Supplementary Material

This appendix is intended to provide a more detailed description of our data, analysis, and results. We hope this allows others to better interpret and replicate our findings.

1. Data

Table S1 includes all the variables examined in our spatial analysis and their descriptive statistics. The Long Shoal-Gull Shoal watershed was omitted because it is an outlier in area. This watershed is 1026.64 km^2, approximately twice as large as the next largest watershed at 532 km^2. Watersheds that cross state boundaries were also clipped so only the area under DMS's jurisdiction was included. Income and College variables were not included in our models due to their high VIFs and correlation with other variables. College was highly correlated with population change ($r = 0.41$), population density ($r = 0.52$), tier ($r = 0.42$), income ($r = 0.35$), and home value ($r = 0.55$). Income was highly correlated home value ($r = 0.83$). We chose to remove income rather than home value because home value can also be understood as a proxy for land prices, which may be an important variable in determining the availability of land for mitigation projects.

Table S1. Variables and descriptive statistics.

Variable	Definition	Range	Mean	Std Dev	VIF
Projects	Number of wetland mitigation projects in each watershed	0 – 6	0.31	0.75	
Projects Binary	No mitigation project in watershed =0; mitigation project in watershed = 1	0 – 1	0.19	0.40	
Watershed Area	Area of watershed in square kilometers	2.35 – 532.60	81.11	34.2	1.03
Watershed Modification	Is the watershed modified? No modification = 0; any modification = 1 For example does the watershed have a dam or other obstruction to flow?	0 – 1	0.21	0.40	1.13
TLW	Is the watershed categorized as a targeted local watershed? Yes = 1; No = 0	0 – 1	0.41	0.49	1.17
LWP	Is the watershed covered in a local watershed? Yes = 1; No = 0	0 – 1	0.09	0.29	1.16
Zoning	Does the watershed fall completely in a county with zoning of unincorporated areas? Yes = 1; No = 0	0 – 1	0.52	0.50	1.98
Partial zoning	Does the watershed have partial county zoning? (either falls partially in county with zoning of unincorporated areas or partial county zoning) Yes = 1; No = 0	0 –1	0.27	0.44	1.79
Population Change	Percent change in population between 2000 and 2010; Calculated: Area Weighted Mean (AWM) of 2000 population – AWM 2010 Pop / AWM 2000 POP	-1 – 7.28	0.15	0.30	1.41

Variable	Definition	Range	Mean	Std Dev	VIF
College	Percent population over 25 years of age with bachelors, masters, professional or doctoral degree	0.02 – 0.75	0.17	0.11	2.59
Population Density	Area weighted mean of population density per square mile	0 – 3326.37	191.57	342.09	1.90
Income	Log of area weighted mean of median household income in 2010 Inflation adjusted dollars	2.85 – 11.8	10.62	0.48	2.84
Home Value	Log of area weighted mean of median home value of owner occupied homes	3.89 – 13.63	11.69	0.56	3.99
Rent	Area weighted mean of median gross rent of renter–occupied housing	0 – 2001.0	587.6	208.83	1.35
Tier	Area weighted mean of tier designation, 1 = 40 most distressed counties, categorical only possible values are 1, 2, or 3	1 – 3	1.83	0.76	1.49
Years as TLW	Number of years watershed has been a targeted local watershed	0 – 11	2.73	4.00	1.22
Years as LWP	Number of years watershed has had a local watershed plan	0 – 9	0.51	1.87	1.15

We are interested in two variables in particular - *targeted local watersheds* and *local watershed planning*. Targeted local watersheds are DMS's implementation of the watershed approach, using watershed information they target specific watersheds for restoration. The DMS attempts to have 25 – 40% of each 8-digit watershed designated as targeted local watersheds. In our dataset, 710 watersheds are targeted local watersheds or 40.8% of the watersheds. Each year based on need a number of targeted local watersheds are selected to undergo local watershed planning. Areas expected to have the most wetland and stream impacts and associated mitigation need are prioritized for planning. There are 147 watersheds with a local watershed plan or 9% of watersheds.

In addition to the dataset used to spatially compare watersheds, we created a temporal dataset to evaluate causal links between targeted local watersheds and local watershed plans and mitigation projects. For each year from 1998 – 2012 we recorded: 1) the number of projects established in every watershed; 2) whether the watershed was a targeted local watershed; and 3) whether the watershed was included in a local watershed plan. Watersheds that did not have complete date records (i.e., missing were the year of local watershed plan, targeted local watershed, or project establishment) were not included in this analysis. There were 480 projects established over 419 watershed-years; there are 5388 targeted local watershed-years and 1026 local watershed plan watershed-years. We used the earliest date available for each project recorded in the DMS Project Documents (NCDENR, 2013a). Most of these dates correspond to the acquisition of a project site and project design. Dates for targeted local watershed establishment were collected from the 39 River Basin Restoration Priority Plans (NCDENR, 2013b). We assumed that once targeted local watersheds are established they remained targeted areas. The year each local watershed plan was completed is included in the DMS GIS dataset (NCDENR, 2013a). Nearly all watersheds become targeted local watersheds before the establishment of local watershed plans; however, data for 33 watersheds indicate that planning was completed before they became targeted. These watersheds were used in all the analyses except the comparison of rates between pre-treatment, targeted local watershed and local watershed plans because they did not fit the required data structure (we had no measure of rate when the watershed was only targeted).

2. Analysis and Results

Our first analysis explores the influence of watershed characteristics on the presence of mitigation projects. We first ran a logistic regression with all of the variables in our dataset without prohibitively high VIF or correlations (results shown in Table S2). In the logistic regression, only targeted local watershed, population density, and tier were found to be statistically significant. Consequently, the other socio-demographic variables are dropped in future analyses. The more parsimonious model, containing only targeted local watershed, local watershed plan, population density, and Tier, has better model fit as indicated by AIC (1606.9) and BIC (1634.2).

Table S2. Coefficient estimates for logistic regression using all variables to predict if a watershed has a mitigation project (N = 1740).

	Coefficient Estimate
Intercept	-1.298 (1.40)
Watershed Area	0.003 (0.001)
Watershed Modification	-0.036 (0.161)
Targeted Local Watershed	0.823 (0.137)*
Local Watershed Plan	0.337 (0.203)
Zoning	-0.053 (0.176)
Partial Zoning	-0.099 (0.195)
Population Change	-0.079 (0.217)
Population Density	0.0004 (0.0002)*
Home Value	-0.126 (0.127)
Rent	0.0004 (0.0004)
Tier	0.212 (0.096) *
AIC	1613.7
BIC	1673.8

Due to the spatial nature of our questions, spatial-autocorrelation is a concern. We used two methods to test for spatial autocorrelation. First, we calculated the Moran's I using ArcMap's spatial statistics tool. Projects were more clustered than would be expected at random (Moran's I = 0.188, $p < 0.001$). The spatial clustering of projects was unsurprising considering the clustering of variables that are likely to influence the location of projects such as population density (Moran's I = 0.646, $p = 0.001$), tier (Moran's I = 0.872, $p = 0.001$), and home value (Moran's I = 0.713, $p = 0.001$). We chose not to model spatial autocorrelation in our primary models after performing a secondary analysis, which included the longitude and latitude of watersheds centroids in a binary logistic model to determine if location within the state influences the presence of mitigation projects in the watershed. Neither longitude, latitude, nor the interaction between them was found to be statistically significant ($p = 0.441$, $p = 0.416$, $p = 0.401$ respectively), demonstrating that projects are dispersed throughout the state.

The high number of zeros in our data makes modeling difficult. While the error rate of the logistic models used was relatively low- the model using dummy variables for targeted local watersheds and local watershed plans only misclassifies 18.8% of watersheds and the model using years as TLW and LWP misclassifies 18.4% (i.e., model predicts 0 projects when it in fact has 1, or model predicts 1 project when it in fact has none). However, due to the large number of watersheds with no projects a null model (a model predicting all watersheds have no projects) also only has an error rate of 18.9%.

Table S3. Coefficient estimates for logistic regression including longitude and latitude to predict if a watershed has a mitigation project. (N = 1740).

	Model 1	Model 2
Intercept	-140.6 (190)	145.6 (183.2)
Watershed Modification	-0.024 (0.163)	
Targeted Local Watershed	0.874 (0.137)*	0.877 (0.137)*
Local Watershed Plan	0.227 (0.203)	0.229 (0.201)
Zoning	-0.045 (0.179)	
Partial Zoning	-0.008 (0.214)	
Population Change	-0.008 (0.214)	
Population Density	0.0004 (0.0002)*	0.0004 (0.0002)*
Home Value	-0.159 (0.125)	
Rent	0.0005 (0.0004)	
Tier	0.253 (0.102) *	0.247 (0.092)*
Latitude	1.958 (2.410)	2.042 (2.321)
Longitude	-4.124 (5.348)	-4.283 (5.148)
Latitude*Longitude	-0.057 (0.068)	-0.059 (0.065)
AIC	1613.7	1596.8
BIC	1673.8	1640.5

The high number of zeros in our data further influences the fit of count models. Figure S1 shows the distribution of the number of projects in our data, predicted by the negative binomial model, and predicted by the hurdle Poisson model. The negative binomial and hurdle Poisson model produce very similar results - both do a relatively good job of predicting the number of watersheds with 0 and 1 project but underestimate the number of watersheds with more projects. We additionally tested a Poisson and zero-inflated Poisson, neither performed better than the negative binomial and hurdle Poisson (which we chose to focus on due to their congruence with our assumptions about differing processes between establishment of the first and additional projects in a watershed).

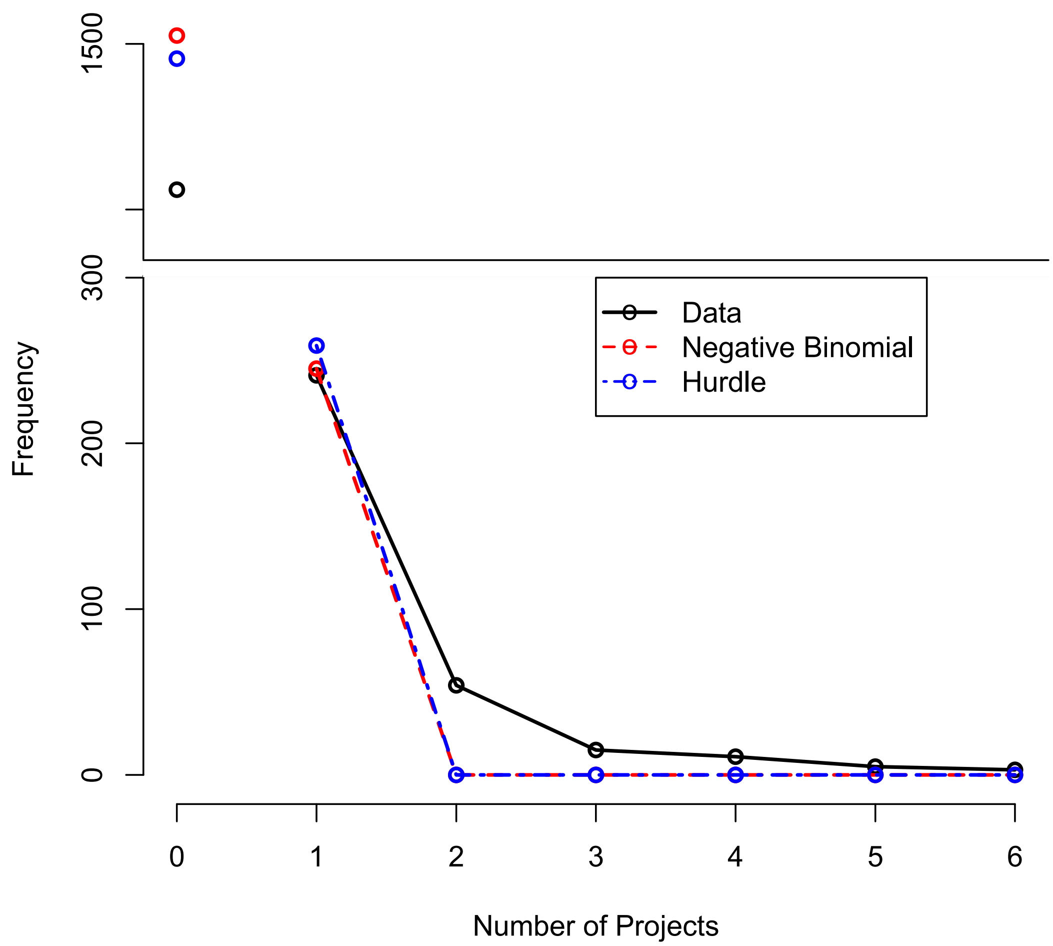

Figure S1. Negative binomial and hurdle model estimates of number of projects compared to data.

This issue is magnified in the analysis of project establishment over time. Although there are 480 mitigation projects in the dataset, there are only 419 watershed-years of 25,980 when a project is established. Only 1.6% of watershed-years have project establishment and, consequently, a null model predicting that projects are never established only has an error rate of 1.6%. In fact, our multi-level model predicts that no watershed-years will have a project established (it is in essence the null model). However, evaluating a model based on error rate fails to explore the variation in propensity, which ranges in our model from 0.004 to 0.203. While our variables of interest help explain this large increase in propensity for project establishment, even the watershed-year with the greatest propensity for a project is unlikely to have one. Gary King, Michael Tomz and Langche Zeng (King & Zeng 2001) have developed the rare events logistic regression to address binary data heavily skewed to zeros, however, to the best of our knowledge no such correction exists for multilevel models.

While the multilevel logistic regression struggles to account for the high prevalence of zeros in the data, there is strong evidence that the multilevel approach is appropriate. Before fully specifying our model, we compared a null single level logistic regression (i.e., intercept only) to a multilevel model that allowed the intercept for each watershed to vary using a likelihood ratio test. Allowing the intercepts to vary greatly improves model fit ($p < 0.0001$). An advantage of the multilevel model is the ability to calculate the total residual variance that is due to between-watershed variation. The variance partition coefficient is calculated:

VPC = level 2 residual variance/ (level 2 residual variance + level 1 residual variance)

In a logistic model, it is assumed that level 1 residual variance is fixed at 3.29 because the binary dependent variable is represented by a latent variable that in this case represents the propensity of a watershed to have a project in a given year. To set the scale of the latent variable the residual variance is fixed (Steele 2009).

Due to the limitations of the multilevel model, comparing the rates of project establishment between each treatment is the most powerful test to determine causality. Figure S2 shows the smoothed probability density function for the rate of project establishment, or in other words, the function that gives the relative likelihood of any given rate for a treatment. In all treatments most watersheds will likely have a very low rate of project establishment, however, the sample after local watershed planning occurs clearly have a different distribution with more watersheds with higher rates of project establishment. While this provides evidence that targeted local watersheds and local watershed plans cause an increase project establishment, it is difficult to parse out the individual effect of these two treatments.

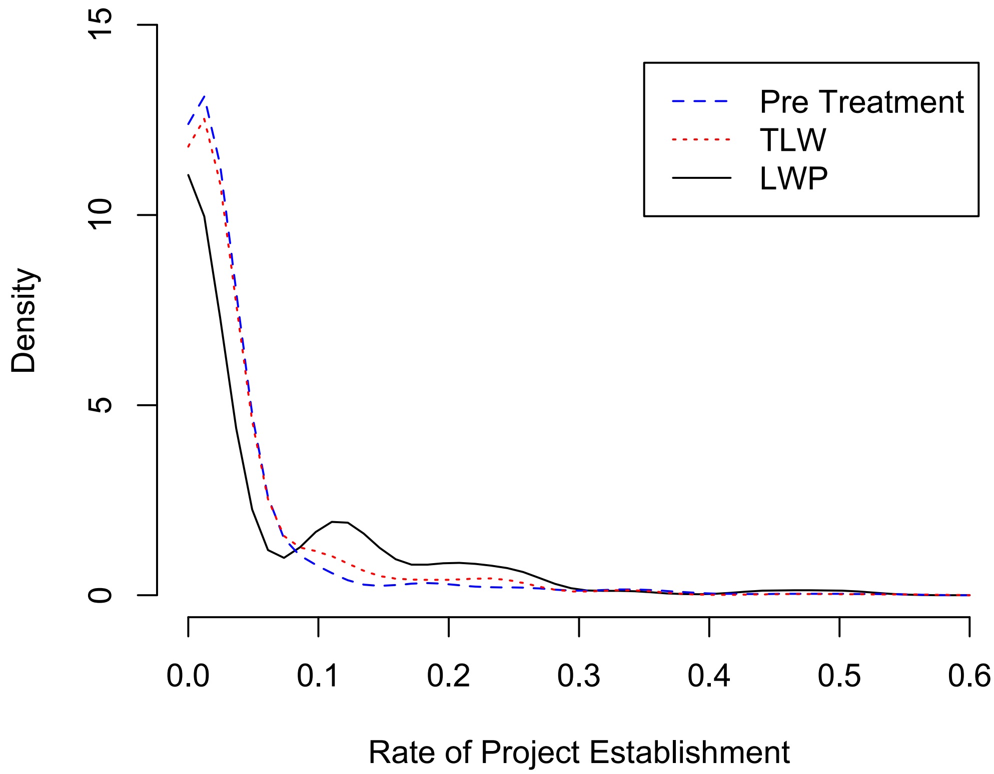

Figure S2. Distribution (relative likelihood) of the rate of project establishment given each treatment.