EXTERNAL MEMORANDUM

TO:	Sadie Drescher, Chesapeake Bay Trust
FROM:	Roxolana Kashuba and Melanie Edwards, Exponent Inc.
DATE:	January 24, 2020
PROJECT:	1901001.000/Restoration Research Award 16925
SUBJECT:	Technical Memorandum #3; USGS Sampling Analysis

The intent of this memorandum is to report and explain progress on the third phase of Exponent's FY2019 Restoration Research Program awarded contract. After data were collected, aggregated, and processed in the first phase, the effect of best management practice (BMP) implementation on annual pollutant loads was statistically evaluated in the second phase. The third phase of this project involved thinning high-frequency U.S. Geological Survey (USGS) data (15-minute measurements) at several sites within the Chesapeake Bay Watershed (described and provided in Deliverable #1) to assess the effect of sampling frequency on the uncertainty of pollutant load estimates.

Applying these analyses tested the second and third hypotheses stated in Exponent's proposal (dated February 20, 2019) and clarified in Exponent's response to reviewer comments (dated March 27, 2019):

2. Stream water sampling at seven-hour frequencies using an automated sampler is sufficient to reduce maximum load estimate error rates to 15%.
3. The uncertainty of pollutant loads estimated at different sampling frequencies significantly differs with watershed size, land use, area of impervious surface, rainfall, and hydrology.

1 Methods - Data Analysis

The methods for analysis of the high-frequency data (watersheds shown in Figure 1) are unchanged from Exponent's proposal to the Chesapeake Bay Trust. USGS did not sample high-frequency measurements of suspended sediment concentration (SSC) and total phosphorus (TP); therefore, predictive models based on turbidity provided estimated values for these pollutants. Using turbidity as a surrogate measurement for SSC and TP is a widely applied method where SSC and TP samples are temporally limited. Coefficients from the most explanatory (highest adjusted R-square) regression model, robust or standard least squares, were used to estimate near-continuous SSC and TP data from the measured near-continuous turbidity data (reported in Table 1 and Table 2. Nitrate was measured *in situ* at near-continuous intervals, therefore measured results were used for analysis without need for surrogate predictions.

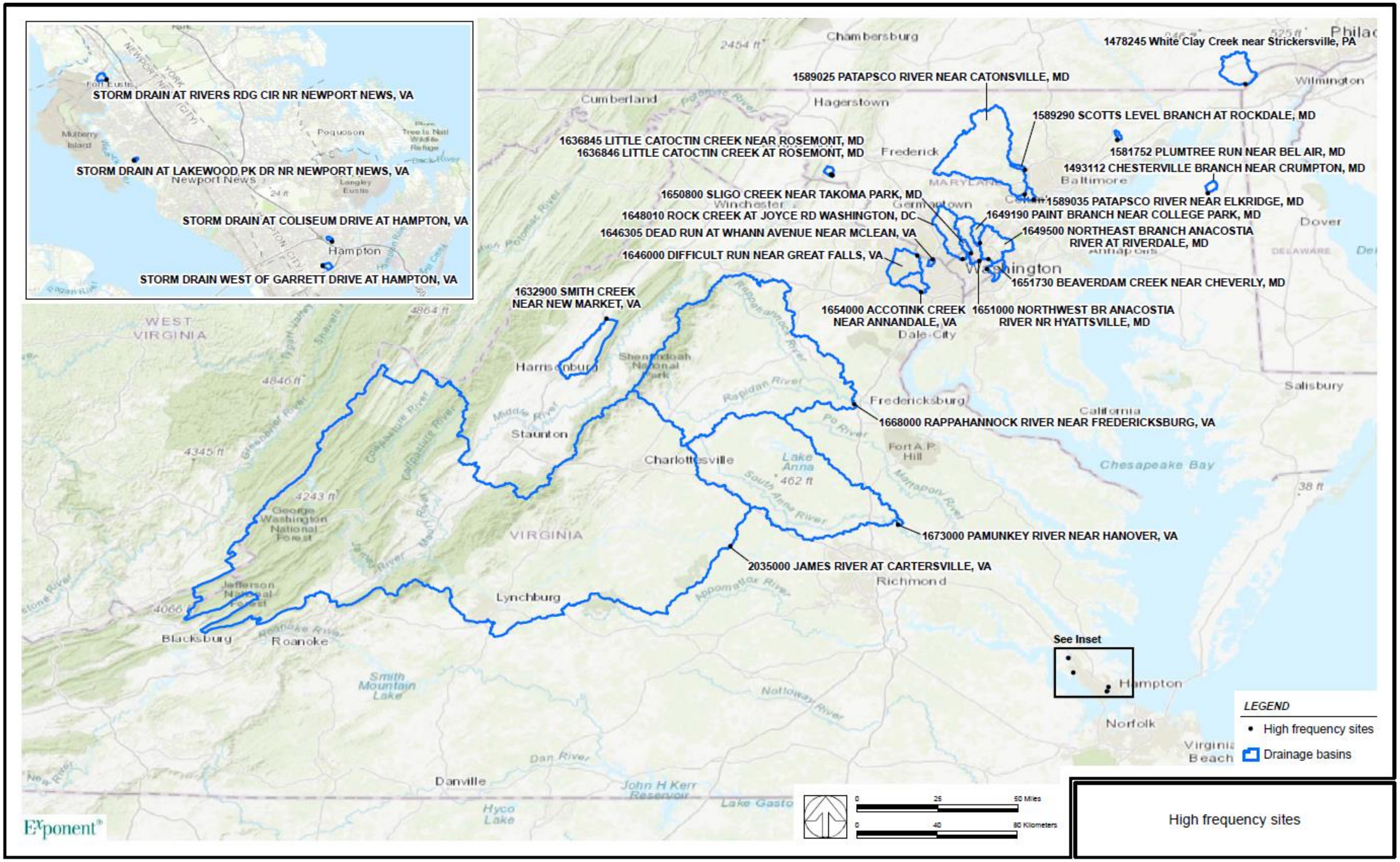

Figure 1. Locations of the watersheds used in Phase III analysis for which USGS sampled high-frequency data

Table 1. Predictive models based on turbidity for suspended sediment concentrations

Site Number	Intercept	Slope	Adjusted R-square	Regression Type
01478245	-221.52	2.992	68%	Standard
01493112	3.92	0.707	100%	Robust
01581752	1.66	0.728	100%	Robust
01589025	53.09	1.307	87%	Robust
01589035	50.81	1.993	89%	Robust
01589290	*Adjusted R-square < 50% for both models*			
01632900	0.17	1.325	100%	Robust
01636845	12.91	0.849	90%	Robust
01636846	21.46	1.017	87%	Robust
01646000	-6.88	2.466	99%	Robust
01646305	-9.87	2.266	93%	Robust
01648010	10.18	1.871	93%	Robust
01649190	27.49	0.825	77%	Robust
01649500	-1.97	1.302	94%	Robust
01650800	1.72	1.408	91%	Robust
01651000	158.70	1.406	65%	Standard
01651730	*Adjusted R-square < 50% for both models*			
01654000	-14.18	2.185	93%	Standard
01668000	2.26	1.181	99%	Robust
01673000	1.66	1.455	95%	Robust
02035000	1.38	1.444	96%	Robust

Table 2. Predictive models based on turbidity for total phosphorus

Site Number	Intercept	Slope	Adjusted R-square	Regression Type
01478245		*Adjusted R-square < 50% for both models*		
01493112	8.759	-0.01648	59%	Robust
01581752		*Adjusted R-square < 50% for both models*		
01632900		*Adjusted R-square < 50% for both models*		
01636845		*Adjusted R-square < 50% for both models*		
01636846		*Adjusted R-square < 50% for both models*		
01646000		*Adjusted R-square < 50% for both models*		
01646305		*Adjusted R-square < 50% for both models*		
01648010	1.217	0.00545	70%	Robust
01649190	1.416	0.00333	60%	Robust
01649500	1.054	0.00557	74%	Robust
01650800	1.732	0.01044	55%	Robust
01654000		*Adjusted R-square < 50% for both models*		
01668000	0.629	0.00533	77.2%	Standard
01673000	0.602	0.00515	73%	Robust
02035000	0.370	0.00573	88%	Standard
0167889257		*Adjusted R-square < 50% for both models*		
0167891721		*Adjusted R-square < 50% for both models*		
0204279245	0.912	0.00810	67%	Robust
0204279294		*Adjusted R-square < 50% for both models*		

Resampling from the near-continuous annual time series for each pollutant resulted in thinned data representing the sampling method frequencies described in Table 3. The resulting data were used to assess the effect of sampling frequency on the uncertainty of load estimates calculated using the load estimation algorithm Equation 1 of Technical Memorandum #1. The R scripts used to conduct this analysis are provided in the folder “R-scripts.”

Table 3. Descriptions of sampling methods assessed

Sampling Method	Number of Permutations	Description
Monthly	1,000	Sample taken every month (7 am–6.30 pm, M–F)
Weekly	1,000	Sample taken every week (7 am–6.30 pm, M–F)
Weekly + Storm	1,000	Sample taken every week (7 am–6.30 pm, M–F) with a daily sample taken when flow >10^{th} percentile.
Seven-Hour	1,000	Sample taken every 7 hours.
Flow-Paced	600	Sample taken when flow exceeds cumulative threshold (threshold set to yield an average 80 sampling pumps per week and aggregated into weekly composite).
MDE MS4 Permittee Requirements	1,000	Samples taken from 12 storms with monthly samples taken during episodes of extended low flow.

MDE – Maryland Department of the Environment; MS4 - NPDES Municipal Separate Storm Sewer System

For each sampling method, stepwise generalized linear regression models were fit to load estimates from the thinned sampling data (outlined above) to identify watershed characteristics that best predicted the variability in load estimate uncertainty (e.g., error). Error was quantified as the absolute difference between the log-transformed median load estimates based on the thinned resampling data compared to the equivalent load estimate based on the full dataset. The explanatory characteristics considered were watershed size, discharge, baseflow index, flashiness index, percentage of watershed area comprising low- and high-intensity developed land, developed open space, impervious surface, and woody wetlands.

2 Results - Data Analysis

2.1 Sampling Analysis

Figure 2, Figure 3, and Figure 4 show how load estimate uncertainty for SSC, TP, and nitrate, respectively, changes between sampling frequency methods, using boxplots compared to a red dashed line representing zero error (i.e., true load estimate). Errors above the red dashed line indicate an over-prediction bias (predicting load estimates greater than the true load estimate) and errors below the red dashed line indicate an under-prediction bias (predicting load estimates less than the true load estimate). Note that for SSC and TP, the variability is larger than for nitrate. This is expected given these pollutant values encompass variability in the predictive model based on turbidity. Furthermore, the variability in nitrate at different stages of flow is low compared to SSC and TP, and it is therefore expected that nitrate load estimates would be less sensitive to sampling frequency method.

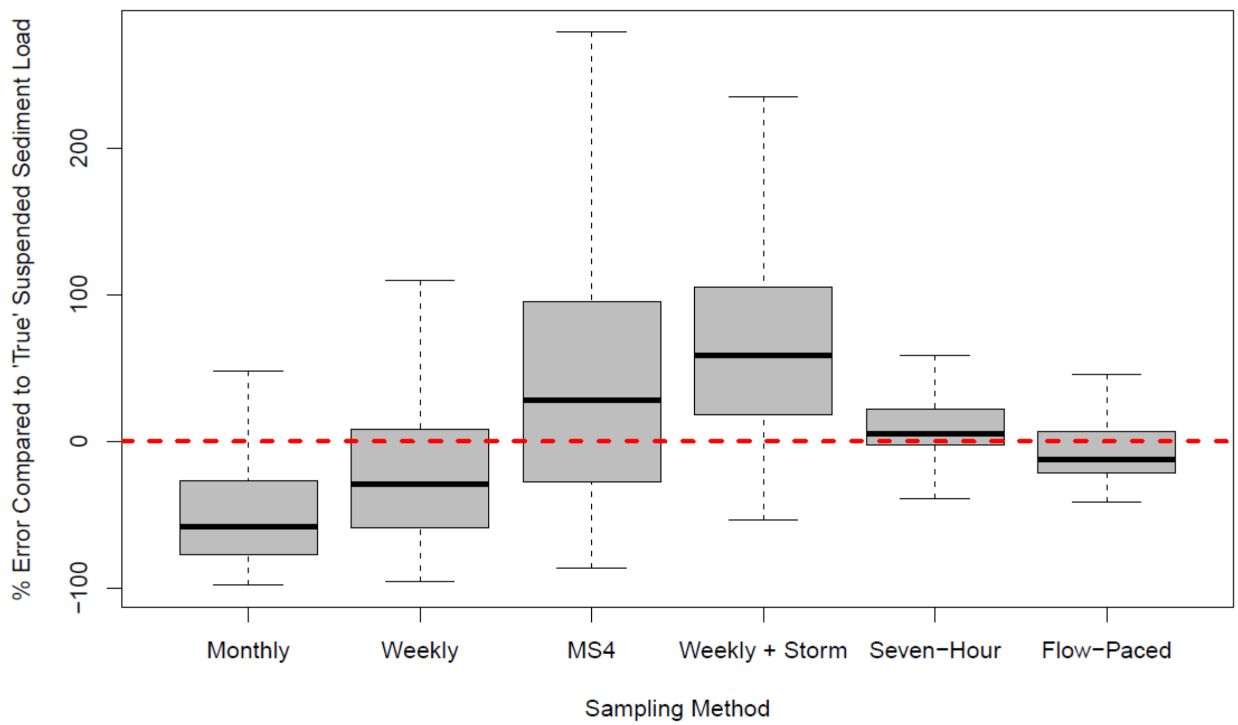

Figure 2. Relative error in SSC estimates based on various sampling frequencies compared with the load calculated from high-frequency data

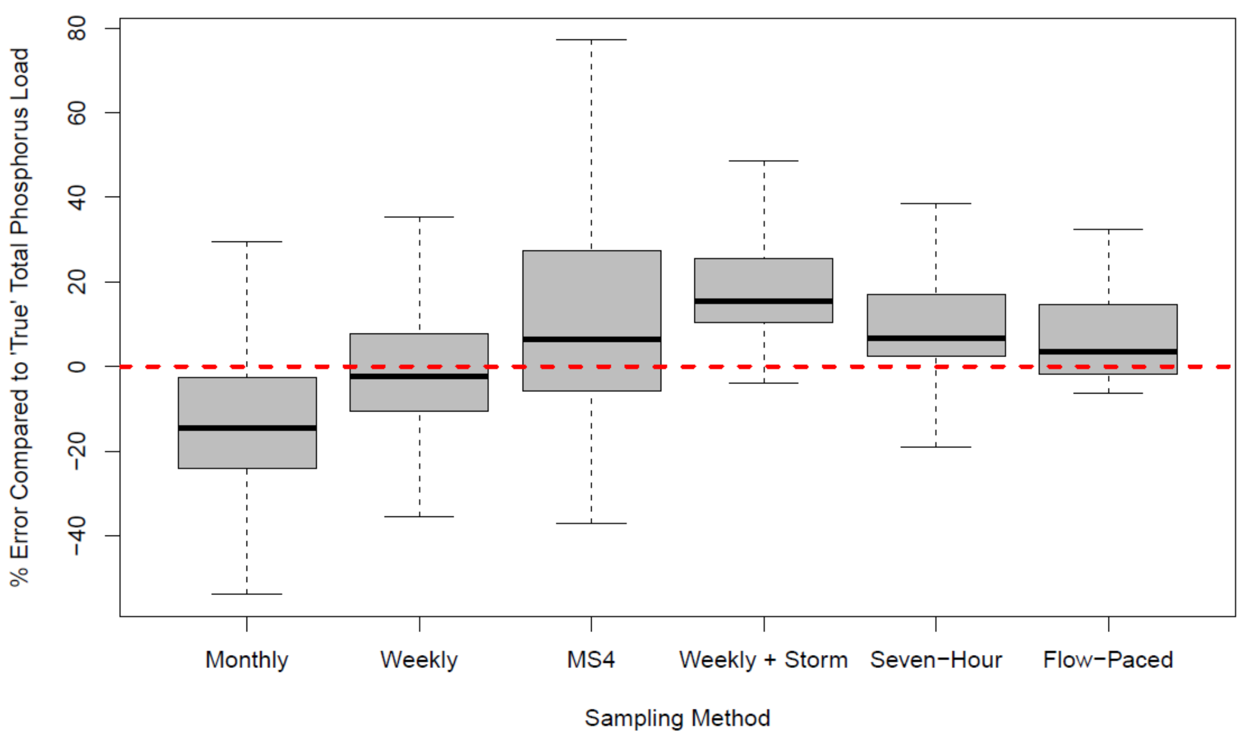

Figure 3. Relative error in TP load estimates based on various sampling frequencies compared with the load calculated from high-frequency data

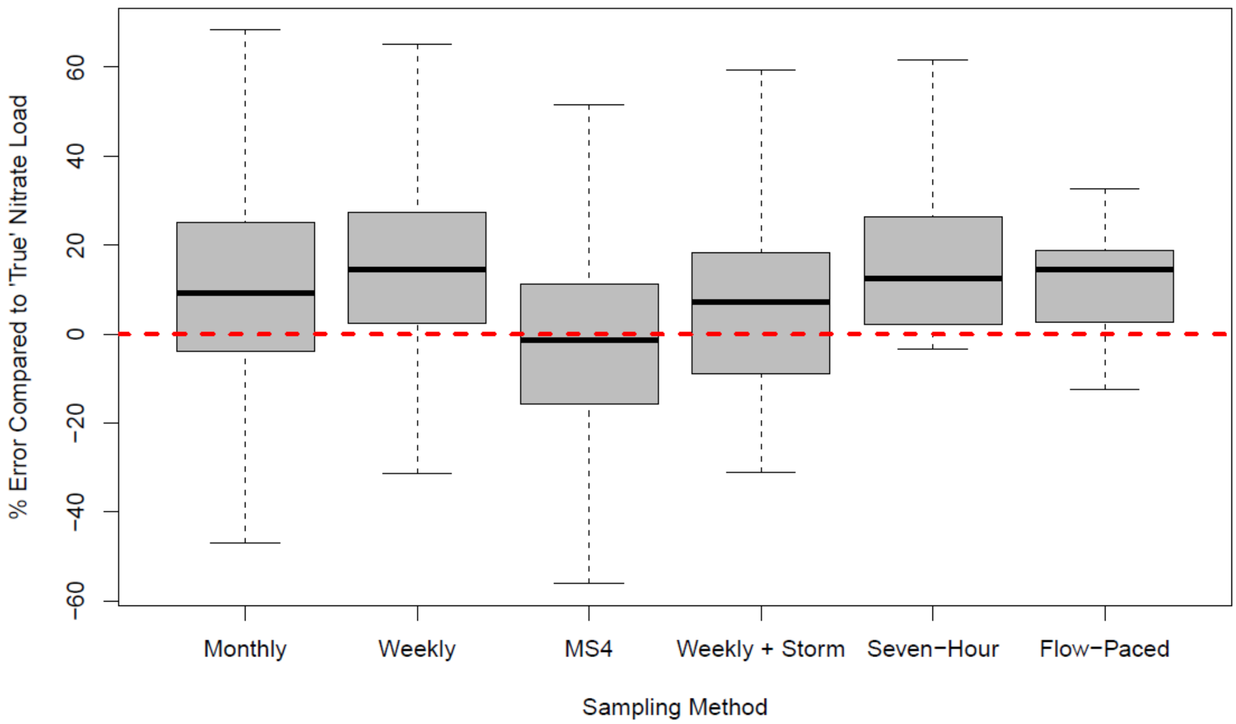

Figure 4. Relative error in nitrate load estimates based on various sampling frequencies compared with the load calculated from high-frequency data

Error magnitude was greater for SSC (-100 to almost 300%) than for TP and nitrate (-60 to 80%). For SSC and TP, MS4 and Weekly + Storm sampling strategies overestimated true load while monthly sampling underestimated load. Weekly sampling underestimated SSC load as well while producing symmetric error for TP load. Seven-hour and flow-paced sampling produced generally accurate SSC load estimates with the smallest relative error spread for SSC (around -40 to 40%) but slightly overestimated TP load with error magnitudes similar in spread to those for Weekly + Storm sampling. Meanwhile, for nitrate, MS4 sampling was the only sampling strategy that accurately estimated nitrate load, albeit with large error (-60 to 50%). All other sampling methods overestimated nitrate load; however, error bars were the smallest for flow-paced sampling (-20 to 30%).

2.2 Watershed Characteristics Associated with Uncertainty in Load Estimates

Table 4, Table 5, and Table 6 show the results of the stepwise generalized linear models that assess the predictive association between various watershed characteristics and the error in median load estimates for SSC, TP, and nitrate, respectively. Note that these tables show the most likely characteristics for the sampling method for each pollutant, but the final fitted model of characteristics may include characteristics that are not statistically significant at the standard 0.05 significance level ($P<0.05$). For each pollutant, the characteristics associated with error in load estimates varies between sampling methods, with no watershed characteristic predictive for all sampling methods. This suggests that different watershed characteristics have different impacts on the performance of each sampling method. This could be because of non-linear effects of watershed characteristics on load estimates at different sampling frequencies.

It is difficult to interpret a pattern between types of watershed characteristics that best predict SSC, TP, and nitrate load error. For example, watershed size is predictive of monthly sampling error for SSC and nitrate but not TP. Baseflow index is predictive of weekly + storm sampling error for SSC and TP but not nitrate. Discharge, % impervious surface, and % developed open space is predictive of error in flow-paced sampling for SSC, nitrate, and TP, respectively. There is not one characteristic consistently influencing the success of a method systematically, but there are characteristics that dictate the success of the method statistically.

Table 4. Results of stepwise generalized linear regression models for suspended sediment concentrations

Model Fit	Monthly		Weekly		Weekly+Storm		Seven-Hour		Flow-Paced		MDE MS4	
AIC	-14.96		29.56		126.39		111.58		98.17		77.91	
Null Deviance, DF	4.141	43	13.573	43	43.224	43	44.487	43	28.152	43	14.011	43
Residual Deviance, DF	1.461	40	3.840	39	39.746	42	27.121	41	20.927	42	13.205	42
Improvement (Dev.diff / df diff)	0.89		2.43		3.48		8.68		7.23		0.81	
Model Terms	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**
Intercept	0.0647	0.5438	-1.0037	0.0036	-1.9465	0.0017	-0.7616	0.0055	7.1003	0.0046	-1.1145	0.0018
Log Watershed Size	-0.0674	0.0001	-0.1655	<0.0001	--		-0.2167	0.0002	--		--	
Log Discharge	--		--		--		--		-3.1229	0.0005	--	
Baseflow Index	--		2.2574	0.0030	3.1928	0.0620	--		--		1.5373	0.1168
Flashiness Index	--		0.0215	0.0283	--		--		--		--	
% Developed Low-Intensity	-0.5250	0.1000	--		--		--		--		--	
% Developed High-Intensity	--		--		--		--		--		--	
% Developed Open Space	--		--		--		--		--		--	
% Impervious Surface	--		--		--		--		--		--	
% Woody Wetlands	-13.366	<0.0001	-18.151	0.0000	--		-16.697	0.0816	--		--	

Table 5. Results of stepwise generalized linear regression models for total phosphorus

Model Fit	Monthly		Weekly		Weekly+Storm		Seven-Hour		Flow-Paced		MDE MS4	
AIC	31.63		66.99		80.47		75.22		89.68		65.74	
Null Deviance, DF	5.779	28	17.296	28	26.285	28	27.944	28	33.400	28	16.455	28
Residual Deviance, DF	3.835	26	13.910	27	20.664	26	15.018	24	30.419	27	11.605	25
Improvement (Dev.diff / df diff)	0.97		3.39		2.81		3.23		2.98		1.62	
Model Terms	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**
Intercept	-3.7921	0.0058	-1.9629	0.0000	-2.2739	0.0002	-5.4862	0.0000	-3.4493	0.0000	-2.0838	0.0002
Log Watershed Size			--		--		-0.1247	0.1308	--		--	
Log Discharge	0.0828	0.0744	--		--		--		--		--	
Baseflow Index	--		--		3.3062	0.0855	6.4425	0.0158	--		2.6755	0.0780
Flashiness Index	--		--		--		0.0452	0.0462	--		--	
% Developed Low-Intensity	--		--		--		--		--		--	
% Developed High-Intensity	--		--		--		--		--		-10.520	0.1413
% Developed Open Space	--		--		--		3.2589	0.0322	2.4975	0.1150	--	
% Impervious Surface	--		--		--		--		--		--	
% Woody Wetlands	-15.777	0.0015	-20.021	0.0162	-25.068	0.0216	--		--		-21.319	0.0130

Table 6. Results of stepwise generalized linear regression models for nitrate

Model Fit	Monthly		Weekly		Weekly+Storm		Seven-Hour		Flow-Paced		MDE MS4	
AIC	47.98		64.56		64.01		84.34		65.64		56.15	
Null Deviance, DF	13.164	22	25.327	22	22.134	22	44.283	22	23.874	22	15.661	22
Residual Deviance, DF	7.658	20	14.435	19	14.097	19	44.383	22	15.133	19	10.927	20
Improvement (Dev.diff / df diff)	2.75		3.63		2.68		0.00		2.91		2.37	
Model Terms	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**	**Coef.**	**P-value**
Intercept	-1.7773	0.0004	-0.6730	0.3963	-0.9068	0.0560	-2.4596	<0.0001	-0.9845	0.2421	-0.1448	0.7940
Log Watershed Size	-0.1116	0.0948	-0.2747	0.0142	--		--		-0.1953	0.0869	-0.2358	0.0082
Log Discharge	--		--		--		--		--		--	
Baseflow Index	--		--		--		--		--		--	
Flashiness Index	--		--		-0.0596	0.0143	--		--		--	
% Developed Low-Intensity	--		--		--		--		--		--	
% Developed High-Intensity	--		--		27.331	0.1007	--		--		--	
% Developed Open Space	2.5429	0.0219	2.1587	0.1521	--		--		--		--	
% Impervious Surface	--		--		--		--		2.8452	0.2412	--	
% Woody Wetlands	--		-15.520	0.1640	-21.025	0.0374	--		-26.600	0.0280	-14.702	0.1202

3 Discussion and Conclusions

In general, flow-paced sampling appears to produce accurate load estimates with lower error rates, while monthly sampling overestimates nitrate loads and underestimates SSC and TP loads. Weekly sampling is slightly more accurate. The MS4 sampling method produces inaccurate load estimates with very large error estimates, which may explain why we did not observe significant relationships between BMP implementation and annual loads in Phase II of this project. This analysis has confirmed that more frequent sampling is required to accurately predict annual loads with less error. This has important implications for assessing the cumulative effects of BMP implementation in watersheds draining to the Chesapeake Bay.

To answer Hypothesis 2, seven-hour sampling produces load estimates with up to 60% error, so the suggestion that error is at maximum 15% is refuted. To answer Hypothesis 3, the uncertainty of load estimates does differ significantly with watershed characteristics and therefore can be predicted by watershed size, land use, area of impervious surface, rainfall, and hydrology.

4 Approach for Next Deliverable

The road map for developing an interactive tool to help stakeholders explore what load estimation error rates can be expected from different sampling frequencies and watershed characteristics is unchanged from Exponent's proposal. This tool will be developed in R Shiny by implementing regression analysis results presented in this memo, submitted for CBT feedback and suggestions, and revised as a final project product.

5 Summary

This Phase III analysis found relationships between sampling frequency and precision and accuracy of SSC, TP, and nitrate load estimates. In general, more frequent sampling decreases load estimation error. This error can be predicted from watershed characteristics, and these relationships may guide watershed managers in selecting the correct sampling methods to reduce bias and error that sampling methods can introduce. The reduction of uncertainty in annual estimates of pollutant loads will allow a more robust analysis of the effect of BMP implementation on pollutant export. Ultimately, this will help jurisdictions make better science-based decisions on which BMPs to implement in a given watershed to provide the greatest water quality improvement.

After quantifying the impacts of sampling frequency on annual load estimates using high-frequency USGS data, and how watershed characteristics influence these estimates, the remainder of the project focuses on developing an interactive interface to relate sampling uncertainty to watershed characteristics. This tool will recommend sampling frequencies that will be able to statistically detect expected reductions in load. This will assist the Chesapeake Bay Trust and other stakeholders in comparing the costs and benefits of different sampling methods in terms of the accuracy of their ability to calculate true annual loads.