# Development of Regional Pollution Export Coefficients For the N-SPECT Model: Use of USGS National Water Quality Assessment (NAWQA) Summary Data To Prototype A Total Nitrogen Regional Coefficient

Michael Schenk (NWS), Nina Garfield (NOS), Dave Eslinger(NOS), Jamie Carter(NOS)

## Introduction

The National Ocean Service's Coastal Services Center has published the Non-Point Source Pollution and Erosion Control Tool (N-SPECT)[4], an extension of ArcGIS, as a screening tool to estimate the water pollution and sediment impacts from changes in land use /land cover. N-SPECT uses pollutant export coefficients (PECs) to quantify the relationship between land use/land cover and pollutant amounts. It is capable of estimating the change in pollutant amount in response to a change in land use/land cover; "what if" scenarios can be conducted by altering the land use GIS layer. N-SPECT provides a default set of PECs derived from studies of various regions. However, N-SPECT users need a method to derive locally-relevant coefficients in order to calibrate the model for a given watershed and increase its value to coastal managers. The National Estuarine Research Reserve System (NERRS) seeks to use this model to test impacts from various scenarios of land use/land cover change.

This study addresses two questions related to development of these local export coefficients:

**Is there an accessible method which N-SPECT users can employ to develop local pollution export coefficients?**

Linear regression methods are relatively simple and can be performed using most common statistical applications.

**Can these coefficients be developed on a regional scale?**

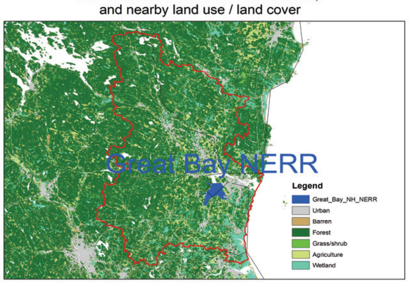
A regional scale increases the sample size and improves results, within limits.

Due to time constaints, this study tested these questions in a limited area of the Northeastern US.

## Study Area

The Great Bay, New Hampshire NERR was selected as a development area, because of the availability of data and the estuary's relative isolation from external hydrologic influences. Coastal Change Analysis Program (C-CAP) land use/land cover data were obtained for the Great Bay watershed and surrounding area. To facilitate analysis, these data were simplified to equivalent Anderson Level 1 categories using a GIS reclassification tool.

Great Bay (NH) National Estuarine Research Reserve, with associated watershed of influence, and nearby land use / land cover




## Water Quality Data for Developing Coefficients

**Required**

- A sufficiently large set of water quality (WQ) data from which a mean concentration of the subject pollutant can be calculated.
- Associated data on stream flow during base flow and storm events and the distribution of land uses in the area which drains to the sampling point.
- Nationally consistent protocols, methods, and quality control standards for water quality sample collection and analysis. National consistency in standards will facilitate future comparisons between NERRs.

**Available**

Many online databases such as NWIS (US Geological Survey) and STORET ( US Environmental Protection Agency) contain large amounts of information on discrete water quality samples taken at a given site. These databases contain "raw" information, which must then be associated with runoff events and land use/land cover distributions in order to develop PECs.

The USGS **National Water Quality Assessment (NAWQA)** program has published a summary[3] of an intensive study of nutrients in 51 basins nationwide. This summary reports annual mean flow, concentration, load and yield for several species of nitrogen and phosphorus, along with land use distribution, associated sources of nutrients, and other ancillary data, for nearly 500 locations. Each sampling site for which data is reported represents multiple discrete WQ samples. The NAWQA summary data facilitate rapid development of PECs, since runoff, flow, and land use data are provided with annual mean nutrient data.

## Statistical Methods

Export coefficients were developed using a simple multiple linear regression model. The predictors are the proportion of each land use category; stream water concentration of total nitrogen ([TN]) is the outcome variable. The y intercept was set to zero on the principle that zero land area should yield zero TN.

Eq. 1 $$[TN] = \beta_1 A_1 + \beta_2 A_2 + \ldots + \beta_n A_n$$

Where [TN] = total nitrogen concentration, A is the fraction of land category in the sampled watershed, and $\beta$ is the export coefficient for that land category.

The goal was to develop a simple method to estimate export coefficients for use in N-SPECT, not to provide a robust predictive model. Therefore, no attempt was made to normalize the data. Overall fit of the equation and number of significant variables were used to judge the success of the equation. The NAWQA summary employs six land use categories. Only five categories are considered here, since the Range category is unrepresented on the East Coast. In many cases, the Forest variable was small and statistically insignificant in many cases. A small coefficient for Forest makes physical sense in that this category will likely contribute little nitrogen; therefore, when Forest was small and not significant that coefficient was included in testing. In some instances, Forest was found to be negative, indicating that Forest may be removing nitrogen. In those cases, zero was substituted, since N-SPECT does not accept negative coefficients.

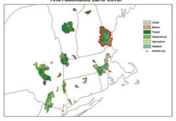
## Does this method generate useable coefficients?

With a sufficient sample size, and with the y intercept set to zero, the regression model produced coefficients which are well within the ranges published in the literature. An examination of 27 Northeastern sampling sites was conducted, with large and nested watersheds. In addition, the watersheds for these 27 sites were reclassified to six Anderson Level 1[1] land categories, which are similar to C-CAP categories. Within limits, increasing the sample size by expanding the spatial extent of the sampled area improved the $R^2$ of the regression and increased the number of significant variables. At this sample size, there was little difference in performance using five NAWQA or six Level 1 land categories.

Regression results for 27 Northeastern sites, $R^2$=0.81

| Category | NAWQA Coef. | Standard Error | P | N-SPECT default |
|---|---|---|---|---|
| Urban | 2.080 | 0.558 | 0.001 | 2.220 |
| Cropland | 12.936 | 3.387 | 0.001 | 2.680 |
| Pasture | 2.542 | 1.414 | 0.086 | 2.480 |
| Forest | -0.561 | 0.472 | 0.247 | |
| Other | 8.412 | 8.093 | 0.310 | |

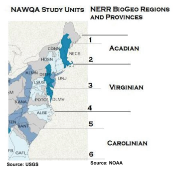
27 NAWQA Sites in New England Area And Associated Land Cover



Although this regression value for Cropland is almost five times the default of 2.68, the regression results are within an order of magnitude, and the rank order Crop>Pasture>Developed is the same. The results for this very small sample are thus near some accepted values, indicating that this method's results are at least approximately correct. The regression equation should be improved and tested. Increasing the sample size is a first step toward this.

## Regional scale: expanding area improves results - within limits

We increased our sample size by cumulatively adding coastal NAWQA study units to the regression, proceeding southward. As study units were added the number of significant variables increased, and $R^2$ stayed roughly the same, up to a point. After ALBE, the number of significant variables and the $R^2$ decreased. A separate set of regressions which included the near-coastal POTO (Potomac) study unit showed a decline in results once POTO was included. It is possible that the decline in significant variables at SANT may be associated with the upland headwaters of SANT, which have similarities with POTO. As an additional test, we derived a set of coefficients for all six NAWQA categories for 360 national sampling sites. The coefficients for Urban, Crop, and Pasture were near the values found for ALBE.




Cumulative regression results

| Study Unit | n | $R^2$ | Sig Vars |
|---|---|---|---|
| NECB | 8 | 0.93 | 0 |
| CONN | 18 | 0.82 | 2 |
| HDSN | 27 | 0.81 | 2 |
| LINJ | 34 | 0.83 | 3 |
| DELR | 42 | 0.81 | 3 |
| DLMV | 44 | 0.84 | 4 |
| ALBE | 50 | 0.83 | 5 |
| SANT | 56 | 0.81 | 4 |
| GAFL | 63 | 0.73 | 3 |

This suggests that there is an optimal extent to which the region of consideration can be expanded, before local differences in the characteristics of the land categories considered exceed the benefits of increasing sample size.

## Discussion

Given time constraints, the study concluded with a general test of regression methods and an examination on the East Coast of regression equation performance versus spatial scale.

These results suggest that multiple linear regression methods can be employed by users of N-SPECT to compute local pollution export coefficients. Data used in multiple linear regression should be of the largest possible sample size and be normally distributed. The variables selected for the regression equation also should not be correlated with each other. Plots of the residuals from these regressions showed that the distribution of many of the variables was somewhat non-normal. Also, a correlation matrix showed that Forest and Urban were highly negatively correlated. Despite these departures from the assumptions of regression, and the small sample size, the coefficients obtained were generally within the range of values reported in the literature. One reason for this may be that the NAWQA summary data are particularly robust. Each data point from the summary represents extensive processing and quality control performed on many discrete WQ data samples. These data points are then associated with land category data. The annualized mean flow for each of sampling sites is also derived by well-proven methods. In all cases, the WQ sample data was associated with flow and land cover data of the approriate time span. Use of the NAWQA summary data precludes the necessity of estimating storm event related concentrations from disparate data sets, and derivation of land category distributions by GIS analysis, thus conserving a large portion of the effort required to estimate export coefficients.

In this study we increased sample size by increasing the spatial domain under consideration. As an additional test, we derived a set of coefficients for all six NAWQA categories for 360 national sampling sites. The coefficients for Urban, Crop, and Pasture were near the values found for ALBE. We encountered a limit where the expansion of the spatial domain eventually introduced additional variance to the regression, which we attribute to local diffferences in the attributes of land categories. For example, does southern Forest process nitrogen the same as northern Forest ? This spatial variance may be linked to latitude, climate zone, physiographic province, species distributions, local land practices, or other factors or combinations of factors. Is the change in the regression relationships at SANT linked to the latitude of the study unit or the elevation of its headwaters? Since including POTO in the cumulative regression had a similar effect, the latter is suggested. It would be worth pursuing the expansion of spatial domain by selecting individual NAWQA sampling sites which share geographic attributes (e.g., location in low elevation coastal areas) rather than by the quicker method of adding NAWQA study units to the regression.

The sample size can also be increased by increasing the density of sampled points within a given area. To date, USGS has completed Cycle 1 of the NAWQA, but sampling of watersheds which were not included in Cycle 1 is in progress. It is reasonable to expect that future NAWQA summaries will have a spatially denser set of data points, which should increase the accuracy and precision of these export coefficient estimates. Also, intensive regional studies such as the New England SPARROW (Spatially-Referenced Regression On Water) model nutrient study[2] may have similarly summarized data which coud be incorporated.

When interpreting the results of our regressions, the assumption was made that the regression which produced a high $R^2$ and the most significant variables would also produce the most accurate export coefficients. This assumption needs to be tested.

## Further Steps

The most important step is to test regression-derived export coefficients against independent accurate data. Are these coefficients better than the default coefficients, or better than those derived using the national NAWQA summary set? Also, the coefficients for several sotail extents should be tested, rather than the set which appears to be optimal. This testing should be accomplished by deploying the N-SPECT model in one or more watersheds which have been extensively characterized for nitrogen behavior. One possibility would be to remove individual NAWQA stations from the regression, and run N-SPECT on these. Another possibility would be to compare the results of an independent study such as the New England SPARROW nitrogen study with N-SPECT's predictions using default, national, and regional coefficients.

Optimization of the regression equation could also improve accuracy. The New England SPARROW model achieved high accuracy ($R^2$=0.95). In this model, atmospheric deposition of nitrogen and municipal point sources were significant. Adding variables such as these to the regression should be tested.

Different geographic attributes used to group data for regional coefficients , such as physiographic province, should be tested to assess the optimal grouping of sampling sites for regional coefficient development.

### Literature cited

1. Anderson, James R., et al, 1976. A Land Use And Land Cover Classification System For Use With Remote Sensor Data, Geological Survey Professional Paper 964
2. Moore, Richard Bridge, et al, 2004, Estimation of Total Nitrogen and Phosphorus in New England Streams Using Spatially Referenced Regression Models, US Geological Survey Scientific Investigations Report 2004-5012
3. Mueller, David K., and Spahr, Norman E., 2006, Nutrients in streams and rivers across the Nation—1992–2001: U.S. Geological Survey Scientific Investigations Report 2006–5107
4. NOAA Coastal Services Center, http://www.csc.noaa.gov/crs/cwq/nspect.html

### Acknowledgements

The authors would like to thank David K. Mueller, Norman E. Spahr, and Naomi Nagataki of the US Geological Survey for their kind assistance and encouragement, Valery Dagostaro (NWS) and Wilson Shaffer (NWS) for reviews and comments, Shan Burkhalter (NOS) for N-SPECT consultation, and NERR staff for suggestions as to data sources..

This work was performed as a NOAA Rotational Assignment Program task. The NOAA National Weather Service provided time and resources for this assignment.

The background image is a mosaic of 30 meter resolution Northeastern US CCAP data, accessed September 2008 from http://www.csc.noaa.gov/crs/lca/northeast.html . The colors have been reduced to 40% saturation.