the plant Indicator Values (IVs, (Diekmann, 2002)) for acidity (mR, from German *Reaktionszahl*) and moisture regime (mF, from German *Feuchtezahl*) as vegetation properties through which soil pH and MSL are estimated. At 32 locations within the study site, the floristic composition within a 2 × 2 m vegetation plot was assessed during August 2012 by an experienced field biologist. Plots were located where species presence was a clear expression of site factors, i.e. excluding forests (most trees in The Netherlands are planted) and wide tolerance species such as *Molinia caerulea*. Observed species were coupled to a previously compiled list of mR and mF for each plant species in The Netherlands (Witte et al., 2007) describing the preference of that species towards acidity (mR, ranging from 1 = acid to 3 = alkaline) and moisture regime (mF, ranging from 1 = open water to 4 = completely dry). Note that IV scales are continuous, in contrast to the discrete values of IVs of e.g. Ellenberg (1992). Subsequently, a plot-average value for mR and mF was calculated. From here onwards, the terms mR and mF and their collective noun 'IVs', refer to plot averaged values, rather than the value for individual species.

An image of the study site was acquired on 30 June 2012 with the Airborne Prism Experiment (APEX) sensor, containing spectral information in 228 bands from 507 to 2410 nm, excluding spectral regions with severe atmospheric absorption 1347–1434 nm and 1804–1957 nm. Ground resolution was 2.1 m. Top of canopy reflectance was generated by correcting the data for atmosphere and geometry by the Flemish Institute for Technological Research (Biesemans et al., 2010).

The image acquisition date (30 June) did not coincide with the MSL calculation period (March–April). However, although the selection of plants is susceptible to environmental conditions at the start of the growing season (March–April) when plants germinate and start to grow (Runhaar et al., 1997), the visual expression of this selection is most prominent during the growing season (end of June–August).

The study site was extracted from the image (and slightly reduced due to few clouds) using topographic map and field observations to delineate forests and tall shrubs. Spectral data was extracted for each vegetation plot as a weighted average of all pixels intersecting the extent of the vegetation plot. The image covered all but one plot. This was compensated by measuring canopy reflectance with an ASD (Analytical Spectral Devices, Inc., Boulder, CO, USA) FieldSpec Pro FR spectrometer in early August 2012 (see Roelofsen et al. (2013) for a detailed description of the sampling procedure). The APEX spectral signature was reconstructed by selecting FieldSpec bands that most closely resemble APEX spectral bands. Resemblance between APEX and FieldSpec canopy reflectance was assessed from several other plots with joint measurements, revealing the FieldSpec reflectance slightly lower in the Near Infrared range (NIR, 700–1500 nm) but otherwise similar (data not shown). From this, we were confident to incorporate the FieldSpec data into the spectral dataset.

The Dutch national elevation dataset (AHN-2, acquired over the study site February–March 2009, (van der Sande et al., 2010)) was used to extract the following topographic variables: elevation, slope, aspect, exposition, profile and planofile curvature. It was expected that these variables could enhance IV estimation, especially mF might be related to elevation (Ecker et al., 2010). Spatial resolution was 0.5 m, re-sampled using nearest-neighbour technique to match the APEX resolution. Topographic data was recovered for each plot as the weighted average of the pixels intersecting the plot extent.

IVs were modelled with partial least squares regression (PLSR) using the observed IVs as dependent and the reflectance and topography data as explanatory variables. PLSR is a multi-linear regression technique that handles the often strongly correlated spectral predictor bands well (Wold et al., 2001). It does so by reducing the dimensionality of the explanatory variables by projecting them into a certain number of new, orthogonal latent variables. PLSR models are calibrated by fitting the model to the training data, and validated by omitting one of the training points in the calibration phase and estimating the dependent variable for that point (Leave One Out, LOO validation). Model accuracy is given by the coefficient of determination (r^2) and root mean square error (RMSE) during both the calibration (subscript 'cal') and validation (subscript 'val') phase. An optimal number of latent variables (NLV) minimises r^2_{cal} to prevent over fitting and maximises r^2_{val} to achieve a generic model.

For each IV, a final PLSR model that maximised r^2_{val} was obtained after iteratively cropping the explanatory variables to those with highest predictive power, following the procedure described in Feilhauer et al. (2011). The modelling was done in R (R Core Team, 2013) using the pls package (Mevik and Wehrens, 2007) and scripts adapted from Feilhauer et al. (2010).

The models were applied pixel-wise to the study site, resulting in full spatial coverage mF and mR estimates. Because PLSR is a linear model, IV estimates outside the theoretical boundaries (1–3 for mR and 1–4 for mF) are possible and occurred for maximum 1% of the pixels in the study site. These pixels were set to no data.

2.3.2. Translating IVs into pH and MSL

Correlations between plot IVs and site factors have been intensively studied. Here, we use relations between mR and soil pH (Cirkel et al., 2012; Witte et al., 2014) and between mF and MSL (Bartholomeus et al., 2012). Both studies derived the relation from vegetation plots from throughout The Netherlands comprising a wide range of measured abiotic conditions and vegetation types. Therefore, the relations are considered generally applicable in The Netherlands, contributing to the generality of our approach.

The pH–mR relation differentiates between moist to wet (mF < 2.25, Eq. (1), n = 32, r^2 = 0.45, RMSE = 19.3%) and moist to dry (mF > 2.25, Eq. (2), n = 58, r^2 = 0.79, RMSE = 13.1%) sites, as plant species adapted to high soil moisture content are less sensitive to soil acidity (Cirkel et al., 2012). Note that Eqs. (1) and (2) are calibrated on the vegetation plot data in Cirkel et al. (2012) and not on the plot data from this study.

$$\text{pH}_{\text{mf}<2.25} = 1.1258 \exp^{0.6164\ \text{mR}} \tag{1}$$

$$\text{pH}_{\text{mf}<2.25} = 1.1189 \exp^{0.6258\ \text{mR}} \tag{2}$$

The data of Bartholomeus et al. (2012) consisted of MSL values over the period 1971–2000 for 145 vegetation plots with known mF. From this a non-linear model between MSL and mF was derived (Eq. (3), n = 145, r^2 = 0.76, RMSE = 12.4%).

$$\text{MSL} = \frac{\ln\left(\frac{3.5-\text{mF}}{\text{mF}-1.9}\right) - 1.6}{4.6} \tag{3}$$

Eq. (3) is asymptotic for mF approaching 1.9 and 3.5, expressing convergence towards aquatic vegetation and vegetation independent of the groundwater table with deeper ground water levels, respectively. In theory, the relation holds for 1.9 < mF < 3.5, but uncertainties in MSL increase as the relation approaches its asymptote. We therefore constrained MSL estimates to pixels with mF estimates from 2.0 to 3.5 (approximately 95% of all study site pixels). Pixels with mF values outside this range were left blank in the MSL estimation. The pH–mR and MSL–mF relations were applied pixel wise to the IVs estimates in order to generate full spatial coverage pH and MSL estimates (referred to as pH_{RS} and MLS_{RS}).

2.4. Validation

2.4.1. Validation data

Validation data represent actual measurements of the site factor, which are, in contrast to traditional source estimates and RS estimates, point data and are measured in situ at the study site.

At the same 32 locations in the study site where the floristic composition had been determined, soil pH-H_2O (referred to as pH_{field}) was measured with a pH electrode (Hannah instruments, IJsselstein, The Netherlands) during July–August 2012. Five measurements were made within a 2 × 2 m plot, at approximately 5 cm below surface level and averaged per plot. Soil pH varies slightly during the growing season (Cirkel et al., 2012). Because the pH measurements were spread over time, so variation is carried into the dataset. While this creates some additional noise, we expect no structural bias because Eqs. (1) and (2), as well as the soil type – pH relation (Section 2.2.1) are derived from plots that were sampled over different periods.

We converted pH_{field} to pH-KCL to align with pH_{trad} and pH_{RS}, using a linear function between both expressions of pH from database of vegetation plots with both pH-H_2O and pH-KCl measurements (Wamelink et al., 2002) (n = 2219, r^2 = 0.89, RMSE = 5.28%). The final number of pH validation plots was reduced to 30: the one plot outside the APEX image extent was discarded and another plot lacked a pH_{trad} value because its location was charted as open water in the soil map.

Groundwater level observations in 33 observation wells (OW) were intermittently available over the period 1980–2013, with an average of 19 consecutive years per OW. OW locations cover the entire study site, but few are located in the drier north and a few clusters in the south create an over representation of expected shallow MSL values.

Each OW is representative for a larger area around it due to gradual changes in the groundwater table around the OW. As such, OWs are well suited to validate the small scale (2 m) MSL estimates from RS. MSL variation within the extent of each 250 m NHI grid cell will be exposed by OWs located within this range. These OWs are thus useful in exposing how NHI is insensitive to small scale MLS variation.

Two OWs are located in the former agricultural parcels converted to nature. See Fig. 3 for OW locations and note that OW locations do not correspond to vegetation plot locations.

To obtain a representative, long term MSL value that is comparable between different OWs, the groundwater level observations were extrapolated to a common period. This was done by modelling the response of the local water level to a an impulse of precipitation using Menyanthes software (Von Asmuth et al., 2012). Using historical precipitation and evaporation data from nearest meteorological station to each OW, MSL could be modelled for

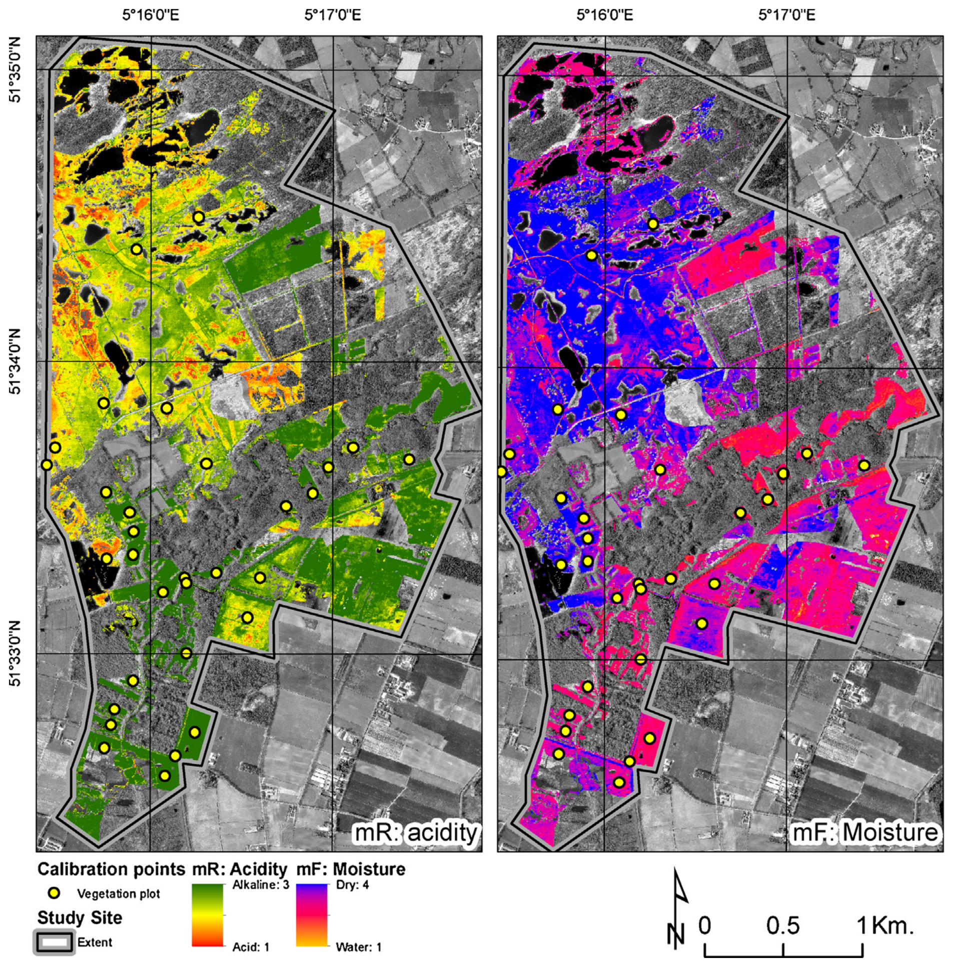

Fig. 1. IV estimates for the study site, resulting from partial least squares regression between APEX hyperspectral and topographical data as explanatory variables and observed Indicator Values as dependant variable. Background: high resolution aerial photograph.

two periods: 1966–1995 for validation MSL_{trad} and 1971–2000 for validation MSL_{RS}. These two series were highly similar: Pearson product moment correlation coefficient $r = 0.99$, $p < 0.001$. The MSL validation data are referred to as MSL_{field}.

2.4.2. Accuracy assessment

The accuracy of all estimates was assessed by point-wise comparison with the validation data. pH_{trad} and pH_{RS} and MSL_{trad} and MSL_{RS} were extracted at the location of 30 vegetation plots and the 33 OWs respectively. Equality between the observed and estimated site factors was assessed by a paired *t*-test, Nash–Sutcliffe coefficient (NSE) (Nash and Sutcliffe, 1970) and the root mean squared error (RMSE).

Two additional validation exercises were performed. Firstly, we applied Eqs. (1) and (2) to the observed mR values at the vegetation plots, yielding pH_{IVobs}. This should reveal how pH estimates are influenced by the PLSR modelling step. Note that this cannot be applied to MSL, because at no point in the study site do MSL_{field} observations coincide with observed mF.

Secondly, the spatial variance in pH and MSL was determined to check for spatial autocorrelation. An omnidirectional sample variogram was calculated for each two estimates of each site factor as well as the validation measurements. The distance between two plots and OWs, respectively, was grouped in distance intervals of 100 m up to a maximum distance of 1000 m. Since the two series of validation MSL data were nearly identical, we used just one of these for the variogram generation. All analyses were performed in R (R Core Team, 2013) and ArcGIS (ESRI, Redlands CA. U.S.A.).

3. Results

3.1. IV estimates from remote sensing data

A large variety for mF and mR was observed for the 32 vegetation plots in the study site, indicative for waterlogged (mF = 1.8) to nearly completely dry (mF = 3.9) sites and fully alkaline (mR = 1.1) to acidic (mR = 2.6).

mR related strongly to the spectral data, with $r^2_{val} = 0.81$. None of the topographic predictors was retained during the band selection and 28 spectral predictors in the NIR and Short Wave Infrared (SWIR, 1500–2410 nm) regions related best to the mR values. This confirms NIR spectra being sensitive to leaf orientation and leaf area index (LAI) (Asner, 1998), which are among the plant properties that change between acidic and alkaline conditions.

The residuals of the estimated mR values at the 32 plots show little to no relation to the observed mR values (Supplementary data 1), indicating that structural deviation is absent.

mF was moderately accurately modelled with $r^2_{val} = 0.62$. Terrain elevation was the most important predictor for mF, suggesting moisture regime is greatly influenced by the sites' elevation, and thus, (roughly) by groundwater tables. In addition, spectral bands located in the SWIR region were influential in the mF model. This conforms to SWIR spectra being sensitive for water content (Asner,

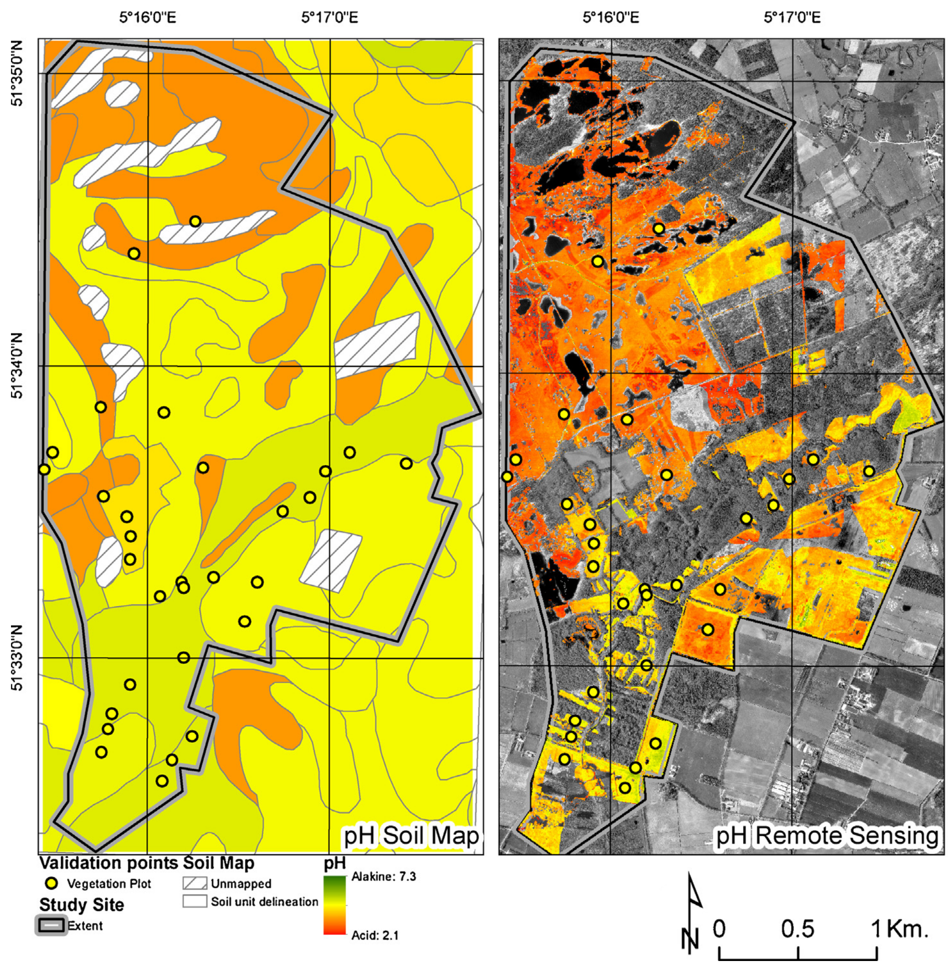

Fig. 2. pH Estimates from the soil map (left) and from remote sensing (right). Note that the same colour scale is used for both maps. Yellow dots are vegetation plots where pH was measured in order to validate both estimates.

1998). For most of the dryer plots (high mF), mF was underestimated (Supplementary data 1).

Applying the PLSR models to the spectral and topographic data of the study site gives a spatially explicit indication of mR and mF (Fig. 1). The spatial pattern of mR (left panel) reflects major vegetation types, with dry heather in the north being correctly labelled as acidic and oligotrophic moist grasslands in the south having alkaline soils. The northerly sand dune ridges are correctly assigned high mF values, reflecting the dry conditions there. Likewise, in the grasslands surrounding the brook valley in south part of the study site, lower mF values are encountered.

Looking at the estimated IVs for the vegetation plots, mR and mF are significantly correlated ($r = -0.65$, $p < 0.05$), meaning that low mR–high mF (acidic-dry) and high mR–low mF (alkaline-wet) often coincide. This pattern is confirmed to some extent in the spatially explicit estimates (Fig. 1).

3.2. pH Estimates from remote sensing and traditional sources

The soil map, from which pH_{trad} was obtained, had polygons that did not reveal pH variation on a finer scale. It is also clear that the pH value can remain unchanged between soil types, resulting in equal pH over several adjacent polygons (Fig. 2, left panel). pH values in the soil map range from 3.4 to 5.2, with lower values observed north and west and a large swath of alkaline values following the rivulet flow south to north east.

The range of pH_{RS} estimates, as derived from the pH–mR relation is wider: pH 2.1–7.9. A broad distinction between acid soil in the north and alkaline soils in the south is again apparent. Small local acidic patches are detected in the south–east, as well as more alkaline conditions in the north east (Fig. 2, right panel). Note that the colour scale is identical for both maps in Fig. 2. This clearly reveals that remote sensing estimated more extreme pH values.

3.3. MSL estimates from remote sensing and traditional sources

MSL_{trad} values for the study site were extracted from the NHI model, ranging from 0.043 m to 4.36 m (Fig. 3, left panel). A single NHI pixel had no value assigned due to open water bodies on that location. Especially in the southern tip of the study site low MSL values are detected. This is in line with the recent wet nature development schemes in that area. The general extent of the small rivulet is evident from the lower MSL values. Higher MSL values are found below the sand dune ridges in the north. Numerous pools are present in the northern dry and moist heather area, even though the NHI suggests deep MSL values here. This indicates that the pools are presumably predominantly rainwater fed.

MSL_{RS} does not attain as large MSL values as the NHI (maximum MSL = 3.21 m) but do estimate up to slightly above soil surface (−0.24 m). Off all pixels in the study site, 9.3% have negative MSL values estimated, compared to none in the NHI. Due to the constraints on the mF values that were legible to translate to MSL ($2.0 < mF < 3.5$), the spatial coverage of estimated MSL values was approximately 5% smaller compared to the coverage of mF estimates. MSL_{RS} estimates attain the deepest values North West in the study site, amidst the pools (Fig. 3, right panel). In the grasslands

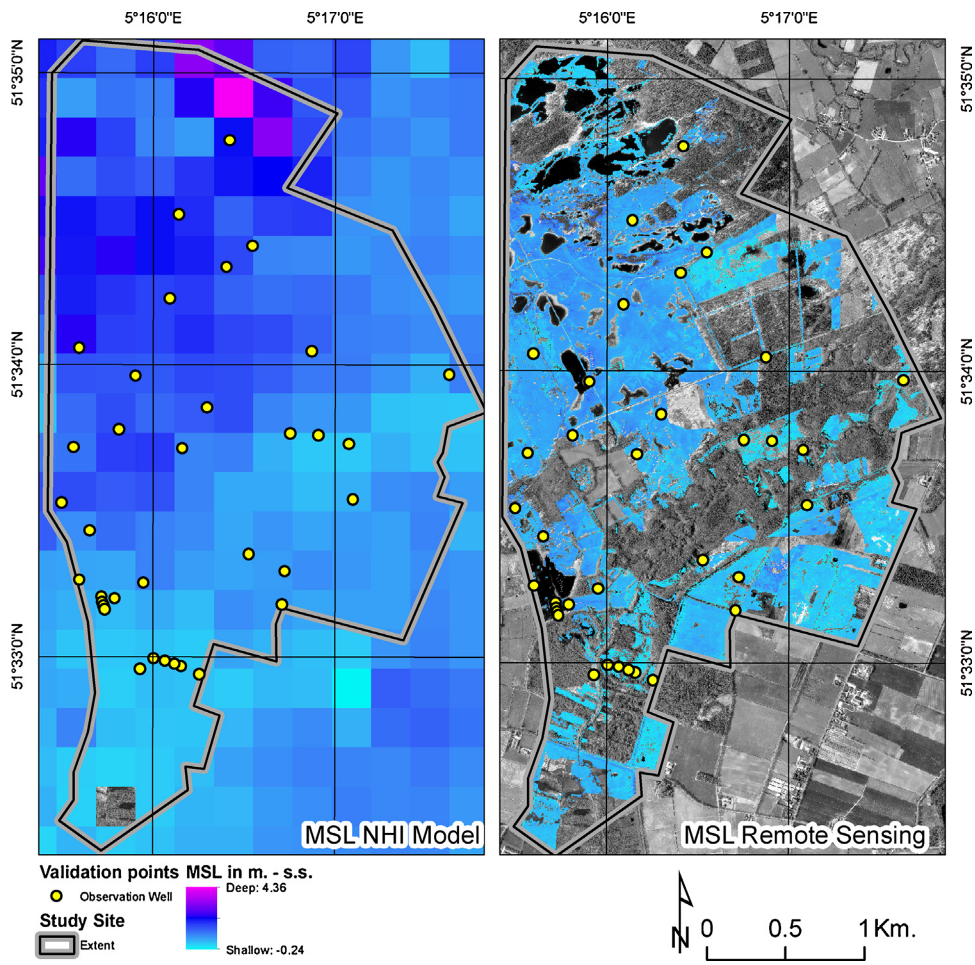

Fig. 3. NHI estimates of MSL (left) and remote sensing based MSL estimates (right). Yellow dots are observation wells where MSL was measured to validate both estimates.

following the rivulet the lighter blue colours indicate a shallow MSL. Note again that the colour scale is identical for both maps in Fig. 3, clearly visualising the wider range of MSL estimates from the NHI.

3.4. Validation pH estimates

A complete overview of all validation exercises is provided in Table 2. According to the validation, pH_{trad} estimates differ significantly from the observed values (paired t-test, $p < 0.05$) and NSE between the observed and estimated pH values is 0.26 (Fig. 4, left panel). This is mainly due to the little variation observed for the soil map pH estimates at the 30 plots. In fact, only seven unique pH values are observed, clustered in a group of plots with mildly alkaline soil with pH ~4.8 and plots with acid soil with pH ~3.5. Most plots are overestimated compared to the pH observations.

pH_{RS} shows more variation compared to pH_{trad} (see also Fig. 3). In general, pH_{RS} follows the 1:1 line with field observations, although large deviations are observed for individual plots. The mean estimated values are not significantly different from the observations (paired t-test, $p = 0.43$), but the NSE is low at 0.26.

The inset graph (Fig. 5, left panel) shows the additional validation of pH_{IVObs} values plotted against pH_{field}. It shows how the pH estimates would relate to the field measurements if the PLSR modelling introduced zero error. Surprisingly, the RMSE of this relation is highly similar to the RMSE of the main plot where pH_{field} and pH_{RS} are compared. This suggests that the PLSR modelling did not introduce additional error.

For both sources of pH estimates, it appears that the highest and lowest observed pH values are often under and overestimated, therefore, the full range of observed pH is not matched by any estimation source. The semivariance values of the soil map estimated pH values are consistently lower than the semivariance of the observed and RS based estimates (Fig. 5). This also points to much less variability in the soil map estimates. Also, the semivariance values do not rise steeply until about 600 m distance. This could point to a polygons size of several km within which correlation is high. The semivariance of pH_{field} and pH_{RS} is highly comparable throughout all distances, except for the smallest distance where a nugget effect (i.e. relatively high semivariance at the minimum distance) is observed for pH_{field}. Such short scale pH variations are apparently not recognised by the RS estimates (for which semivariance is nearly 0 at the shortest distance), but over increasing distances the RS estimates match the actual spatial variation in pH.

3.5. Validation of MSL estimates

The two series of MSL_{field} were nearly identical, although small differences can be observed in the right panel of Fig. 4. MSL_{trad} tends to overestimate MSL values (i.e. too deep ground water levels), with values up to 2.2 m, while the observations only just exceed 1 m. MSL_{trad} differed significantly from the observations (paired t-test, $p < 0.05$). The NSE is <0 (Table 2), which indicates that mean MSL_{field} exceeds the accuracy of the NHI as estimator of MSL. MSL_{RS} does not attain large MSL values (Supplementary data 1) as these require high estimated mF values. Therefore, most relatively high observed MSL values are underestimated by MSL_{RS}. Negative MSL_{RS} estimates are obtained when mF < 2.17 and this is often the case,

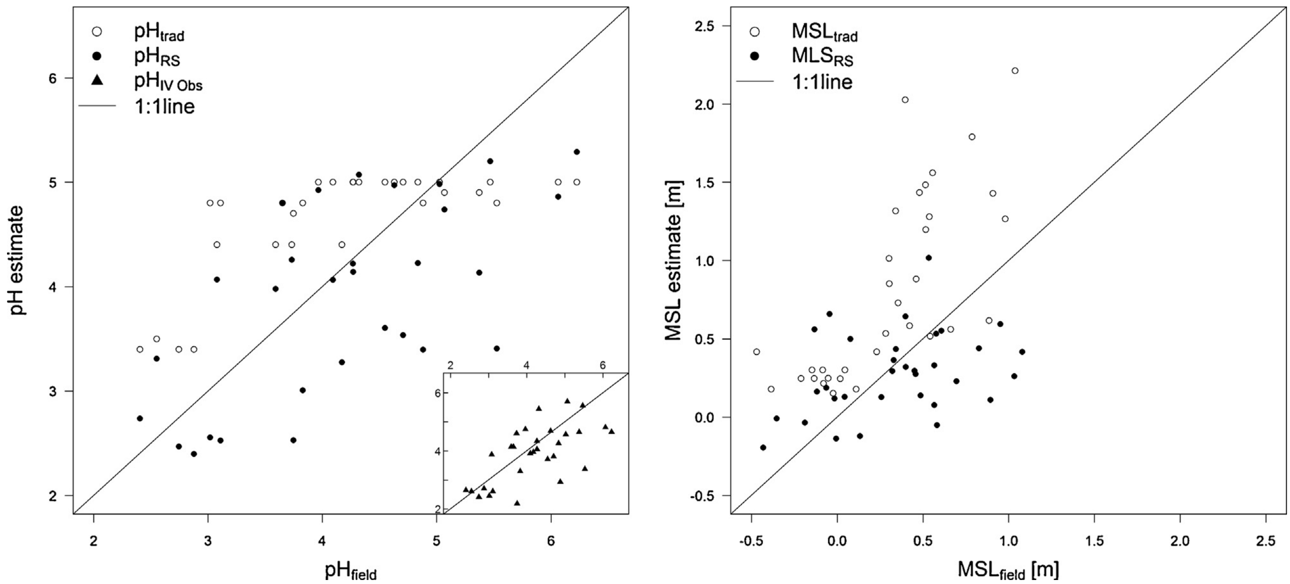

Fig. 4. Scatterplot of observed and estimated pH (left panel) and MSL (right panel). Open dots are estimates from the traditional source; black dots are estimates from remote sensing. The inset graph on the left shows again field measured pH on the horizontal axis and on the vertical axis a pH estimate (pHIVObs) acquired from applying the mR-pH relation directly to observed mR values, i.e. pH estimates without the remote sensing component.

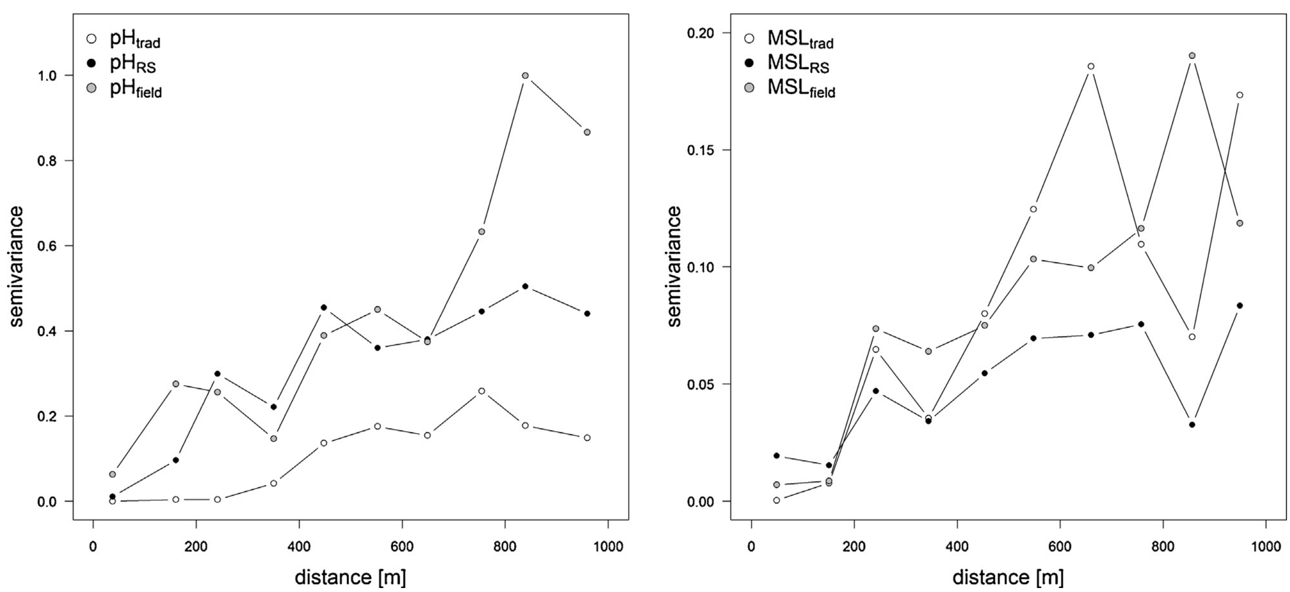

Fig. 5. Omnidirectionalvariogram of pH (left panel) and MSL (right panel) as derived from traditional source (open dot), remote sensing (black dot) and field observations (grey dot).

even for OWs where the observed MSL is nearly 1 m. Overall, the MSL_{RS} tend to increase in value with increasing MSL_{field} and does not differ significantly from MSL_{field} (paired *t*-test, $p = 0.19$). However, the deviations of the estimates from the observed values are large, leading to a very low NSE (0.01).

The semivariance of the MSL estimates and observations is overall much lower compared to pH. For distances >200 m the spatial correlation is highest for MSL_{RS}, which is indicative for the retarded range of MSL estimates. At the same time, MSL_{RS} shows a nugget effect, meaning MSLRS estimates can differ over short distances. The MSL_{field} variogram also shows a small nugget effect. Spatial correlation is observed for all three MSL sources up to about 180 m, beyond which the semivariance strongly increases.

4. Discussion

4.1. Accuracy of site factor estimation

Spatially explicit estimates of two site factors (pH and MSL) from both traditional sources (a soil map and the NHI hydrolog-

Table 1
Summary statistics of PLSR model calibration (cal) and validation (val). Sig. bands = significant bands, NLV = number of latent variables and RMSE = Root Mean Square Error.

Indicator value	# Plots	# Bands	# Sig. bands	NLV	r^2_{cal}	r^2_{val}	$RMSE_{cal}$	$RMSE_{val}$
mR	32	28	23	3	0.86	0.81	0.16	0.19
mF	32	189	39	4	0.79	0.62	0.27	0.36

Table 2
Overview of the validations that were performed on the various estimates of pH and mean spring groundwater level (MSL). NSE = Nash–Sutclife coefficient, RMSE = root mean square error and p = paired t-test and NL = the Netherlands.

Site factor	Estimation source	Validation source	# Points	Type and origin validation point	NSE	RMSE	p
pH_{trad}	Soil map	Field measurement	30	Vegetation plot in study site	0.26	0.86	<0.05
pH_{RS}	Remote sensing	Field measurement	30	Vegetation plot in study site	0.26	0.86	0.43
pH_{IVObs}	Eq. (1) and 2 applied to observed mR	Field measurement	32	Vegetation plot in study site	0.24	0.87	0.06
MSL_{trad}	NHI model	Modelled MSL 1966–1995	36	Observation well in the study site	−1.82	0.64	<0.05
MSL_{RS}	Remote sensing	Modelled MSL 1971–2000	36	Observation well in the study site	0.01	0.39	0.19

ical model) and from a RS – topographical data mix were validated against field measurements of pH and MSL. We found for pH that both sources were about equally accurate, while for MSL the RS estimates more closely resembled observed MSL.

We can reasonably identify the sources of errors of the traditional sources. The soil map suffered from a temporal mismatch between the map surveying and validation measurements, as well as the low spatial resolution due the large polygons and identical pH values assigned to different soil types. The NHI model was intended for national use and thus unable to include small local features that influence hydrology. While it models the hydrological situation on a national scale accurately, large inaccuracies are inevitable on a small scale as this study indicated. These would be revealed when several validation OWs are located inside a single NHI cell, as was the case for two cells in the southern end of the study site. These OWs appear in Fig. 4 as points on a horizontal line, with identical MSL_{trad} but different MSL_{field} values.

RS estimates also suffered from methodological flaws. Their spatial coverage was reduced to that type of vegetation for which IVs are truly indicative of the site conditions and thus omits forests, tall shrubs, agricultural land and intensely managed land. This weakens the idea that RS offers full coverage estimates. Also, weather conditions, flight schedules, as well as the associated costs can constrain acquisition of suitable RS data and thus impede the implementation of RS approach.

The correlation of estimated mR and mF ($r = -0.65$, $p < 0.05$) is also a drawback of the RS approach, because it demonstrates that a few mR–mF combinations are overly favored by the PLSR model. In this case acidic-dry and alkaline-wet situations. Possibly, wet-acidic situations are underrepresented in the PLSR calibration data, or the model is poor in estimating extreme values of both IVs simultaneously. The latter is a known issue of linear models (Roelofsen et al., 2014; van Bodegom et al., 2013).

Neither the spatial resolution of the RS nor that of the traditional sources appeared to match the spatial correlation pattern of the field observations. pH_{trad} is highly correlated for distances up to about 250 m (Fig. 5), while pH_{field} already varies considerably by then. Naturally, small scale pH variations such as in the alkaline rectangular area north in the study site are not detected in the soil map. pH_{RS} estimates covariance is so high that the small nugget effect in the pH_{field} is overlooked, but follows the spatial correlation of pH_{field} well for larger distances. Likewise, the variogram of MSL_{field} shows that up to about 180 m distance, MSL has high covariance (Fig. 5). Estimating MSL then every 2.5 m, as with the RS image, may be superfluous. On the other hand, the 250 m NHI pixels are prone to overlook smaller details, such as the particularly drier region in the south east of the study site that is detected by the RS data.

The RS estimates are at least as accurate (pH) or more accurate (MSL) than the traditional sources. This is likely due to the relative highly accurate IV estimates that underlie the pH and MSL estimates. For mF r^2_{val} was only slightly below r^2_{val} reported in Roelofsen et al. (2014) and both mF and mR estimation accuracy exceeded that of Schmidtlein (2005). Local vegetation plot data surveyed close to image acquisition was a key ingredient for an accurate relation. Another component was the long gradients in mR and mF (Schmidtlein, 2005) of which the extreme points leveraged the predictive power of PLSR. The gradient of mF was from low mF (wet) to high mF (dry) from low elevation along the small river to high elevation along the northern sand ridges. Vegetation properties that modulate canopy reflectance such as chlorophyll and water content will have covaried with the mF gradient. Likewise, correlation between mR and spectral data may be attributed to co-variance of mR with vegetation structure. Acidic soils hosted moist and dry heather vegetation with small shrub like and open canopy structure, in contrast to low and densely grown graminoid and forb vegetation found on more alkaline soils. Leaf orientation and surface area, as well as presence of woody components and increased internal canopy scattering may have been the underlying vegetation properties that created a distinct spectral signature for acidic and alkaline vegetation.

Despite the accuracy of the IV estimates, the final accuracy was relatively low: the accuracy of the pH estimates did not exceed 0.86 pH units, which is approximately 22.5% of the range of pH_{field}. This bandwidth of the pH estimate is sufficient to distinguish acidic from alkaline sites, but prevents more accurate and quantitative acidity variations. Whether the obtained accuracy satisfies the demands on estimation accuracy depends on the vegetation. The mesotrophic moist grasslands around the rivulet are characterised by a narrow tolerance towards acidity fluctuations. Over or underestimating nearly a full pH unit may thus lead to an unrealistic assessment of the suitability of an area for such vegetation. On the other hand, Molinia caerulea tolerates widely varying environmental conditions. When determining the occurrence potential for such vegetation, a mis-estimation of 0.86 pH is thus of less consequence.

Some uncertainties arise from the mR–pH relationship that we used data from a national study (Cirkel et al., 2012). This relation is calibrated on vegetation plots from throughout The Netherlands and can thus be implemented anywhere in the country. As such, we avoided locally calibrated solutions and ensured that our estimated pH values are comparable with other implementations of this relation. The mR–pH relationship of Cirkel et al. (2012) did, however, had a RMSE = 1.17 for moist to wet sites (mF < 2.25) and RMSE = 0.66 for dry sites (mF > 2.25), which is comparable to the $pH_{IVObs} \sim pH_{field}$ calibration (RMSE = 0.87). A much better fit could thus not be obtained.

The applicability of the mF–MSL relation derived from national database of Bartholomeus et al. (2012) could not be determined in a similar fashion. The accuracy of MSL_{RS} (RSME = 0.39) was below the accuracy of the mF–MSL relation (RMSE = 0.24, data not shown), but it is reasonable to assume that the RS modelling of mF is at least partly responsible for this. The mF–MSL relation is tailored to the range of MSL values were the groundwater level still directly influences the vegetation, so the relation does not hold for groundwater independent vegetation. Judging by the RS estimated mF values, 8% of the study site was out of line for the mF–MSL relation. This strengthens the legitimacy of using this relation. By constraining application of Eq. (3) to mF not exceeding 3.5, the deepest possible estimated MSL was approximately 2.95 m. With MSL_{field} not exceeding 1.1 m, the mF–MSL relation was in theory suitable for estimates in this site. Instead, and unlike the situation for pH, the high uncertainties seem to be due to uncertainties in the PLSR

(Table 1). Lower and higher end values of mF were hardly predicted. Still, despite the uncertainties associated with our RS-approach, its RMSE was clearly better than the one associated with the national hydrology model.

MSL was estimated through mF in earlier work (Kaiser et al., 2012), but we cannot benchmark MSL_{RS} estimates this work because MSL was not validated in Kaiser et al. (2012).

Whether inaccuracies in MSL estimation are a problem depends on the specifics of the vegetation. When the groundwater table is so deep that it does not affect the vegetation, an error of 0.39 m in estimated MSL is irrelevant. In those cases, vegetation occurrence and ecological processes are modulated by other site factors such as e.g. acidity, soil porosity and grazing. The more specific the demands made by a plant species to its site conditions, the less likely the site factor will be estimated with sufficient accuracy. Consider the oligotrophic moist grassland along the rivulet, whose optimal growing conditions consist of water tables up to near soil surface. An error of 0.39 m between estimated and true MSL can decide the change of survival of such valuable species. In such situations, calibration of a local hydrological model may be the best solution.

4.2. Implications for ecological assessments

This research demonstrated that acquiring full spatial coverage site factor estimates is fraught with certain inaccuracies. We indicated that the severity of these errors depends on the ecological application at hand and the magnitude of the error depends on the source of the site factor estimate. If such site factor estimates are to be used in subsequent ecological assessments, it essential to take these errors into account. Among applications of our work are using site factors to model floral (Witte et al., 2007) and faunal (Guisan and Thuiller, 2005) diversity, as well as benchmarking site factors against vegetation type requirements or regulations.

Although this was not quantified by actually carrying the estimated site factors forward as model input, it is likely that their large uncertainty would create an ambiguous model outcome at best. In another application (Roelofsen et al., 2014), the low accuracy of an ecological modelling exercise was indeed attributed to the errors in the remote sensing derived IVs that had been used as model input. Because these errors were spatially known, areas with particularly reduced accuracy could be identified.

In ecological assessments and species distribution modelling, input data are often accepted as is and any inaccuracy in the outcome is attributed to the model or process itself. Indeed, only the importance of biological causes of prediction errors is recognised (Guisan and Thuiller, 2005) while the accuracy of input data such as land cover seems to be neglected (Thuiller et al., 2004). This research shows that inaccuracies are as likely in the input data as in any model itself and should be taken into consideration. Preferably, the spatial variation of the accuracy itself is quantified so that closer inspection can be targeted to specific locations.

In our approach, modelling of site factors from RS was a two tier process that was validated at every stage. Therefore, overall accuracy could be retraced to each step so that potential additional data collection or refinement can be targeted. The flexibility and accountability of our RS approach is an advantage compared to the 'as is' nature of the traditional sources. This accountability is essential for an integrated evaluation of the error and uncertainty accumulation in ecological assessment studies and to undertake remedial response such as applying an accuracy threshold.

Our approach is in principle replicable in any natural area where the vegetation is an expression of site factors and can be described by the IV system. IVs were accurately modelled here, but it is unrealistic to carry these PLRS models to another study area (Feilhauer and Schmidtlein, 2011). A local solution where local vegetation plots leverage gradients in vegetation and site factors is then required to replicate our approach. High accessibility of natural areas in The Netherlands makes it easy to acquire new vegetation plots for PLSR model fitting. Areas exempt from our approach are those where variation in site factors and consequently vegetation is poor, or where vegetation with wide tolerance towards site factors dominates. In such areas, IVs fail to describe site factors and the lack of variation dilutes the spectral expression of IVs. In absence of local vegetation plots, legacy plots can be used (Roelofsen et al., 2014) whilst risking that the floristic composition has changed since the image acquisition.

5. Conclusions

We found that full spatial coverage information on site factors is in demand for ecological assessments, such as species modelling or benchmarking a natural area towards species requirements or regulations. However, the accuracy of traditional sources of such information is hardly known. RS processing techniques may provide an alternative method to derive the desired information. In all, more accurate estimates of the site factors were derived from spectral and topographical data compared to traditional data sources. In addition, this approach allowed us to quantify the uncertainties in the estimates. The uncertainties suggest that the absolute estimation accuracy of any of the data sources may fall short, for example if the MSL estimation error exceeds the narrow tolerance range of species towards MSL. Whether or not RS provides a viable alternative can thus only be decided in the light of the application at hand. The investment in acquiring and processing RS data, the spatial and temporal scale, the desired data type and the availability of traditional sources should be considered to come to a balanced decision on how to obtain the desired site factor information. If the RS methodology is applied based on local vegetation plots in an area with strong site factor variation, we think that our approach can alleviate some of the downsides of traditional sources namely their low spatial resolution and possible temporal mismatch. Whatever the outcome, however, we recommend quantifying and integrating the accuracy of the source data, preferably by use of validation measurements.

Acknowledgements

We would like to thank colleagues that shared their data and assisted with the data analysis: Inke Leunk (KWR), Ruud Bartholomeus (KWR), Gijsbert Cirkel (KWR), Martijn Visser (Deltares), Martin de Haan (KWR) and Folkert de Vries (Alterra). The fieldwork campaign was conducted with help of Erwin de Hoop (Natuurmonumenten), Rutger Barendse and several field assistants. The comments of three anonymous reviewers greatly improved this manuscript. This research was funded by the joint research program of the Dutch drinking water utilities.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2015.05.005

References

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64 (3), 234–253.

Bartholomeus, H.M., Epema, G., Schaepman, M.E., 2007. Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 9 (2), 194–203.

Bartholomeus, R.P., Witte, J.P.M., van Bodegom, P.M., Dam, J.C., Becker, P., Aerts, R., 2012. Process-based proxy of oxygen stress surpasses indirect ones in predicting vegetation characteristics. Ecohydrology 5 (6), 746–758.

Bartholomeus, R., Witte, J.P.M., van Bodegom, P.M., Aerts, R., 2008. The need of data harmonization to derive robust empirical relationships between soil conditions and vegetation. J. Veg. Sci. 19, 799–808.
Bastiaanssen, W.G.M., Noordman, E.J.M., Pelgrum, H., Davids, G., Thoreson, B.P., Allen, R.G., 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. J. Irrig. Drain. Eng. 131 (1), 85–93.
Bell, D., Menges, C., Ahmad, W., Van Zyl, J.J., 2001. The application of dielectric retrieval algorithms for mapping soil salinity in a tropical coastal environment using airborne polarimetric SAR. Remote Sens. Environ. 75 (3), 375–384.
Biesemans, J., Horsten, W., Verbeke, B., Vanderstraete, T., van der Linden, S., van Camp, N., VITO, 2010. Image orthorectification and image mosaicing: algorithm theoretical base and validation. VITO reports. Mol, BE: VITO.
Buringh, P., Steur, G.G.L., Vink, A.P.A., 1962. Some techniques and methods of soil survey in the Netherlands. Neth. J. Agric. Sci. 10, 157–178.
Cirkel, D.G., Witte, J.P.M., Nijp, J.N., van Bodegom, P.M., van der Zee, S.E.A.T.M., 2012. The influence of spatiotemporal variability and adaptations to hypoxia on empirical relationships between soil acidity and vegetation. Ecohydrology 7 (1), 21–32.
Cousins, S.A.O., Lindborg, R., 2004. Assessing changes in plant distribution patterns – indicator species versus plant functional types. Ecol. Indic. 4 (1), 17–27.
Damgaard, C., Strandberg, M., Kristiansen, S.M., Nielsen, K.E., Bak, J.L., 2014. Is Erica tetralix abundance on wet heathlands controlled by nitrogen deposition or soil acidification. Environ. Pollut. 184, 1–8.
Damoiseaux, J.H., Teunissen van Manen, T.C., 1984. Bodemkaart van Nederland, 1:50,000 51 West EINDHOVEN. Kanters, H.L., (Ed.). Bodemkaart van Nederland 1:50,000. Wageningen, the Netherlands: Stichting voor Bodemkartering.
de Lange, W.J., Prinsen, G.F., Hoogewoud, J.C., Veldhuizen, A.A., Verkaik, J., Oude Essink, G.H., van Walsum, P.E.V., Delsman, J.R., Hunink, J.C., Massop, H.T.L., Kroon, T., 2014. The Netherlands Hydrological Instrument: an operational, multi-scale, multi-model system for consensus-based, integrated water management and policy analysis. Environ. Modell. Software 59, 98–108.
de Vries, F., 1999. Karakterisering van Nederlandse Gronden naar Fysisch-Chemische Kenmerken. DLO-Staring Centrum, Wageningen, the Netherlands.
Diekmann, M., 2002. Species indicator values as an important tool in applied plant ecology – a review. Basic Appl. Ecol. 4 (6), 493–506.
Doughty, C.E., Asner, G.P., Martin, R.E., 2010. Predicting tropical plant physiology from leaf and canopy spectroscopy. Oecologia 165 (2), 289–299.
Ecker, K., Waser, L.T., Küchler, M., 2010. Contribution of multi-source remote sensing data to predictive mapping of plant-indicator gradients within Swiss mire habitats. Bot. Helvet. 120 (1), 29–42.
Ellenberg, H., 1992. Zeigerwerte der Gefäßpflanzen (ohne Rubus). Zeigerwertevon Pflanzen in Mitteleuropa 18, 9–166.
Feilhauer, H., Asner, G.P., Martin, R.E., Schmidtlein, S., 2010. Brightness-normalized partial least squares regression for hyperspectral data. J. Quant. Spectrosc. Radiat. Transfer 111 (12), 1947–1957.
Feilhauer, H., Faude, U., Schmidtlein, S., 2011. Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. Remote Sens. Environ. 115 (10), 2513–2524.
Feilhauer, H., Schmidtlein, S., 2011. On variable relations between vegetation patterns and canopy reflectance. Ecol. Inf. 6 (2), 83–92.
Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8 (9), 993–1009.
Hannerz, M., Hånell, B., 1997. Effects on the flora in Norway spruce forests following clearcutting and shelterwood cutting. For. Ecol. Manage. 90 (1), 29–49.
Harbaugh, A., Banta, E.R., Hill, M.C., McDonald, M.G., 2000. MODFLOW 2000. In: The U.S. Geological Survey Modular Ground Water Model User Guide to Modularization Concepts and the Ground Water Flow Process. US Geological Survey, Reston, Virginia.
Kaiser, T., Wehrhan, M., Werner, A., Sommer, M., 2012. Regionalizing ecological moisture levels and groundwater levels in grassland areas using thermal remote sensing. Grassland Sci. 58 (1), 42–52.
Knapp, A.K., Beier, C., Briske, D.D., Classen, A.T., Luo, Y., Reichstein, M., Smith, M.D., Smith, S.D., Bell, J.E., Fay, P.A., Heisler, J.L., Leavitt, S.W., Sherry, R., Smith, B., Weng, E., 2008. Consequences of more extreme precipitation regimes for terrestrial ecosystems. Bioscience 58 (9), 811–821.
Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. J. Stat. Software 18 (2), 1–24.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—a discussion of principles. J. Hydrol.
Olde Venterink, H., Wassen, M.J., 1997. A comparison of six models predicting vegetation response to hydrological habitat change. Ecol. Modell. 101 (2), 347–361.
Ordoñez, J.C., van Bodegom, P.M., Witte, J.P.M., Wright, I.J., Reich, P.B., Aerts, R., 2009. A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. Global Ecol. Biogeogr. 18 (2), 137–149.
Panciera, R., Walker, J.P., Kalma, J.D., Kim, E.J., Saleh, K., Wigneron, J.-P., 2009. Evaluation of the SMOS L-MEB passive microwave soil moisture retrieval algorithm. Remote Sens. Environ. 113 (2), 435–444.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R foundation for statistical computing, Vienna, Austria.
Reich, P.B., Wright, I.J., Lusk, C., 2007. Predicting leaf physiology from simple plant and climate attributes: a global GLOPNET analysis. Ecol. Appl. 17 (7), 1982–1988.
Roelofsen, H.D., Kooistra, L., van Bodegom, P.M., Verrelst, J., Krol, J., Witte, J.P.M., 2014. Mapping a priori defined plant associations using remotely sensed vegetation characteristics. Remote Sens. Environ. 140, 639–651.
Roelofsen, H.D., van Bodegom, P.M., Kooistra, L., Witte, J.P.M., 2013. Trait estimation in herbaceous plant assemblages from in situ canopy spectra. Remote Sens. 5 (12), 6323–6345.
Runhaar, J., Jalink, M.H., Hunneman, H., Witte, J.P.M., Hennekens, S.M., 2009. Ecologische vereisten habitattypen. KWR Watercycle Research Institute, Nieuwegein, NL.
Runhaar, J., Witte, J.P.M., Verburg, P., 1997. Ground-water level, moisture supply, and vegetation in the Netherlands. Wetlands 17 (4), 528–538.
Schmidtlein, S., 2005. Imaging spectroscopy as a tool for mapping Ellenberg indicator values. J. Appl. Ecol. 42 (5), 966–974.
Schmidtlein, S., Feilhauer, H., Bruelheide, H., 2011. Mapping plant strategy types using remote sensing. J. Veg. Sci. 23 (3), 395–405.
Soudzilovskaia, N.A., Elumeeva, T.G., Onipchenko, V.G., Shidakov, I.I., Salpagarova, F.S., Khubiev, A., Tekeev, D.K., Cornelissen, J.H.C., 2013. Functional traits predict relationship between plant abundance dynamic and long-term climate warming. Proc. Natl. Acad. Sci. U. S. A 110 (45), 18180–18184.
Steur, G.G.L., Heijink, W., 1991. Bodemkaart Van Nederland, Schaal 1:50,000 Algemene Begrippen En Indelingen, 4th ed. Staring Centrum, Wageningen, the Netherlands.
Thuiller, W., Araújo, M., Lavorel, S., 2004. Do we need land-cover data to model species distributions in Europe. J. Biogeogr. 31 (3), 353–361.
Tüxen, R., 1954. Pflanzengesellschaften und Grundwasser-Ganglinien. Angew. Pflanzensoziol. 8, 64–97.
van Bodegom, P.M., Douma, J.C., Verheijen, L.M., 2013. A fully traits-based approach to modeling global vegetation distribution. Proc. Natl. Acad. Sci. U. S. A. 111 (38), 13733–13738.
van der Sande, C., Soudarissanane, S., Khoshelham, K., 2010. Assessment of relative accuracy of AHN-2 laser scanning data using planar features. Sensors 10 (9), 8198–8214.
van Walsum, P.E.V., Veldhuizen, A.A., 2011. Integration of models using shared state variables: Implementation in the regional hydrologic modelling system SIMGRO. J. Hydrol. 409 (1–2), 363–370.
Vermulst, J.A.P.H., Kroon, T., de Lange, W.J., 1998. Modelling the hydrology of the Netherlands on a nation wide scale. In: Wheater, H., Kirby, C. (Eds.), Hydrology in a Changing Environment. John Wiley and Sons Ltd., p. 710.
Von Asmuth, J.R., Maas, K., Knotters, M., Bierkens, M.F.P., Bakker, M., Olsthoorn, T.N., Cirkel, D.G., Leunk, I., Schaars, F., Von Asmuth, D.C., 2012. Software for hydrogeologic time series analysis, interfacing data with physical insight. Environ. Modell. Software 38, 178–190.
Wamelink, G.W.W., Joosten, V., van Dobben, H.F., Berendse, F., 2002. Validity of Ellenberg indicator values judged from physico-chemical field measurements. J. Veg. Sci. 13 (2), 269–278.
Wigneron, J.-P., Kerr, Y.H., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., De Rosnay, P., Gurney, R., Calvet, J.-C., 2007. L-band Microwave Emission of the Biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields. Remote Sens. Environ. 107 (4), 639–655.
Witte, J.P.M., Bartholomeus, R.P., Cirkel, D.G., Doomernik, E., Fujita, Y., Runhaar, J., 2014. Manual and description of ESTAR, version 01; A software tool to analyse vegetation plots. KWR Watercycle Research Institute Nieuwegein, NL.
Witte, J.P.M., Wójcik, R.B., Torfs, P.J.J.F., De Haan, M.W.H., Hennekens, S.M., 2007. Bayesian classification of vegetation types with Gaussian mixture density fitting to indicator values. J. Veg. Sci. 18 (4), 605–612.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58 (2), 109–130.