



# Large-scale regionalization of water table depth in peatlands optimized for greenhouse gas emission upscaling

M. Bechtold<sup>1</sup>, B. Tiemeyer<sup>1</sup>, A. Laggner<sup>1</sup>, T. Leppelt<sup>1</sup>, E. Frahm<sup>1,\*</sup>, and S. Belting<sup>1,2</sup>

<sup>1</sup>Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany

<sup>2</sup>Belting Umweltplanung, Quernheim, Germany

\* now at: Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

Correspondence to: M. Bechtold (michel.bechtold@ti.bund.de)

Received: 18 March 2014 – Published in Hydrol. Earth Syst. Sci. Discuss.: 7 April 2014

Revised: 2 July 2014 – Accepted: 20 July 2014 – Published: 1 September 2014

**Abstract.** Fluxes of the three main greenhouse gases (GHG) CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O from peat and other soils with high organic carbon contents are strongly controlled by water table depth. Information about the spatial distribution of water level is thus a crucial input parameter when upscaling GHG emissions to large scales. Here, we investigate the potential of statistical modeling for the regionalization of water levels in organic soils when data covers only a small fraction of the peatlands of the final map. Our study area is Germany. Phreatic water level data from 53 peatlands in Germany were compiled in a new data set comprising 1094 dip wells and 7155 years of data. For each dip well, numerous possible predictor variables were determined using nationally available data sources, which included information about land cover, ditch network, protected areas, topography, peatland characteristics and climatic boundary conditions. We applied boosted regression trees to identify dependencies between predictor variables and dip-well-specific long-term annual mean water level (WL) as well as a transformed form (WL<sub>t</sub>). The latter was obtained by assuming a hypothetical GHG transfer function and is linearly related to GHG emissions. Our results demonstrate that model calibration on WL<sub>t</sub> is superior. It increases the explained variance of the water level in the sensitive range for GHG emissions and avoids model bias in subsequent GHG upscaling. The final model explained 45 % of WL<sub>t</sub> variance and was built on nine predictor variables that are based on information about land cover, peatland characteristics, drainage network, topography and climatic boundary conditions. Their individual effects on WL<sub>t</sub> and the observed parameter interactions provide insight into natural and anthropogenic boundary conditions that control

water levels in organic soils. Our study also demonstrates that a large fraction of the observed WL<sub>t</sub> variance cannot be explained by nationally available predictor variables and that predictors with stronger WL<sub>t</sub> indication, relying, for example, on detailed water management maps and remote sensing products, are needed to substantially improve model predictive performance.

## 1 Introduction

Greenhouse gas (GHG) emissions from organic soils can be high compared to mineral soils. In Germany, the fraction of organic soils classified as peatland covers only 5 % of the land surface, but does account for 40 % of GHG emissions in the reporting categories “agriculture” and “land use, land use change and forestry” of the UN Framework Convention on Climate Change (UNFCCC) (UBA, 2012). Also, other organic soils with a lower soil organic carbon content (SOC) but still meeting the definition of organic soils according to IPCC (2006) are important sources of persistently high GHG emissions (Leiber-Sauheitl et al., 2014). In our study, we also consider these soils. For simplification, we will refer in the following to the total of peatlands and “other organic soils” as organic soils. Current estimates of GHG emissions from organic soils are fairly uncertain and reporting of most countries relies on IPCC default emission factors (EF) for CO<sub>2</sub> emissions which are stratified for land use and climatic region, e.g., 10 t C ha<sup>-1</sup> year<sup>-1</sup> for arable land in the warm temperate zone.

Artificial drainage turns the function of former natural peatlands from a C sink into a C source. Experimental work with organic soils during the last 2 decades showed that the aerated soil pore space above the water level is one of the key variables explaining the amount of CO<sub>2</sub> emissions (Moore and Dalva, 1993). Frequently, the water level relative to soil surface (further simply referred to as “water level”, with negative values below ground) is used as proxy for air-filled porosity, given the simplicity and availability of water level measurements. Additionally, low water levels and oxygen availability are also key drivers of nitrous oxide (N<sub>2</sub>O) production in organic soils (Regina et al., 1996), which increases the relevance of organic soils for climate change mitigation policy. During anaerobic conditions when water levels are at or above the land surface, substantial methane (CH<sub>4</sub>) emissions can occur (Levy et al., 2012).

It is postulated that the GHG budget – the sum of the CO<sub>2</sub>-equivalents of the three main greenhouse gases (CO<sub>2</sub>, N<sub>2</sub>O, CH<sub>4</sub>) – is at a minimum for annual mean water levels (annual mean further defined by the variable name WL) at about –0.05 to –0.1 m (Drösler et al., 2011). Following atmospheric sign convention, a positive sign stands for net emissions, while a negative sign indicates a net uptake of GHGs. Other parameters, such as physical and chemical soil properties and vegetation, also influence the amount of the emissions and thus weaken the relation between total GHG budget and WL.

If available, information about the spatial distribution of WL can identify GHG hot spot regions and improve the accuracy of the total GHG budgets at large scales. The application of transfer functions that relate GHG emissions to WL and potential other influencing site characteristics can refine the estimates derived from simple application of IPCC default EFs. However, in many countries and regions, as for example Germany and Europe, a map of WL in organic soils does not exist. The spatial availability of measured WL is much higher than that of measured GHG fluxes, which suggests the use of WL as scaling parameter for upscaling GHG emissions.

Several methods were applied in the past to produce WL maps. Their suitability is strongly related to data availability, which very often decreases in quality and spatial density with increasing scale of the study area. Spatially distributed process-based modeling (Thompson et al., 2009) and semi-physical statistical approaches (Bierkens and Stroet, 2007) are able to reproduce well the water level dynamics in wetlands environments including peatlands. However, they heavily rely on spatial information about the system’s physical properties and boundary conditions (peat hydraulic properties, hydraulic conductivity of peat base, drainage system), data that is often only available with sufficient detail at a regional scale (Limpens et al., 2008). Despite this difficulty there are studies in which process-based models were applied to model peatland water levels at a large scale (national or continental). Gong et al. (2012) adopted a common

soil–vegetation–atmosphere transfer model to account for the differing hydrological processes in pristine fens, pristine bogs and drained peatlands, and modeled water level fluctuations in boreal peatlands for all Finland. But calibration and validation with data from only three mires does not allow for conclusions about the accuracy and general applicability of the model. Numerous large-scale hydrological wetland models are often developed with a focus on delineating wetland extent (Melton et al., 2013). TOPMODEL-based schemes (Ju et al., 2006) and more advanced large-scale hydrologic frameworks (Fan and Miguez-Macho, 2011) are suited to model WL but do not account for anthropogenic drainage and thus are only applicable to pristine (or nearly pristine) peatland systems.

When detailed physical model input that is needed for a physically based approach is lacking, statistical or machine-learning tools represent a promising alternative (Finke et al., 2004). Potential predictor variables that are available at the final map scale are determined for each location with water level data and the algorithm identifies dependencies between potential predictors and target variables, such as WL or other statistical values that describe water level dynamics. For areas rich in water level data, e.g., the Netherlands, residuals of the statistical model can afterwards be analyzed for spatial correlation. If this is present, it can be used to correct for spatially correlated model bias by kriging. This scheme has been applied to agricultural areas by Finke et al. (2004) and to nature conservation areas by Hoogland et al. (2010). Spatial interpolation approaches can include ancillary data such as mapped geophysical parameters (Buchanan and Triantafyllis, 2009). Statistical approaches strongly rely on both the quantity and quality of the data on the target variable itself, i.e., the water level data. An important quality criterion for water level data from organic soils is the measurement depth. It is crucial that there is little or no hydraulic resistance by a low conductive layer between the perforated part of the monitoring well and the fluctuating water level. If hydraulic resistance is too high, the monitoring well acts as a piezometer and water levels may substantially differ from the actual phreatic level as shown for peatlands by van der Gaast et al. (2009). If such piezometer data is part of a data set and interpreted as phreatic water level data during model calibration, this can lead to an under- or overestimation of predicted water levels in organic soils. An underestimation of water level predictions (too dry) is discussed for Dutch modeling studies in van der Gaast et al. (2009).

At present in Germany, a map of water levels in organic soils that could be used for GHG upscaling is lacking. This fact and current efforts on improving GHG emission estimates for German organic soils were the main drivers for our study. Thus, the major goal of this study was the development of a model concept that produces a water level map at the scale of all organic soils in Germany that is specifically optimized for water level ranges to which GHG emissions react sensitively. We emphasize that the objective of our study

was to regionalize annual mean water levels and not the GHG emissions themselves. The latter are influenced by more site characteristics, in particular soil properties. Furthermore, we suppose that annual mean water level is probably not the only or optimal statistical measure to describe the water level effect on annual GHG emissions. However, we are not aware of well-established information about transfer functions that relate more complex statistical measures of water level dynamics to GHG emissions. Therefore, we here focused on the simple and frequently applied “annual mean water level”.

As a first step, we compiled a new data set of phreatic water level time series of organic soils with contributions from numerous data sources. Based on this data, we developed a modeling approach for the annual mean water level that follows the basic idea of the statistical regionalization presented in Finke et al. (2004). However, the data coverage in our study substantially differed from their study. Our data covers only a small fraction of the peatlands of the final map and spatial interpolation of residuals was not possible. We thus extended their approach by:

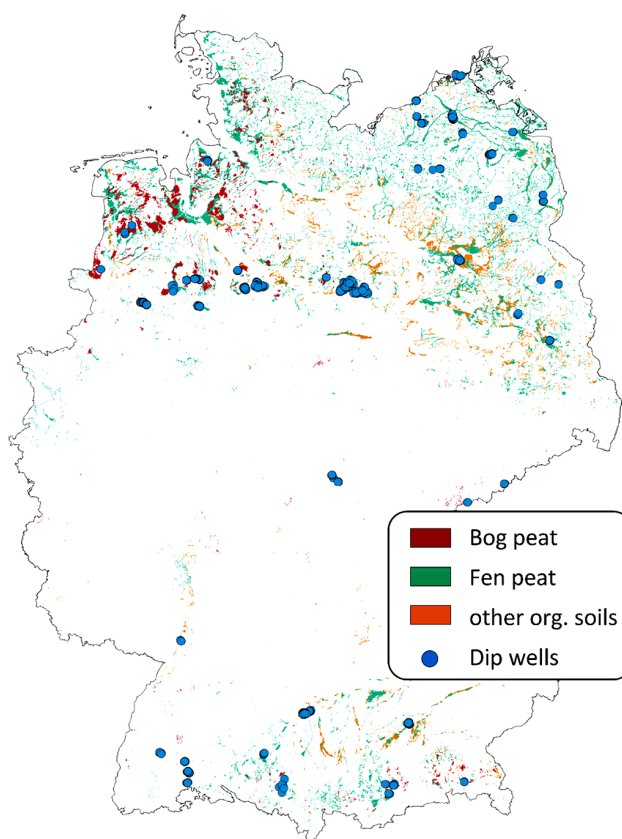
- including additional possible predictor variables,
- using boosted regression trees as a modeling tool to identify the influence of both numerical and categorical variables simultaneously,
- applying a new weighting scheme that balances out heterogeneous water level data sets with highly variable spatial data density,
- transforming the annual mean water level,  $WL$ , into a transformed annual mean water level,  $WL_t$ , that shows a linear relationship with the GHG budget and optimizes model calibration for the  $WL$  range relevant for GHG emissions and
- restricting the water level regionalization to phreatic water levels of organic soils.

We present a detailed analysis of the influence of the individual predictor variables on water levels of organic soils as well as their interactions. Furthermore, the manuscript includes the estimation of model uncertainty and possible paths of future model improvement. Finally, the calibrated model is used to derive a map of  $WL_t$  for all organic soils in Germany, and the regionalization results are presented.

## 2 Data set and methods

### 2.1 Data set of phreatic water levels in organic soils

Available data of phreatic water levels in organic soils are scarce. In contrast to data of rather deeply drilled observation wells of official groundwater monitoring networks, short peatland observation wells of only 1 or 2 meters length that



**Figure 1.** Locations of the 1094 dip wells of the data set. Base map (geological map 1 : 200 000, BGR) shows the distribution of bog and fen peat, and other organic soils.

measure the phreatic water level of the peat layer are currently not collected in central data management systems in Germany or any of its federal states. With a comprehensive questionnaire started in 2011, we collected water level time series of organic soils from local agencies, non-governmental organizations, universities, consultants and other sources, and combined this data with water level data from our projects. Time series included manual and automatic measurements. Years with less than six measurements or data gaps of more than 3 months were excluded. Water level time series of each dip well were visually checked on plausible dynamics by comparing with data from neighboring dip wells and weather data time series. Based on auxiliary data and local knowledge, we further identified dip wells that reached down to the underlying aquifer. If dip wells failed these quality checks, they were removed from the data set.

The final data set comprised 7155 years of data from 53 German peatlands and 1094 dip wells. On average time series ranged over 7 years. All time series were collected at some period between the years 1988 to 2012. Data are well distributed over most of the German peatland regions and cover the three major types of organic soils (Fig. 1). Compared with the distribution of the types of organic soils in

Germany, the fraction of dip wells on bogs is overrepresented in the data set by the factor of 2.5, while dip wells on fens and other organic soils are slightly underrepresented. Data also cover the common land-use types (for data sources see Table 1). However, dip wells on organic soils that are neither used for agriculture, forestry or peat mining, further referred to as “unused peatlands”, are overrepresented in the data set by a factor of 6 as data was collected more frequently and in higher spatial data density in the frame of conservation projects. The fraction of unused peatlands of the German organic soils is 6 % and the fraction in the data set is 36 %. In contrast, dip wells on arable land are underrepresented in the data set by a factor of 6. The fraction of arable land on German organic soils is 24 %, and the fraction in the data set is 4 %. The other two key land-use types of organic soils in Germany, grassland and forest, are well represented in the data set. The misbalance of the land-use types in the data set is accounted for in the weighting of data (see Sect. 2.3.2).

If land use changed within the measurement period of a dip well, the time series was split at the moment when the land-use record indicates the transition. For each segment the annual mean water level, WL (here with negative values defined as water levels below ground), was calculated as the multi-year average value over the whole measurement period of the specific land use.

The primary application of the WL map produced in this study is for the upscaling of long-term GHG emissions as emission reporting may only reflect anthropogenic effects, but not interannual climatic effects. As GHG transfer functions are developed on annual data, their application requires both the long-term annual mean water level, as well as its interannual variability. Due to the non-linear dependence of GHG emissions on WL, single years with extreme water levels can strongly influence long-term average GHG fluxes. This study is focused on the regionalization of the long-term annual mean water levels. For this objective, model building should be based on long-term water level time series to average out the effect of weather variation within a complete climatic period (commonly 30 years). The existing nationally available data on water level time series of organic soils, however, does not comprise a single time series with complete data coverage over the last 30 years. Due to the lack of sufficient long-term water level time series, we included all time series in the model building process. Average climatic boundary conditions (precipitation, reference evapotranspiration, water balance) of the specific measurement period of each dip well are part of the predictor variables (see Sect. 2.2), and thus are supposed to partly account for the effect of specific weather conditions on WL in case of short measurement periods.

## 2.2 Predictor variables

Spatial coverage of phreatic water level data of organic soils is too low to obtain WL maps by simple spatial interpolation

(Fig. 1). Additional spatial data is needed as basis for regionalization. Ancillary information that covers fully or at least most of the extent of the final map are necessary. They can be used as predictor variables. A comprehensive set of variables (numerical and categorical) with potential indication for the hydrological condition of an organic soil were determined for each dip well (Fig. 2 and Table 1).

The predictor variables, which can partly be found also in Finke et al. (2004), can be divided into seven groups:

### 2.2.1 Land cover

As certain land use and vegetation require and reflect certain WL, such information can be used as an indicator for the average drainage level around the dip well. Land-use and vegetation information is based on the German Digital Landscape Model (ATKIS Basis-DLM), which is updated continuously by aerial photos as well as sporadic ground mapping and has a temporal accuracy of 3 months to 5 years. It is provided as fine-scaled polygons and represents the best uniform land cover information available in Germany. It contains information on primary land-use type, few optional vegetation attributes and whether “wet soil” has been observed during mapping. As we noticed that the use of a large number of categorical variables lowers the performance of boosted regression trees, we further aggregated the three information types (i) land use, (ii) vegetation and (iii) wet soil into a set of nine combined land cover classes (Table 1). These land cover classes were a trade-off between fine differentiation and the number of replicates in each class. For grasslands, a “wet grassland” class was separated when grassland was overlaid with wet soil and/or tree or shrubs vegetation, which may indicate a less intensive management. Forests overlaid with wet soil were separated as “wet forest”. Further, unused peatlands overlaid with wet soil and showing no coverage with tree attributes were characterized by higher water levels and were thus separated as “wet unused peatland”. The very few dip wells classified as open water ( $n = 2$ ) and peat cutting ( $n = 5$ ) were merged to the reed and arable land cover class, respectively. Land-use type and land cover class were extracted at the dip well (point extraction) and as fractions in various buffers around the dip well (Table 1). As using too many weak predictor variables lowers model performance and increases overfitting, the numerous land cover fractions were further aggregated into two classes: the fraction of dry (arable and grassland) and wet (reed, wet grassland, wet forest and wet unused peatland) land cover on organic soils. For the calculation of the fraction of dry land cover, we tested various factors for the reduction of the contribution of grassland compared to arable land as the grassland class also includes wetter grasslands that could not be detected with the available land cover catalogue. A factor of 0.5 was an optimal value, which was then set fixed.

### 2.2.2 Drainage network

Locations of ditches that are included as lines in the Digital Landscape Model were used to obtain information about the drainage network. The total ditch length was calculated for various buffer sizes. Further, the distance to the next ditch was calculated for each dip well. A short distance to the next ditch may indicate either lower or higher water levels, depending on whether the ditches are used for drainage or already blocked and used for rewetting measures. Similarly, the indication of the total length of ditches is not unique. Therefore, we defined two different sets of ditch variables. A first set, for which we calculated values for all land cover classes and a second one, for which we only calculated values for land cover classes for which ditches are undoubtedly used for drainage, i.e., arable and grassland.

### 2.2.3 Peatland characteristics

The geological map of Germany (scale 1 : 200 000) defined the area for which WL predictions were modeled. It is also the basis for topological peatland predictor variables, i.e., the fraction of organic soils in different buffer sizes as well as the dip well distance to the edge of the peatland. Information about the peatland type and the substrate at the peat base is presented in more detail in a newly compiled raster map of organic soils (Roßkopf et al., 2014) and was thus extracted from this map. Peatland types were aggregated into five classes: lowland bog (North German Plains and Alpine Forelands), upland bog (Central Uplands and Alps), fen neighboring surface water, fen without neighboring surface water, and a class of “other organic soils” that do not fulfill the C content and thickness criteria to be classified as peatland. Substrates at the peat base included loose unconsolidated rock (alluvial sand and gravel deposits), consolidated rock (bedrock) and peat clay layer. The first type may indicate the occurrence of seepage (positive or negative), whereas the latter two types may indicate rather a hydraulic decoupling from the aquifer hydraulic head.

### 2.2.4 Climatic boundary conditions

Climatic boundary conditions directly influence water level. On the one hand, the typical long-term climatic boundary conditions may indicate the general vulnerability of peatlands in a specific region. On the other hand, given the different lengths of measurement periods of the time series in this study, climatic boundary condition predictor variables may account for the effect of a climatically wetter or drier measurement period, compared to the long-term averages, on the water level. Climatic boundary conditions were extracted from a 1 × 1 km raster from the German Weather Service. Annual, summer and winter precipitation, FAO56 Penman–Monteith reference evapotranspiration and climatic water balance (difference between precipitation and

reference evapotranspiration) were determined for the individual measurement period of each dip well and as long-term averages (30 years).

### 2.2.5 Relative altitude

Relative altitude was calculated by subtracting the median altitude of various buffer sizes from the absolute altitude at each dip well in the digital elevation model (DEM). Relative altitude is expected to have two different indications depending on the applied buffer size: (i) in many peatlands, the former smooth peatland relief at the scale of approximately > 5 m has been disturbed due to peat cutting and differences in drainage and mineralization rate. As a consequence, the rather smooth phreatic surface often does not follow the uneven and patchy terrain. Relative altitude with respect to smaller buffer sizes (< 250 m) may therefore explain part of the WL variation, e.g., a dip well that is located at a surface much higher than the surroundings may indicate deeper water levels; (ii) for large buffer sizes (> 250 m) relative altitude indicates whether the peatland lies in a larger morphological depression or elevation, and thus may indicate whether large-scale lateral inflow of water can be expected or not. Similar indication is provided by the topographic index (see below). The accuracy of relative altitude values depends on the resolution and accuracy of the DEM. The nation-wide available DEM is based on data sets of varying quality, which may lower the influence of this variable.

### 2.2.6 Topographic wetness index

The topographic wetness index is a common wetness indicator used in hydrology (Beven and Kirby, 1979). It is a combined measure of catchment area and slope at a given point and indicates the extent of flow accumulation. High values indicate wetter conditions. If calculated at larger scales, higher values may hint at the occurrence of positive seepage, i.e., upward flow of water from the aquifer. Topographic wetness index was calculated for various DEM resolutions using the GRASS 7 module `r.watershed`.

### 2.2.7 Protection status

The protection status of a peatland area may reflect hydrological conditions. Therefore we checked for seven protection status at each dip well (see Table 1 for details).

## 2.3 Model building scheme

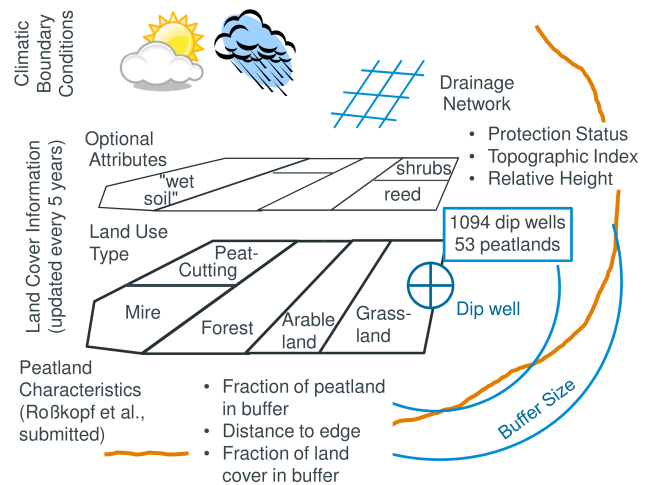
Model building was performed using boosted regression trees (BRT), implemented in the two R packages “`gbm`” (Ridgeway, 2013) and “`dismo`” (Hijmans, 2013). BRT is a machine-learning algorithm, in which the final model is derived from the data. Functions that relate target to predictor variables are not predetermined but freely developed. BRT is based on the decision (or regression) tree concept. In the



decision tree concept, the parameter space is searched sequentially for the best split that results in the lowest model mean squared error. The mean responses of the groups that result from the various splits, and correspond to certain parameter ranges, represent the model. The common procedure is the growth of a large tree which is subsequently simplified by dropping weak links that are identified with cross-validation. Growing only a single tree has several disadvantages such as uneven functions that are very sensitive to the specific sample of the data. Therefore, ensemble techniques have been combined with the decision tree concept. These were, first, the development of multiple models by bootstrapping of the samples (bagging technique) and the random creation of subsets of predictors at each split (random forest technique). Later, with the “boosting” technique of BRT, a sequential procedure was developed in which data is re-weighted after each tree to increase emphasis on data that is poorly modeled by the existing collection of trees (Elith et al., 2008).

BRT modeling is increasingly applied in spatial modeling of species or numerical environmental variables (Elith et al., 2008, Martin et al., 2011), thereby often showing superior performance compared to other machine-learning algorithms. The increasing application of BRT is related to several of its favorable characteristics: the strength of this method lies in the ability to fit complex functional dependencies including non-linear relationships and interactions between predictor variables. Based on its flexibility, BRT is invariant to monotonic transformations of predictors. Furthermore, BRT allows for missing values in the predictor variables, thus predictor variable information does not necessarily need to fully cover the total map extent. The *gbm* package handles missing values in predictor variables by introducing surrogate splits. The mean target value belonging to the missing predictor values is attributed to these surrogate splits during model building. We observed that the contribution of a predictor variable to the final model decreases with an increasing number of missing values. This is intuitive as target observations of missing predictor values are mostly supposed to scatter strongly. BRT is further fairly insensitive to outliers and allows estimating the relative contribution of each predictor variable to the model. Due to these characteristics we expected BRT to be very well suited to the very heterogeneous data set of this study.

BRT model calibration is prone to overfitting, and there are various options to reduce this behavior. Due to the overfitting behavior, cross validation is generally part of the model building process. However, cross validation can be performed in several ways and, if performed carelessly, can lead to overly optimistic model performance (De'ath, 2007). Here, cross validation was performed by leaving out whole peatland areas instead of a random set of dip wells. This represents a stricter cross validation, and we noticed that it strongly reduced overfitting of the water level data, and thus contributed to the development of a more robust model.



**Figure 2.** Illustration of the predictor variables determined for each dip well based on available national maps (see Table 1).

Another option to avoid overfitting is to impose monotonic slopes on the effects of individual parameters, which can even lead to improved prediction performance (De'ath, 2007). For all our numerical variables, we expected monotonic slopes rather than optimum functions. To avoid pre-defining any expected direction, all numerical variables were added twice to the set of predictors, constraining the slope to a monotonic increase and decrease. We let the model decide whether monotonic increase or decrease has higher predictive power.

Models were calibrated using a Gaussian response type, aimed at minimizing deviance (squared error) (Ridgeway, 2013). In all calibration runs, we applied the *gbm.step* function of the *dismo* package, which assesses the optimal number of boosting trees using cross validation. We tested various learning rates (0.001–0.01), bag fractions (0.1–0.8) and levels of tree complexity (3 to 7), i.e., the number of nodes in a tree. By trial and error we determined the most effective algorithm parameters for our data set being 0.005 for the learning rate, 0.6 for the bag fraction and 5 for the tree complexity.

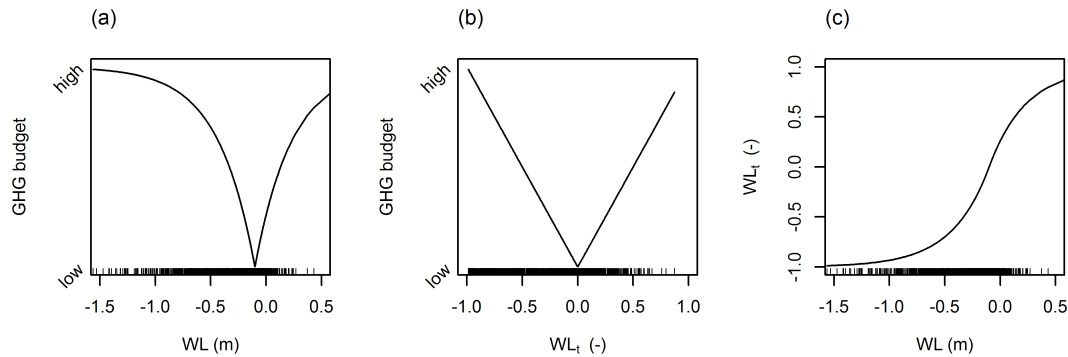
The final BRT model building is commonly performed as a two-step procedure (Elith et al., 2008) which we basically also followed in our study:

- i. In the first step, the whole set of predictor variables is used to calibrate a BRT model.
- ii. In a second step, the number of parameters is reduced sequentially to avoid overfitting and to derive a more parsimonious model. We tracked predictive performance criteria during the simplification process. As various variables were calculated for different buffer sizes, our predictors included a large number of correlated variables. Correlation coefficients between predictor variables of  $> 0.7$  are known to severely distort model estimation and subsequent prediction (Dormann

**Table 1.** Overview on predictor variables.

Predictor variable	Variable name	Values	Point/buffers (m)	Data source
Land-use type		Arable, grassland, forest, shrubs, peat-mining, unused peatland, swamp, open water	Point, 100, 500, 1000, 2500	Digital Landscape Model <sup>1</sup>
Vegetation attributes (optional)		Deciduous forest, mixed forest, coniferous forest, reed, shrubs, grass	Point	Digital Landscape Model <sup>1</sup>
“Wet soil observed”		Yes, no	Point	Digital Landscape Model <sup>1</sup>
Combined land cover information (land-use type, veg. and wet soil attr.)	lc	Arable, grassland, wet grassland, deciduous including mixed forest, wet forest, coniferous forest, reed, unused peatland, wet unused peatland	Point, 100, 500, 1000, 2500	Digital Landscape Model <sup>1</sup>
Dry land cover fraction	$f_{dry}(X)$	Arable, $0.5 \times$ grassland on organic soil area; 0 to 1	100, 500, 1000, 2500	Digital Landscape Model <sup>1</sup>
Wet land cover fraction		Reed, wet grassland, wet forest and wet unused peatland on organic soil area; 0 to 1	100, 500, 1000, 2500	Digital Landscape Model <sup>1</sup>
Total length of ditches for all lc and only for arable and grassland (subscr.: “dry”)	$d_{len,dry}(X)$	$\geq 0$ m	Point, 50, 250, 1000, 2500	Digital Landscape Model <sup>1</sup>
Distance to next ditch		$\geq 0$ m	Point	Digital Landscape Model <sup>1</sup>
Peatland type	$P_{type}$	Lowland bog, upland bog, fen neighboring surface water, fen without neighboring surface water, other “low-C” organic soil	Point	Map of organic soils <sup>2</sup>
Material at peat base	$P_{base}$	Unconsolidated rock, peat clay layer, rock, no information	Point	Map of organic soils <sup>2</sup>
Peatland fraction	$f_{peat}(X)$	0 to 1	Point, 500, 1000, 2500	Geological map (BGR) <sup>3</sup>
Distance to edge of peatland		$> 0$ m		Geological map (BGR) <sup>3</sup>
Ratio of $d_{peat}/f_{peat}$		$> 0$	2500	Geological map (BGR) <sup>3</sup>
Precipitation		$\geq 0$ mm	Point	Raster map $1 \times 1$ km (DWD) <sup>4</sup>
Evapotranspiration		$\geq 0$ mm	Point	Raster map $1 \times 1$ km (DWD) <sup>4</sup>
Climatic water balance	$wb_{summer}$	$< 0$ and $\geq 0$ mm	Point	Raster map $1 \times 1$ km (DWD) <sup>4</sup>
Relative height	$h_{rel}(X)$	$< 0$ and $\geq 0$ m	Point – median 25, 50, 100, 250, 500, 1000	Digital Elevation Model <sup>5</sup>
Topographic index	$t_{irasR}(X)$	$> 0$	Point and 1000 buffer for 10, 25, 250, 1000 raster values	Digital Elevation Model <sup>5</sup>
Protection status		Nature conservation area, special areas of conservation, special protection area for wild birds, UNESCO biosphere reserve, nature park, national park, landscape protection area	Point	Maps of protected areas <sup>6</sup>

<sup>1</sup> ATKIS Basis DLM, Federal Agency for Cartography and Geodesy; BKG; <sup>2</sup> map of organic soils (Rohkopf et al., 2014, Humboldt University of Berlin); <sup>3</sup> geological map 1:200000 (GUEK 200, BGR – Federal Institute for Geosciences and Natural Resources); <sup>4</sup> raster map  $1 \times 1$  km of weather data (German Weather Service); <sup>5</sup> BKG; variable name indicated for the nine variables in the final model with (X) indicating buf. size and R indicating raster resolution; <sup>6</sup> Federal Agency for Nature Conservation (BfN).



**Figure 3.** Illustration of the annual mean water level (WL) transformation. Hypothetical transfer function relating GHG budget to WL (m) (a). GHG budget vs. the transformed water level ( $WL_t$ ) (b).  $WL_t$  vs. WL. The lines along the  $x$  axes indicate the data quantiles of the analyzed data set (c).

et al., 2013). Thus, we performed this simplification process by first dropping those parameters with a correlation  $> 0.7$  (either Pearson or Spearman type) to another parameter with a higher contribution (Clapcott et al., 2011). This ensured that two highly correlated parameters would not remain in the parameter set longer than the last parameter of another group of variables, which may contribute less compared to the two highly correlated parameters but provides extra information that is not covered by the other parameters. After all highly correlated parameters have been dropped, further parameters with low contribution were dropped progressively.

Predictor contributions are calculated as proportional contributions to the total error reduction and can be considered as a measure for the influence of the individual predictors. Additionally, a BRT model allows the derivation of partial dependence plots which indicate how the response is affected by a certain predictor after accounting for the average effects of all other predictors in the model (Elith et al., 2008). These plots do not show the full effect of each parameter on the model response due to interactions with other parameters that are fixed to derive these plots as well as due to parameter correlation. However, they can be used for interpreting model behavior (Elith et al., 2008).

### 2.3.1 $WL_t$ : transformation of WL

The map of water levels of this study was developed to improve the upscaling of greenhouse gas emissions from organic soils. Therefore, the final map should provide the highest accuracy for the water level range for which the highest differences of greenhouse gas emissions occur. This can be achieved by transforming WL into a transformed variable  $WL_t$ , which shows a linear relationship with GHG emissions. The sensitivity of greenhouse gas emissions to water level has been analyzed in several laboratory and field experimental and monitoring studies (Berglund and Berglund,

2011; Drösler et al., 2011; Hahn-Schöfl et al., 2011; Leiber-Sauheitl et al., 2014; Moore and Roulet, 1993; Moore and Dalva, 1993; van den Akker et al., 2012). General trends are a strong increase of methane ( $CH_4$ ) emissions for annual mean water levels of approximately  $> -0.1$  m and an increase of  $CO_2$  emissions for water levels  $< -0.1$  m with a trend similar to a saturation function that levels out approximately between  $-0.4$  and  $-0.8$  m (Fig. 3a). While studies agree over these general trends, the exact shape of the transfer function and the maximum levels of emissions as well as their dependence on soil properties and other environmental parameters are still controversial. Here, we assume a hypothetical transfer function, relating the normalized GHG budget, ranging from 0 to 1, to the water level (see also Fig. 3),

$$\text{GHG Balance} = \begin{cases} -e^{3(WL+0.1)} + 1 & WL \leq -0.1 \\ 1 - e^{-3(WL+0.1)} & WL > -0.1 \end{cases}. \quad (1)$$

As the GHG budget can be positive for both low and high WL, we introduced the transformed water level,  $WL_t$ , as (Fig. 3),

$$WL_t = \begin{cases} e^{3(WL+0.1)} - 1 & WL \leq -0.1 \\ 1 - e^{-3(WL+0.1)} & WL > -0.1 \end{cases}. \quad (2)$$

By calibrating the model to both WL and  $WL_t$ , we test if the optimization of  $WL_t$  provides the highest model accuracy for the water level range relevant for GHG emissions and if it optimizes the map for application to GHG upscaling.

### 2.3.2 Weighting scheme

When considering possible data weighting schemes, it is worth emphasizing, at this point, that the goal of this study is the development of a statistical model that can explain both the water level variability within a peatland as well as among different peatlands. The data of target and predictor variables for building this model is highly heterogeneous. First, the target variable data set contains peatland areas that strongly differ in their spatial extent and in the number of installed dip



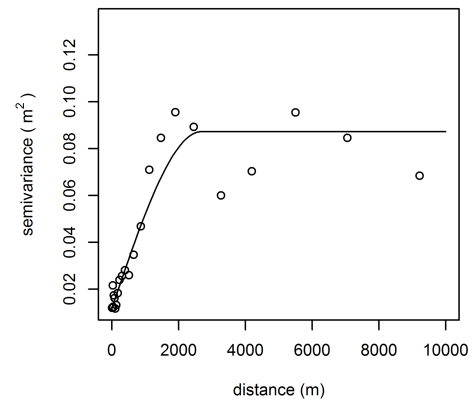
wells. Second, the predictor variable data set contains categorical and numerical data, and part of the predictor variables predominantly vary from peatland to peatland (e.g., climatic boundary conditions, large-scale topographic wetness index, peatland characteristics) whereas others also show within-peatland variability (e.g., land use, small-scale topographic wetness index, drainage network). As the influence of the individual predictor variables on our target  $WL_t$  is expected to be rather diffuse due to abundant interactions with other site characteristics, the robustness of derived dependencies will strongly depend on the number of different peatlands in the data set.

There are no universal data weighting rules for similarly heterogeneous data situations and some degree of expert judgment and subjectivity is inevitable involved when developing an appropriate scheme (Francis, 2011). The need for introducing a data weighting scheme is obvious, as without data weighting during calibration, too much influence would be given to small and well-studied peatlands, which will reduce predictive model performance for large, less-well-studied peatland areas. To avoid this in a simple manner, weight could be reduced by the number of dip wells in each peatland, which results in each peatland being equally weighted. This scheme, however, does not sufficiently use the high information content provided by well-studied large peatlands, which should have a higher impact on model calibration than a small peatland with only few dip wells.

Here, we propose a new weighting scheme that takes into account both factors, peatland size and local density of dip wells, to derive dip well specific weighting factors. It is based on principles of data uncertainty reduction, by repeated measurements, and of geostatistics. First, we consider our data situation as an analogue of meta-analysis with grouped data. It has been shown for homogeneous problems (all data from same population) that optimal group weights for meta-analysis is  $1/SE^2$  (Hedges and Olkin, 1985) with SE being the standard error of each group,

$$SE = \frac{\sigma_e}{\sqrt{N}}, \quad (3)$$

where  $\sigma_e$  is the error standard deviation of a measurement and  $N$  is the number of measurements in a group. For homogeneous problems and uniform  $\sigma_e$ , this results in weights that are linearly dependent on  $N$ , which we here call the first end member of weighting. Heterogeneity (within-group variance) reduces the variation of the group weights which can be shown by random effect models (Cumming, 2012). As with second end member of weighting, when heterogeneity totally dominates within-group variance, optimal group weights are uniform for all groups, i.e., weights are independent of  $N$ . We are not aware of a method that allows the estimation of the degree of heterogeneity for the complex target and predictor data situation in this study, including data (spatial and temporal variability, measurement error) and model errors (missing parameters). As a trade-off



**Figure 4.** Sample semivariogram and fitted semivariogram model of the annual mean water level data, WL.

between  $1/SE^2$  (homogeneous end member) and 1 (heterogeneous end member), we decided on a group weight that is the inverse of the standard error,  $1/SE$ , which is, for example, often used in econometric studies (Dickens, 1990). We emphasize that this is a subjective decision.

The group weight,  $1/SE$ , is the basis for the geostatistical part of our weighting scheme. There are two reasons why we cannot directly treat our peatlands as groups. First, there is within-peatland variability, which is related to changing site characteristics. It is one objective of our study to describe this variability by statistical modeling. Thus, dip wells must be treated individually and data cannot be aggregated at the peatland level. Second, we expect the model to learn more when the same number of dip wells is installed in a larger peatland. In a small peatland, spatial autocorrelation between dip wells is higher, i.e., the information content is lower than for large peatlands. As a consequence of the first point, we do not aggregate and keep all dip wells in the target variable data set by attributing to each dip well the fraction  $1/N$  of its group weight, so that the relative weights of the groups remain constant. As a consequence of the second point, we use principles of geostatistics in our weighting scheme. We replace the group size  $N$  (positive integer number) by the “statistical” group size  $n$  (positive continuous number being  $> 1$ ), which we derive from the spatial autocorrelation among the dip wells.

Therefore, we analyze the spatial autocorrelation structure of the data set. A single spherical variogram model was fitted to the sample variogram of all data (Fig. 4 in Sect. 3.1). Variogram models allow the differentiation of the total data variance (called “sill”) into a spatially uncorrelated variance (called “nugget”) and a spatially correlated variance (called “structural variance” and defined as sill–nugget) (Wackernagel, 2003). The variogram model allows for derivation for any distance between two locations the average squared difference of values, here defined as  $\gamma$ . By definition, at distance 0, the average squared difference equals the nugget, and at distances greater than that called the “range” of spatial

autocorrelation, the average squared difference equals the sill. Accordingly, the autocorrelated fraction,  $f$ , of the average squared difference between two dip wells  $i$  and  $j$  is

$$f_{i,j} = \frac{\text{sill} - \gamma_{i,j}}{\text{sill} - \text{nugget}}. \quad (4)$$

We now define the “statistical” group size  $n$  of each dip well  $i$  to be the sum of one plus the autocorrelated fractions  $f_{i,j}$  of all dip wells that are within the range of spatial autocorrelation of  $i$ ,

$$n_i = 1 + \sum_{j=1}^m \frac{\text{sill} - \gamma_{i,j}}{\text{sill} - \text{nugget}}. \quad (5)$$

According to the discussion above, dip-well-specific weights can then be calculated with

$$w_i = \frac{1}{n_i \text{SE}_i} = \frac{1}{\sigma_{e,i} \sqrt{n_i}}, \quad (6)$$

where  $n_i$  is derived from Eq. (5). The equation shows that with increasing “statistical” group size  $n$ , i.e., with increasing spatial data density, the weight of an individual dip well is “down-weighted” to some degree, a behavior that corresponds to our initial intention to lower the influence of small peatlands compared to large ones. The error standard deviation  $\sigma_e$  is dependent on several factors, e.g., the length of the time series, the temporal measurement density and the microtopography around the dip well. For simplicity, we here assumed  $\sigma_e$  to be uniform for all dip wells, which simplifies Eq. (6) to  $w_i = \frac{1}{\sqrt{n_i}}$ .

Only dip wells with the same land-use type were summed up with Eq. (5), which avoids the down-weighting by dip wells that have different land-use types. The latter are mostly characterized by fairly different  $\text{WL}_t$  and thus by rather low spatial autocorrelation to dip well  $i$ .

After spatial correlation has been accounted for, the sum of the weights of all dip wells of each land-use type were adjusted that they correspond to the fractions of this land-use type in Germany. This adjustment accounts for the overrepresentation in the data set of dip wells in unused peatlands and underrepresentation of dip wells in arable land.

### 2.3.3 Model performance criteria

Model fit and predictive performance after cross-validation were quantified by the weighted root mean square error,

$$\text{RMSE} = \sqrt{\frac{1}{\sum_{i=1}^m w_i} \sum_{i=1}^m (w_i (x_{o,i} - x_{s,i})^2)}, \quad (7)$$

where  $m$  is the number of dip wells,  $x_{o,i}$  is observed WL or  $\text{WL}_t$  of dip well  $i$ ,  $x_{s,i}$  is simulated WL or  $\text{WL}_t$  of dip well  $i$  and  $w_i$  is the data weight of dip well  $i$  (see below). We

refer to the root mean square error of the predicted data of cross validation as  $\text{RMSE}_{\text{cv}}$ . Model performance was further quantified by Nash–Sutcliffe efficiency (NSE),

$$\text{NSE} = 1 - \frac{\sum_{i=1}^m w_i (x_{o,i} - x_{s,i})^2}{\sum_{i=1}^m w_i (x_{o,i} - \bar{x}_o)^2}, \quad (8)$$

where  $\bar{x}_o$  is the mean of all observed WL or  $\text{WL}_t$ . It indicates how well observed vs. predicted values match the 1 : 1 line. NSE is a good overall indicator of predictive performance because it combines scatter and bias (common offset and/or slope difference from 1:1 line) (Nash and Sutcliffe, 1970). Values greater than 0 signify a model that is better than the reference model based on the data mean. We refer to the NSE of the training data as  $\text{NSE}_{\text{cal}}$ , and of the predicted data of cross validation as  $\text{NSE}_{\text{cv}}$ .

Systematic errors were quantified by calculating the model bias, here defined as,

$$\text{BIAS} = \sum_{i=1}^m (w_i x_{o,i} - w_i x_{s,i}). \quad (9)$$

### 2.4 Model uncertainty and stability evaluation

Uncertainty of the model predictions was assessed by bootstrapping, cross-validation and residual analysis.

For the bootstrapping analysis, we followed the procedure of Leathwick et al. (2006). We estimated the confidence intervals around the predictions and the fitted functions by taking 1000 bootstrap samples of the 53 peatlands. The number of peatlands in each sample was equivalent to the data set, but peatlands were selected randomly with replacement. Using the predictor variables of the final model, a BRT model was fitted to each sample. Cross validation was again performed on peatlands, thus a peatland in the calibration data set was not part of the cross-validation data set to avoid overly optimistic results. Variances of the predictions and of the fitted functions of the 1000 models were evaluated.

If data sets are relatively small (e.g.,  $n < 1000$ , De’ath, 2007) then the small size of the training and test data sets lowers model accuracy. Given the fairly small number of peatlands in the data set and the partly high spatial correlation of dip wells within these peatlands, we decided not to split the data set into a training and test data set. Estimates of model accuracy can then be based on cross-validation, thereby making effective use of all the data (De’ath, 2007). The prediction uncertainty of the final model is estimated by the root mean square error of prediction ( $\text{RMSE}_{\text{cv}}$ , see above) for each land cover class. After testing for near-normal distribution of the residuals,  $\text{RMSE}_{\text{cv}}$  can be used to derive the 68 and 95 % confidence intervals of the predictions with  $\text{RMSE}_{\text{cv}}$  and  $2 \times \text{RMSE}_{\text{cv}}$ , respectively.

Finally, additional residual analysis was performed to evaluate whether the predictions are biased for different land cover classes or geographical regions.

## 2.5 Regionalization

In the final regionalization step, the predictor variables contributing to the final model were determined at a  $25 \times 25$  m raster for all organic soil in Germany. Predictor variables were determined with the same map input that was used for model building. Land cover information including information on ditches was based on the data from year 2012 and the climatic data was based on the average of the last 30 years. The fine spatial resolution of  $25 \times 25$  m was not chosen to fool the reader with a highly spatially accurate model. Rather this fairly fine scale was necessary to map the relatively small-scale effects of the topography, land use and peatland geometry variables. The final model was then used to make a prediction for each of these raster cells.

## 3 Results and discussion

### 3.1 Spatial correlation structure of the data set

The variogram model fitted to the sample variogram provided a nugget ( $0.012 \text{ m}^2$ ; 0.11 m), a sill ( $0.09 \text{ m}^2$ ; 0.3 m) and a range of spatial correlation (2700 m) for our data set of WL (Fig. 4). The nugget represents the very small-scale soil hydraulic variability and micro-topography effects on WL (van der Ploeg et al., 2012) and measurement error, e.g., by differences in the determination of the ground surface and in the timing of the manual measurements. Furthermore, micro-topography (e.g., hummocks) and oscillating peat surfaces of wet peatlands pose a challenge for an accurate determination of both ground surface and water level. The water level time series in the data set were of different lengths and ranged from 1 to 20 years. Interannual variability of water levels can be large (e.g., Knotters and van Walsum, 1997). For simplicity, in our analysis, data were not harmonized by extrapolating WL time series using weather data to a 30-year period. Thus, the nugget also includes errors that are introduced by dip wells with different measurement periods that are located in the range of spatial correlation. In consideration of these error sources, the fitted nugget of 0.11 m appears to be a realistic value. At 0.3 m, the fitted sill matched nearly perfectly with the standard deviation of the data (0.31 m), which indicates consistency between semivariogram model and data set. The fitted range of spatial correlation of 2700 m reflects both physical effects, i.e., the average range of lateral flow due to hydraulic gradients, as well as the effect of average land-use patterns in Germany on the spatial correlation of WL. Fitted values were used in the calculation of the dip-well-specific weights using Eq. (6).

### 3.2 Typical water levels for land-use types in German organic soils

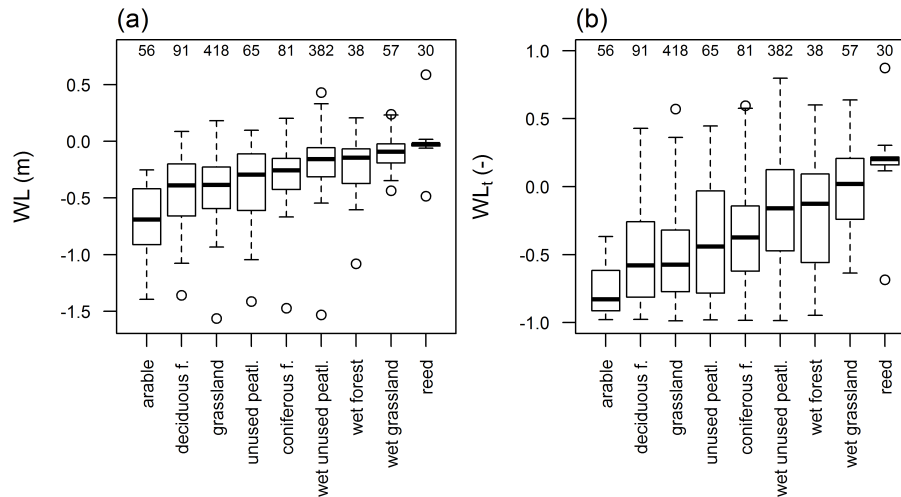
The land cover classes are characterized by plausible mean and median water levels, which show consistent differences between each other (Table 2 and Fig. 5a). The mean values of arable land and grassland agree with what can be expected for their agronomic requirements, with slightly lower water levels for arable land. The high variability observed for both classes may be related to the variability of the efficiency of installed drainage systems, as for example the presence and condition of tile drains and the depth of ditches. Grasslands can be managed with very variable intensity, which is partly reflected in different water levels. Figure 5a further shows that deciduous forests seem to dominate slightly drier organic soils compared to coniferous forests, which dominate under wetter conditions. A high variability of water levels is observed for the land cover class unused peatland. On the one hand, post peat-cutting topography increases the variability of WL over short distances. It probably contributes to the high variance observed for this class. On the other hand, this class comprises both rather dry unused peatlands and wetter peatlands in which re-wetting measures already took place, which however do not show yet a wet soil attribute in the ATKIS Digital Landscape Model. This may also cause part of the variance observed in the grassland and forest land cover class. All wet land cover classes (reed, wet grassland, wet forest and wet unused peatland) that were separated by wetness indication clearly show higher water levels, showing the wetness attribute of the Digital Landscape Model is a useful attribute.

Figure 5b shows the transformed water level for all classes. It can be observed that the variances of the wetter land cover increase relative to the variances of the dry land cover classes. This is due to the highest sensitivity of GHG emissions in the wet range of water levels ( $> -0.5$  m). Consequently, the rather high variance of WL for arable land corresponds to a rather low variance of  $WL_t$ , i.e., to a rather low assumed effect of WL variability on the GHG budget.

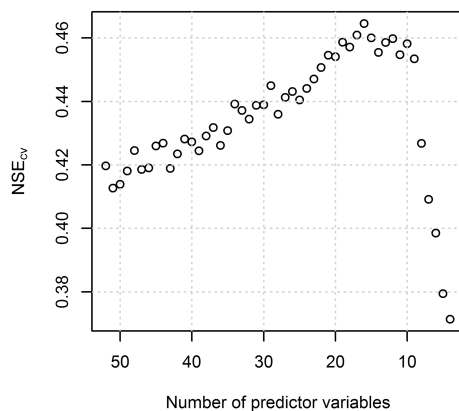
### 3.3 BRT model calibration and validation: WL vs. $WL_t$

In contrast to land cover class, the other predictor variables showed, if at all, only weak relations to WL and  $WL_t$  when evaluating them with box plots, 2-D cross plots and simple correlation matrices. Here, we expected BRT to detect the strongest predictor interactions and to identify the most informative predictors.

After model calibration with all predictors, subsequent model simplification successively dropped those parameters with correlation  $> 0.7$  and the lowest contribution. For both, WL and  $WL_t$ , model performance improved during this simplification. For  $WL_t$ , the highest values of  $NSE_{cv}$  of approximately 0.46 were achieved with 21 to 9 model parameters. The development of  $NSE_{cv}$  for the last 50 parameters is



**Figure 5.** Water level relative to ground surface, WL (m), and transformed water level,  $WL_t$  (–), by land cover class illustrated as a weighted box plot.  $WL_t = -1$  corresponds to maximum  $CO_2$  emissions and  $WL_t = 1$  to maximum  $CH_4$  emissions. In the top horizontal axes, the number of dip wells in each class is indicated.



**Figure 6.**  $NSE_{CV}$  as a function of number of predictor variables used in the model of  $WL_t$  during model simplification and shown for the last 50 parameter drops.

shown in Fig. 6. Further elimination of parameters led to a pronounced decline of model performance. Similar behavior was observed for the calibration on WL. In favor of a more parsimonious model, we chose the model with the lowest number of parameters before the pronounced decline of model performance occurred. For the calibration on  $WL_t$ , this corresponded to the model with lowest number of parameters that still achieved  $NSE_{CV}$  values of  $> 0.45$  (Fig. 6). The final  $WL_t$  model comprised nine predictor variables, and the final WL model, seven parameters. The percentages of parameter contributions to the final model and their individual influences are discussed for  $WL_t$  in Sect. 3.4.

Table 3 summarizes the statistical performances of the models calibrated on WL and  $WL_t$ . For both models  $NSE_{cal}$  is considerably higher than  $NSE_{CV}$  and shows the commonly

**Table 2.** Weighted mean and standard deviation of WL and  $WL_t$  data, and of the  $WL_t$  map presented in Sect. 3.6, for the nine land cover classes.

$WL_t$ (–),	WL (m)	$WL_t$ (–)	
	Mean $\pm$ sd	Mean $\pm$ sd	Map mean $\pm$ sd
Arable land	$-0.69 \pm 0.30$	$-0.76 \pm 0.17$	$-0.66 \pm 0.22$
Deciduous f.	$-0.45 \pm 0.34$	$-0.49 \pm 0.37$	$-0.47 \pm 0.35$
Grassland	$-0.44 \pm 0.29$	$-0.52 \pm 0.32$	$-0.49 \pm 0.30$
Unused peatl.	$-0.39 \pm 0.36$	$-0.39 \pm 0.41$	$-0.37 \pm 0.40$
Coniferous f.	$-0.36 \pm 0.36$	$-0.37 \pm 0.37$	$-0.46 \pm 0.35$
Wet unused peatl.	$-0.22 \pm 0.27$	$-0.18 \pm 0.40$	$-0.17 \pm 0.36$
Wet forest	$-0.22 \pm 0.29$	$-0.17 \pm 0.43$	$-0.21 \pm 0.39$
Wet grassland	$-0.10 \pm 0.14$	$-0.00 \pm 0.31$	$-0.15 \pm 0.39$
Reed	$-0.01 \pm 0.17$	$0.20 \pm 0.29$	$-0.06 \pm 0.32$

observed overfitting behavior of BRT models. The different measures that we conducted to minimize overfitting (cross-validation on peatlands, restriction to monotonic responses and model simplification including elimination of highly correlated variables) lowered the difference between  $NSE_{cal}$  and  $NSE_{CV}$  but could not totally avoid overfitting.  $NSE_{CV}$  of the  $WL_t$  model (0.453) indicates higher predictive model performance compared to the WL model (0.381). However, as the data ranges differ due to the transformation, this comparison may be misleading. Therefore, we transformed the predictions of the WL model to obtain  $WL_t$  values from this model and equally calculated the performance criteria (Table 3, second column). Then,  $NSE_{CV}$  is slightly increased (0.397), but does not achieve the values of the model that was calibrated on  $WL_t$ . A better predictive model performance of the model calibrated on  $WL_t$  is also visible for the  $RMSE_{CV}$  values. The total  $RMSE_{CV}$ , as well as the  $RMSE_{CV}$  values for the

**Table 3.** Performance criteria of the different models; dry range defined as  $WL < -0.3$  m and wet range as  $WL > -0.3$  m.

	WL (m) (calibrated on WL)	WL <sub>t</sub> (-) (calibrated on WL)	WL <sub>t</sub> (-) (calibrated on WL <sub>t</sub> )
NSE <sub>cal</sub>	0.627	0.559	0.642
NSE <sub>cv</sub>	0.381	0.397	0.453
RMSE <sub>cv</sub>	0.269	0.299	0.284
RMSE <sub>cv,dry</sub>	0.284	0.263	0.259
RMSE <sub>cv,wet</sub>	0.222	0.382	0.355
Bias	-0.003	0.083	0.002
Bias <sub>dry</sub>	-0.012	0.070	0.003
Bias <sub>wet</sub>	0.021	0.120	0.000

dry ( $WL < -0.3$  m) and wet range ( $WL > -0.3$  m), show slightly lower values for the WL<sub>t</sub> model compared to WL<sub>t</sub> values from the model calibrated on WL. Given our hypothetical transfer function (Fig. 3) in which the GHG budget is linearly related to WL<sub>t</sub>, the higher accuracy of WL<sub>t</sub> predictions directly corresponds to a higher accuracy of GHG budget predictions.

Superior model performance is also evident when evaluating model bias. Only when calibrating directly on WL<sub>t</sub> are the WL<sub>t</sub> predictions bias-free. Calibration on WL and subsequent transformation to WL<sub>t</sub> introduces a model bias towards systematically lower WL<sub>t</sub> values. In subsequent applications to GHG emission upscaling, lower WL<sub>t</sub> values would lead to an overestimation of CO<sub>2</sub> emissions and to an underestimation of CH<sub>4</sub> emissions.

### 3.4 Influence of predictor variables on WL<sub>t</sub>

Given the beneficial characteristics of the model calibrated on WL<sub>t</sub> for GHG upscaling, presentation and discussion of further model results is restricted to the WL<sub>t</sub> model.

The BRT method allows the analysis of the parameter contributions to, and influences on, the model (Elith et al., 2008) and thus may contribute to system understanding. The percentages of the contributions of the nine predictor variables to the final model ranged from 25.2 to 5.6 % (Fig. 7). Except protection status, at least one parameter of each of the seven parameter groups contributed to the final model. All protection status information was dropped early during the simplification process due to low contribution, although WL showed slightly higher values for data from nature protection or special areas of conservation. However, other parameters seem to be able to fully compensate the information that is lost by dropping this predictor.

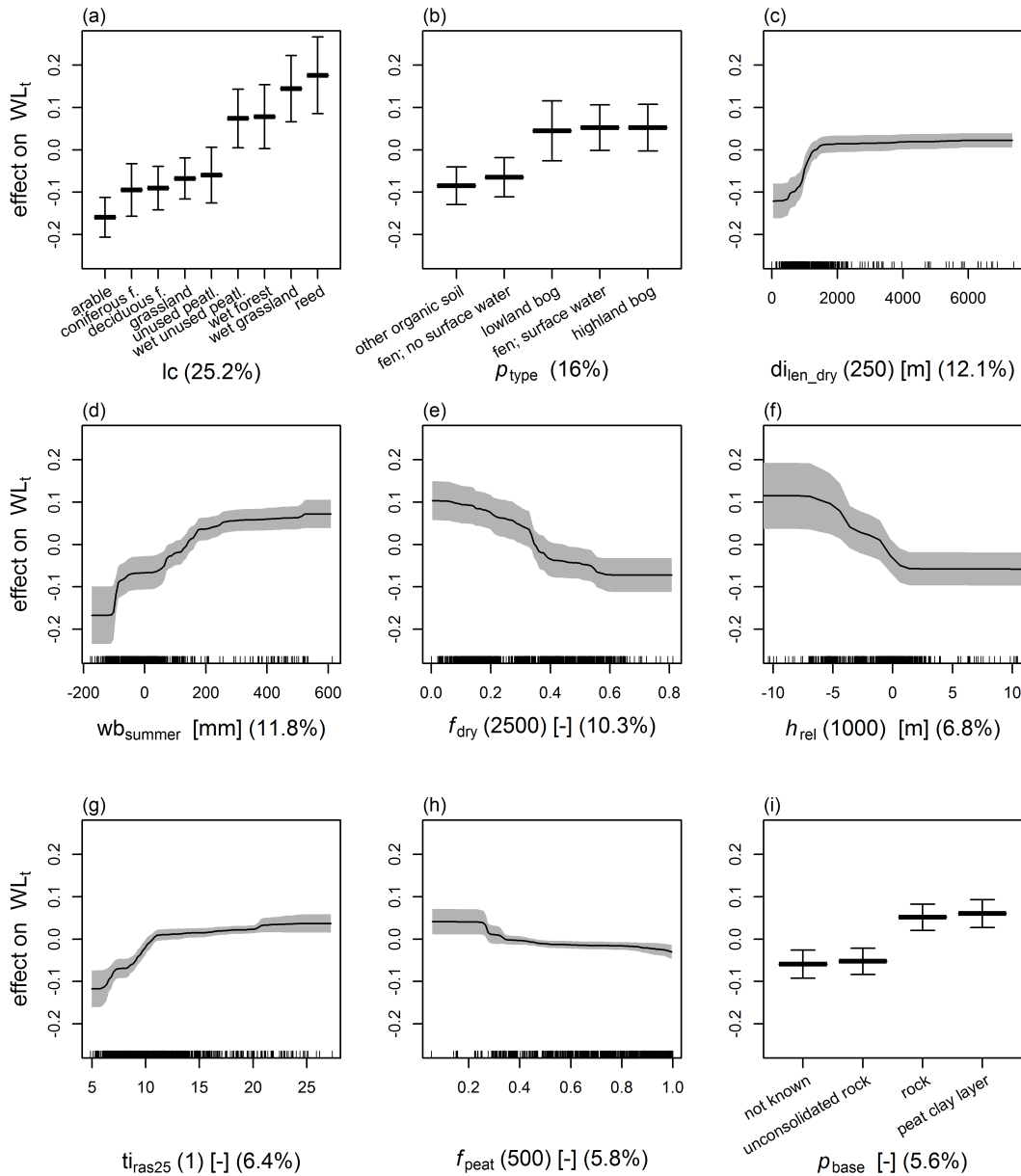
Land cover class, lc, at the dip well was the parameter with strongest contribution (25.2 %). It basically follows the trend illustrated in Fig. 5b. The bootstrap error plotted as standard deviation (Fig. 7) shows the variation of this influence over the 1000 bootstrap models. A second land cover parameter, the fraction of dry land cover classes on organic soils in a buffer of 2500 m radius,  $f_{dry}$  (2500), contributed to the model

with 10.3 %. The monotonic decrease of WL<sub>t</sub> with increasing  $f_{dry}$  (2500) is plausible as higher values reflect intensive land use in the surroundings of the dip well and thus indicate intensive artificial drainage. Together both parameter contributed 35.5 % and thus land cover represents the parameter group with the strongest model contribution.

Peatland characteristics are the second most important parameter group. The peatland type contributed 16 %. The model indicates that peatlands without any connection to surface water bodies (river or lake) and the class of other organic soils are characterized by lower WL<sub>t</sub> compared to the peatland types lowland bog, upland bog and fen neighboring surface water. As the class of other organic soils is generally expected to reflect lower water levels and as surface water may have a stabilization effect on water levels of organic soils, the influence of the peatland type can be considered plausible. Besides peatland type, the substrate of the peat base contributes 5.6 %. Here, organic soils overlying peat clay layers (e.g., limnic sediments such as calcareous gyt-tja) or basement rock are characterized by higher WL<sub>t</sub> compared to organic soils overlying unconsolidated rock. This can be explained by the lower drainage resistance of unconsolidated rocks. This may cause an increased efficiency of anthropogenic drainage and/or a general higher vulnerability to seepage losses. Finally, slightly lower WL<sub>t</sub> values are indicated by a high fraction of organic soils for the 500 m buffer,  $f_{peat}$  (500). This may reflect the higher land-use pressure on large peatlands compared to rather small peatlands, which tentatively are more easily preserved by nature protection efforts.

The remaining four parameter groups are represented in the model by only one parameter each. The third most influential parameter was the length of ditches on arable land and grassland for the 250 m buffer,  $d_{len,dry}$  (250). At first glance, it may be surprising that with increasing ditch density, WL<sub>t</sub> values tend to be higher as ditches are supposed to drain the water when land is used as arable land and grassland. The fact that the model identifies a rather strong effect in the opposite direction may be caused by incomplete information about the drainage network. There is not detailed information about the spatial distribution of tile drains. Based on expert knowledge, agricultural areas with a lower ditch density are more likely to have tile drains. As these drains, easily installed with a narrow drain spacing, are more effective at draining organic soils, low WL<sub>t</sub> values for arable land and grassland may be related to low ditch densities. Furthermore, ditches were originally dug at narrow spacing in especially wet areas of organic soils but there is no information available whether these ditches still function properly.

The parameters  $wb_{summer}$ ,  $h_{rel}$  and  $ti_{ras25}$  all show expected trends. The model predicts higher WL<sub>t</sub> for increasing climatic water balance in the summer period (May to October),  $wb_{summer}$ , for dip wells located in depressions (low values of  $h_{rel}$ ) and for higher small-scale topographic wetness indices calculated on the  $25 \times 25$  digital elevation model ( $ti_{ras25}$ ).



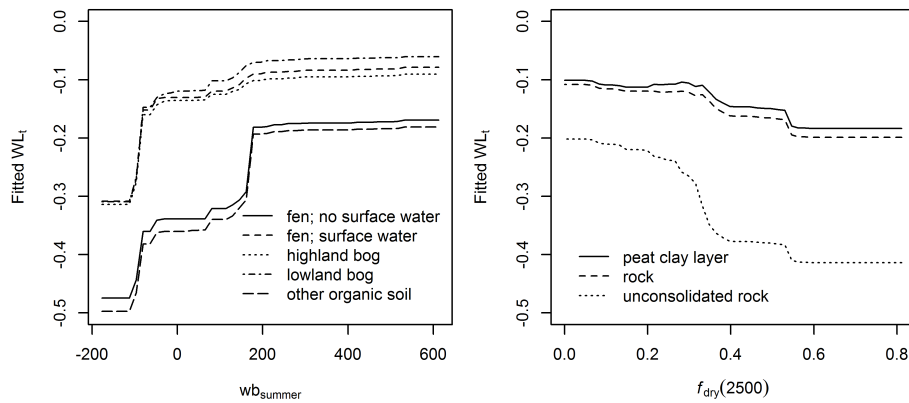
**Figure 7.** Partial dependence plots for the predictor variables. For an explanation of variables see Table 1. The y axes are on  $WL_t$  scale and are centered around the mean  $WL_t$ . Error bars and grey area indicate standard deviation of the response of over 1000 bootstrap models. The relative contribution of each predictor is indicated as percentage. The lines along the x axes of each plot show distribution of data across that variable, in deciles.

The fact that all parameters show expected or explainable responses in the model corroborates the reliability of the calibrated  $WL_t$  model. The standard deviation of the predictor responses based on the bootstrap samples shows the stability of the observed responses.

Further insight into model behavior can be obtained by analyzing parameter interactions. This is obtained by changing two parameters simultaneously while keeping mean values for all other parameters (Elith et al., 2008). Figure 8 shows the two strongest parameter interactions. Parameter  $wb_{summer}$

strongly interacts with  $p_{type}$ . The generally lower values of  $WL_t$  of fens without surface water connection and other organic soils show a stronger dependency on the summer climatic water balance. While a summer climatic water balance of  $> -80$  mm shows a rather weak effect on  $WL_t$  for the wetter peatland types, in contrast to the two drier peatland types there is still a strong effect with increasing  $wb_{summer}$ . The trend for  $wb_{summer} > 130$  mm for the dry peatland types is supported by seven different peatlands.





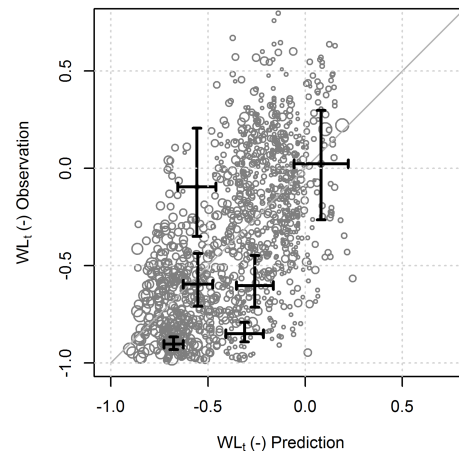
**Figure 8.** Partial dependence plots representing the two strongest interactions in the model: (a) between  $p_{\text{type}}$  and  $wb_{\text{summer}}$  and (b) between  $p_{\text{base}}$  and  $f_{\text{dry}}$ . Fitted  $WL_t$  is plotted on the y axis which is obtained after accounting for the average effect of all other predictor variables.

Another strong interaction is observed for  $p_{\text{base}}$  and  $f_{\text{dry}}$  (2500). While a rather weak effect of the fraction of arable land and grassland is observed for organic soils overlying basement rock and peat clay layer, a strong effect is observed for organic soils overlying unconsolidated rock. This interaction reflects the higher lateral range of drainage effects for organic soils with little flow resistance at the peat base. In these organic soils, intensive land use lowers the water level over large areas.

### 3.5 Discussion of model uncertainty

Plotting observed vs. predicted  $WL_t$  from cross-validation (Fig. 9) illustrates the rather large residual variance that cannot be explained by the model. As indicated by the higher  $RMSE_{\text{cv}}$  for the wet range (Table 3), scatter increases with increasing  $WL_t$ . Error bars in the y direction indicate data error derived from the nugget of the variogram. It is shown for a few data points as an example. Due to transformation, data error increases for higher  $WL_t$ . Figure 9 demonstrates that the fraction of unexplainable variance related to data error is much higher for the wet than for the dry range. Bootstrap error indicating the variation of the model predictions for 1000 bootstrap samples is shown in the x direction for the same data points. Bootstrap error is lower than the data error for the wet range and slightly higher for the dry range.

Bootstrap errors demonstrate the sensitivity of model predictions to changes of the data set used for calibration. When a model possesses structural deficits such as missing predictor variables, bootstrap errors should not be used to define confidence intervals for the model predictions. Figure 10 shows residuals from cross-validation and standard deviation of bootstrap predictions for all land cover classes. The residuals of each land cover class show near-normal distributions. For five of the nine land cover classes (wet forest, wet unused peatland, arable land, coniferous forest and reed), the Shapiro–Wilk test of normality is positive ( $p > 0.05$ ). Figure 10a further indicates that residuals of each land cover

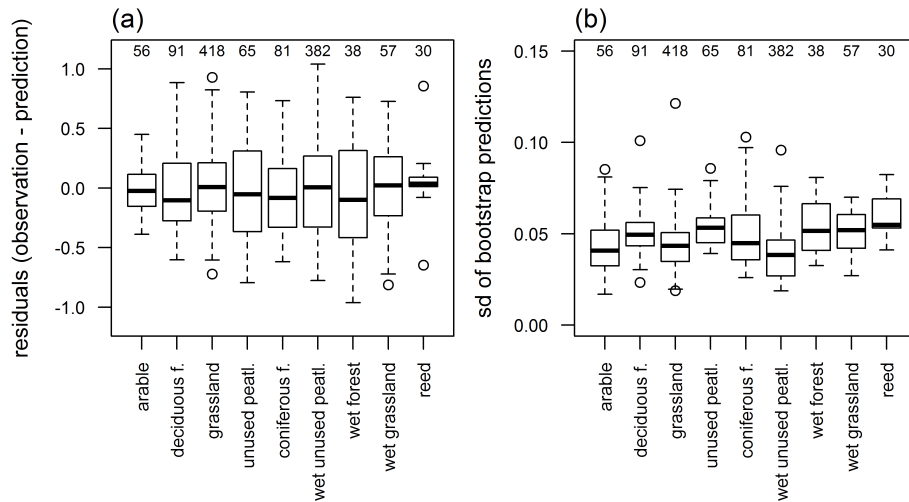


**Figure 9.** Observed vs. predicted transformed annual mean water level ( $WL_t$ ) from cross-validation results. Error bars show selected data and bootstrap model errors as standard deviation. Data points are scaled by their weights.

scatter fairly well around zero, indicating low bias for the various land cover classes. Land-cover-class-specific confidence intervals of model predictions can thus be derived from the  $RMSE_{\text{cv}}$  of each land cover class, e.g.,  $2 \times RMSE_{\text{cv}}$  representing the 95 % confidence interval.

The prediction uncertainty derived from cross-validation is much higher than the bootstrap prediction uncertainty obtained from the bootstrap standard deviation (sd), with  $2 \times \text{sd}$  corresponding to the 95 % confidence interval (Fig. 10). The large difference between these values indicates that the model has structural deficits that can be attributed to several error sources:

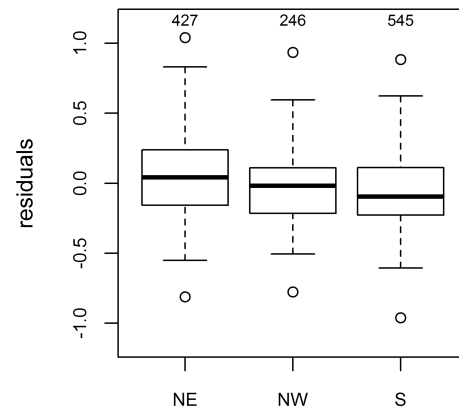
- i. Key influences on  $WL_t$  are missing in the set of predictor variables. None of the predictor variables indicate whether and to which extent water level increase due to re-wetting measures took place in the last



**Figure 10.** (a) Residuals (observation–prediction) of  $WL_t$  predictions and (b) standard deviation (sd) of bootstrap predictions shown for the nine land cover classes. In the upper part, the number of dip wells in each class is indicated.

years. Wetness indicators (wet soil and/or vegetation attributes) that are obtained from the Digital Landscape Model probably react with a delay of several years. Thus, we expect the occurrence of several observed high  $WL_t$  values that cannot be explained by any of the predictor variables.

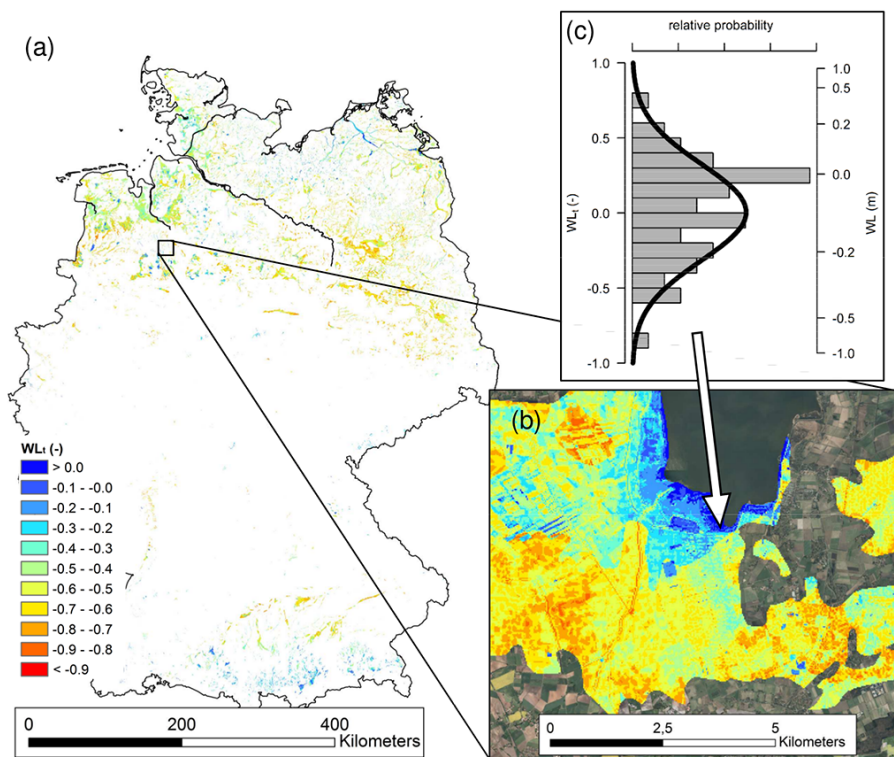
- ii. Small-scale topography that is not represented with sufficient detail and accuracy in the DEM may cause several predictions to strongly differ from what would be expected from the other predictor variables. A common example may be a dip well that is located on a narrow peat ridge, which remained after peat-cutting and is absent in the DEM, and that is situated in an area classified as wet soil by the Digital Landscape Model. Then, the model indicates a  $WL_t$  that is much higher than the observed  $WL_t$ ; as for the observed value, the reference surface was the surface of the peat ridge.
- iii. Consistent information about tile drains is missing and only exists on the regional scale (Tetzlaff et al., 2009). At the national scale, however, there are no maps on tile drains. Tile drains are known to have a strong effect on  $WL_t$  for arable land and grassland. As explained above, we expect parameter  $di_{\text{fen,dry}}$  (250) to partially compensate for this missing information.
- iv. Another source of prediction uncertainty may comprise inconsistent and erroneous land cover classification of the Digital Landscape Model due to the high degree of subjectivity for many of the attributes. Furthermore, the temporal accuracy of the Digital Landscape Model may be as inaccurate as 5 years which can cause time series with land-use change to be split at the wrong date, and vegetation and wetness attributes to be not yet updated to the current conditions.



**Figure 11.** Residuals (observation–prediction) of  $WL_t$  predictions for the three major geographical peatland regions of Germany. In the upper part, the number of dip wells in each class is indicated.

- v. The water balance of fens strongly depends on the size and the hydraulic head of the groundwater catchment, i.e., of the aquifer underlying the peat layer. Unfortunately, there is no consistent map of hydraulic heads or groundwater catchments for all Germany.

We checked model predictions for geographical bias. Geographical location was not one of the model parameters. However, the history and policy of land use on organic soils, current ditch water management and climate do show large-scale geographical trends. We divided our data set into the three major German peatland regions (NE, NW and S) and evaluated the model residuals (Fig. 11) to see whether our model is biased due to important missing geographical effects. A serious bias for any of the three major German peatland regions cannot be identified.



**Figure 12.** Map of predictions of transformed annual mean water level ( $WL_t$ ) for all German organic soils (a) and an enlarged map section (b). Probability distribution in (c) indicates the uncertainty of a specific point prediction for wet grassland as an example. Here, predicted value is approximately  $WL_t = 0$ , but note that wet grassland predictions do vary in space depending on the values of the other model parameters. The histogram shows the residuals from cross-validation for wet grassland, to which the probability distribution was fitted.

When applying calibrated statistical models during regionalization, it is important to check model behavior for extrapolation outside the range of the parameter space that is covered by the data upon which the model was built. BRT always extrapolates at a constant value from the most extreme environmental value in the training data. In contrast to other types of statistical models, e.g., generalized linear models, BRT does not continue the fitted trend beyond the last observation. Regarding the categorical variables, the data set covers all classes occurring in Germany with several peatlands. The data set also covers the major range of values occurring in Germany for the numerical predictor variables. Furthermore, Fig. 7 indicates that the constant values, at which the model extrapolates the influence of the variables, do not raise major concern for any extreme predictions outside the parameter range.

### 3.6 Regionalization

The map of  $WL_t$  resulting from the application of the fitted  $WL_t$  model to all grid cells shows gradients at the regional scale (Fig. 12a). In the south of Germany, for example, a gradient from wet to dry can be observed for the pre-alpine upland bogs and the peatlands of the moraine plain. In the north of Germany, the map indicates that organic soils in the

very NE are wetter than the rest. For the rest of the north, a slight gradient can be observed from less dry to dry from NW to E, which is mainly driven by the higher summer climatic water balance in the NW. As both categorical and numerical predictor variables do also vary at the sub-regional scale, the resulting map also shows gradients within peatland areas, e.g., due to small-scale land-use ditch density gradients and topography effects (Fig. 12b).

We calculated  $WL_t$  averages of the land cover classes using the regionalized  $WL_t$  from the map (Table 2, column 3). The given standard deviation comprises both the variability within a land cover class that is explained by the model as well as the uncertainty of each prediction. Resulting means and standard deviations slightly differ from the corresponding values of the data set. The land-cover-specific  $WL_t$  values obtained from the map can be considered as being more representative as the regionalization procedure is supposed to partly account for potential bias in the data set.

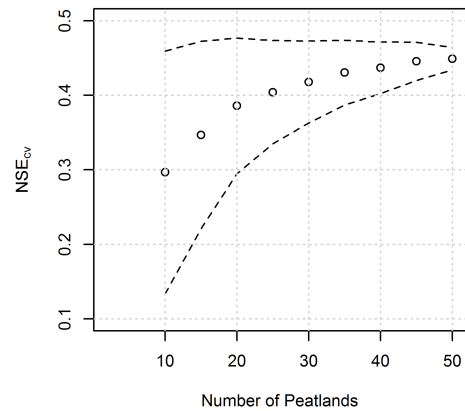
When applying this map and its predicted  $WL_t$  values in subsequent GHG upscaling, it is crucial that model uncertainty is propagated properly. An example demonstrates the necessity of uncertainty propagation. For a grid cell classified as wet grassland, the probability distribution of  $WL_t$  is shown based on a normal distribution that was fitted to the

residuals of this land cover class (Fig. 12c). Without propagating the uncertainty and when only translating the predicted  $WL_t$  (eventually in combination with other parameters, e.g., soil properties) into a GHG budget, GHG budget is strongly underestimated as the  $WL_t$  prediction is close to zero, indicating neither large  $CO_2$  nor  $CH_4$  emissions. When translating the full distribution of  $WL_t$  into a GHG budget, the resulting GHG budget would be much higher as at both sides of the predicted  $WL_t$ , the GHG budget increases.

### 3.7 Possible paths for model improvement

The model performance that is achieved by the statistical approach presented in our study raises the question whether collecting more WL data can improve model performance or whether the factor that is constraining the model performance is the limited strength of the nation-wide available predictor variables. To assess this question, additional “hold-out models” were developed by fitting the BRT model to various random sets of data with a limited number of peatland areas (from 10 to 50 peatlands). For each number of peatland areas, 500 random selections were calibrated and model performance was evaluated with  $NSE_{cv}$ . As expected, results indicate an increase of model performance with increasing number of peatlands used in the model building process (Fig. 13). Results also indicate a substantial flattening of the learning curve. Thus, further collection of WL data may only lead to a substantial model improvement when including many more peatlands into the data set. More promising would be the specific collection of more data on the weakly represented and/or important land cover classes, arable land and grassland.

Another path to achieve a stronger model is the development of new predictor variables. In the future, the availability of a more accurate DEM based on laser-scanning data, which is already available at full coverage for some federal states of Germany, may strongly increase the predictability of the observed WL data. Additionally, a nation-wide map of water management and of the distribution of tile drains would have great potential to explain large parts of the residual variance and/or even allow setting up a large-scale physically based model that includes water management. Furthermore, data harmonization by extrapolating the water level time series of our data set with the climatic boundary conditions of the last 30 years may lower the unexplainable variance of the data set due to short measurement periods (Bartholomeus et al., 2008), an effort that has been successfully conducted in Finke et al. (2004) using the transfer noise model of Bierkens et al. (1999). Finally, we believe that the inclusion of remote sensing products in our statistical model approach, as e.g., spaceborne microwave soil moisture observations (Sutanudjaja et al., 2013), may hold large potential to improve model performance as moisture differences due to varying water levels are high for organic soils.



**Figure 13.** NSE of cross-validation vs. number of randomly selected peatland areas. Dashed lines indicate  $NSE_{cv} \pm sd$ .

## 4 Conclusions

Our study demonstrates the potential of statistical modeling for the regionalization of water levels in organic soils when data covers only a small fraction of peatlands of the final map and thus spatial interpolation is not possible. With the available data set of target and predictor variables, it was possible to predict 45 % of the GHG relevant water level variance in the data set in a cross-validation scheme. The variance is explained by nine predictor variables. With the analysis of their effect on the water level it was possible to gain insight into natural and anthropogenic boundary conditions that control water levels of organic soils in Germany.

Based on a hypothetical GHG transfer function relating GHG emissions to annual mean water levels (WL), we showed the advantages of transforming the annual mean water level into a new variable ( $WL_t$ ) to which GHG emissions linearly depend on. The transformation improved model accuracy, increased the explained variance of the water level range that is relevant for GHG emissions and avoided model bias.

The presented approach is transparent and allows successive improvement when new input data and predictor variables become available. Our results show that model improvement by increasing number of  $WL_t$  data, however, seems to be limited. If efforts are made, data collection should be concentrated on agriculturally used organic soils for which relatively few data is available. We believe that the constraining factor of model performance is rather the weakness of the predictor variables that are currently available at large scales. The development of new, more informative predictor variables, as for example water management maps and remote sensing products, may be the more promising path for model improvement.

The proposed regionalization approach is suited to application to any other country where similar data of target and predictor variables is available. It is important that the spatial

resolution of the predictor variables is high enough (Finke et al., 2004). If predictor variables like land use and peatland type are only available at a much coarser scale and provided as percentages for grid cells, the dependency between predictor variables and the rather local WL will probably be lost for most of the predictor variables.

Our work must be considered as one piece of a broader framework for the regionalization of GHG emissions that includes other site characteristics and must be further developed in future research. For example, if for specific regions detailed information on peat properties becomes available and its effect on GHG emissions can be estimated by the use of multivariate transfer functions, the map of transformed water levels ( $WL_t$ ) can be used as an input for this follow-up regionalization.

*Acknowledgements.* Several institutions made their data available for this synthesis study. We gratefully thank ARGE Schwäbisches Donaumoos, Biologische Station Steinfurt, Biosphärenreservat Vessertal, BUND Diepholzer Moorniederung, Deutscher Wetterdienst (DWD), Bezirksregierung Detmold, Förderverein Feldberg-Uckermärkische Seenlandschaft, Hochschule für Wirtschaft und Umwelt Nürtingen (Institut für Angewandte Forschung), Hochschule Weihenstephan-Triesdorf (Professur für Vegetationsökologie), Humboldt Universität zu Berlin (Fachgebiet Bodenkunde und Standortlehre), Landkreis Gifhorn, LBEG Hannover (Referat Boden- und Grundwassermonitoring), LUNG Mecklenburg-Vorpommern, Eberhard Gärtner, IGB Berlin (Zentrales Chemielabor), Landkreis Gifhorn, NABU Minden-Lübbecke, Naturpark Drömling, Naturpark Erzgebirge/Vogtland, Region Hannover, Naturpark Nossentiner/Schwinzer Heide, Naturschutzfond Brandenburg, Ökologische Station Steinhuder Meer, Technische Universität München (Lehrstuhl für Vegetationsökologie), Universität Hohenheim (Institut für Bodenkunde und Standortlehre), Johannes Gutenberg-Universität Mainz (Geographisches Institut, Bodenkunde), Universität Rostock (Professur für Bodenphysik und Ressourcenschutz, Professur für Landschaftsökologie und Standortkunde, Professur für Hydrologie), Kees Vegelin, ZALF (Institut für Bodenlandschaftsforschung, Institut für Landschaftsbiogeochemie), Werner Kutsch, Christian Brümmer, and Miriam Hurkuck. Furthermore, we acknowledge Maik Hunziger and Sören Gebbert for field and GRASS support, Katharina Leiber-Sauheitl, Stefan Frank, Ullrich Dettmann and René Dechow for data processing support, Annette Freibauer for reviewing the paper draft and three anonymous reviewers for providing helpful comments during the revision process. The study was financially supported by the joint research project “Organic soils” funded by the Thünen Institute.

Edited by: J. Liu

## References

- Bartholomeus, R., Witte, J. P. M., van Bodegom, P. M., and Aerts, R.: The need of data harmonization to derive robust empirical relationships between soil conditions and vegetation, *J. Veg. Sci.*, 19, 799–808, doi:10.3170/2008-8-18450, 2008.
- Berglund, O. and Berglund, K.: Influence of water table level and soil properties on emissions of greenhouse gases from cultivated peat soil, *Soil Biol. Biochem.*, 43, 5, 923–931, doi:10.1016/j.soilbio.2011.01.002, 2011.
- Beven, K. J. and Kirby, M.: A physically based variable contributing area model of catchment hydrology, *Hydrol. Sci. Bull.*, 24, 43–69, 1979.
- Bierkens, M. F. P. and Stroet, C. B. M. T.: Modelling non-linear water table dynamics and specific discharge through landscape analysis, *J. Hydrol.*, 332, 412–426, doi:10.1016/j.jhydrol.2006.07.011, 2007.
- Bierkens, M. F. P., Knotters, M., and van Geer, F. C.: Calibration of transfer function-noise models to sparsely or irregularly observed time series, *Water Resour. Res.*, 35, 1741–1750, doi:10.1029/1999wr900083, 1999.
- Buchanan, S. and Triantafyllis, J.: Mapping Water Table Depth Using Geophysical and Environmental Variables, *Groundwater*, 47, 80–96, doi:10.1111/j.1745-6584.2008.00490.x, 2009.
- Clapcott, J., Young, R., Goodwin, E., Leathwick, J., and Kelly, D.: Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity, Department of Conservation, DOC Research and Development series 326, Wellington, New Zealand, 2011.
- Cumming, G.: *Understanding The New Statistics*, Routledge, New York, USA, 535 pp., 2012.
- De’ath, G.: Boosted trees for ecological modeling and prediction, *Ecology*, 88, 243–251, doi:10.1890/0012-9658(2007)88[243:Btfema]2.0.Co;2, 2007.
- Dickens, W. T.: Error components in grouped data: Is it ever worth weighting?, *Rev. Econ. Stat.*, 72, 328–333, 1990.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J. R. G., Gruber, B., Lafourcade, B., Leitao, P. J., Munkemuller, T., McClean, C., Osborne, P. E., Reineking, B., Schroder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, 27–46, doi:10.1111/j.1600-0587.2012.07348.x, 2013.
- Drösler, M., Freibauer, A., Adelman, W., Augustin, J., Bergmann, L., Beyer, C., Chojnicki, B., Förster, C., Giebels, M., Görlitz, S., Höper, H., Kantelhardt, J., Liebersbach, H., Hahn-Schöfl, M., Minke, M., Petschow, U., Pfadenhauer, J., Schaller, L., Schägner, P., Sommer, M., Thuille, A., and Wehrhan, M.: Klimaschutz durch Moorschutz in der Praxis, Ergebnisse aus dem BMBF-Verbundprojekt Klimaschutz – Moornutzungsstrategien 2006–2010, vTI-Arbeitsberichte 4/2011, Johann Heinrich von Thünen-Institut, Braunschweig, Germany, 2011.
- Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, *J. Anim. Ecol.*, 77, 802–813, doi:10.1111/j.1365-2656.2008.01390.x, 2008.
- Fan, Y. and Miguez-Macho, G.: A simple hydrologic framework for simulating wetlands in climate and earth system models, *Clim. Dynam.*, 37, 253–278, doi:10.1007/s00382-010-0829-8, 2011.



- Finke, P. A., Brus, D. J., Bierkens, M. F. P., Hoogland, T., Knotters, M., and de Vries, F.: Mapping groundwater dynamics using multiple sources of exhaustive high resolution data, *Geoderma*, 123, 23–39, doi:10.1016/j.geoderma.2004.01.025, 2004.
- Francis, R. I. C. C.: Data weighting in statistical fisheries stock assessment models, *Can. J. Fish. Aquat. Sci.*, 68, 1124–1138, doi:10.1139/F2011-025, 2011.
- Gong, J. N., Wang, K. Y., Kellomaki, S., Zhang, C., Martikainen, P. J., and Shurpali, N.: Modeling water table changes in boreal peatlands of Finland under changing climate conditions, *Ecol. Model.*, 244, 65–78, doi:10.1016/j.ecolmodel.2012.06.031, 2012.
- Hahn-Schöfl, M., Zak, D., Minke, M., Gelbrecht, J., Augustin, J., and Freibauer, A.: Organic sediment formed during inundation of a degraded fen grassland emits large fluxes of CH<sub>4</sub> and CO<sub>2</sub>, *Biogeosciences*, 8, 1539–1550, doi:10.5194/bg-8-1539-2011, 2011.
- Hedges, L. V. and Olkin, I.: *Statistical Methods for Meta-Analysis*, Academic Press, Orlando, USA, 369 pp., 1985.
- Hijmans, R. J.: Species distribution modeling, Documentation on the R Package “dismo”, version 0.9-3, <http://cran.r-project.org/web/packages/dismo/dismo.pdf> (last access: February 2014), 2013.
- Hoogland, T., Heuvelink, G. B. M., and Knotters, M.: Mapping Water-Table Depths Over Time to Assess Desiccation of Groundwater-Dependent Ecosystems in the Netherlands, *Wetlands*, 30, 137–147, doi:10.1007/s13157-009-0011-4, 2010.
- IPCC: IPCC guidelines for national greenhouse gas inventories, edited by: Eggleston, H. S., Buendia, L., Miwa, K., and Ngara, T., IGES, Japan, 2006.
- Ju, W. M., Chen, J. M., Black, T. A., Barr, A. G., McCaughey, H., and Roulet, N. T.: Hydrological effects on carbon cycles of Canada’s forests and wetlands, *Tellus B*, 58, 16–30, doi:10.1111/j.1600-0889.2005.00168.x, 2006.
- Knotters, M. and van Walsum, P. E. V.: Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component, *J. Hydrol.*, 197, 25–46, doi:10.1016/S0022-1694(96)03278-7, 1997.
- Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., and Taylor, P.: Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees, *Mar. Ecol.-Prog. Ser.*, 321, 267–281, doi:10.3354/Meps321267, 2006.
- Leiber-Sauheitl, K., Fuß, R., Voigt, C., and Freibauer, A.: High CO<sub>2</sub> fluxes from grassland on histic Gleysol along soil carbon and drainage gradients, *Biogeosciences*, 11, 749–761, doi:10.5194/bg-11-749-2014, 2014.
- Levy, P. E., Burden, A., Cooper, M. D. A., Dinsmore, K. J., Drewer, J., Evans, C., Fowler, D., Gaiawyn, J., Gray, A., Jones, S. K., Jones, T., Mcnamara, N. P., Mills, R., Ostle, N., Sheppard, L. J., Skiba, U., Sowerby, A., Ward, S. E., and Zielinski, P.: Methane emissions from soils: synthesis and analysis of a large UK data set, *Global Change Biol.*, 18, 1657–1669, doi:10.1111/j.1365-2486.2011.02616.x, 2012.
- Limpens, J., Berendse, F., Blodau, C., Canadell, J. G., Freeman, C., Holden, J., Roulet, N., Rydin, H., and Schaepman-Strub, G.: Peatlands and the carbon cycle: from local processes to global implications – a synthesis, *Biogeosciences*, 5, 1475–1491, doi:10.5194/bg-5-1475-2008, 2008.
- Martin, M. P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., and Arrouays, D.: Spatial distribution of soil organic carbon stocks in France, *Biogeosciences*, 8, 5, 1053–1065, doi:10.5194/bg-8-1053-2011, 2011.
- Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), *Biogeosciences*, 10, 753–788, doi:10.5194/bg-10-753-2013, 2013.
- Moore, T. R. and Dalva, M.: The Influence of Temperature and Water-Table Position on Carbon-Dioxide and Methane Emissions from Laboratory Columns of Peatland Soils, *J. Soil Sci.*, 44, 651–664, doi:10.1111/j.1365-2389.1993.tb02330.x, 1993.
- Moore, T. R. and Roulet, N. T.: Methane Flux – Water-Table Relations in Northern Wetlands, *Geophys. Res. Lett.*, 20, 587–590, doi:10.1029/93gl00208, 1993.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.
- Regina, K., Nykänen, H., Silvola, J., and Martikainen, P. J.: Fluxes of nitrous oxide from boreal peatlands as affected by peatland type, water table level and nitrification capacity, *Biogeochemistry*, 35, 401–418, doi:10.1007/BF02183033, 1996.
- Ridgeway, G.: Generalized boosted regression models, Documentation on the R Package “gbm”, version 2.1, <http://cran.r-project.org/web/packages/gbm/gbm.pdf> (last access: February 2014), 2013.
- Roßkopf, N., Fell, H., and Zeitz, J.: Organic soils in Germany, their distribution and carbon stocks, *Catena*, in review, 2014.
- Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Using ERS spaceborne microwave soil moisture observations to predict groundwater head in space and time, *Remote Sens. Environ.*, 138, 172–188, doi:10.1016/j.rse.2013.07.022, 2013.
- Tetzlaff, B., Kuhr, P., and Wendland, F.: A New Method for Creating Maps of Artificially Drained Areas in Large River Basins Based on Aerial Photographs and Geodata, *Irrig. Drain.*, 58, 569–585, doi:10.1002/Ird.426, 2009.
- Thompson, J. R., Gavin, H., Refsgaard, A., Sorenson, H. R., and Gowing, D. J.: Modelling the hydrological impacts of climate change on UK lowland wet grassland, *Wetl. Ecol. Manage.*, 17, 503–523, doi:10.1007/s11273-008-9127-1, 2009.
- UBA: National Inventory Report for the German Greenhouse Gas Inventory 1990–2008, Submission under the United Nations Framework Convention on Climate Change and the Kyoto Protocol 2012, Dessau, Germany, 2012.
- van den Akker, J. J. H., Jansen, P. C., Hendriks, R. F. A., Hoving, I., and Pleijter, M.: Submerged infiltration to halve subsidence and GHG emissions of agricultural peat soils, Proceedings of the 14th International Peat Congress, Stockholm, Sweden, 2012.
- van der Gaast, J. W. J., Massop, H. T. L., and Vroon, H. R. J.: Actuele grondwaterstandsituatie in natuurgebieden: Een Pilotstudie, Wettelijke Onderzoekstaken Natuur & Milieu, WOt-rapport 94, Wageningen, 134 pp., 2009.



van der Ploeg, M. J., Appels, W. M., Cirkel, D. G., Oosterwoud, M. R., Witte, J. P. M., and van der Zee, S. E. A. T. M.: Microtopography as a Driving Mechanism for Ecohydrological Processes in Shallow Groundwater Systems, *Vadose Zone J.*, 11, 52–62, doi:10.2136/Vzj2011.0098, 2012.

Wackernagel, H.: *Multivariate Geostatistics*, Springer, Berlin, Germany, 387 pp., 2003.