

SYSTEMATIC REVIEW

Open Access



How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? A systematic review

Magnus Land^{1*}, Wilhelm Granéli^{2,3,5}, Anders Grimvall⁴, Carl Christian Hoffmann⁵, William J. Mitsch³, Karin S. Tonderski⁶ and Jos T. A. Verhoeven⁷

Abstract

Background: Eutrophication of aquatic environments is a major environmental problem in large parts of the world. In Europe, EU legislation (the Water Framework Directive and the Marine Strategy Framework Directive), international conventions (OSPAR, HELCOM) and national environmental objectives emphasize the need to reduce the input of nutrients to freshwater and marine environments. A widely used method to achieve this is to allow water to pass through a created or restored wetland. However, the large variation in measured nutrient removal rates in such wetlands calls for a systematic review.

Methods: Searches for primary studies were performed in electronic databases and on the internet. One author performed the screening of all retrieved articles at the title and abstract level. To check that the screening was consistent and complied with the agreed inclusion/exclusion criteria, subsets of 100 articles were screened by the other authors. When screening at full-text level the articles were evenly distributed among the authors. Kappa tests were used to evaluate screening consistency. Relevant articles remaining after screening were critically appraised and assigned to three quality categories, from two of which data were extracted. Quantitative synthesis consists of meta-analyses and response surface analyses. Regressions were performed using generalized additive models that can handle nonlinear relationships and interaction effects.

Results: Searches generated 5853 unique records. After screening on relevance and critical appraisal, 93 articles including 203 wetlands were used for data extraction. Most of the wetlands were situated in Europe and North America. The removal rate of both total nitrogen (TN) and total phosphorus (TP) is highly dependent on the loading rate. Significant relationships were also found for annual average air temperature (T) and wetland area (A). Median removal rates of TN and TP were 93 and 1.2 g m⁻² year⁻¹, respectively. Removal efficiency for TN was significantly correlated with hydrologic loading rate (HLR) and T, and the median was 37 %, with a 95 % confidence interval of 29–44 %. Removal efficiency for TP was significantly correlated with inlet TP concentration, HLR, T, and A. Median TP removal efficiency was 46 % with a 95 % confidence interval of 37–55 %. Although there are small differences in average values between the two quality categories, the variation is considerably smaller among high quality studies compared to studies with lower quality. This suggests that part of the large variation between studies may be explained by less rigorous study designs.

Conclusions: On average, created and restored wetlands significantly reduce the transport of TN and TP in treated wastewater and urban and agricultural runoff, and may thus be effective in efforts to counteract eutrophication. However, restored wetlands on former farmland were significantly less efficient than other wetlands at TP removal.

*Correspondence: magnus.land@eviem.se

¹ Mistra EviEM, Stockholm Environment Institute, Box 24218, 10451 Stockholm, Sweden

Full list of author information is available at the end of the article

In addition, wetlands with precipitation-driven HLRs and/or hydrologic pulsing show significantly lower TP removal efficiencies compared to wetlands with controlled HLRs. Loading rate (inlet concentrations \times hydraulic loading rates) needs to be carefully estimated as part of the wetland design. More research is needed on the effects of hydrologic pulsing on wetlands. There is also a lack of evidence for long-term (>20 years) performance of wetlands.

Keywords: Nitrogen, Phosphorus, Nutrient, Removal rate, Removal efficiency, Wetland creation, Restored wetland, Constructed wetland, Pond, Eutrophication

Background

In Europe, like in many other parts of the world, nutrient enrichment of water bodies is a major environmental problem [1]. Several EU directives emphasize the need to reduce the input of nutrients to both freshwater and marine ecosystems (e.g., the Water Framework Directive, the Marine Strategy Framework Directive and the Nitrate Directive). This is also an important part of the Helsinki Commission (HELCOM) Baltic Sea Action Plan, which contains several suggested measures targeting nutrient losses from agricultural land. Wetland creation is one of these, as it is known that the biogeochemical transformations that occur in wetlands generally result in a reduction in the nutrient content of the water flow.

Wetland creation has been practiced, e.g., in Sweden, where wetlands have been constructed and restored on a fairly large scale since the 1990s—initially focused on nitrogen removal and biodiversity enhancement. Nitrogen was usually assumed to limit primary production in marine ecosystems [2, 3], and also in the brackish-water Baltic Sea [4] which catchment includes most of Sweden. However, this is a disputed assumption, and some scientists have the opinion that phosphorus is ultimately limiting production in the Baltic Sea [5–7]. In freshwater bodies eutrophication is usually thought to be controlled by phosphorus inputs only [6], although this is also somewhat controversial and some scientists argue that nitrogen inputs to lakes have to be reduced as well [5]. Thus, singling out one nutrient or the other as limiting in the marine and freshwater system, respectively, may simplify reality too far. In many systems both nitrogen and phosphorus are limiting depending on time of year and location [8]. Therefore, whether the major concern is the marine environment or freshwater ecosystems, quantifying the effect of measures to remove both nitrogen and phosphorus from water is very relevant.

Created wetlands can be of different types, and are usually classified as free water surface constructed wetlands (FWS), horizontal subsurface flow constructed wetlands (HSF), and vertical flow constructed wetlands (VF) [9]. FWS wetlands are usually between 0.1 and 2 m deep, with a plant community that can be composed of algae, and submersed, floating or emergent wetland plants. HSF constructed wetlands are typically designed with a

permeable filter material (“soil”) planted with emergent wetland plants. Water flows horizontally in and beneath their rhizosphere, which creates a mix of saturated anaerobic, and unsaturated aerobic zones. A VF wetland is similarly constructed, but water is applied on the surface of the filtering material, and percolates through the rhizosphere. This results in a typically unsaturated aerated “soil”. When wetlands are restored, interventions are typically made to recreate previously drained, or by other means altered, natural wetlands. There are numerous studies of the physical and biogeochemical processes involved in the removal of both nitrogen and phosphorus. These are therefore relatively well known. When comparing different studies, the removal efficiency varies considerably, which makes it difficult to assess the extent to which wetland creation is an efficient measure to reduce eutrophication. Several previous studies [10–18] (cited in the protocol for this systematic review [19]) have indicated that removal differences between wetlands are related to a number of different factors such as inflow concentration, load variations, hydraulic retention time, temperature, hydraulic efficiency, and type of wetland. This calls for a systematic review of removal rates and how they are influenced by the wetland characteristics, loading differences, and environmental factors.

In Sweden, thousands of hectares of wetlands have been financed through various governmental funds and created as a means of reaching the Environmental objectives “Thriving wetlands” and “Zero eutrophication”, and to meet the commitments made in the Baltic Sea Action Plan (BSAP) [20]. The nutrient removal effects of those investments have been estimated using both statistical models and a catchment hydro-chemical model [21–24]. In general, surprisingly low mean removal rates have been reported, and this is mainly attributed to the fact that many wetlands were created as part of biodiversity conservation efforts and thus receive low nutrient loads.

In a recent evaluation, Weisner et al. [25] assessed the nutrient removal effects of around 5300 ha wetlands created within the Rural Development Program 2007–2013. Previously used calculation models were modified based on new monitoring data, and it was estimated that those wetlands would remove $0.35\text{--}0.6\text{ g m}^{-2}\text{ year}^{-1}$ of phosphorus and $3.2\text{--}4.6\text{ g m}^{-2}\text{ year}^{-1}$ of nitrogen. If only

looking at the group of wetlands that were created with nutrient removal as the main goal, the removal was about ten times higher or $3\text{--}4.5\text{ g m}^{-2}\text{ year}^{-1}$ for phosphorus and $30\text{--}34\text{ g m}^{-2}\text{ year}^{-1}$ for nitrogen. Based on the relatively small number of monitoring data sets that exist, the authors concluded that it would be possible to achieve a removal of around $10\text{ g m}^{-2}\text{ year}^{-1}$ of phosphorus and $100\text{ g m}^{-2}\text{ year}^{-1}$ of nitrogen in individual wetlands, provided they are located in optimal locations and with a design adapted to achieve high nutrient removal rates. For non-point source (mostly agricultural and urban runoff) nutrient removal in eastern USA, Mitsch et al. [26] suggested that sustainable removal rates would be $10\text{--}40\text{ g m}^{-2}\text{ year}^{-1}$ for nitrogen and $0.5\text{--}5\text{ g m}^{-2}\text{ year}^{-1}$ for phosphorus, and based this on wetland studies in the last quarter of the 20th century.

The different models used in the cited evaluations are all based on assumed relationships between removal rates and loads. Although not directly comparable, the results clearly indicate that the removal rates for nitrogen and phosphorus in created or restored wetlands span over a wide range. Published research studies also show that even though there is a generally positive relationship between removal rate and loading rate, the variation is quite large between wetlands with similar loads. It is thus unclear to what degree created wetlands will contribute to fulfilling the different Swedish environmental goals related to eutrophication. In this context, the Swedish Board of Agriculture and the Swedish Agency for Marine and Water Management were interested in obtaining a comprehensive evaluation of measured retention rates in individual wetlands. They were also interested in obtaining a coherent picture of how different wetlands function in a variety of conditions to facilitate planning of more effective water pollution control. In addition, the wastewater treatment industries are important stakeholders, as they will most likely be expected to have to comply with even stricter regulations on nutrient emissions than today. Created wetlands may prove to be a cost-efficient polishing step to reach those regulation demands.

Objective of the review

The objective of this review is to quantify observed removal rates of nutrients in created or restored wetlands, and to examine the distribution of these rates and quantify the variation between different studies. The primary question this review seeks an answer to is *“How effective are created or restored freshwater wetlands for nitrogen removal and phosphorus retention?”* This question implicitly includes the relationship between removal and load, as effectiveness can be expressed both as removal rates ($\text{g m}^{-2}\text{ year}^{-1}$) and relative removal (% of load). In this review, removal refers to a reduced amount

of nitrogen and phosphorus in the water phase. Removal processes imply transformation of the nutrients to other forms; for nitrogen this often means nitrogen gas and a smaller proportion of the greenhouse gas nitrous oxide, which will be emitted to the atmosphere.

Secondary questions are related to how various effect modifiers, such as environmental conditions and wetland characteristics, influence the nutrient removal rates. For that reason this review covers a fairly wide range of climatic conditions, although the performance of wetlands in temperate and boreal regions is most relevant to the stakeholders in Sweden. The review will not engage in detailed investigations of various removal processes and mechanisms but rather treat each wetland as a “black box”. This of course introduces some uncertainties. However, when assessing study quality, studies presenting complete nutrient budgets where removal by individual processes had been quantified were rated higher than studies merely providing inlet/outlet data. A fairly balanced budget may indicate that the numbers are reasonably accurate and/or that no major source or sink have been overlooked. The structure of the primary question is further discussed in the section Study Criteria, where information on relevant subjects, interventions, comparator and outcomes is given in more detail.

Methods

Searches

Searches for literature were made in ten different literature databases. The searched fields and search date are shown for each database in Table 1. Specific search strings for each database is shown in Additional file 1. No particular time, document type or language constraints were applied. However, at a later stage it was decided that articles in Chinese should be excluded due to the lack of translation resources.

Grey literature was searched for using the search engine Google where simplified search strings were used. In addition to searches where English search terms were used, searches were also performed using Swedish, Danish, and Dutch search terms. Search terms used for the Google search engine are shown in Additional file 1. Searches were performed 2014-03-03.

In addition, the websites of relevant specialist organisations (listed below) were also searched. Where possible the same search string as for Google was used (in relevant language). Generally, the first 100 hits were examined in the searches using Google and on specialist websites.

- Swedish Environmental Protection Agency (SEPA)
- Swedish Board of Agriculture
- The Swedish Agency for Marine and Water Management

Table 1 Electronic databases used for searching

Database	Searched field	No of hits	Date
ISI web of science	Topic	3249	2013-02-26
Georef and geobase	Subject/title/abstract	2249	2013-02-26
Scopus	Title, abstract and keywords	2842	2013-02-26
Agricola	Keyword anywhere	524	2013-02-26
ASFA	All fields (no full text)	1933 ^a	2013-02-27
Academic search	Title/abstract/subject/keyword	1131	2013-02-27
Biological abstracts	Keywords	401	2013-02-27
Wiley online library	Abstract or title or keywords	174	2013-02-27
Directory of open access journals	All fields	84 ^a	2013-03-05
ScienceDirect	Abstract or title or keywords	876	2013-07-02

^a After removal of duplicates

- Swedish directory of Master thesis (DiVA)
- South Florida Water Management District
- U.S. Environmental Protection Agency (EPA)
- North American Data Base (NADB)
- U.S. Department of Agriculture (USDA)
- Foundation for Applied Water Research (STOWA)
- Ekologgruppen i Landskrona AB
- Norwegian Institute for Agricultural and Environmental Research (Bioforsk)
- Danish Centre for Environment and Energy (DCE)
- European Environment Agency (EEA)
- Wetland Solutions Inc.
- Wetlands International
- Finnish Environment Institute (SYKE)
- Federal Environment Agency (UmweltBundesAmt, Germany)
- Stichting Toegepast Waterbeheer (STOWA, The Netherlands)

To test the comprehensiveness of the searches, bibliographies of review articles were examined and compared with the search results.

Study inclusion criteria

In this review some constraints regarding the type of water entering the wetlands have been applied. Untreated wastewater was not considered since it is not permissible to discharge such water into the environment in most European countries. Industrial or agricultural wastewater can vary considerably in composition, and was therefore also excluded. Farmyard runoff was in most cases classified as agricultural wastewater, and thus excluded, since it is often mixed with untreated parlour washings and silage/farmyard manure effluents, among other things.

The inclusion criteria below were developed in collaboration with stakeholders during development of the review protocol [19].

- *Relevant subject*: Secondary or tertiary treated domestic wastewater, urban storm water, stream/river water, freshwater aquaculture effluents, and runoff from agricultural fields. To be able to make a distinction between treated and untreated wastewater, a guideline value of 100 mg/l was used for the highest acceptable concentration of BOD in the water entering the wetland.
- *Types of intervention*: Creation or restoration of wetlands. In this review, creation of a wetland refers to the construction of a wetland on a site that never was a wetland, regardless of the main purpose of the wetland. Sometimes the term Constructed wetland is used for a wetland with the specific aim of treating wastewater, storm water, acid mine drainage, or agricultural runoff. Such wetlands are thus regarded as Created wetlands. Created wetlands include both horizontal and vertical subsurface flow systems and free water surface systems. Restoration refers to recovery of ecological and hydrological processes as well as geomorphology in areas where natural wetlands previously have been drained or by other means altered. In this review the term restoration does not imply any specific purpose of the wetland. To be included the created or restored wetlands must host some type of vegetation.
- *Types of comparator*: No intervention (inlet conditions can serve as control).
- *Types of outcome*: Mass removal of total nitrogen (TN) or total phosphorus (TP) from the water body per unit wetland area and year. Removal efficiency of TN or TP (% of load).
- *Types of study*: The most common way to evaluate the overall retention rate in a wetland is to compare the nitrogen or phosphorus loads in the inlet and outlet water, respectively. Quite often the retention in wetlands is evaluated in experiments where effect

modifiers such as loading rate or vegetation type are varied. This is a version of a control-impact (CI) study where inlet conditions serve as control. In rare cases, nutrient loads in a river or stream have been recorded both before and after the establishment of a wetland, which corresponds to a typical before–after (BA) study. Both types of studies are eligible.

The removal rates and efficiencies may show large seasonal variation. Therefore, to be included in the review, it is a prerequisite that the wetland is established in field conditions and exposed to the ambient climate. This means that laboratory and greenhouse studies were excluded and that each study must cover at least one complete annual cycle. Also, in order to reflect realistic conditions the wetland must be of a reasonable size. While typical microcosm studies were excluded, mesocosm studies were included since they potentially provide valuable information on the variability of the outcomes based on true replicates. A cut-off wetland size of 1 m² was applied.

This systematic review is focused on boreal and temperate regions, but for comparison sub-tropical regions have also been included. In the Köppen-Geiger climate classification system [27] this corresponds roughly to group D (snow climates), group C (warm temperate climates) and parts of group A (Equatorial climates with one dry season, i.e., As and Aw). Studies on wetlands located in other climates were excluded. Furthermore, to be included the studies must have taken all removal processes into account. Studies that only report results for selected processes were excluded.

Wetlands may be created or restored for purposes other than nutrient removal. Although some wetlands serve multiple purposes [28], the main purpose is in some cases to promote biodiversity or reduce flood risks. In this review wetlands have been considered regardless of the main purpose of the wetland, i.e., inclusion and exclusion was not based on the reasons for constructing or restoring the wetlands. However, the main purpose of the wetland was recorded during data extraction.

Screening process

Articles found in the searches were checked for relevance at (1) title and abstract, and (2) full text levels. At the title and abstract level, the first author of this review performed the screening of all articles. To check that the screening was consistent and complied with the agreed inclusion/exclusion criteria, a subset of 106 articles was also screened by another author. A second subset of 106 articles was screened by two other authors, and a third subset of 107 articles was screened by a third pair of authors. In this way 319 articles were double-screened.

Thus, we checked the consistency between the main screener and the other authors as well as between the authors within each screening pair. To evaluate the consistency Kappa tests were used.

Full-text articles were randomly and evenly distributed for screening among six authors. However, before full-text screening at full scale, three subsets of 34 articles in each were double-screened in the same manner as at the title and abstract level. Again, Kappa tests were used to test the consistency between the authors. Kappa values of 0.6 or higher were considered acceptable both at title and abstract screening and at full-text screening.

Potential effect modifiers and reasons for heterogeneity

The nutrient retention may vary considerably between different studies. The anticipated large variation is easy to understand in the light of the fact that the removal rate is a result of several independent processes. Nitrogen removal takes place through: (1) sedimentation and sediment accretion, (2) plant uptake, and (3) denitrification and volatilization. The processes involved in phosphorus removal are: (1) sedimentation and sediment accretion, (2) plant uptake (3) sorption and (4) precipitation/coprecipitation. The success of each of these mechanisms may depend on factors such as:

Loading characteristics

Hydraulic loading rate (HLR)

Concentration and speciation of nitrogen and phosphorus at the inlet

Wetland characteristics

Type of wetland

Size and shape (area, depth, length)

Flow pattern and hydraulic efficiency

Hydroperiod

Age

Sediment/soil type

Oxygen concentration and redox potential

Vegetation type and coverage

Fauna

Management methods and frequency

Climate characteristics

Mean temperature

Ice coverage

Study quality assessment

Studies included after the full text screening were subject to critical appraisal. Bilotta et al. [29] suggested a

systematic method for critical appraisal using an Environmental-Risk of Bias Tool for assessing the internal validity, and an Environmental-GRADE tool for assessing the overall quality of a study. Although the major part of our critical appraisal was performed before the article by Bilotta et al. [29] was published, we have applied a fundamentally similar approach. Concerning risk of bias the following conditions specific to this review should be noted. (1) Wetland studies included in the review have been designed in such a way that selection bias due to inadequate randomisation or inadequate allocation concealment is not an issue. However, another type of selection bias could potentially be introduced if samples were to be taken only at certain favourable (or unfavourable) conditions, e.g., at low hydrological loading rates or during growing seasons. To assess the risk for this type of selection bias the study length and sampling frequency were evaluated. Studies should cover complete annual cycles and, as a guideline value, include at least 12 sampling occasions. Furthermore, it has to be taken into account that the investigated wetlands do not form a random sample of a well-defined population of potential created wetlands. (2) Performance bias is mainly caused by exposure to factors other than the intervention and may be related to e.g., hydrological flow paths not accounted for or use of chemicals added to the water or soil in order to promote certain processes in the wetland. To assess the risk of bias caused by hydrological processes the hydrological mass balance was evaluated. (3) Detection bias is mainly related to sampling and analytical methods, and using different methods for treated and untreated water is most unusual. (4) Attrition bias may occur if there for some reason are fewer samples of either treated or untreated water compared to the other group. In some cases differing number of samples may however be justified. For example, during dry conditions it is not unusual that the hydraulic loading rate at the outlet is zero due to e.g., evapotranspiration while there still is an active water inflow to the wetland. This could potentially result in fewer samples of treated water than of untreated water. Provided that the water mass balance adds up reasonably well, attrition bias may not occur in such cases since the outcome measure is based on total nutrient mass transport during complete annual cycles. (5) The risk of selective reporting bias is fairly easy to assess by comparing e.g., study length with the number of years reported, or the measured outcomes with the reported outcomes.

When assessing the overall quality of the studies they were assigned to either of three quality categories: (1) Does not meet the quality criteria, (2) Acceptable, and (3) High standard. Studies in category 1 did not qualify for data extraction and quantitative synthesis. All studies relevant for this review (i.e., passed the screening at

full text level) were observational studies, predominantly control-impact (CI) studies where inlet conditions served as control. Based on this study design, all studies were by default assumed to be of acceptable quality (assigned to category 2). A set of quality criteria (see Table 2) were then used to justify upgrading to category 3, downgrading to category 1, or keeping the study in category 2.

True replication is unusual in studies of nutrient removal in wetlands. In most cases just one wetland, or a set of different wetlands, were studied. However, repeating the measurements in the same wetland during several complete years may be regarded as quasi-replication. We regarded this form of quasi-replication acceptable for meta-analysis.

The criteria for assigning the studies to either of the three quality categories are shown in Table 3.

Four reviewers performed the critical appraisal. To check the consistency between the reviewers a small number of articles were critically appraised by all reviewers. Differences were then discussed by all reviewers and, as a result of that discussion, the quality criteria were clarified. During the remainder of the critical appraisal the reviewers had an option to code studies as “uncertain”. All studies with that code were then discussed at a meeting where consensus was reached.

Table 2 Quality criteria and requirements for fulfillment

Quality criteria	Requirements
1 Moderate risk of bias	The study covered at least one complete annual cycle and included at least 12 sampling occasions evenly distributed in time The hydrologic data obtained in the study was adequate for calculating hydraulic loads at inlet and outlet There were no obvious deficiencies in the water balance
2 High sampling frequency	Samples were flow proportional composites or grab samples with a reasonably high sampling frequency (at least twice a month)
3 Hydraulic control	All major water flow paths were quantified and considered in the calculations of the outcomes
4 Removal processes quantified	At least one important removal process was quantified independently, and the result was reasonable in relation to the inlet/outlet difference
5 Effect modifiers	Sufficient information related to potential effect modifiers was reported
6 Replication	Replicate wetlands were studied, or the same wetland was studied for two or more complete annual cycles. If not reported it should be possible to calculate the variance of the replicates

Table 3 Quality categories and criteria for assignment

Quality categories	Criteria
Category 1	High risk of bias (quality criteria 1 is not fulfilled)
Category 2	Moderate risk of bias (quality criteria 1 is fulfilled)
Category 3	(1) Moderate risk of bias (quality criteria 1 is fulfilled) and (2) At least four of the six quality criteria listed in Table 2 are fulfilled and (3) Quality criteria 3 and 6 are fulfilled

Data extraction strategy

The outcomes evaluated in this review are the removal rate and removal efficiency of total nitrogen and phosphorus; typically the results are reported quantitatively as $\text{g m}^{-2} \text{ year}^{-1}$ and as % of load, respectively. Results reported in other units were recalculated where possible. In cases where multiple-year studies just reported an overall average without the inter-annual variance, we extracted data for each sampling occasion and calculated annual values for each separate year and the inter-annual variance where possible.

In order to assess the quality of the studies and to be able to evaluate the importance of various effect modifiers, additional data was recorded as well (see Table 4). Not all studies provided information on all parameters shown in Table 4. For instance, a very small number of studies reported data on fauna in the wetland. It is

however quite possible that nutrient cycling and retention are influenced by, e.g., benthic organisms through bioturbation [30] or by birds [31, 32]. Also, planktivorous fish species can feed intensively on zooplankton and thereby protect phytoplankton from being grazed, leading to turbid water [33]. The parameters for which data was extracted from all included studies are indicated in Table 4 with underscored text.

The four reviewers who performed the critical appraisal extracted data related to wetland characteristics. A fifth reviewer extracted all other data. To make the data extraction as consistent as possible the data was entered into a pre-designed Excel spreadsheet. All reviewers tested the spreadsheet and after some minor modifications it was used throughout the entire data extraction process.

Data synthesis and presentation

The studies with true or temporal replication were subjected to meta-analyses. Log response ratios ($\ln R$) where $R = \text{Load}_{\text{out}}/\text{Load}_{\text{in}}$ were used to quantify effect sizes, and random effects models [34] were used to calculate summary effects and uncertainty bounds of such effects. The between-study variance (τ^2) was estimated by calculation of T^2 using the DerSimonian and Laird method, and to estimate the ratio of true heterogeneity to total variance in observed effects, the I^2 statistic was calculated [35]. In subgroup analyses, separate estimates of τ^2 were made for each individual subgroup. The results were presented

Table 4 Type of extracted data (underscored parameters were recorded for all included studies)

Loading characteristics	Wetland characteristics	Climate characteristics	Study design parameters	Results
Type of water	Type of wetland	Annual mean air temperature ^b	Study length	Annual average removal rates ($\text{g m}^{-2} \text{ year}^{-1}$)
Concentrations of TN and/or TP at inlet	Length, width, depth	Annual mean water temperature	Number of replicates	Annual average removal efficiency (% of load)
Speciation of N and/or P at inlet	Area	Ice coverage	Sampling frequency	Variance of annual removal rate
Average nutrient loading rates	Hydroperiod (d year^{-1})	GPS coordinates	Flow proportional sampling (Y/N)	Variance of annual removal efficiency
Average hydraulic loading rates	Dominant plant species and coverage	Climate type ^c		
Average residence time (d)	Sediment/soil type			
Hydrologic regime	Land use history and wetland age			
pH	O ₂ concentration and redox potential			
	Fauna			
	Management methods			
	Main purpose of wetland			
	Other characteristics ^a			

^a Examples are, e.g., specific substrates, soil amendments, configuration of multiple wetlands, or other means of enhancing the outcome

^b Extracted from WorldClim database (<http://www.worldclim.org/>)

^c According to the Köppen-Geiger climate classification. This type of data was in most cases not provided by the primary studies and was therefore extracted from the map database referred to by [26]

in forest plots and in tables where the log response ratios had been back-transformed and recalculated to median removal efficiencies and confidence intervals of median removal efficiencies.

To examine how the removal of nitrogen or phosphorus was influenced by effect modifiers all included studies, even those without replication, were subjected to response surface analyses using various regression models. The removal of nitrogen and phosphorus, expressed as removal efficiency and removal rate, were the primary target variables. The removal efficiency was defined as

$$\begin{aligned} \text{Removal}_{\text{efficiency}} &= 100 \times \frac{\text{Substance}_{\text{flow}_{\text{in}}} - \text{Substance}_{\text{flow}_{\text{out}}}}{\text{Substance}_{\text{flow}_{\text{in}}}} \\ &= 100 \times \left(1 - \frac{\text{Substance}_{\text{flow}_{\text{out}}}}{\text{Substance}_{\text{flow}_{\text{in}}}} \right) \end{aligned}$$

which is a monotonic function of the response ratio R .

The removal rate was defined as

$$\text{Removal}_{\text{rate}} = \text{Load}_{\text{in}} - \text{Load}_{\text{out}}$$

i.e., the mass removed per unit time and per wetland area.

The hydraulic loading rate and concentrations of nitrogen and phosphorus in the inflow to the wetlands were considered to be the primary predictors or effect modifiers. Attention was also paid to type of wetland, type of inflow, climate zone, average air temperature, and wetland area. Cross-validations were not carried out because the number of wetlands included in our analysis turned out to be smaller than expected.

The relationships between the mean output-input ratio and potential predictors were assumed to have a multiplicative structure, e.g.,

$$\begin{aligned} \text{Output}_{\text{input}}_{\text{ratio}} \\ = a(\text{Hydraulic}_{\text{loading}}) b(\text{Concentration}_{\text{in}}) c(\text{Climate}) \end{aligned}$$

that after taking logarithms can be rewritten as a general additive model (GAM) of the form

$$\begin{aligned} \text{LOG}(\text{Output}_{\text{input}}_{\text{ratio}}) &= f(\text{LOG}(\text{Hydraulic}_{\text{loading}})) \\ &\quad + g(\text{LOG}(\text{Concentration}_{\text{in}})) \\ &\quad + h(\text{Climate}) \end{aligned}$$

where a , b , f and g are assumed to be smooth functions, and $c(\text{Climate})$ and $h(\text{Climate})$ are functions of climate zone indicators or average air temperature. The error terms on the log scale for different wetlands were assumed to be statistically independent and normally distributed with mean zero and constant variance.

The removal rate was assumed to be a smooth function of the magnitude of the hydraulic loading and the concentration of the chemical element or species under consideration. The basic model had an additive structure

$$\begin{aligned} \text{Removal}_{\text{rate}} &= f(\text{LOG}(\text{hydraulic}_{\text{loading}})) \\ &\quad + g(\text{LOG}(\text{inflow}_{\text{concentration}})) \\ &\quad + h(\text{airtemperature}) \end{aligned}$$

Potential pairwise interaction effects of the predictors were taken into account by allowing thin plate splines (TPS) in the GAM models. Such splines encompass a very large class of smooth functions (response surfaces) that enable a very flexible description of both main effects and interaction effects of any pair of predictors. The error terms for different wetlands were assumed to be statistically independent and normally distributed with mean zero and constant variance.

GAM models with or without thin plate splines were fitted to the collected data using standard least squares algorithms in the software package SAS. Fitted values of the removal efficiency were obtained by first using $\text{LOG}(\text{Output}_{\text{input}}_{\text{ratio}})$ as target variable and then back-transforming the fitted values to the *Percentage_{removed}*. Because this transformation is nonlinear the back-transformed surfaces should be interpreted as estimates of the median removal efficiency. Fitted values of the removal rate were obtained by directly fitting GAM models to this target variable and various sets of effect modifiers.

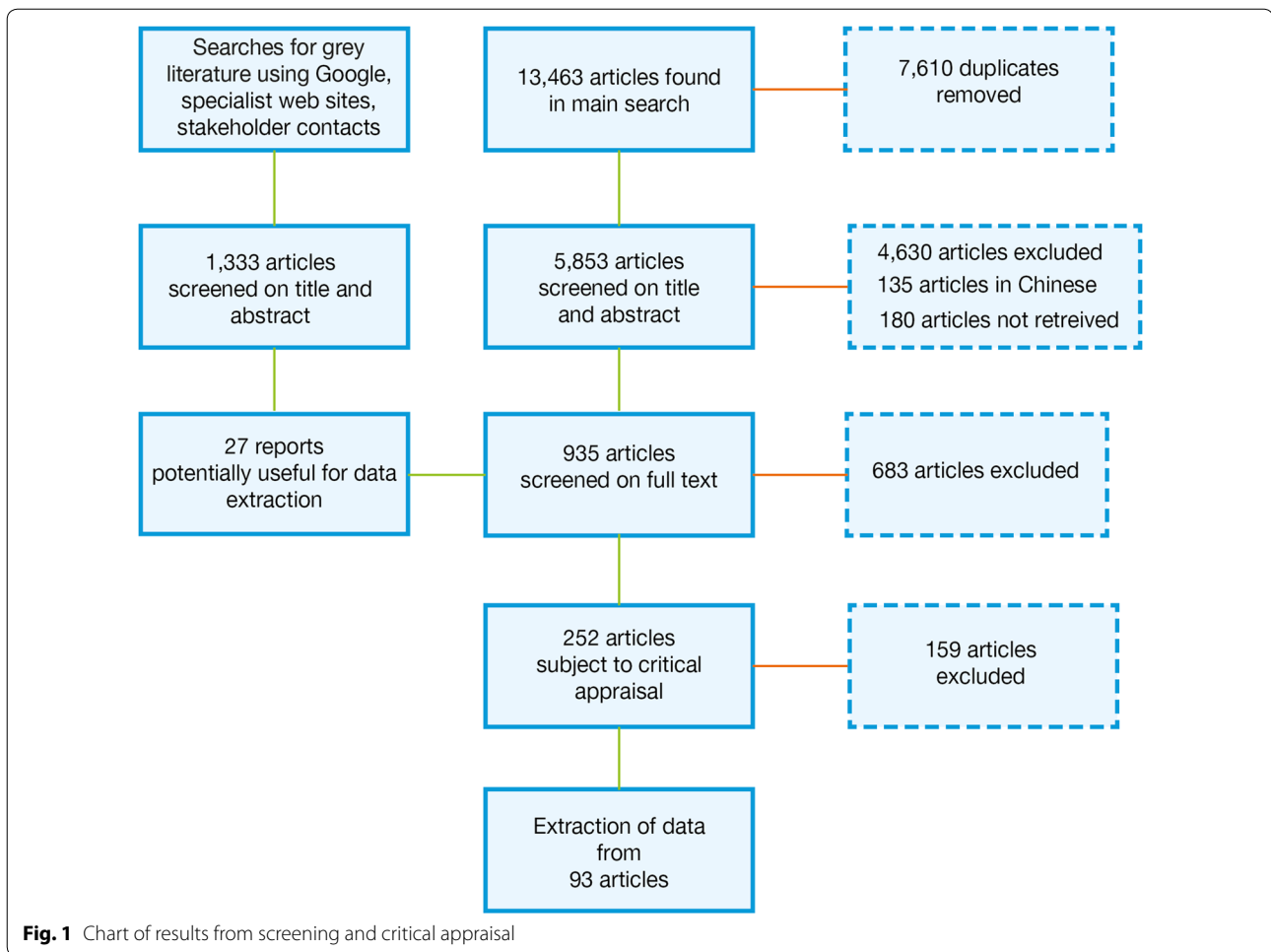
Unless otherwise stated, the target variables as well as the predictors in the response surfaces analyses were temporal mean values for each of the studied wetlands. This simplified the modelling and justified the assumption that the underlying data were statistically independent. Separate analyses of studies with temporal replicates were carried out to reveal potential drawbacks of GAM modelling.

Results

Review descriptive statistics

The searches were performed in February 2013. A flow chart of the screening process is shown in Fig. 1. The search in literature databases generated 13,463 records, of which 5853 records were unique.

At the title and abstract level 4630 articles were excluded while 1223 articles were included for full text screening. However, 135 of the included articles were in Chinese and had to be excluded due to lack of translation resources. Also, 180 of the articles could not be retrieved due to limitations in library resources. This resulted in 908 articles available for full text screening. Searches for grey literature using Google, specialist websites, and stakeholder contacts added another 27 reports that were screened at full text level. As a result of full text screening 685 articles were excluded. A list of these articles, with an indication of the reason for exclusion, is shown in Additional file 2. Note however that only one reason for exclusion is shown for each article, although in many cases



there were actually multiple reasons. The most common reason for exclusion was that the studied subject (type of water) did not conform to the inclusion criteria, followed by a lack of desired outcome data (Table 5). A total of 252 articles were subject to critical appraisal, and 93 of these passed to full data extraction.

The oldest article included in this review was published in 1981. The number of articles from each year was fairly constant between 1993 and 1999, after which it started to increase from 1 to 2 articles per year to around 5–9 articles per year (Fig. 2).

In total, 203 wetlands are included in this systematic review. Consecutive wetlands in series were treated as separate wetlands if removal rates and efficiencies were reported or could be calculated for each individual wetland. In other cases the train of wetlands was treated as one single wetland. Most of the wetlands are located in USA ($n = 110$) and Europe ($n = 64$). Seventeen wetlands are located in Sweden. The locations of the wetlands are shown in Fig. 3, and in Additional file 3 the number of

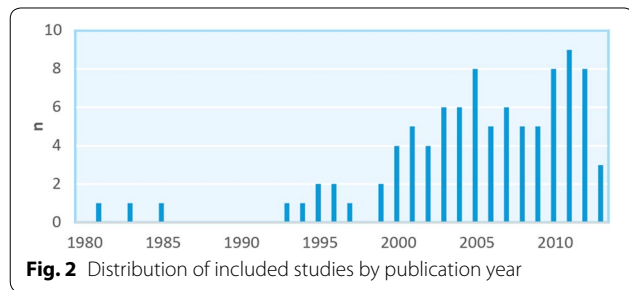
wetlands in each state or country is shown for nitrogen and phosphorus separately. The climate at these locations ranges from subtropical (climate zone Aw) in Florida to snow climate in parts of Scandinavia, northern USA and Canada (zone Df) and South Korea (zone Dwa).

The number of specific wetland types, inlet water types, water regimes, vegetation types, and climate zones are shown in Tables 6, 7, 8, 9 and 10. The most common wetland in this review is a free water surface (FWS) wetland with emergent vegetation treating agricultural runoff with a variable hydraulic loading rate. All of the included wetlands were primarily created or restored for the purpose of nutrient removal, although a small number of of them were multi-purpose wetlands.

The size of the included wetlands ranges from 1 to 10^7 m². As shown in Fig. 4, most of the studied wetlands were between 10^3 and 10^5 m², but smaller wetlands in the range 1–10 m² are also well represented.

Table 5 Number of articles (n) excluded at title and abstract screening and full text screening

Title and abstract screening	n	Full text screening	n
Exclude on intervention	1655	Exclude on subject	191
Exclude on outcome	1432	Exclude on outcome	164
Exclude on subject	719	Exclude on study length	143
Exclude on process study	467	Exclude on secondary data	81
Exclude on study length	156	Exclude on intervention	69
Exclude on secondary data	116	Exclude on process study	22
Exclude on location	75	Exclude on location	16
Exclude on no annual data	9	Abstract only	9
		Exclude on no annual data	4



Study quality assessment

After full-text screening 252 articles remained for quality appraisal (Fig. 1). Some (15) articles described the same wetland study more than once, adding more data after the first study was published. In such cases the most

recent and comprehensive article was chosen for quality assessment and potential data extraction.

Critical appraisal resulted in 143 articles in Category 1, i.e., articles that were not used for data extraction. From some of these articles it was not possible to calculate an annual total P and/or total N removal, expressed as mass/unit area/year. Even if removal rates were presented, or could be calculated, several articles were assigned to the lowest quality category due to deficiencies in the water budget. We deemed a high quality water budget (no large water flows unaccounted for) to be a necessary prerequisite for the study to be useful for data extraction. This is because calculations of nutrient removal are sensitive to uncertainties in water flow measurements. Another important reason to reject studies was measurements lasting less than a year, or more than a year, but without possibilities to break down measurements into a full year or several individual years. To be able to compare the annual efficiency of wetlands as nutrient traps, at least one annual cycle was demanded. The annual removal capacity is of main interest for stakeholders.

One of the reasons why quite a number of studies had to be excluded was the lack of a detailed description of the study design and sampling methodologies. Studies, which at first glance were expected to be suitable for extraction, proved to have insufficient detail for a proper quality appraisal and could not be included in the highest quality classes. Such shortcomings in methodological descriptions have been reported commonly in other systematic reviews and metadata analyses as well [36]. A more critical attitude of authors, reviewers and editors of scientific publications is desirable.

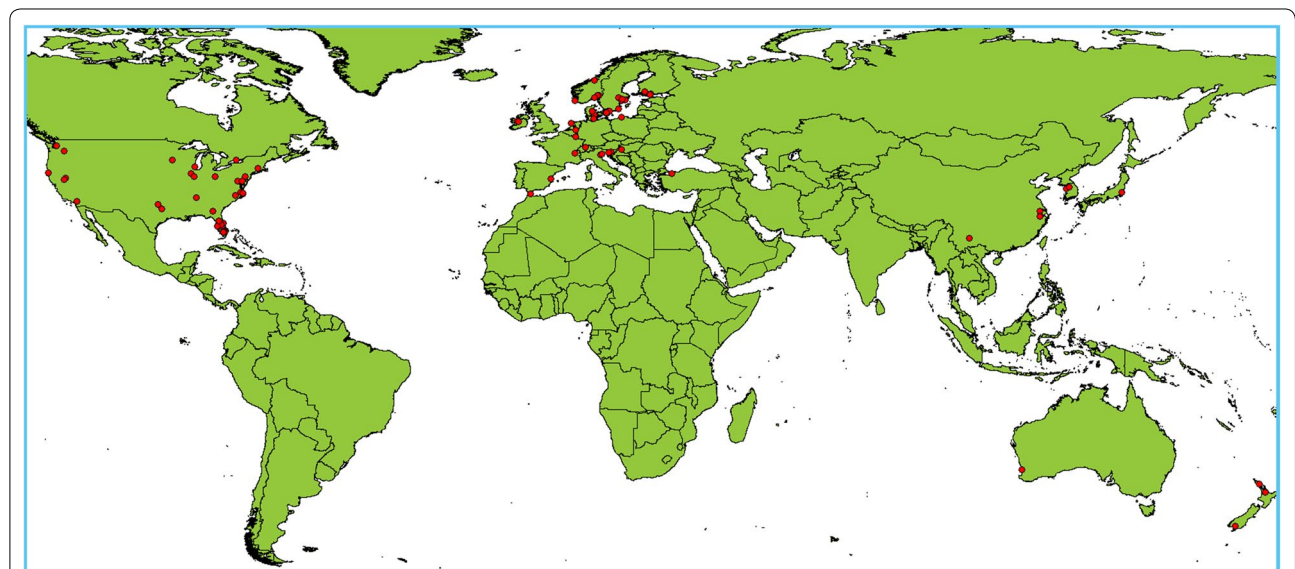


Fig. 3 Location of included wetland studies

Table 6 Number of included wetlands by wetland type

Wetland type	Description	n
Free water surface	Wetland with visible water flowing horizontally	163
Horizontal subsurface flow	Wetland where water flows horizontally below the soil/substrate surface	17
Vertical subsurface down	Wetland where water flows vertically downwards through a soil/substrate	9
Combined horizontal	FWS wetland combined with other units such as sedimentation pond or overflow unit	8
Riparian	Any wetland at the interface between land and a river or stream	5
Vertical subsurface up	Wetland where water flows vertically upwards through a soil/substrate	1

Table 7 Number of included wetlands by inflow type

Inflow type	n
Agricultural runoff	100
River/lake water	38
Secondary domestic wastewater	29
Tertiary domestic wastewater	24
Urban storm water/runoff	7
Agricultural + urban runoff	4
Secondary domestic wastewater + urban runoff	1

Table 8 Number of included wetlands by water regime

Water regime	Description	n
Continuous, variable	Continuous flow with varying hydrologic loading rate (HLR)	72
Precipitation-driven	HLR driven by precipitation, continuous or intermittent but always variable	61
Constant	Continuous and constant HLR	43
Intermittent, constant	Periods with HLR = 0. When HLR > 0 it is constant and fixed at one value	14
Intermittent, variable	Periods with HLR = 0. When HLR > 0 it is variable	13

Table 9 Number of included wetlands by vegetation type

Vegetation type	n
Emergent	117
Mixed	28
Submerged	24
Unspecified	23
Trees	6
Filamentous algae	4
Floating	1

The difference between quality category 2 and 3 lies mainly in the demand for either two or more full years of measurements, or replicate wetlands, and that all major water flows should be quantified. For the studies in category 3 an annual mean value for nutrient removal, with standard deviation, is thus possible to calculate. A small number of studies (4) reported data from replicate wetlands, thus the majority of figures on variation in nutrient removal come from multiple year studies. This is a kind of pseudo replication, and variation in nutrient removal for the same wetland during different years may of course not reflect the true variation between replicate wetlands. On the other hand, compared to a 1-year study with true replicate wetlands, this pseudo replication is more likely to reflect changes in climatic conditions (e.g., temperature and precipitation), which may be equally important. Unfortunately, due to practical constraints, replication has only been done for mesocosm-size or experimental wetlands in wetland research parks. There were 41 studies of wetlands ≤ 10 m². For these, variation among replicates is likely to be smaller than for replicate “full-size” wetlands, i.e., wetlands restored or created not primarily for research purposes, but to reduce nutrient transport in a catchment. Such “full-size” wetlands are expected to produce more realistic figures on nutrient removal, but will lack replication. Some studies report major individual nutrient removal processes. If these match the total removal, based on a black box approach, then the quality of the study is strengthened. However, there were only a small number of studies (≈ 10) providing full nutrient budgets. Thus, from this review we cannot draw conclusions with respect to nutrient removal mechanisms, only the magnitude of nutrient removal, using a black box approach. Some indirect evidence on removal mechanisms may be indicated through statistical evaluations of effect modifiers, e.g., temperature and vegetation.

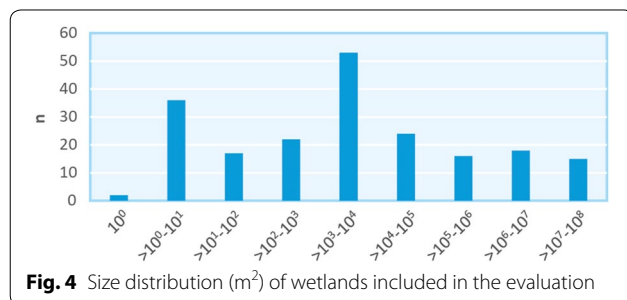
All studies included at the quality appraisal stage had at least one potential effect modifier reported, or such a modifier could be quantified from other sources (e.g., mean temperature from climate databases). However, many of the potential effect modifiers were reported in only a small number of articles, or were only semi quantitative. Thus a limited number of potential effect modifiers (c.f. Table 4) could be used in the final statistical meta-analysis and response surface analysis.

Of the 93 articles placed in category 2 and 3, 39 were in the highest quality category (3) and 54 in the second highest (2). Critical appraisal of individual studies, including assessments of internal validity (risk of bias) is reported in Additional file 4.

Although there remained a reasonable number (39) of studies in quality category 3 with low susceptibility to

Table 10 Number of included wetlands in different climate zones

Climate zone	Description	Example location	n
Aw	Equatorial savannah with dry winter	Southern Florida	58
Cfa	Warm temperate climate, fully humid, hot summer	Northern Florida, Washington DC, Bucharest	28
Cfb	Warm temperate climate, fully humid, warm summer	London, Paris, Berlin, Southern Sweden	50
Csa	Warm temperate climate with dry summer, hot summer	Barcelona, Rome, Athens (Mediterranean)	20
Csb	Warm temperate climate with dry summer, warm summer	Ankara, Genoa, Seattle, San Francisco	10
Cwb	Warm temperate climate with dry winter, warm summer	Kunming	4
Dfa	Snow climate, fully humid, hot summer	Chicago, Boston	18
Dfb	Snow climate, fully humid, warm summer	Stockholm, Ottawa, Moscow	13
Dfc	Snow climate, fully humid, cool summer and cool winter	Northern Scandinavia, Anchorage	1
Dwa	Snow climate, dry winter, hot summer	Beijing, Pyongyang	1

**Fig. 4** Size distribution (m²) of wetlands included in the evaluation

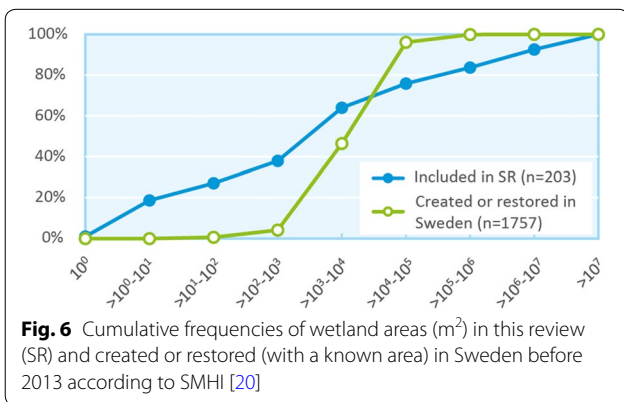
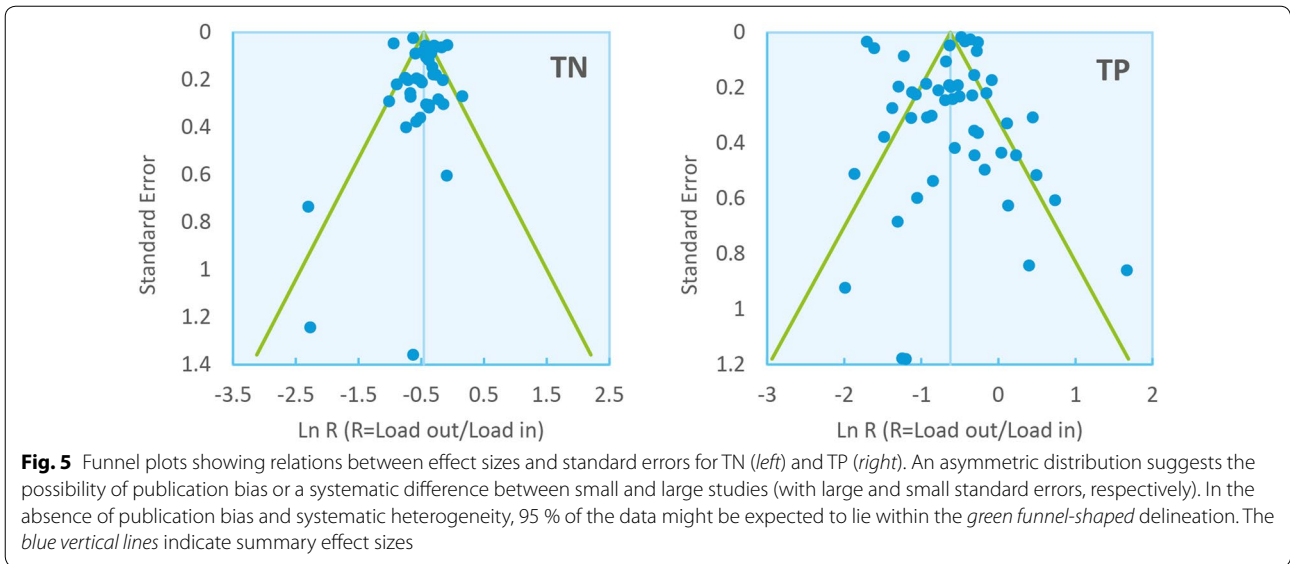
bias (high internal validity), there is a risk of various types of bias of the body of evidence, affecting conclusions from this review. One of these is publication bias which may arise if results in a particular direction are less likely to be published. Figure 5 shows funnel plots for TN and TP. For TN there may be a small number of studies missing in the lower right corner, indicating that some studies showing net releases of TN from the wetland, albeit most probably insignificant, may have been overlooked or not reported. The distribution for TP is more symmetric but does also reveal a small number of studies showing very large effect sizes with small standard errors. Thus, in the present case publication bias seems not to be a major concern. This finding is not surprising since both positive and negative results are scientifically interesting, and also of importance for stakeholders.

Another source of bias affecting the external validity, or the generalizability of the included studies, is uneven geographical distribution of wetlands (Fig. 3). Indeed studies are clustered in two geographical regions: most studies are from Europe and North America. In so far as the bias is towards North Europe, especially Scandinavia, North America and areas with similar climate, geographical bias may, from a Swedish stakeholder point of view, not be a problem. However, a fairly large number of studies, especially for phosphorus, have been conducted

in Florida and other states in Southern USA, which have a climate quite different from Sweden (Additional file 3).

The size distribution of included wetlands may also be biased. Whereas most created or restored wetlands in Sweden range from 10² to 10⁵ m² (Fig. 6), the wetlands included in this review show a much broader distribution with more of both smaller and larger wetlands. Almost 40 % of the included wetlands in this systematic review range between 1 and 10² m², and 17 % of the wetlands are between 10⁵ and 10⁷ m². The largest included wetland is 6.7 × 10⁷ m².

Most studies of nutrient removal in wetlands have been made during the years following wetland restoration or creation. Median wetland age at the start of study periods was 1 year for the included wetlands, whereas the median age at the end of the studies was 3 years. Thus our systematic review may be biased towards short-term nutrient removal effects. The performance of wetlands as nutrient traps after several decades (the expected minimum lifespan of a restored or created wetland) in comparison to the first years after restoration/construction has been studied in a very small number of cases. It might be that nutrient removal changes over time. It has, e.g., been reported that P removal decreases [17] and may even become negative with time. In a more recent study, Mitsch et al. [18] combined the data reported by Mitsch et al. [17] with data for two subsequent years and showed that although there was a significant declining trend for the entire period (1994–2010), there was also a significant improving trend at the end of the period (2003–2010). One explanation for this pattern could be that the wetlands were created on former agricultural soil and that it took a decade or so to wash the accumulated P out of the soil. Chen et al. [37] studied long-term (up to 17 years) TP removal in wetlands in Florida. The authors did not report any temporal trend but concluded that performance, in terms of outflow TP concentration and/or k value (first order removal constant, which could be



interpreted as the settling rate for TP removal), depended primarily on HLR, inflow TP concentration, and TP loading rate. In addition, the impacts of these variables on P removal are often confounded by soil and vegetation conditions, regional rainfall, management activities, and other factors. In another study conducted in Florida, Moustafa et al. [38] showed that the TP removal efficiency remained relatively constant during their 9-year study. Also N removal may be expected to change, e.g., due to succession in aquatic vegetation and accumulation of organic matter in wetland sediments.

Given the large number of studies fulfilling our rigorous quality criteria, we feel confident that the review has answered the primary question about wetland effectiveness (see section “Objective of the review”) in a general sense. However, the heterogeneity is high, so that it is not possible to use the graphs to estimate nutrient removal for an individual wetland. This was not unexpected since we have included studies from subtropical to cold

climates, sizes of wetlands from one to more than 10⁵ m², various types of created wetlands, wetlands receiving a wide variety of influent water quality, etc. On the other hand the statistically significant general trends found for all these different types of wetlands should represent a high quality outcome from the review.

With respect to the secondary questions about influence of effect modifiers, the results of the review for some effect modifiers are statistically solid, e.g., wetland size, hydraulic loading, temperature, water type, and substance concentration. However, for other potential effect modifiers (e.g., soil type, vegetation type and coverage, harvest, and fauna) there were not enough studies to make solid predictions (models).

We have chosen to include a rather wide variety of wetlands and water types. Thus the material is very heterogeneous, which increases variation. An alternative approach could have been to review only one wetland type and only one source of water in a more restricted geographical area, e.g., wetlands in agricultural areas in N Europe, receiving drainage from agricultural fields. This would probably have produced less variation, although the number of studies qualifying for the highest quality category would have been very small. Also, the possibility to generalize would have become reduced.

Narrative synthesis

Most studies on nutrient removal in wetlands do not report any variance in annual removal rates or efficiencies. In some cases this is a consequence of the fact that only one wetland was studied and that the study only lasted for one year. There is thus neither any true replication nor any quasi-replication. In other cases the study

of one wetland lasted for several years but only a long-term average was reported with no information about the inter-annual variance. Only four of the included studies investigated multiple wetlands that were similar enough to be regarded as replicates and also reported the results in such a way that it was possible to calculate the variance, while 60 wetlands were quasi-replicated through measurements for more than one year and reported in such a way that the inter-annual variance could be calculated. Total nitrogen was measured in two of the studies with replicated wetlands and TP was measured in three of them. Among the quasi-replicated wetlands, TN and TP were measured in 37 and 49 wetlands, respectively.

Removal of TN

The annual loading rates of TN in the included wetlands ranged from 2.1 to 2486 g m⁻² year⁻¹, and averaged 505 g m⁻² year⁻¹. The average removal rate of TN was 181 g m⁻² year⁻¹, whereas the average removal efficiency was 39 %. Summary statistics for included wetlands are shown in Table 11. Results and data for individual wetlands are shown in Additional file 5. The ranges in loading and removal rates between wetlands are quite wide, and the distributions are skewed to the right, i.e., the median values are lower than the arithmetic means. The distribution of removal efficiencies is more likely to be normally distributed. Although there is no significant difference in average removal efficiencies between category 2 studies and category 3 studies, the variability is smaller among category 3 studies. It is worth noting that none of the wetlands among the category 3 studies had negative removal rates.

One included wetland showed a small (non-significant) negative TN removal rate [39]. This was a multi-purpose FWS wetland that had been restored on formerly drained cropland. Other studies, e.g., by Koskiaho et al. [40], have also shown negative TN removal rates but these were judged to be highly susceptible to bias and assigned to quality category 1.

Removal of TP

As in the case of TN, the spans in the loading and removal rates of TP are quite large (Table 12). The

average loading rate and removal rate were 36 and 13 g m⁻² year⁻¹, respectively. The average loading rate was considerably lower in category 3 studies compared to category 2 studies, even though the range was similar in both categories. There is no significant difference in average TP removal efficiencies between category 2 studies and category 3 studies but, as with TN, the variability between wetlands was smaller among the category 3 studies compared to the category 2 studies.

Negative removal rates of TP were reported in 17 of the 146 wetlands. Six of these were created on former cropland and one on previous cattle pasture [41]. The soil in such areas is usually rich in phosphorus that may be released after construction of the wetland. The other ten wetlands with reported negative TP removal rates were all FWS wetlands created 0–2 years before the start of the study, which means that release of P from initially P rich sediments could have contributed to the results. For example, Bass, Evans [42] argued that mineralization of phosphorus from formerly anoxic organic layers uncovered during the excavation could have caused the negative TP removal rate. Release of phosphorus associated with iron complexes under anaerobic conditions can also contribute to low or negative removal rates, as suggested by Healy, Cawley [43] as an explanation for the observed low TP removal rates.

Quantitative synthesis

Meta-analysis

Data from studies with replication were subjected to meta-analyses. The results obtained for individual wetlands are summarized in Figs. 7 and 8, which show forest plots of log response ratios (Ln (load out/load in)) for TN and TP, respectively.

The forest plot for TN removal shows an overall net removal with reasonably narrow confidence limits (Fig. 7). Only three out of 38 studies reported a strong variability, and it can be noted that two of them also had the highest removal rates. For 21 cases, the confidence limits indicate a statistically significant TN removal. For the remaining 17 cases, the confidence limits encompass the zero-effect line (including one case where the average indicate a release of TN). The overall average summary

Table 11 Summary statistics for TN

Quality category	n ^a	TN loading rate (g m ⁻² year ⁻¹)		TN removal rate (g m ⁻² year ⁻¹)		TN removal Efficiency (%)	
		Av. ± 1SE (range)	Med	Av. ± 1SE (range)	Med.	Av. ± 1SE (range)	Med.
2 and 3	112	505 ± 579 (2.12–2486)	255	181 ± 251 (–0.3–1270)	93	39 ± 21 (–12.8–93)	38
2	88	578 ± 608 (2.12–2486)	378	207 ± 269 (–0.3–1270)	120	39 ± 23 (–12.8–93)	38
3	24	237 ± 353 (12.3–1424)	113	85 ± 128 (1.7–584)	43	39 ± 16 (14.8–87)	36

n number of wetlands, Av. arithmetic mean, SE standard error, Med median

^a Some articles have reported results for more than one wetland

Table 12 Summary statistics for TP

Quality category	n ^a	TP loading rate (g m ⁻² year ⁻¹)		TP removal rate (g m ⁻² year ⁻¹)		TP removal efficiency (%)	
		Av. ± 1SE (range)	Median	Av. ± 1SE (range)	Median	Av. ± 1SE (range)	Median
2 and 3	146	36 ± 73 (0.03–373)	3.2	13 ± 38 (–16.8–240)	1.2	41 ± 52 (–422–99)	50
2	85	54 ± 83 (0.08–373)	14	19 ± 46 (–16.8–240)	2.9	39 ± 60 (–422–98)	51
3	61	10 ± 43 (0.03–307)	1.4	4 ± 18 (–7.3–133)	0.63	44 ± 38 (–104–99)	49

n number of wetlands, Av. arithmetic mean, SE standard error, Med median

^a Some articles have reported results for more than one wetland

effect ± 1 SE is -0.46 ± 0.05 . The between-study variance (T^2) was estimated to 0.06, and the I^2 statistic was 86 %. The heterogeneity of the evidence base may thus be regarded as high. The overall average summary effect represents a median TN removal ratio (R) of about 0.63. This means that the median TN load reduction, or removal efficiency, is 37 %, with a 95 % confidence interval of 29–44 %.

The forest plot for TP removal generally shows wider confidence intervals, and a higher number of cases were reported with an average net release rather than removal of TP (Fig. 8). For 29 out of 51 wetlands a significant net removal was reported. Among the remaining cases, 13 exhibited a non-significant net TP removal and nine a non-significant net TP release. The overall summary effect size is highly significant with rather narrow confidence limits. The average ± 1 SE is -0.62 ± 0.08 , which is even lower than that for TN. Similar to TN, the heterogeneity among the studies is high also for TP. The estimated between-study variance (T^2) was 0.24, while the I^2 statistic was 97 %. The overall average summary effect represents a median TP removal ratio (R) of about 0.54, which can be recalculated to a median TP removal efficiency of 46 %, with a 95 % confidence interval of 37–55 %.

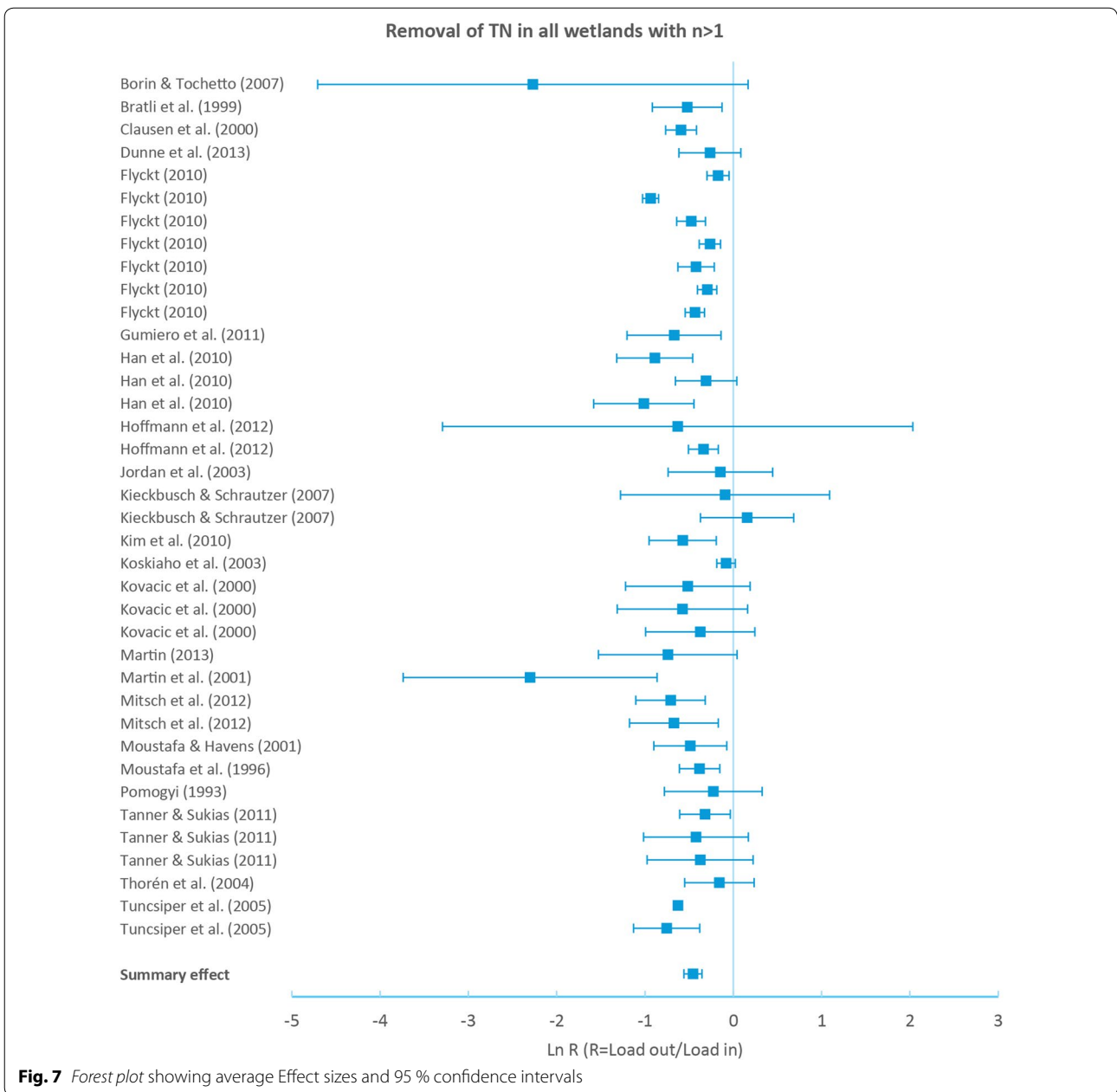
Summarizing, there is strong evidence that created wetlands generally remove TN and TP and that the overall removal efficiency is roughly 40 %. An annual release rather than removal of TN has been shown in a very small number of studies, whereas negative TP removal rates are less uncommon. Also, TP removal ratios generally show a larger variance.

The high heterogeneity of the evidence base calls for subgroup analyses that could potentially identify effect modifiers. Results of such subgroup analyses, i.e., where the studies were divided into different climate zones, wetland types, water types, hydrologic regimes etc., are shown in Figs. 9 and 10. The forest plot for TN removal with wetland studies grouped per climate zone shows that TN removal is significantly different from zero in all climate zones (Fig. 9a). The removal tends to be more efficient (i.e., the effect size tends to be more negative) for wetlands in groups with hot summers, although the 95 %

confidence intervals overlap each other. It should also be noted that some groups include only a small number of studies, and hence the confidence intervals are quite broad. Subdivided by wetland type (Fig. 9b), the pattern is remarkably similar for the four types of wetlands: the summary effects all differ significantly from zero and the averages are relatively close to each other with confidence levels showing a strong overlap. Separated by water type, all four averages are again significantly different from zero, while river/lake water and agricultural runoff show quite similar averages and confidence limits (Fig. 9c). The secondarily treated domestic wastewater showed a higher removal efficiency as well as wider confidence interval.

The subgroup analyses demonstrate that wetlands have a robust capacity to remove TN from through-flowing water. Except for the observation that the removal of TN seems to be more efficient for secondarily treated wastewater than for tertiary treated wastewater, no significant effect modifier could be identified in the subgroup analysis. Numerical values of TN removal efficiency for each subgroup are shown in Table 13.

The forest plots for TP for the various wetland groupings show different patterns than those for TN. Two out of the six climate zones had wetlands with TP removal non-significantly different from zero, i.e., Mediterranean (Csa) and Snow climate with hot summers (Dfa) (Fig. 10a). However, the effect sizes in these climates are based on only 1 and 3 studies, respectively, and the 95 % confidence intervals are broad. Subdivided by water type, there was a TP removal significantly different from zero in wetlands receiving four out of five water types (Fig. 10b). The averages are not significantly different from each other, and the narrowest confidence interval was observed for agricultural runoff (also by far represented by the highest number of studies). The grouping based on wetland history shows a TP removal significantly different from zero in all cases except for restored wetlands on formerly drained cropland, for which there was an insignificant net release of phosphorus. Wetlands in all the other history groups had a significant removal, with no further differences among the types (Fig. 10c). Also, the grouping by water regime suggests that wetlands with precipitation-driven



HLR are less effective than wetlands with other water regimes, although all subgroups significantly differed from zero. This is true also when the restored wetlands on former drained cropland are removed (Fig. 10d). If such wetlands are included the difference between precipitation-driven and other wetlands would appear to be even larger (data not shown). Inclusion or exclusion of restored wetlands on formerly drained cropland does not alter the general patterns shown in the other subgroup analyses. Numerical values of TP removal efficiency for each subgroup are shown in Table 14.

The subgroup analyses show that wetlands have a robust capacity to remove TP, although the 95 % confidence intervals of the means are generally wider, and there are more cases not significantly different from zero, than for TN. There is evidence that cases with a net TP release do occur, but mostly in wetlands that were restored on formerly drained cropland without excavating or isolating the soil. Water regime seems to be an additional significant effect modifier. Wetlands with a precipitation-driven HLR are less efficient than wetlands with a controlled HLR. Wetlands in warmer

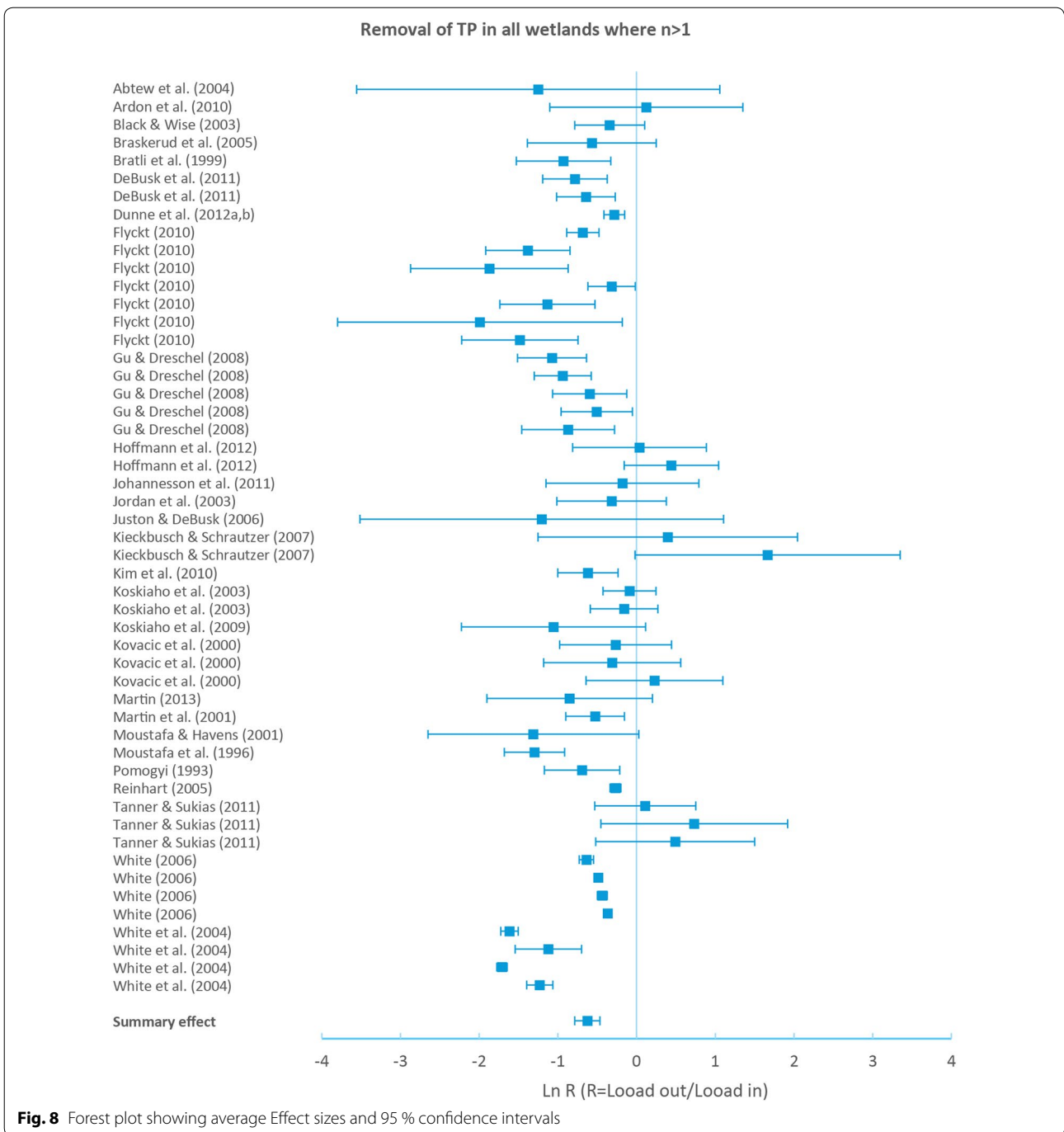


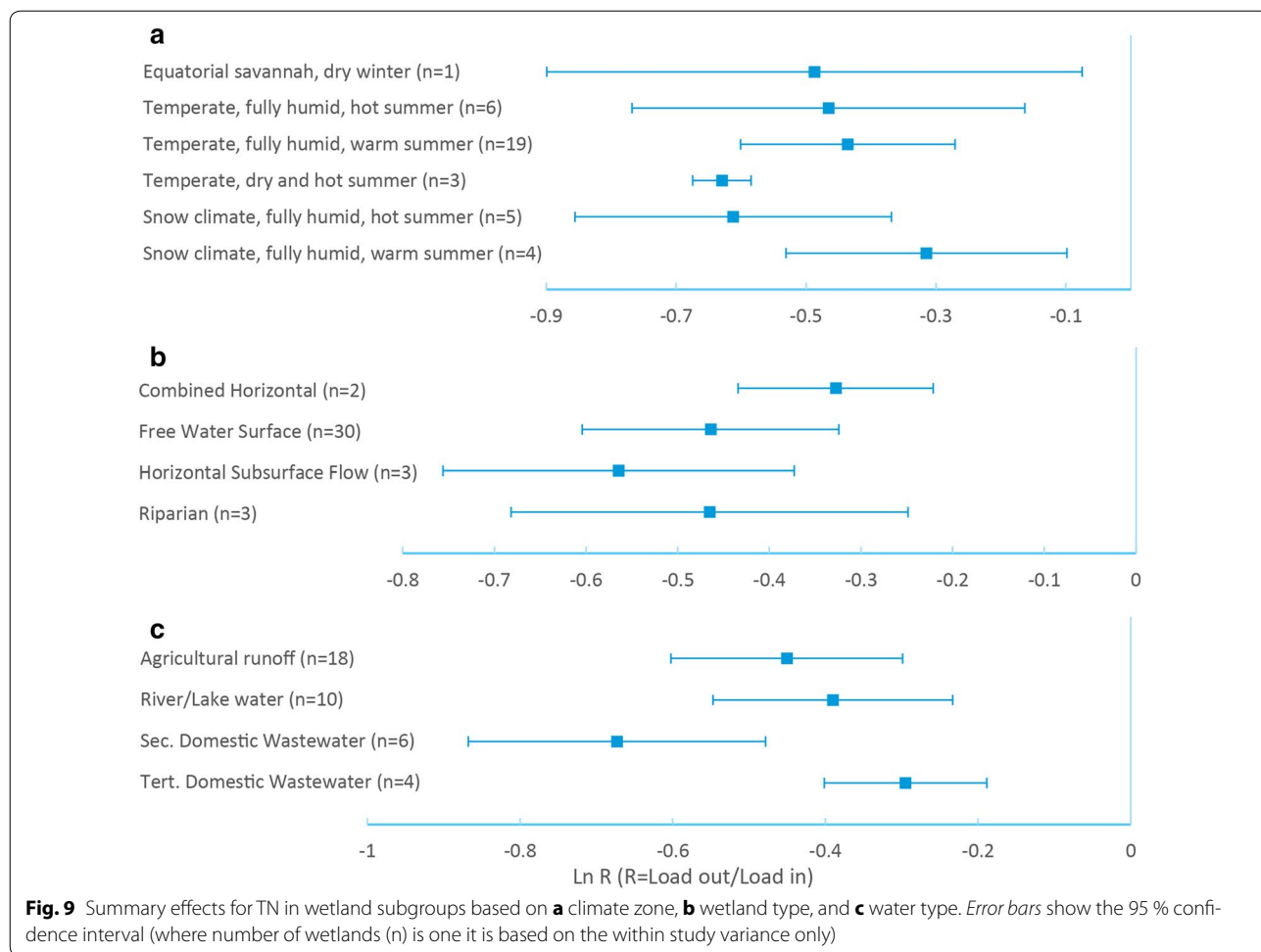
Fig. 8 Forest plot showing average Effect sizes and 95 % confidence intervals

climates (tropical savanna and warm temperate) tend to have a more reliable TP removal than in colder climates, although the 95 % confidence intervals overlap each other.

Response surface analysis

In this section, the results obtained by response surface analyses are presented. This type of regression analyses

was based on mean values per wetland study, and the response surfaces derived illustrate how estimates of median removal efficiency and median removal rate are influenced by various effect modifiers. Preliminary analyses demonstrated that, after both main effects and possible interaction effects of the two most relevant effect modifiers had been taken into account, introduction of additional explanatory variables or interaction usually



had a rather low impact on the goodness-of-fit of the model. Moreover, we noticed that the statistical significance of individual components of complex response surface models could be strongly influenced by the inclusion/exclusion of a small number of observations representing rather unusual levels of the moderators. Therefore, we decided to focus on parsimonious models and present the results of a forward selection of explanatory variables and interactions. Detailed statistical outputs for different models are shown in Additional file 6.

TN removal efficiency (% load reduction) was significantly negatively related to hydraulic loading rate. According to a combined linear/spline model (Model 2, see Additional file 6) the linear component was strongly significant ($p < 0.0001$), whereas the non-linear component was less significant ($p = 0.034$). TN removal efficiency was also found to be positively correlated with annual average air temperature (Model 4). Other investigated predictors showed non-significant ($p > 0.05$) relationships to TN removal efficiency.

Using both hydraulic loading rate and air temperature as predictors in a general additive model (GAM) (Model 7) improved the model fit (reduced the deviance) and demonstrated that the linear response to air temperature was significant also in the presence of a function of hydraulic loading. The model fit was further improved when the one-dimensional splines in log hydraulic loading and air temperature, respectively, were substituted for a thin plate spline that allowed interaction effects between hydraulic loading and air temperature without changing the degrees of freedom of the model (Model 10). The fitted removal efficiency according to model 10 is shown in Fig. 11. Adding a cubic spline for wetland area resulted in an even better fit (Model 12) and demonstrated that the linear response to log wetland area was statistically significant also in the presence of a thin plate spline in log hydraulic loading and air temperature.

The TN removal rate expressed as $g\ m^{-2}\ day^{-1}$ was found to be positively correlated with the inflow concentration, with a steeper increase in removal rate at

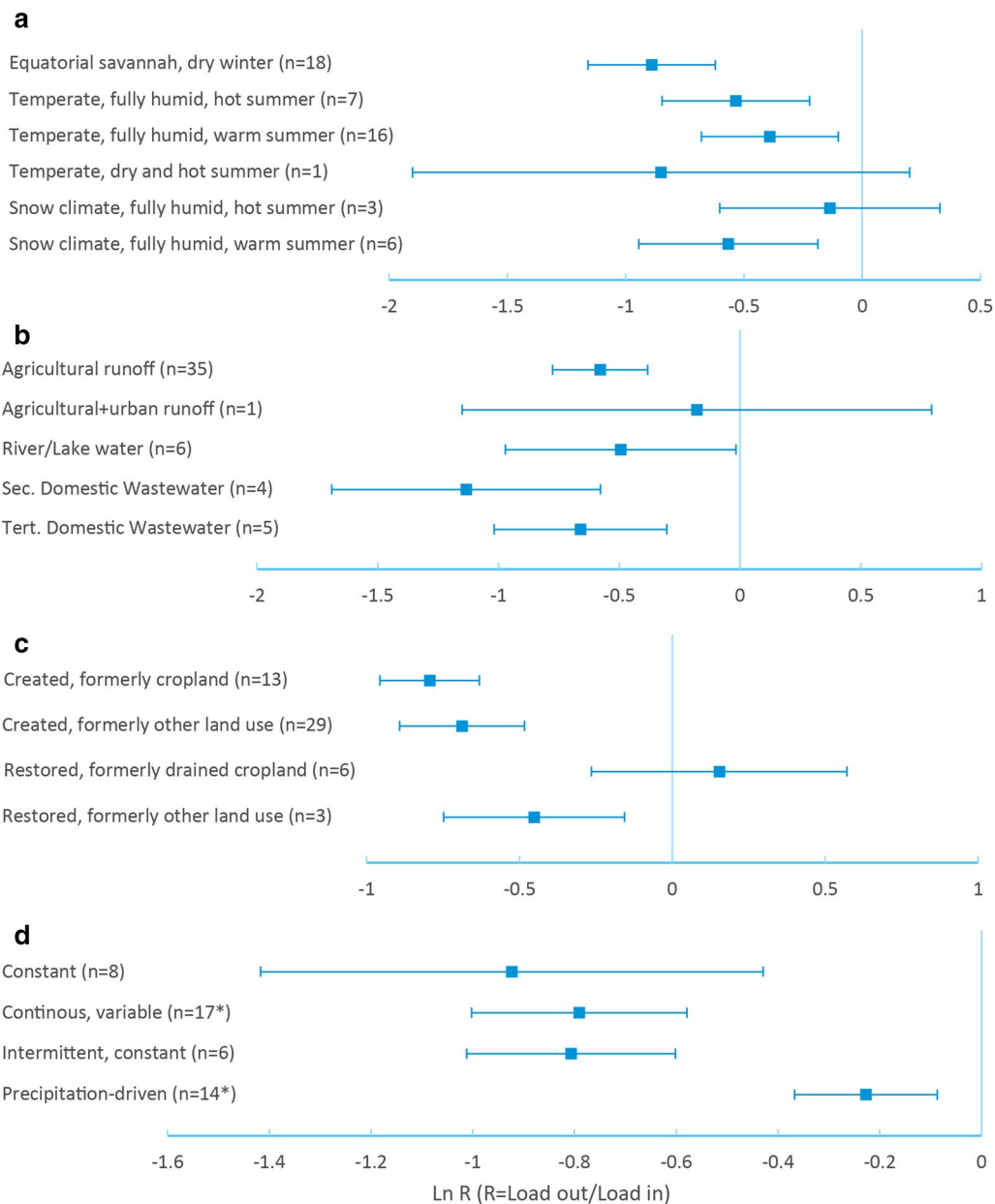


Fig. 10 Summary effects for TP in wetland subgroups based on **a** climate zone, **b** water type, **c** wetland history, and **d** hydrologic regime. Error bars show the 95 % confidence interval (where number of wetlands (n) is one it is based on the within study variance only). *Restored wetlands on formerly drained cropland are not included (five precipitation-driven and one wetland with continuous flow and variable HLR)

concentrations higher than about 18 mg/l (Model 14). The TN removal rate was also positively correlated with hydraulic loading, at least up to about 650 l m⁻² day⁻¹ at which level the TN removal rate started to slowly decline (Model 15). However, the non-linear component was not quite statistically significant (p = 0.080). Furthermore, the TN removal rate was negatively correlated with wetland area, but the decline in removal rate with

wetland size appeared to be somewhat lower at areas above approximately 1 ha (Model 16).

When both hydraulic loading and TN concentration at inlet were used as predictors in a GAM the deviance was substantially reduced (Model 18), and a further reduction was achieved when the two one-dimensional splines were substituted for a thin plate spline allowing interaction effects without increasing the degrees of freedom

Table 13 Results of subgroup analyses for TN

	n ^a	Ln R ± 1 SE	R	Median removal efficiency (%)	95 % confidence interval
All wetlands	38	-0.46 ± 0.05	0.63	37	29–43
Climate zone					
Aw (equatorial savannah, dry winter)	1	-0.49 ± 0.21	0.61	39	7–60
Cfa (warm, fully humid, hot summer)	6	-0.47 ± 0.15	0.63	37	15–54
Cfb (warm, fully humid, warm summer)	19	-0.44 ± 0.08	0.64	36	23–46
Csa (warm, dry and hot summer)	3	-0.63 ± 0.02	0.53	47	44–50
Dfa (snow, fully humid, hot summer)	5	-0.61 ± 0.12	0.54	46	30–58
Dfb (Snow, fully humid, warm summer)	4	-0.31 ± 0.11	0.73	27	9–42
Wetland type					
Combined horizontal	2	-0.33 ± 0.05	0.72	28	19–36
Free water surface	30	-0.46 ± 0.07	0.63	37	27–46
Horizontal subsurface flow	3	-0.56 ± 0.1	0.57	43	31–54
Riparian	3	-0.47 ± 0.11	0.63	37	22–50
Water type					
Agricultural runoff	18	-0.45 ± 0.08	0.64	36	25–46
River/lake water	10	-0.39 ± 0.08	0.68	32	20–43
Sec. domestic Wastewater	6	-0.67 ± 0.1	0.51	49	38–59
Tert. domestic Wastewater	4	-0.29 ± 0.05	0.75	25	17–34
Wetland history					
Created, formerly cropland	3	-0.45 ± 0.05	0.64	36	29–43
Created, formerly other land use	25	-0.49 ± 0.06	0.61	39	30–46
Restored, formerly drained cropland	6	-0.33 ± 0.12	0.72	28	9–44
Restored, formerly other land use	3	-0.49 ± 0.36	0.61	39	-23–70
Water regime					
Continuous, variable	21	-0.54 ± 0.06	0.58	42	33–49
Intermittent, variable	1	-0.67 ± 0.27	0.51	49	12–70
Precipitation-driven	15	-0.43 ± 0.09	0.65	35	22–46
Intermittent, constant	1	-0.42 ± 0.1	0.66	34	19–47

R = load out/load in

^a Where n = 1 the SE is based on the within study variance only

of the model (Model 21). A plot of predicted removal rates according to this model is shown in Fig. 12, and the overall positive response to hydraulic loading and inflow concentration is clearly visible. When cubic splines for wetland area and temperature were added on top of the thin plate spline in inflow concentration and hydraulic loading the deviance was further reduced and the response to wetland size (log area) was statistically significant (Model 26).

According to combined linear/spline regression models, the removal efficiency of TP was influenced by all four of the investigated predictors, i.e., TP inlet concentration, hydraulic loading, wetland area, and air temperature (Models 27–30, respectively). More specifically, the response to the log-transformed values of inlet concentration, hydraulic loading, and wetland area had both linear and non-linear components, whereas the response

to inlet concentration was primarily nonlinear. Closer examination of the removal efficiency indicated that it had a maximum for intermediate concentrations (0.05–0.5 mg/l) at wetland inlet.

When GAM models with two predictors were examined the best fit (lowest deviance) was obtained for a thin plate spline model with log inlet concentration and log hydraulic loading rate (Model 36; Fig. 13). However, the removal efficiency exhibited a substantial random variation and the deviance of model 36 was only slightly lower than that of model 28, which was the best one-dimensional spline model. Neither, was there any statistically significant linear response to air temperature or log area when model 36 was extended with one-dimensional splines in these variables.

The TP removal rate was positively correlated with TP concentration at inlet with a steeper increase in removal

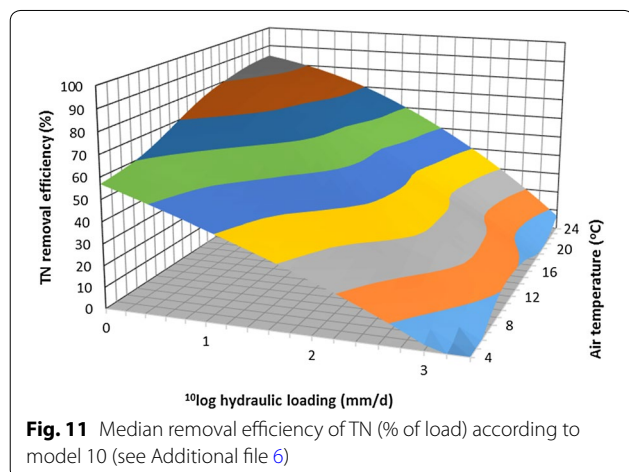
Table 14 Results of subgroup analyses for TP

	n ^a	Ln R ± 1 SE	R	Median removal efficiency (%)	95 % confidence interval
All wetlands	51	-0.62 ± 0.08	0.54	46	37-55
Climate zone					
Aw (equatorial savannah, dry winter)	18	-0.89 ± 0.14	0.41	59	46-69
Cfa (warm, fully humid, hot summer)	7	-0.53 ± 0.16	0.59	41	19-58
Cfb (warm, fully humid, warm summer)	16	-0.39 ± 0.15	0.68	32	9-50
Csa (warm, dry and hot summer)	1	-0.85 ± 0.54	0.43	57	-22 to 86
Dfa (snow, fully humid, hot summer)	3	-0.14 ± 0.24	0.87	13	-38 to 46
Dfb (snow, fully humid, warm summer)	6	-0.57 ± 0.19	0.57	43	17-62
Wetland type					
Combined horizontal	2	-1.22 ± 0.29	0.3	70	47-84
Free water surface	47	-0.64 ± 0.08	0.53	47	37-56
Riparian	2	0.31 ± 0.25	1.36	-36	-122 to 17
Water type					
Agricultural runoff	35	-0.58 ± 0.1	0.56	44	31-54
Agricultural + urban runoff	1	-0.18 ± 0.5	0.84	16	-120 to 69
River/lake water	6	-0.49 ± 0.24	0.61	39	1-63
Sec. domestic wastewater	4	-1.13 ± 0.28	0.32	68	43-82
Tert. domestic wastewater	5	-0.66 ± 0.18	0.52	48	26-64
Wetland history					
Created, formerly cropland	13	-0.79 ± 0.08	0.45	55	46-62
Created, formerly other land use	29	-0.69 ± 0.10	0.50	50	38-60
Restored, formerly drained cropland	6	0.15 ± 0.21	1.16	-16	-77 to 24
Restored, formerly other land use	3	-0.45 ± 0.15	0.64	36	14-53
Water regime ^b					
Constant	8	-0.92 ± 0.25	0.4	60	34-76
Continuous, variable	17	-0.79 ± 0.11	0.45	55	43-64
Intermittent, constant	6	-0.81 ± 0.1	0.44	56	45-64
Precipitation-driven	14	-0.23 ± 0.07	0.79	21	8-31

R = load out/load in

^a Where n = 1 the SE is based on the within study variance only

^b Restored wetlands on former drained cropland have been removed from these calculations



rate at concentrations above approximately 0.55 mg/l (model 40). In contrast, the TP removal rate was negatively correlated with wetland area, especially at areas below $2 \times 10^4 \text{ m}^2$, above which the removal rate was fairly constant (model 42). A statistically significant spline function was found for air temperature (model 43), and a maximum in removal efficiency appeared at intermediate annual average temperatures (approximately 14–19 °C).

When both inlet concentration and hydraulic loading rate were used as predictors of removal rate and interaction effects between these predictors were taken into account using a thin plate spline function (model 47) an ordinary F-test indicated that the deviance was significantly lower than in the best one-dimensional spline

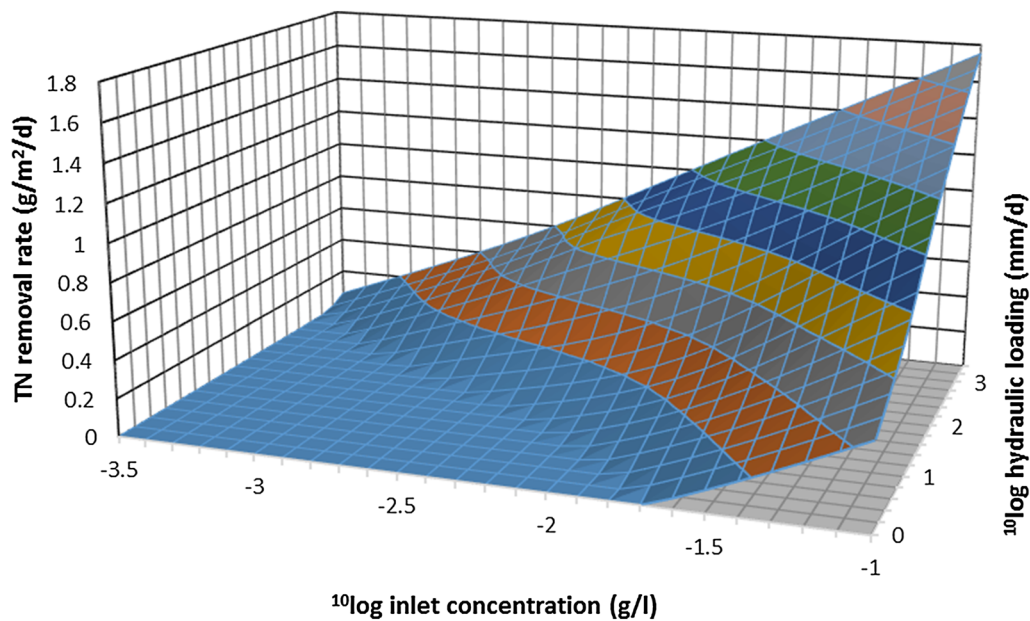


Fig. 12 Median removal rate of TN ($\text{g m}^{-2} \text{day}^{-1}$) according to model 21 (see Additional file 6). The surface has been truncated at $0 \text{ g m}^{-2} \text{day}^{-1}$

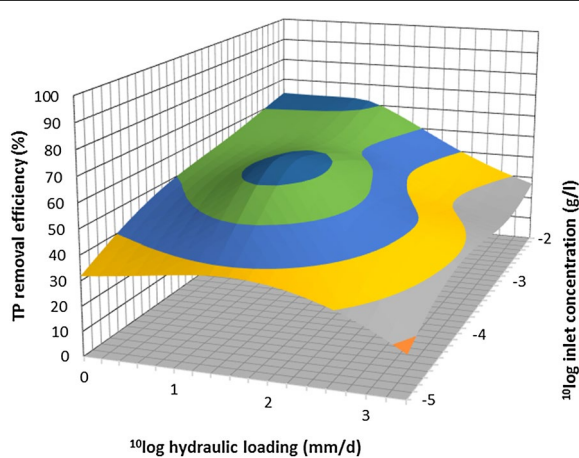


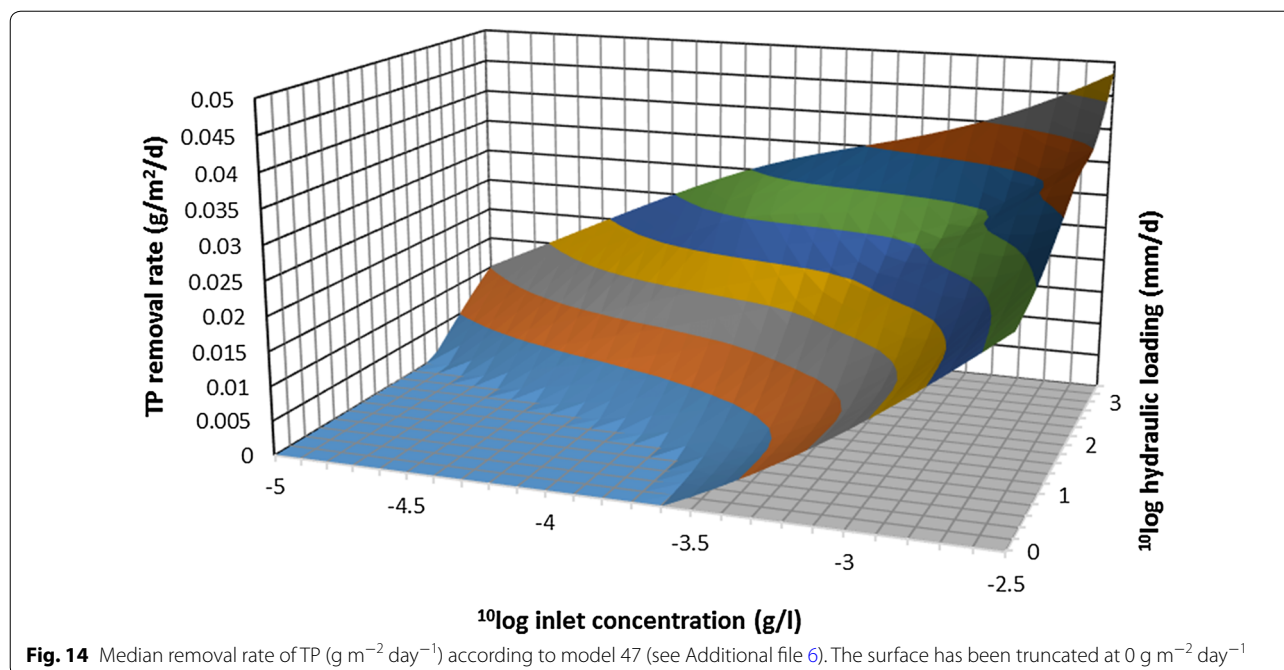
Fig. 13 Median removal efficiency of TP (% of load) according to model 36 (see Additional file 6)

model (model 40). Fitted TP removal rates according to model 47 are shown in Fig. 14. Further analysis showed that adding a one-dimensional spline for air temperature made the deviance even lower and that there was a statistically significant non-linear component of air temperature in this extended model (model 50).

Figures 12 and 14 suggest that the removal rates are very low at low nutrient concentrations at the inlet and low HLRs. To obtain an appreciable removal rate either the inlet concentration or the HLR (or both) need to be increased. On the other hand, the HLR should be

increased with some caution since the removal efficiency decreases with increasing HLR (Figs. 11, 13). When a wetland is being designed, a balance should thus be found between an HLR that is high enough to allow for a meaningful removal rate at a given inlet concentration, and an HLR that is low enough to keep the removal efficiency sufficiently high to make a significant difference to the total transport of nutrients. However, periods with none or very low removal rates will inevitably occur in wetlands with precipitation driven, intermittent, and to some extent variable continuous water flow. In all cases the removal rate increases with increasing concentration at inlet. The removal efficiency for TP generally increases with increasing inlet concentrations, primarily at low to intermediate HLRs, whereas the removal efficiency for TN is less influenced by the inlet concentration.

Because the number of wetlands included in our review is too small to allow sophisticated modelling of mean values per wetland, it is tempting to extract several effect sizes from the same study, for example by using data for individual years as model inputs. In principle, such data can be analysed by using hierarchical generalized linear models. However, the data collected in the present study were far from ideal for such models. First, less than half of the included wetland studies had any temporal replicates. Second, it was difficult to identify a suitable covariance structure for the study-specific random components; some of the studies exhibited a large inter-annual variation, whereas others exhibited a very small within-study variation.



Discussion

Reasons for heterogeneity and review limitations

Although the water quality criteria set for this review are narrowed to secondarily treated domestic wastewater, tertiary treated domestic wastewater, urban water runoff, agricultural drainage water, and river or lake water, these water types differ considerably in composition and thus also the processes needed for removal of nitrogen and phosphorus differs between the water types. It would have been desirable to look at the different N and P species in connection with these removal and retention processes, as e.g., nitrate in connection with denitrification or particulate-P in connection with retention of phosphorus. This might have improved the analysis of water type divided into subgroups (Figs. 11, 13). However, in many cases the articles only reported TN and TP results, and therefore further analysis was impossible.

Especially for phosphorus, it might have been beneficial to look into different P-species because it has been documented that some created wetlands retain large amounts of particulate P (PP) when there are high loads of this species due to the upland characteristics [16]. Furthermore, phosphorus retention in wetlands is dependent on several factors, such as age, past history of P (i.e., P load, P fertilizer addition, P saturation in soil), redox conditions, climate conditions (e.g., frost–thaw), which are not always included in the information.

Excluded from this review due to limitations in sampling frequency are restored and created floodplains and other riparian areas subjected to flooding and inundation.

In these areas sedimentation of suspended solids and particulate phosphorus is often high [15, 44–47] making sedimentation the most important process for retention of phosphorus in such wetlands.

This review is limited to only one function of created and restored wetlands—, namely their role in catchment nutrient losses. This function should be seen only as complementary to all the measures that can be undertaken to improve nutrient use and management on agricultural land and in fields. In reality, created and restored wetlands provide multiple ecosystem services, such as biodiversity enhancement, reservoirs for water and recreation that are difficult to value. Wetlands are also important ecosystems in the global cycles of greenhouse gases, as potential sources of CH_4 , CO_2 and N_2O emissions and for their role in carbon sequestration and as possible sinks for N_2O in drainage water from farmland. These aspects are outside the scope of this review, but are potential subjects for other systematic reviews.

Hydrological processes and especially hydraulic loading are inadequately measured in many papers: 45 out of the 143 category 1 papers only included inlet measurements, had incomplete water balances, or lacked hydrological data, making it impossible or too uncertain to calculate mass balances. For created wetlands with a lining and constant load, it is relatively simple to set up a water balance. In contrast, for wetlands with varying or event driven hydraulic loading the measurements need to be much more comprehensive and require much higher temporal resolution to cover the

variation. For restored wetlands it is difficult to set up and measure all variables in the water balance (surface water flow, groundwater flow, precipitation, evapotranspiration). In some wetland studies there are large deficits in the water balance due to unaccounted contributions from e.g., groundwater inflow and/or because all budget terms in the water balance have not been measured. Thus especially for restored or recreated wetlands it may be needed to carry out measurements with high spatial and temporal replication. For example, to calculate groundwater flow through a riparian meadow it is necessary to measure hydraulic potentials, hydraulic conductivity in different soil layers or at least to assign hydraulic conductivities according to soil profile descriptions.

In individual studies, measurement errors of the predictor loading rate may make the response variable removal efficiency correlated with the predictor irrespective of any underlying causality. This is one of the short-comings of treating a studied wetland as a black box (which most of the included studies do). From this point of view it would have been better if a larger number of studies used multiple measurements of the loading rate and removal processes (denitrification, plant uptake, chemical adsorption etc.). On the other hand, removal processes are variable in space and time, potentially leading to even higher experimental errors than input–output balance studies and difficulty to scale up for the whole wetland and an entire year. On the contrary, input and output fluxes can be measured with more precision because hydraulic flow measurements combined with frequent concentration measurements of N and P are relatively straightforward, and we have selected the studies for state-of-the-art techniques in our critical appraisal. The input–output approach is therefore quite robust. Furthermore, while measurement errors in loading rates may be in the range 0–30 %, this review has included studies where the loading rates span three orders of magnitude for TN and four orders of magnitude for TP. Measurement errors in loading rates in individual studies should thus play a minor role for the correlation between loading rate and removal efficiency shown by the response surface analyses performed in this systematic review.

Conclusions

Implications for policy/practice

The objective of this review was to quantify observed retention rates of nutrients in created or restored wetlands and to quantify the variability between different studies. We also investigated importance of environmental conditions and wetland characteristics for nutrient removal rates. This study examined wetlands in a wide

range of climatic conditions, although the performance of wetlands in temperate and boreal regions is emphasized. Overall, data from 203 wetlands has been used in our analyses.

- Our survey found that, on average, created or restored wetlands removed $184 \text{ g m}^{-2} \text{ year}^{-1}$ of total nitrogen and $15 \text{ g m}^{-2} \text{ year}^{-1}$ of total phosphorus. These average retentions are three to four times the retention rates suggested 15 years ago as sustainable for nonpoint source treatment wetlands ($10\text{--}40 \text{ g m}^{-2} \text{ year}^{-1}$ for nitrogen and $0.5\text{--}5 \text{ g m}^{-2} \text{ year}^{-1}$ for phosphorus). However, the median values for TN and TP removal rates, respectively, are lower and in line with the earlier recommendations.
- Restored and created wetlands remain appropriate and potentially sustainable ecological engineering approaches for removing nutrients from treated wastewater and urban and agricultural runoff. Loading rates (inlet concentrations \times hydraulic loading rates) need to be carefully estimated as part of the design of these wetlands. In general, high nutrient loading rates result in high removal rates (expressed in $\text{g m}^{-2} \text{ year}^{-1}$). However, high hydraulic loading rates may result in reduced removal efficiency (expressed in %).
- Seventeen of 146 wetlands were shown as phosphorus sources; six of those had been restored on former drained cropland. The studies included in this systematic review suggest that TP removal is less efficient in such wetlands compared to other wetlands. Six out of nine restored wetlands on former drained cropland released more phosphorus than they received. However, one long-term study suggests that the performance of wetlands may be enhanced and that such wetlands can become significant phosphorus sinks after several years of operation.
- Water regime seems to be another factor that can influence phosphorus removal efficiency. Wetlands where the hydraulic loading rate is driven by precipitation show a lower phosphorus removal efficiency than wetlands with a controlled hydraulic loading rate.
- Removal efficiency of total nitrogen in wetlands was positively correlated with average annual air temperature and negatively correlated with hydraulic loading rate. The model fit was better if interaction effects between these variables were allowed. The total nitrogen removal rate was positively correlated with the inflow concentration and was also found to be positively correlated with hydraulic loading.
- The removal efficiency of total phosphorus was correlated with total phosphorus concentrations at the

inlet, hydraulic loading, wetland area, and air temperature. The total phosphorus removal rate was positively correlated with concentration at inlet and hydraulic loading rate. In contrast, the total phosphorus removal rate was negatively correlated with wetland area, especially for wetlands smaller than $2 \times 10^4 \text{ m}^2$.

Implications for research

- Hydrological processes are inadequately measured in many papers: 45 of the papers excluded during critical appraisal only included inlet measurements, had incomplete water balances or lacked hydrological data, making it impossible or too uncertain to calculate mass balances.
- Only total nitrogen or total phosphorus was measured or reported in many studies. This prevented us from evaluating the influence of the speciation of these elements on the removal.
- Long-term performance of wetlands as nutrient sinks is poorly investigated.
- More research is needed on the effects of seasonality, particularly in wet/dry climates, and on hydrologic pulsing on wetlands used to treat agricultural and urban runoff. More research is also needed on the ecosystem services of carbon sequestration and flood mitigation that these wetlands could and do provide in addition to their primary role in water purification.

Additional files

Additional file 1. Searches.

Additional file 2. Articles not used in the review.

Additional file 3. Geographical representation.

Additional file 4. Critical appraisal.

Additional file 5. Narrative synthesis table.

Additional file 6. Response surface analyses.

Authors' contributions

The writing of this report was shared between KT (Introduction), ML (Methods and results), JV (Results), WG (Results), CCH (Discussion), and WM (Conclusions). Searches were performed by ML. Screening of literature was performed by ML, KT, JV, WG, CCH, and WM. Critical appraisal was undertaken by KT, JV, WG, CCH, and WM. Meta-analysis and response surface analysis were performed by ML and AG, respectively. All authors read and approved the final manuscript.

Author details

¹ Mistra EviEM, Stockholm Environment Institute, Box 24218, 10451 Stockholm, Sweden. ² Department of Biology, Lund University, 22362 Lund, Sweden. ³ Everglades Wetland Research Park, Florida Gulf Coast University, 4940 Bayshore Drive, Naples 34112, FL, USA. ⁴ Swedish Institute for the Marine Environment, Box 260, 405 30 Gothenburg, Sweden. ⁵ Department of Bioscience, Aarhus University, 8600 Silkeborg, Denmark. ⁶ Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden. ⁷ Department of Biology, Utrecht University, 3584 CA Utrecht, Netherlands.

Acknowledgements

This systematic review was financed by the Mistra Council for Evidence-based Environmental Management (Mistra EviEM). EviEM is funded by the Swedish Foundation for Strategic Environmental Research (Mistra) and hosted by Stockholm Environment Institute.

Competing interests

The authors declare that they have no competing interests.

Received: 25 November 2015 Accepted: 20 April 2016

Published online: 09 May 2016

References

1. Smith VH. Eutrophication of freshwater and coastal marine ecosystems—a global problem. *Environ Sci Pollut Res*. 2003;10(2):126–39.
2. Nixon SW, Ammerman JW, Atkinson LP, Berounsky VM, Billen G, Boicourt WC, et al. The fate of nitrogen and phosphorus at the land sea margin of the North Atlantic Ocean. *Biogeochemistry*. 1996;35(1):141–80.
3. Ryther JH, Dunstan WM. Nitrogen, phosphorus, and eutrophication in coastal marine environment. *Science*. 1971;171(3975):1008.
4. Graneli E, Wallstrom K, Larsson U, Graneli W, Elmgren R. Nutrient limitation of primary production in the baltic sea area. *Ambio*. 1990;19(3):142–51.
5. Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE, et al. Controlling eutrophication: nitrogen and Phosphorus. *Science*. 2009;323(5917):1014–5.
6. Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ, et al. Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci USA*. 2008;105(32):11254–8.
7. Boesch D, Hecky R, O'Melia C, Schindler D, Seitzinger S. Eutrophication of Swedish Seas. Stockholm: Swedish Environmental Protection Agency, Report 5509; 2006.
8. Danielsson Å, Papush L, Rahm L. Alterations in nutrient limitations—scenarios of a changing Baltic Sea. *J Mar Syst*. 2008;73(3–4):263–83.
9. Kadlec RH, Wallace SD. *Treatment wetlands*. Boca Raton: CRC Press; 2009. p. 965.
10. Mitsch WJ, Day JW, Zhang L, Lane RR. Nitrate-nitrogen retention in wetlands in the Mississippi river basin. *Ecol Eng*. 2005;24(4):267–78.
11. Carleton JN, Grizzard TJ, Godrej AN, Post HE. Factors affecting the performance of stormwater treatment wetlands. *Water Res*. 2001;35(6):1552–62.
12. Vymazal J. Removal of nutrients in various types of constructed wetlands. *Sci Total Environ*. 2007;380(1–3):48–65.
13. Vymazal J, Kroepfelova L. Removal of nitrogen in constructed wetlands with horizontal sub-surface flow: a review. *Wetlands*. 2009;29(4):1114–24.
14. Kadlec RH. Nitrogen farming for pollution control. *J Environ Sci Health Part A-Toxic/Hazard Subst Environ Eng*. 2005;40(6–7):1307–30.
15. Hoffmann CC, Kjaergaard C, Uusi-Kamppa J, Hansen HCB, Kronvang B. Phosphorus retention in riparian buffers: review of their efficiency. *J Environ Qual*. 2009;38(5):1942–55.
16. Braskerud BC, Tonderski KS, Wedding B, Bakke R, Blankenberg AGB, Ulen B, et al. Can constructed wetlands reduce the diffuse phosphorus loads to eutrophic water in cold temperate regions? *J Environ Qual*. 2005;34(6):2145–55.
17. Mitsch WJ, Zhang L, Stefanik KC, Nahlik AM, Anderson CJ, Bernal B, et al. Creating wetlands: primary succession, water quality changes, and self-design over 15 years. *Bioscience*. 2012;62(3):237–50.
18. Mitsch WJ, Zhang L, Waletzko E, Bernal B. Validation of the ecosystem services of created wetlands: two decades of plant succession, nutrient retention, and carbon sequestration in experimental riverine marshes. *Ecol Eng*. 2014;72:11–24.
19. Land M, Graneli W, Grimvall A, Hoffmann CC, Mitsch WJ, Tonderski KS, et al. How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? A systematic review protocol. *Environ Evid*. 2013;2(1):1–8.
20. SMHI Vattenwebb (in Swedish) [database on the internet]. Available from: <http://vattenwebb.smhi.se/wetlands/>. Accessed 25 Feb 2013.
21. Svensson JM, Strand J, Sahlén G, Weisner S. Utvärdering av våtmarker anlagda inom lokala investeringsprogram och med LBU-stöd avseende

- näringsretention och biologisk mångfald. Rikare mångfald och mindre kväve. Utvärdering av våtmarker skapade med stöd av lokala investering-program och landsbyggsutvecklingsstöd (in Swedish). Naturvårdsverket, Rapport 5362; 2004.
22. Tonderski KS, Arheimer B, Pers CB. Modeling the impact of potential wetlands on phosphorus retention in a Swedish catchment. *Ambio*. 2005;34(7):544–51.
 23. Brandt M, Arheimer B, Gustavsson H, Pers C, Rosberg J, Sundström M et al. Uppföljning av effekten av anlagda våtmarker i jordbrukslandskap. Belastning av kväve och fosfor (in Swedish). Naturvårdsverket, Rapport 6309; 2009.
 24. Weisner S, Thiere G. Mindre fosfor och kväve från jordbrukslandskapet (in Swedish). Jordbruksverket, Rapport. 2010;2010:21.
 25. Weisner S, Johannesson K, Tonderski K. Näringsavskiljning i anlagda våtmarker i jordbruket. Analys av mätresultat och effekter av landsbyggsprogrammet (in Swedish). Jordbruksverket, Rapport 2015:7.
 26. Mitsch WJ, Horne AJ, Nairn RW. Nitrogen and phosphorus retention in wetlands-ecological approaches to solving excess nutrient problems. *Ecol Eng*. 2000;14(1–2):1–7.
 27. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World map of the Koppen-Geiger climate classification updated. *Meteorol Z*. 2006;15(3):259–63.
 28. Thiere G, Milenkovski S, Lindgren PE, Sahlen G, Berglund O, Weisner SEB. Wetland creation in agricultural landscapes: biodiversity benefits on local and regional scales. *Biol Conserv*. 2009;142(5):964–73.
 29. Bilotta GS, Milner AM, Boyd IL. Quality assessment tools for evidence from environmental science. *Environ Evid*. 2014;3(1):1–14.
 30. Mermillod-Blondin F, Lemoine D, Boisson JC, Malet E, Montuelle B. Relative influences of submersed macrophytes and bioturbating fauna on biogeochemical processes and microbial activities in freshwater sediments. *Freshw Biol*. 2008;53(10):1969–82.
 31. Hahn S, Bauer S, Klaassen M. Quantification of allochthonous nutrient input into freshwater bodies by herbivorous waterbirds. *Freshw Biol*. 2008;53(1):181–93.
 32. Andersen DC, Sartoris JJ, Thullen JS, Reusch PG. The effects of bird use on nutrient removal in a constructed wastewater-treatment wetland. *Wetlands*. 2003;23(2):423–35.
 33. Bernes C, Carpenter SR, Gårdmark A, Larsson P, Persson L, Skov C, et al. What is the influence of a reduction of planktivorous and benthivorous fish on water quality in temperate eutrophic lakes? A systematic review. *Environ Evid*. 2015;4(1):1–28.
 34. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Random-effects model. In: *Introduction to meta-analysis*. Chichester: Wiley; 2009. p. 69–75.
 35. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Identifying and quantifying heterogeneity. In: *Introduction to meta-analysis*. Chichester: Wiley; 2009. p. 107–25.
 36. Haddaway NR, Verhoeven JTA. Poor methodological detail precludes experimental repeatability and hampers synthesis in ecology. *Ecol Evol*. 2015;5(19):4451–4.
 37. Chen HJ, Ivanoff D, Pietro K. Long-term phosphorus removal in the Everglades stormwater treatment areas of South Florida in the United States. *Ecol Eng*. 2015;79:158–68.
 38. Moustafa MZ, Chimney MJ, Fontaine TD, Shih G, Davis S. The response of a freshwater wetland to long-term “low level” nutrient loads—marsh efficiency. *Ecol Eng*. 1996;7(1):15–33.
 39. Kieckbusch JJ, Schrautzer J. Nitrogen and phosphorus dynamics of a rewetted shallow-flooded peatland. *Sci Total Environ*. 2007;380(1–3):3–12.
 40. Koskiaho J, Ekholm P, Rätty M, Riihimäki J, Puustinen M. Retaining agricultural nutrients in constructed wetlands—experiences under boreal conditions. *Ecol Eng*. 2003;20(1):89.
 41. Kovacic DA, David MB, Gentry LE, Starks KM, Cooke RA. Effectiveness of constructed wetlands in reducing nitrogen and phosphorus export from agricultural tile drainage. *J Environ Qual*. 2000;29(4):1262–74.
 42. Bass KL, Evans RO. Water quality improvement by a small in-stream constructed wetland in North Carolina's Coastal plain. *Watershed Management and Operations Management* 2000. Fort Collins: American Society of Civil Engineers; 2000.
 43. Healy M, Cawley AM. Nutrient processing capacity of a constructed wetland in western Ireland. *J Environ Qual*. 2002;31(5):1739–47.
 44. Walling DE. Linking land use, erosion and sediment yields in river basins. *Hydrobiologia*. 1999;410:223–40.
 45. Walling DE, Owens PN. The role of overbank floodplain sedimentation in catchment contaminant budgets. *Hydrobiologia*. 2003;494(1–3):83–91.
 46. Fink DF, Mitsch WJ. Hydrology and nutrient biogeochemistry in a created river diversion oxbow wetland. *Ecol Eng*. 2007;30(2):93–102.
 47. Mitsch WJ, Zhang L, Fink DF, Hernandez ME, Altor AE, Tuttle CL, et al. Ecological engineering of floodplains. *Ecohydrol Hydrobiol*. 2008;8(2–4):139–47.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

