

A multi-scale approach to prioritize wetland restoration for watershed-level water quality improvement

Neal E. Flanagan · Curtis J. Richardson

Received: 27 October 2009 / Accepted: 20 May 2010 / Published online: 12 June 2010
© Springer Science+Business Media B.V. 2010

Abstract Wetland restoration is commonly presented as an important strategy for maintaining and enhancing the water quality and ecological capital of watershed-scale ecosystems. Prioritizing restoration sites on the landscape is often a haphazard process based on widely held, though often untested, assumptions about relationships between watershed characteristics and water quality. We present a framework to target and prioritize wetland restoration locations using both regional and watershed-level screening models. The regression-tree and random forest models presented in this paper identify watershed variables with the strongest relationships to a given water quality parameter, present a clear hierarchy of variable importance, and present approximate thresholds in watershed area where these variables express the greatest impact on water quality. The proportion of watersheds classified as prior-converted agricultural land was an important predictor of both ortho and total phosphorus. Fortunately because prior-converted agricultural lands were historically wetlands, they are often very suitable for wetland restoration. These sites often have poorly-drained soils requiring artificial drainage to be suitable for agriculture. These drainage systems become conduits for transporting phosphorus from agricultural field

and to area streams and rivers. Maintaining natural land-cover within stream buffers is identified as another important predictor of water quality. This seems to be especially true with regard to $\text{NO}_3\text{--NO}_2$ concentrations. Our model results support specific management recommendations including: (a) exclusion of agricultural land-uses from riparian buffers, (b) maintaining or increasing watershed-level wetland-cover and (c) reducing wetland fragmentation.

Keywords Wetland · Restoration · Watershed · Water-quality · Prioritize · Decision

Introduction

In the last three decades, water quality legislation has substantially decreased point source pollution into the streams and rivers of the United States of America (Malina 1996; Mays 1996). While reductions in point source pollutant loads have had remarkable positive impacts on individual river-systems, many others are still not meeting state and national water quality objectives. Indeed, failure to preserve water quality and quantity may significantly impact human quality of life, and may have potentially irreversible effects on ecosystems. These consequences underlie initiation of major federal programs to address basin scale watershed issues, e.g., (1) the USGS National

N. E. Flanagan (✉) · C. J. Richardson
Duke University Wetland Center, Durham,
NC 27708, USA
e-mail: nflanaga@duke.edu

Water Quality Assessment (NAWQA) Program, and (2) the USEPA Total Maximum Daily Load Program (TMDL). To reach the water quality objectives of these programs, scientists and regulators increasingly focus on non-point sources of pollutants in watersheds. Thus, the watershed has been widely adopted as the logical unit for scientific study and management of non-point sources of water pollution.

Wetland restoration is commonly presented as an important strategy for maintaining and enhancing the water quality and ecological capital of watershed scale ecosystems. While numerous studies have assessed ecosystem function of individual wetland restoration projects (Richardson 1985; Mitsch and Reeder 1991; Mitsch et al. 1995; Bruland and Richardson 2005, 2006), fewer studies have assessed the “cumulative impact” of wetlands on watersheds (Whigham 1988; Childers and Gosselink 1990; Johnston et al. 1990; Weller et al. 1996). None of these studies have examined the importance of watershed-level land-use variables as potential sources of water quality impairment that can be addressed through riparian (i.e., integrated wetland and/or stream) restoration.

The key objectives of our study are: (a) to develop a decision support tool for prioritizing wetland restoration sites within the context of watershed-

level water quality functions, (b) to apply our decision tool to select restoration sites in the Harrison Creek watershed in eastern North Carolina (Fig. 1), and (c) to quantify the effect of restoring high priority sites on in-stream nutrient concentrations in the Harrison Creek watershed.

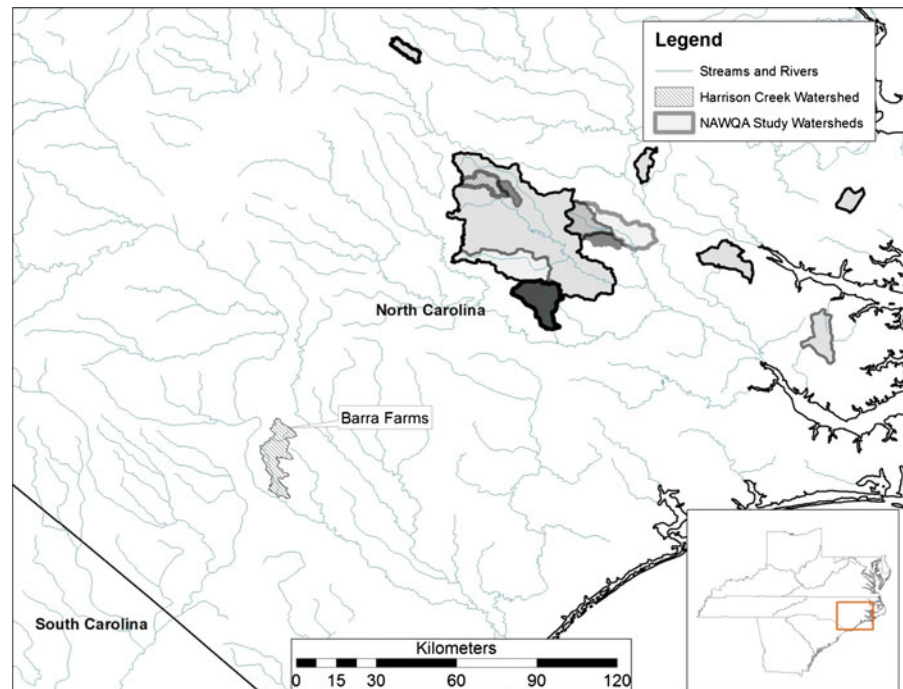
Methods

Statistical modeling

We examine the relationships between land-use/land-cover and water quality on the Coastal Plain of North Carolina, an area of low topographical relief and high annual precipitation. The Coastal Plain has historically supported vast areas of wetlands including pocosins and carolina bays on poorly drained interfluvial areas and bottom-land hardwood forests on the floodplains of major river systems. Many of these wetland areas have proven to be prime agricultural land after the installation of extensive drainage systems (Sharitz and Gresham 1998; Richardson et al. 1981; Richardson and McCarthy 1994; Tiner 2003).

Our models were developed using water quality data from 14 watersheds in eastern North Carolina

Fig. 1 Location of study watersheds in eastern North Carolina, USA



(Fig. 1) that are part of the United States Geological Survey's National Water Quality Assessment (NAWQA) Program. A subset of stations were selected based on location within the coastal plain of North Carolina, and availability of parameters including instantaneous discharge, total phosphorus (TP), ortho-phosphorus (OP), and Nitrate-Nitrite (NOx). The dataset included over 300 individual samples collected between 1992 and 1997. The land-use and land-cover characteristics of the watersheds of each NAWQA station were summarized using the Arc-View[®] GIS system and publicly available spatial data from the United States Geological Survey, the United States Department of Agriculture (USDA), Natural Resource Conservation Service (NRCS), and the United States Fish and Wildlife Service (USFWS). The database includes area of wetlands from the National Wetlands Inventory (USFWS 1975–2005) and other land-uses from the 1992 National Land Cover Dataset (Vogelmann et al. 2001), types of wetlands in the watershed, and continuity of riparian buffers. Codes describing watershed variables are presented in Table 1. These codes have a prefix describing landscape position and a numeric suffix describing land-use. A prefix of WS indicates land-use variable as a percentage of the entire watershed area, and BF indicates percent area of a buffer extending 90 m on both sides of a given stream reach. Prior-converted croplands were defined as polygons with hydric soils, current agricultural land-uses, and intersection by drainage ditches.

The first phase of our study was to identify watershed scale sources of water quality impairment having potential for mitigation through stream and wetland management and/or restoration. We developed regression tree models using the “rpart library” (Therneau and Atkinson 2005) for the R statistical package (R Development Core Team 2008). Rpart produces a graphical tree model that can be thought of as a decision tool relating watershed variables such as land-use/land-cover, soils, buffer continuity, wetland fragmentation, and hydrologic discharge to water quality variables including ortho-phosphorus, total phosphorus, and nitrate-nitrite concentration. Regression tree models are well suited for exploring and modeling complex data sets, which are nonlinear and often display interactions between variables (Breiman et al. 1984; Ripley 1981; De Ath and Fabricius 2000). Tree-based

modeling is an increasingly accepted technique for uncovering structure in data sets because it: (a) is useful for variable screening, (b) can be used for developing prediction rules that are easily evaluated, (c) is invariant to monotone re-expressions of predictor variables, (d) allows for missing values, (e) allows simultaneous use of continuous and categorical variables, (f) is adept at capturing non-additive behavior, and (g) does not have the restrictive assumptions of many other techniques (Qian and Anderson 1999; Qian et al. 2001; Qian 2009).

The model solutions produced by recursive tree procedures are not the only feasible model for a given dependent variable (Breiman et al. 1984). The collinearity of variables creates opportunities for alternative splits that perform nearly as well as the selected split. This is a consequence of the ‘greediness’ of the regression tree algorithms such as rpart (Venables and Ripley 1999). If several predictor variables are very similar in their ability to improve model homogeneity at the first split, it is possible for the model algorithm to select randomly between these competing variables.

The random forest method (Breiman 2001) is an alternative modeling procedure that addresses the shortcomings of rpart. The random forest procedure does not produce a single graphical decision tree, but relies on a “forest” of 500 or more trees developed from bootstrap samples of training data. The product of this procedure is a nutrient concentration prediction for a given combination of predictor variables that is the “consensus” of predictions from the individual trees that make up the forest. Predictions produced by Random Forest models typically demonstrate 30% lower error rates than those of individual tree models (Breiman 2001). The accuracy of model predictions can be summarized with a pseudo- R^2 produced by comparing the nutrient concentration prediction of a single tree to an observed value. The R^2 for the entire forest is then calculated using the equation:

$$R^2 = 1 - \frac{\text{MSE}}{\text{var}(y)} \quad (1)$$

where MSE is the mean squared error between predicted and observed nutrient concentrations averaged from all trees making up the forest, and var(y) is variance in the observed data set (Breiman 2001; Parkhurst et al. 2005).

Table 1 Description of predictor variable codes used in the classification tree models

Type	Prefix	Numerical code ^a	Abbreviation	Description
Land-use/ land- cover	WS, BF	21	RES.LI	Low intensity residential
		22	RES.HI	High intensity residential
		23	COM	Commercial/industrial/transport
		33	TRANS	Transitional
		41	FOR.DEC	Deciduous forest
		42	FOR.EVG	Evergreen forest
		43	FOR.MXD	Mixed forest
		81	AG.PAST	Pasture/hay
		82	AG.ROW	Row crops
		91	WET.WDY	Woody wetlands
		92	WET.EMG	Emergent herbaceous wetlands
	Summations	WS, BF	NA	UNDEV
			FOR.ALL	Portion of watershed or 90 m stream buffer in forest land-cover sum(41, 42, 43)
			AG.ALL	Portion of watershed or 90 m stream buffer in agricultural land-use sum(81, 82)
			WET.ALL	Portion of watershed or 90 m stream buffer in wetland land-cover sum(91, 92)
Other	NA	NA	NWI ^b	Portion of watershed or 90 m stream buffer NWI polygons
			WT.Peri.Area	Average ratio of perimeter to area of NWI wetland polygons (fragmentation)
			DIS.MM	Instantaneous discharge measurements standardized as millimeters per day.
			HYDRIC ^c	Portion of watershed or 90 m stream buffer as hydric soil series
			ANIMAL	Number of confined animal operations within watershed
			PRIOR	% watershed area as prior converted agricultural lands (see text).
			PT.SRC	Number of permitted point source discharge
	PT.DIST	Upstream distance to nearest point source (km)		

Predictors with numeric codes have prefix of WS for those characterizing the entire watershed or BF for those characterizing areas with a 90 m riparian buffer

^a National Land Cover Data Set numerical codes

^b USFWS National Wetland Inventory

^c SSURGO database

A key feature of random forest is an assessment of the importance of explanatory variables in influencing the predictive accuracy of the forest of individual tree models. Importance values utilize out-of-bag samples (i.e., cases excluded from the bootstrap samples used to develop individual trees) to evaluate error of fit of each tree model as explanatory variables are randomly permuted and error of fit is evaluated. The importance of an individual variable increases as a function of the average increase in MSE when the variable is permuted compared with the average MSE when no explanatory variables are

permuted (Liaw and Wiener 2002; Parkhurst et al. 2005).

Site monitoring

We monitored water quality at the Barra Farms Cape Fear River Mitigation Bank located in the headwaters of Harrison Creek (Fig. 2). This monitoring was performed to assess nutrient delivery to the Cape Fear River from prior converted croplands and a restored wetland, and to provide data for validation of our statistical model. The site is located within a Carolina

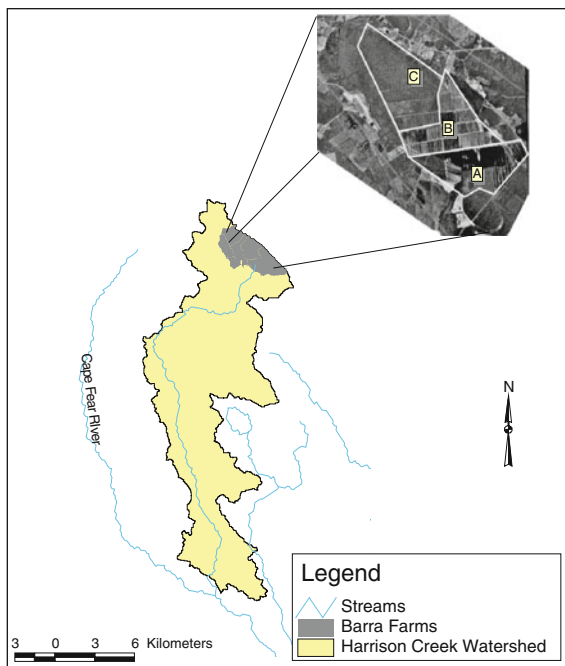


Fig. 2 Location of the Barra Farms study site within the Harrison Creek watershed. Letters on the inset aerial photograph indicate: A) restored wetland, B) active agriculture, and C) reference area

Bay and displays the shallow elliptical basin, organic soils, and broadleaf evergreen plant communities that are characteristic of these wetlands (Bruland et al. 2003). Large areas of this wetland were ditched and converted to agricultural production, but 25% remained undisturbed. In 1997 about 240 hectares of converted agricultural land was restored to wetland, while adjacent agricultural areas remained in production. This spatial arrangement allows for the unique opportunity to monitor areas of Harrison Bay that are largely unimpacted reference areas, areas that remain actively managed for agriculture, and former agricultural areas that have been restored to wetlands (Fig. 2). Flows from these areas were quantifiable because they occur through clearly defined drainage ditches. We delineated the watersheds of the restored, agricultural, and reference areas of Harrison Bay using ArcView[®] GIS. The land-use variables described in Table 1 were compiled for each of these local watersheds. Water samples were collected monthly from the restored, reference, and adjacent agricultural areas for 2 years beginning in the spring of 1999. Each watershed site was gauged to

determine daily streamflow volume between September 1999 and April 2001.

Results and discussion

Regression trees

Natural log transformed ortho-phosphorus (OP) concentration was modeled using 34 potential explanatory watershed variables. The rpart modeling procedure selected a small subset of these explanatory variables for their ability to reduce the variance of the response variable. These included the portion of 90 m wide stream buffers in agriculture (BF.AG.ALL), the percentage of basin-wide area of prior converted croplands (PRIOR), the portion of watershed area in transition from agriculture to forest (WS.TRANS) and instantaneous measure of discharge standardized as millimeters per day (DIS.MM) (Fig. 3). The model has a cross-validated r-square of 0.83. Water samples with the highest OP concentrations (lowest water quality) came from watersheds where prior converted croplands account for more than 4.4% of the total watershed area and where agricultural land uses account for more than 16.9% of buffer areas.

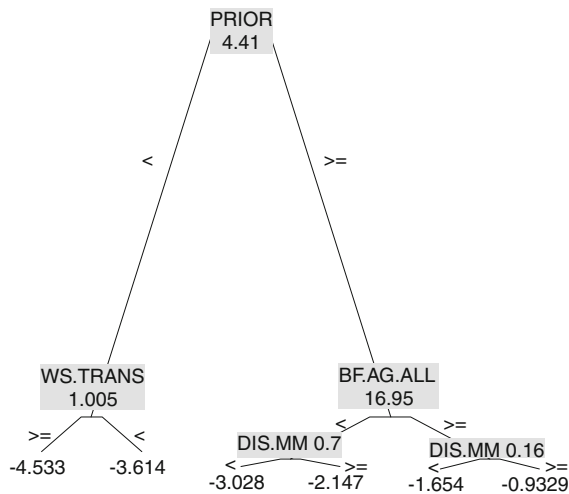


Fig. 3 Regional regression tree of log natural transformed ortho phosphorus concentration. The text above each split shows the variable that is being split and the test below indicates the threshold value identified by the model (% of watershed or buffer area). The text below the terminal nodes shows the natural log of the predicted constituent concentration. See Table 1 for abbreviations

Watersheds with the lowest OP concentrations have small areas of prior converted agricultural lands and the presence of some areas of transitional land-use (WS.TRANS) that often indicate cleared agricultural lands are reverting to forest or wetland due to low commodity prices, or due to conservation set-asides. Thus in targeting restored wetland sites, primary objectives would be to reduce total area of prior converted cropland and increase buffer continuity by locating restored wetlands within stream buffer areas and/or on prior converted croplands.

A regression tree model for natural log transformed total phosphorus (TP) with a cross-validated r-square of 0.71 is presented in Fig. 4. The explanatory variables most able to reduce the variance of the response variable included the proportion of the basin classified as prior converted croplands (PRIOR), wetland fragmentation (WET.PERI.AR), and the portion of stream buffers used for pasture/hay (BF.AG.PAST). The highest TP concentrations occur in watersheds where prior-converted cropland (PRIOR) constitutes more than 4.41% of the basin and substantial portions (>4.78%) of stream buffers are used for pasture/hay (BF.AG.PAST). The lowest total phosphorus concentrations are predicted in watersheds where prior converted cropland constitutes less than 4.41% of the total area, the same as found for ortho-phosphorus model, and wetland fragmentation (WET.PERI.AR) is low (as indicated by perimeter/area ratios of polygons from U.S. Fish

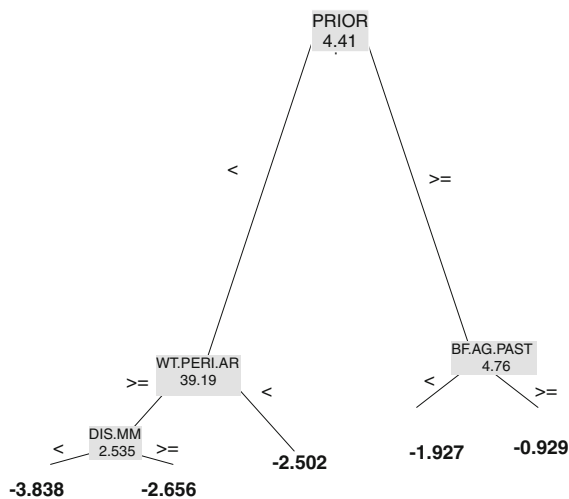


Fig. 4 Regional regression tree model of log natural transformed total phosphorus concentration. See Fig. 3 caption for further explanation

and Wildlife Service National Wetland Inventory). The management implications of this model are much the same as those of the ortho-phosphorus model except greater impetus is given to management actions that reduce wetland fragmentation.

A regression tree model for in stream nitrate-nitrite (NO_x) concentration is presented in Fig. 5. The portion of undeveloped land, i.e., forest and wetland, within the riparian buffer (BF.UNDV) largely explains the variability in stream NO_x concentration. The lowest NO_x concentrations (highest water quality) are predicted for streams where BF.UNDV accounts for greater than 71.66% of the watershed area, evergreen forests (WS.FOR.EVG) coverage is greater than 10.77%, and samples were collected in low flow conditions (DIS.MM < 1.02 mm/day). The highest NO_x concentrations (lowest water quality) occur in watersheds where BF.UNDEV coverage is less than 71.66% and where wetland fragmentation (WT.PERI.AR) is relatively high. The NO_x model presented in Fig. 5 has a cross-validated r-square of 0.71. The management implications of this model support the importance of maintaining riparian buffer continuity and reversing wetland fragmentation in reducing NO_x concentrations in streams. Both of these management directives are readily addressed through stream and wetland restoration.

Random forests

The random forest procedure was used to develop regional models relating watershed explanatory variables to total phosphorus, total nitrogen and nitrate-nitrite concentrations at USGS NAWQA program stations located across the coastal plain of North Carolina. These data were used to “train” our model and water quality data collected from the restored, reference, and agricultural subwatersheds of Harrison Creek were used as “test” data sets. Table 2 lists the *R*² values (%) of predicted versus measured values for both regional (training data set) and local watershed (test data sets) predicting NO_x, TP, and OP concentrations, and the natural log transformed constituents (lnTP, lnOP, lnNO_x). These values ranged from 15 to 94.7%. In our regional models Land-use/Land-cover variables along with discharge explained between 74.81, 75.01, and 83.72 percent of variance in log transformed TP, NO_x, and OP concentrations, respectively. These numbers are quite high for regional

Fig. 5 Regional regression tree model of NO₃–NO₂ concentration. See Fig. 3 caption for further explanation

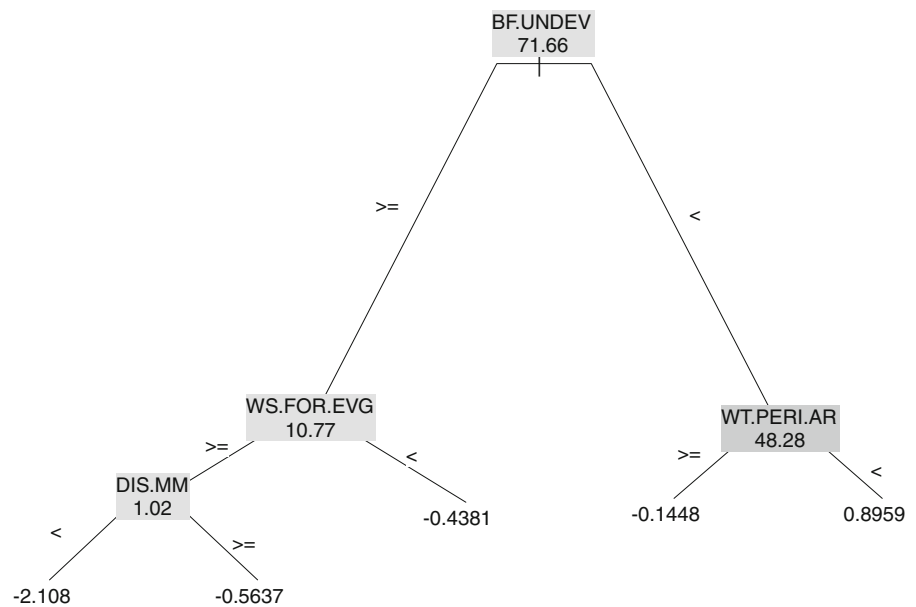


Table 2 The R^2 values (%) between Random Forest model predictions and observed water quality constituent values for both the regional training data set and the local test data sets from the Harrison Creek watershed

Water quality constituent	Regional model	Harrison Creek sub-watershed models		
		Restored	Agricultural	Reference
TP	41.27	58.14	42.47	28.97
lnTP	74.84	60.8	27.6	29.31
OP	42.39	34.19	48.13	35.67
lnOP	83.72	26.18	41.52	45.38
NO _x	76.7	94.73	55.21	8.7
lnNO _x	75.01	72.25	48.33	15.16

TP total phosphorus, OP ortho-phosphorus, NO_x nitrate-nitrite, ln indicates natural log transformation of data

models and suggest our random forest models can provide useful insights about the relationships between watershed land-use/land-cover variables and water quality. Within the local watersheds of Harrison Creek, the random-forest models performed best in the restored watershed where models explained between 26.2 and 94.7% of variance. Accuracy of model predictions were not as strong in the Agricultural and Reference watersheds suggesting that other factors such as human management (fertilizer application) and other stochastic processes played larger roles in these relatively small watersheds.

The random forest procedure generates importance indices describing the predictive value of each potential explanatory variable (Parkhurst et al. 2005; Breiman 2001). While the outputs of the random

forest models lack the graphical representation and the threshold values generated by the rpart procedure, they are expected to be a more robust ranking of the strength of relationship between a given explanatory variable and predicted water quality (Breiman 2001). The 10 explanatory variables with the highest importance values derived from the regional models of OP, TP, and NO_x are shown in Table 3.

The single most important predictor of phosphorus concentrations in the study watersheds, for both OP and TP, was the portion of the watershed classified as prior converted cropland (PRIOR). Several other studies have identified drained agricultural lands as having substantial impacts on water quality in the coastal plain region of North Carolina (Evans et al. 1991, 1995; Skaggs et al. 1994; Dukes and Evans

Table 3 Explanatory variable importance values (IMP; see text) produced by Random Forest analyses of regional TP, OP, and NO_x datasets

NOX		lnTP		lnOP	
Variable	IMP		IMP		IMP
WS.UNDV	28.31	PRIOR	32.76	PRIOR	27.21
BF.NWI	13.04	BF.RES.LI	19.44	BF.WET.EMG	19.36
BF.AG.ALL	12.99	DIS.MM	11.09	WS.WET.EMG	11.30
BF.AG.ROW	10.27	BF.COM	10.08	BF.RES.LI	10.76
WS.AG.ROW	9.51	WS.WET.EM	9.55	DIS.MM	10.68
WS.AG.ALL	9.26	BF.FOR.ALL	9.25	BF.COM	7.18
DIS.MM	8.75	BF.RES.HI	8.45	ANIMAL	7.11
BF.AG.PAST	8.41	BF.WET.EM	8.04	WS.FOR.MXD	6.00
BF.WET.EMG	8.24	WS.TRANS	7.33	WS.TRANS	5.59
WS.WET.EMG	7.68	WS.FOR.MXD	6.14	WS.RES.HI	4.86

The table includes the ten most important explanatory variables from the random forest models presented in Table 2 with the highest R^2 for a given water quality constituent. Variable codes are explained in Table 1

TP total phosphorus, OP ortho-phosphorus, NO_x nitrate-nitrite, ln indicates natural log transformation of data

2006). This is due in part to the enhanced transport of dissolved and sorbed phosphorus to streams due to the presence of drainage systems and erosion of cropland soils.

It has been demonstrated that various systems of controlled drainage can reduce loads of nitrogen and phosphorus from agricultural lands between 30 and 50% (Skaggs et al. 1994; Evans et al. 1995). It can be logically inferred that the controlled drainage resulting from wetland restoration would reduce these nutrient loads by at least 30–50% due to the re-establishment of wetland hydrology, cessation of fertilizer and manure application, reduced soil erosion, increased P sorption, and enhanced denitrification. These results indicate that potential restoration sites located on prior converted croplands should be given the highest priority for restoration. Other high priority explanatory variables for total phosphorus concentration include several urban land-uses. These results were somewhat unexpected given the largely rural setting of the study watersheds. Low intensity residential land-use intrusion into riparian buffers (BF.RES.LI) represented only about 1.7% of the total watershed area and yet was the second most important predictor of TP concentration (lnTP) suggesting that fertilizer application, septic fields, and disruption of riparian buffer continuity had a disproportionate impact on phosphorus concentration. Where appropriate, stream restoration in these areas might present

an opportunity to reestablish riparian buffers and reduce TP loads from these areas.

Emergent wetlands were among the most important predictors of OP concentration (Table 3). Many studies have demonstrated the connection between wetlands and water quality at the watershed scale (Weller et al. 1996; Whigham 1988; Brinson 1988). In our study of watersheds the area of emergent wetlands is strongly related to predicted OP concentrations. It has long been reported that emergent wetlands were very effective in the retention and/or transformation of soluble inorganic phosphorus (Richardson 1985; Kadlec and Knight 1996; Reddy et al. 1999; Vymazal 2007; Mitsch and Gosselink 2007). It is unclear if OP undergoes long-term storage, short-term storage, or transformation in these emergent wetlands (Richardson 1999). However, emergent wetland related explanatory variables (WS.WET.EM, BF.WET.EM) are among the 10 most important predictors of TP concentration (Table 3), suggesting these emergent wetlands are net phosphorus sinks instead of transformers of OP to a organic component of TP (Richardson 1999). A synaptic analysis of GIS coverages suggests that some of these emergent wetland areas may be associated with existing Best Management Practices such as restored wetlands, vegetated swales, and storm water ponds so that wetland restoration practices may already be having a demonstrable affect on water quality in our study area.

The most important predictors of in stream NO_x concentrations were explanatory variables that expressed the relative proportion of disturbed versus natural land-cover within a given watershed. The total proportion of undeveloped land-use in the watershed (WS.UNDEV, see Table 3 for calculations) was the most important predictor of NO_x concentration. The portion of riparian buffers classified as wetland (BF.NWI) was also an important predictor of stream NO_x concentrations. Other studies have linked stream NO_x concentration with the proportion of a watershed classified as natural land-cover (i.e. forest and wetlands) and the continuity of riparian wetlands (Johnston et al. 1990; Gilliam 1994; Groffman 1994; Lowrance et al. 1997; Mitsch et al. 2001). Agricultural land-uses were also important explanatory variables of stream NO_x concentration. These included the proportion of the riparian buffer occupied by all classes of agricultural lands (BF.AG.ALL), and row crops (BF.AG.ROW); and the proportion of the watershed occupied by all agricultural lands (WS.AG.ALL), and row crops (WS.AG.ROW). These results present several management implications for reducing stream NO_x concentrations in the coastal plain of North Carolina. Shifts in watershed land-use away from agricultural to natural land-uses will reduce in stream NO_x concentrations. Such shifts might include conservation set-asides and/or large stream and wetland restoration projects. Reducing the area of agricultural land-uses and increasing wetlands in riparian buffers would decrease predicted in stream NO_x concentrations, and these management goals could be accomplished through the use of integrated stream and wetland restoration projects in the riparian zone (Evans et al. 2007).

Monitoring

The final component of our proposed framework is monitoring the effectiveness of a high priority wetland restoration site to confirm if restoration will, as predicted by our regression tree models, improve water quality on the landscape. The wetland restoration project undertaken at the Barra Farms site was consistent with the site priorities produced by our regional modeling framework in that: (a) the site is a prior converted agricultural area with a well developed system of drainage ditches and (b) there was

extensive disturbance of the riparian buffer. In the spring of 1999, we began monitoring at the Barra Farms Cape Fear Regional Mitigation Bank. Water quality data from both the restored wetland area and from adjacent agricultural areas indicated substantial differences in nutrient concentration. These data suggest that the restored wetland is having a substantial positive impact on the water quality of Harrison Creek, and consequently on the Cape Fear River. For example, boxplots of nutrient concentrations in grab samples collected from the restored wetland, the agricultural area and from a bay forest reference area are shown in Fig. 6. The mean TP, OP and NO₃–NO₂ concentrations were consistently lowest at the reference site, intermediate at the restored site, and highest at the agricultural site. There is noticeably greater variability in nutrient concentrations from the agricultural sites. Importantly, we found that nutrient runoff concentrations for NO₃–NO₂ N and total phosphorus data collected during a hurricane related storm-peak were consistently 5–7 times lower in the restored wetland than in agricultural area runoff (Fig. 7). These results provide a basis to validate the watershed-level site prioritization produced by our tree models, and

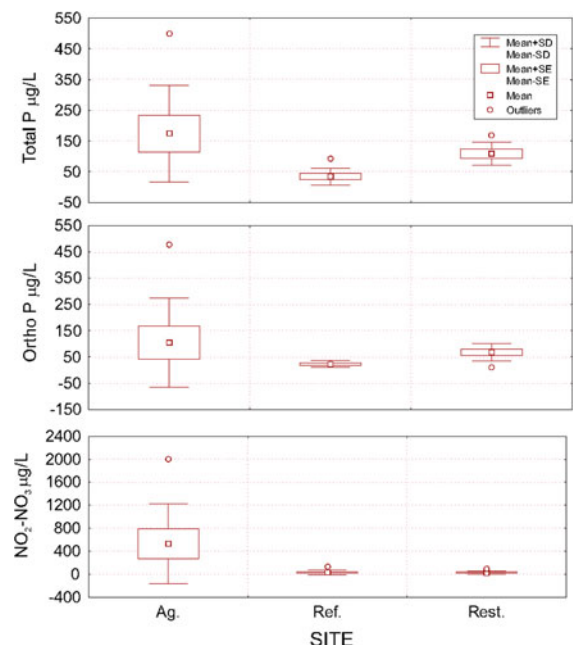


Fig. 6 Boxplots of long-term water quality data collected between April, 1999 and April 2001 from the agricultural, reference and restored areas of Harrison Bay (Barra Farms site)

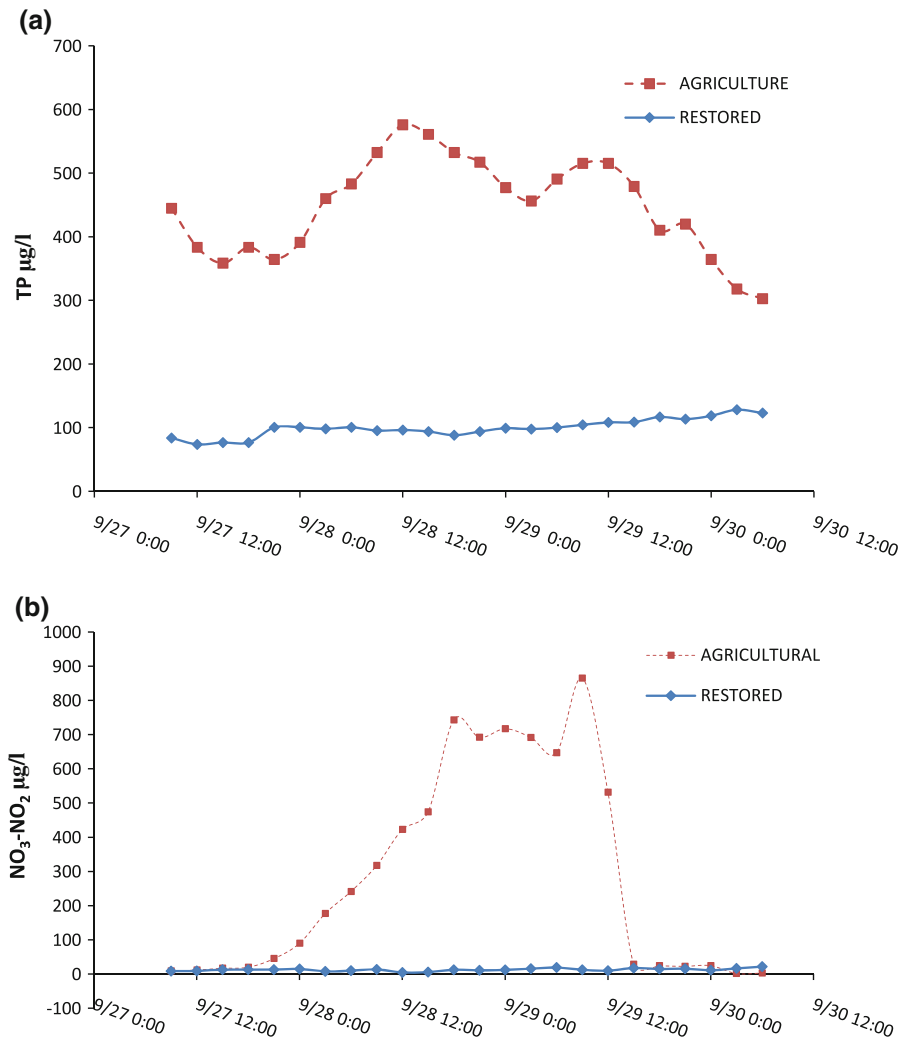


Fig. 7 Nitrate-nitrite nitrogen and total phosphorus concentrations in storm samples from taken between 9/24/99 and 9/27/99. Stations are located in the reference (REF), restored (REST) and agricultural (AG) areas of the Barra Farms site. See Fig. 1

demonstrate the water quality effects of the ensuing restored wetlands. Restoration activities at the Barra Farms Cape Fear River Mitigation Bank resulted in conversion of 250 ha of drained agriculture land to restored wetland (See area A of inset in Fig. 2), mitigating the water quality impacts fertilizer application to the site before restoration (Bruland et al. 2003). Collectively our findings suggest that a model and prioritization approach which helps select important land use types and related water quality provides a basis for restoring wetlands on the landscape to provide a significant improvement in water quality relative to agricultural and other land-uses.

Conclusions

Prioritizing restoration sites on the landscape is often a haphazard process when water quality improvement is a primary restoration objective. Prioritization is often based on widely held, though often untested, assumptions about relationships between watershed characteristics and water quality. We present a framework to target and prioritize wetland restoration locations using both regional and watershed-level screening models. The regression tree and random forest models presented in this paper identify the variables with the strongest relationships to a given

water quality variable, present a clear hierarchy of variable importance, and present approximate thresholds in watershed area where these variables express the greatest impact on water quality. The tree models could be used as “stand alone” decision tools in watersheds where limited restoration opportunities make less robust and less costly decision making approaches more appropriate. While the Random Forest models do not provide the graphical decision tree produced by the regression-tree procedure nor the estimate of thresholds, they do produce much more accurate predictions of water quality based on land-use/land-cover and discharge. Prior converted agricultural lands were identified as important predictors of both ortho and total phosphorus. These sites have poorly-drained hydric soils that require artificial drainage to be suitable for agriculture. These drainage systems become conduits for transporting dissolved and total phosphorus from agricultural field and to area streams and rivers (Skaggs et al. 1994). Fortunately because these areas were historically wetlands, they are often very suitable for wetland restoration. The continuity of stream buffers is identified as another important predictor of water quality. This seems to be especially true with regard to $\text{NO}_3\text{--NO}_2$ concentrations. Both the rpart and random forest models agree that maintaining natural land-cover within the 90 m buffer is a fundamental strategy to maintaining low $\text{NO}_3\text{--NO}_2$ concentrations. Specifically, it is recommended that: (a) agricultural land-uses are excluded from riparian buffers, (b) wetland land-cover is increased and c) wetland fragmentation is reduced (Fig. 5 and Table 3). All of these objectives can be achieved by conversion of selected agricultural land-cover to wetland through stream and/or wetland restoration. Low intensity residential land-use is also important predictor of phosphorus concentration in streams, likely due to septic system failure and fertilizer application in yards. Reestablishment of riparian buffers associated with stream restoration projects in urban and semi-urban residential areas may considerably reduce OP and TP concentration in streams.

Finally, the models present in this paper were not primarily developed to predict in stream nutrient concentrations, but to identify important wetland-related landscape variables that can be used for selection of restoration criteria where appropriate. We have identified several such variables and have

presented a graphical tool with an easily communicable format for conveying our results to watershed managers.

Acknowledgements This study was funded by grant OEP:9804857 from the United States Department of Agriculture Cooperative State Research, Education, and Extension Service. We would like to thank Troy Rogers, Wes Willis and other faculty, staff and students of the Nicholas School of the Environment at Duke University for their help in completing the field work and laboratory analyses that made this study possible. We also wish to thank our reviewers for their suggested improvements to our manuscript.

References

- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Breiman L, Friedman JH, Olshen R, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA
- Brinson M (1988) Strategies for assessing the cumulative effects of wetland alteration on water quality. *Environ Manage* 2(12):655–662
- Brunland GL, Richardson CJ (2005) Spatial variability of soil properties in created, restored, and paired natural wetlands. *Soil Sci Soc Am J* 69:273–284
- Brunland GL, Richardson CJ (2006) Comparison of soil organic matter in created, restored and paired natural wetlands in North Carolina. *Wetlands Ecol Manage* 14:245–251
- Brunland GL, Hanchey MF, Richardson CJ (2003) Effects of agriculture and wetland restoration on hydrology, soils, and water quality of a Carolina bay complex. *Wetlands Ecol Manage* 11:141–156
- Childers DL, Gosselink JG (1990) Assessment of cumulative impacts to water quality in a forested wetland landscape. *J Environ Qual* 19:455–464
- De Ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81(11):3178–3192
- Dukes MD, Evans RO (2006) Impact of agriculture on water quality in the North Carolina middle coastal plain. *J Irrig Drain Eng* 132(3):250–262
- Evans RO, Skaggs RW, Gilliam JW (1991) Controlled drainage management guidelines of improving water quality. *Bulletin AG-443*. N.C. Cooperative Extension Service, Raleigh, NC
- Evans RO, Gilliam JW, Skaggs RW (1995) Controlled versus conventional drainage effects on water quality. *J Irrig Drain Eng* 121(4):271–276
- Evans RO, Bass KL, Burchell MR, Hinson RD, Johnson R, Doxey M (2007) Management alternatives to enhance water quality and ecological function of channelized streams and drainage canals. *J Soil Water Conserv* 62(4):308–320
- Gilliam JW (1994) Riparian wetlands and water quality. *J Environ Qual* 23:896–900
- Groffman PM (1994) Denitrification in freshwater wetlands. *Curr Top Wetland Biogeochem* 1:15–35

- Johnston CA, Detenback NE, Niemi GJ (1990) The cumulative effect of wetlands on stream water quality and quantity: a landscape approach. *Biogeochemistry* 10(2):105–142
- Kadlec HR, Knight RL (1996) *Treatment wetlands*. Lewis, Boca Raton, FL, p 893
- Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* 2(3):18–22
- Lowrance R, Altier LS, Newbold JD, Schnabel RR, Groffman PM, Denver JM, Correll DL, Gilliam JW, Robinson JL, Brinsfield RB, Staver KW, Lucas W, Todd AH (1997) Water quality functions of riparian forest buffers in Chesapeake Bay watersheds. *Environ Manage* 21(5):687–712
- Malina JF Jr (1996) Water quality. In: Mays L (ed) *Water resources handbook*, Chapter 8. McGraw-Hill, New York
- Mays LW (ed) (1996) *Water resources handbook*. McGraw-Hill, New York
- Mitsch WJ, Gosselink JG (2007) *Wetlands*, 4th edn. Wiley, New York
- Mitsch WJ, Reeder BC (1991) Modeling nutrient retention of a freshwater coastal wetland: estimating the roles of primary productivity, sedimentation, resuspension and hydrology. *Ecol Modell* 54:151–188
- Mitsch WJ, Cronk JK, Wu X, Nairn RW, Hey DI (1995) Phosphorus retention in constructed freshwater riparian marshes. *Ecol Appl* 5(3):830–845
- Mitsch WJ, Day JW, Gilliam JW, Groffman PM, Hey DL, Randall GW, Wang N (2001) Reducing nitrogen loading to the Gulf of Mexico from the Mississippi River Basin: strategies to counter a persistent ecological problem. *Bioscience* 51(5):373–388
- Parkhurst DF, Brenner KP, Dufour AP, Wymer LJ (2005) Indicator bacteria at five swimming beaches—analysis using random forests. *Water Res* 39(7):1354–1360
- Qian SS (2009) *Environmental and ecological statistics*. R. Chapman and Hall/CRC, Boca Raton, FL
- Qian SS, Anderson CW (1999) Exploring factors controlling the variability of pesticide concentrations in the Willamette River Basin using tree-based models. *Environ Sci Technol* 33:3332–3340
- Qian SS, Warren-Hicks W, Keating J, Moore DRJ, Teed RS (2001) A predictive model of mercury fish tissue concentrations for the southeastern United States. *Environ Sci Technol* 35:941–947
- R Development Core Team (2008) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Reddy KR, Kadlec RH, Flaig E, Gale PM (1999) Phosphorus retention in streams and wetlands: a review. *CRC Crit Rev Environ Sci Technol* 29:83–146
- Richardson CJ (1985) Mechanisms controlling phosphorus retention capacity in freshwater wetlands. *Science* 228:1424–1426
- Richardson CJ (1999) The role of wetlands in storage, release, and cycling of phosphorus on the landscape: a 25 year retrospective. In: Reddy KR (ed) *Phosphorus biogeochemistry in sub-tropical ecosystems*. CRC Press/Lewis Publishers, Boca Raton, FL, USA, pp 47–68
- Richardson CJ, McCarthy EJ (1994) Effect of land development and forest management on hydrologic response in southeastern coastal wetlands: a review. *Wetlands* 14: 56–71
- Richardson CJ, Evans R, Carr D (1981) Pocosins: an ecosystem in transition. In Richardson CJ (ed) *Pocosin wetlands*. Hutchinson Ross Publishing Co., Stroudsburg, PA, pp 3–19
- Ripley BD (1981) *Spatial statistics*. Wiley, New York
- Sharitz RR, Gresham CA (1998) Pocosins and Carolina bays. In: Messina MG, Conner WH (eds) *Southern forested wetlands. Ecology and management*. Lewis Publishers, Boca Raton, FL, pp 343–377
- Skaggs RW, Breveacute MA, Gilliam JW (1994) Hydrologic and water quality impacts of agricultural drainage. *Crit Rev Environ Sci Technol* 24(1):1–32
- Therneau TM, Atkinson B (2005) Rpart: recursive partitioning. R package version 3.1–27. S-PLUS 6.x original at <http://mayo-research.mayo.edu/mayo/research/biostat/splusfunctions.cfm>
- Tiner RW (2003) Geographically isolated wetlands of the United States. *Wetlands* 23(3):494–516
- U.S. Fish and Wildlife Service. National Wetlands Inventory Center (1971–2005) National Wetlands Inventory Electronic Map Files. hdl:1902.5/784038 National Archives and Records Administration [Distributor]
- Venables WN, Ripley BD (1999) *Modern applied statistics with S-PLUS*, 3rd edn. Springer, 501 pp
- Vogelmann JE, Howard SM, Yang L, Larson CR, Wylie BK, Van Driel JN (2001) Completion of the 1990's National Land Cover Data Set for the conterminous United States. *Photogramm Eng Rem Sens* 67:650–662
- Vymazal J (2007) Removal of nutrients in various types of constructed wetlands. *Sci Total Environ* 380:48–65
- Weller CM, Watzin MC, Wang D (1996) Role of wetlands in reducing phosphorus loading to surface water in eight watersheds in the Lake Champlain Basin. *Environ Manage* 20:731–739
- Whigham DF (1988) Impacts of freshwater wetlands on water quality: a landscape perspective. *Environ Manage* 12(5): 663–674