Investigating uncertainty of phosphorus loading estimation in the Charles River Watershed, eastern Massachusetts

Authors: Alana Burton Spaetzel,

Boston College Electronic Thesis or Dissertation, 2018

Investigating uncertainty of phosphorus loading estimation in the Charles River Watershed, eastern Massachusetts

Alana Burton Spaetzel

A thesis

submitted to the Faculty of

the department of Earth & Environmental Sciences

in partial fulfillment

of the requirements for the degree of

Master of Science

Boston College
Morrissey College of Arts and Sciences
Graduate School

December 2018

Investigating uncertainty of phosphorus loading estimation in the Charles River Watershed, eastern Massachusetts

Alana Burton Spaetzel

Advisor: Noah P. Snyder, Ph.D.

ABSTRACT

Estimating annual phosphorus (P) loading in impaired fresh water bodies is necessary to identify and prioritize management activities. A variety of monitoring programs and water quality models have been developed to estimate P loading in impaired watersheds. However, uncertainty associated with annual riverine P loads tends to receive less attention. This study addresses this gap by exploring the range in annual total phosphorus (TP) loads from two common load estimation methods using data collected in the Charles River watershed (CRW) in eastern Massachusetts. The CRW has two P Total Maximum Daily Load (TMDL) reports due to impairments with respect to excessive summer algal growth. Three estimation methods are used in this thesis to quantify annual TP loads (L_Y): the concentration-discharge relationship (CQ), the land use coefficient (LUC) method, and the average concentration, continuous discharge (ACQ) method. L_Y derived using the LUC method spanned an average relative percent range of 214% at each site, whereas L_Y results from the concentration-discharge method spanned an average relative percent range of 56%. While results of the CQ method produced a narrower range of L_Y, the CQ relationship is subject to seasonal dependencies and inconsistency through time. Seasonal terms in the LOADEST program, a publicly available and commonly used statistics tool, do allow the model estimates to capture trends through time, an advantage over the LUC method. Results of an interlaboratory

comparison of P concentrations demonstrate the potentially large role of analytical uncertainty in L_Y estimation. Significant discrepancies between the results of each method for a single location and time suggest that loading estimates and consequently management priorities may be dependent on the estimation technique employed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Addressing impaired water bodies is a priority for federal and state environmental agencies. Section 303(d) of the Clean Water Act (CWA) mandates that states identify impaired and threatened water bodies and develop Total Maximum Daily Loads (TMDL) for the pollutants of concern. Furthermore, assessing the response of water bodies to increasing stressors like urbanization and population growth is important to protect ecosystem services (Schindler, 2001; Postel, 1998). Defining current and/or historical nutrient loadings to a water body is critical for developing reduction targets within the associated watershed (Vollenweider, 1968). Nutrient loading, particularly that of phosphorus (P), is important in shaping the trophic status and ecosystem health of fresh water bodies where P is most commonly limited (Vollenweider, 1968). P reductions are typically prioritized in watersheds with eutrophic and mesotrophic water bodies (MassDEP, 2007).

There are several approaches to quantifying total phosphorus (TP) loading. Loads are estimated by directly monitoring discharge and TP concentrations for the period of interest. Discharge (Q) datasets are usually provided at 5 to 15 minute intervals and are developed by a rating curve, which defines the relationship between stage height and instantaneous discharge. Due to analytical requirements for TP concentration, it cannot be monitored in-situ. Continuous TP concentration (C) data may be developed by using a surrogate parameter such as turbidity (Jones et al., 2011; Schenk et al., 2016). TP *C* data may also be collected from automated flow proportional composite samples; however, this kind of sampling cannot be deployed continuously like an in-situ sensor. The most common method of TP data collection is discrete sampling, which can be paired with instantaneous discharge data to calculate instantaneous loads. Due to the analytical constraints of monitoring TP directly, a model is usually employed to estimate loading at the monthly, seasonal, or annual time step.

Land used based loading coefficients (LUCs) and the concentration-discharge (CQ) load estimation technique are two methods that rely on a deterministic relationship between P concentration or load with more readily accessible watershed data (e.g., land cover mapping or discharge; Chen et al., 2015; Janke et al., 2013; Duan et al., 2012; Aulenbach, 2006). These methods may be less data intensive than surface water quality models like Environmental Fluid Dynamics Code (EFDC) (Hamrick, 1992) and Enhanced Stream Water Quality Model with Uncertainty Analysis (QUAL2EU) (USEPA, 1995) which can simulate multiple water quality parameters and eutrophication dynamics as well as predict future changes in water quality (Wang et al., 2013). The LUC and CQ methods are simpler and more readily available; however, the magnitude and nature of uncertainty about the annual P load estimates (L_Y) yielded from these methods are not always appreciated (MassDEP, 2007; CRWA, 2011). To address this, this study explores the range of literature reported LUC values and the full range L_Y outcomes associated with them, as well as the uncertainty about the TP CQ relationship as modeled in Load Estimator (LOADEST) (Runkel et al., 2004). These methods are applied in the Charles River watershed (CRW) in eastern Massachusetts where active nutrient and discharge monitoring occurs and two P TMDLs were developed (MassDEP, 2007; CRWA, 2011). The CRW spans a steep gradient of land uses from the less developed headwaters in Hopkinton, Massachusetts to the more densely populated city and suburbs of Boston.

It is well known that land use influences P loading because impervious cover increases runoff volume and collects pollutants from various sources; urban drainage systems provide conduits for contaminant transport, and anthropogenic activities can increase P availability (e.g., application of fertilizer and P bearing cleaning agents) (e.g., Bannerman et al., 1979; Winter and Duthie, 2000; Goonetilleke et al., 2005; Duan et al., 2012; Valtanen et al., 2014). LUCs, reported in units of kg P $ha^{-1}year^{-1}$, are used to apportion loading to subcatchments based on land cover (e.g., MassDEP, 2007; CRWA, 2011; Liu, 2012b, Johnes, 1996). LUCs are developed by monitoring single land use watersheds for discharge and TP concentration. Table 1 reports studies that have derived LUC values based on monitoring data, and Figure 1 shows the locations of those studies for which the original document could be located. In some cases, LUCs are used to approximate reduction scenarios for watershed users, such as towns and municipalities (MassDEP, 2007). However, LUCs have been applied in environments with different climates, geologies, and management practices (CRWA, 2011) resulting in potentially high uncertainties on L_Y estimates (Liu et al., 2012a, 2012b). Environmental non-governmental organizations and/or government environmental protection agencies developing load estimations based on LUCs have to choose among the available coefficients or some average of these coefficients; this method may be susceptible to uncertainty based on the coefficient selection. P LUC values can range considerably even within a single climate and geologic setting (Table 1), making it unclear which value is appropriate for a given watershed. This suggests an assessment of the variability in land use dependent P LUC values and the resulting annual P load variability is necessary.

The relationship between discrete TP concentration (*C*) and daily river discharge (*Q*) can be used to estimate TP concentration dataset at the daily time step. The CQ method is ideal where continuous *Q* gaging stations are already maintained because with daily pairs of TP *C* and *Q*, daily loads are calculated and summed to estimate L_Y.

This study builds on existing work in the CRW in eastern Massachusetts (Figure 2). Water quality assessments within the past decade have indicated significant impairments due to P loading (MassDEP, 2007; CRWA, 2011). These assessments were intended to evaluate the Charles River TMDL. Reduction scenarios for towns within the CRW are based on LUC values adjusted to match loads derived from field measurements. The adjusted LUC are within ~1% of the initial literature-based values in the Lower CRW, which suggests LUC derived from watersheds in other regions are appropriate for use within CRW (MassDEP, 2007). A similar approach was used to adjust the LUC for application in the middle and upper watersheds. Adjustments made in the TMDL study in middle and upper watershed were higher than those made for the lower watershed, resulting, in some cases, in coefficient values that are nearly an order of magnitude higher than those used in the lower watershed; however, values in the middle and upper watersheds were also intended to capture the effects of in-river sinks of P (CRWA, 2011). A limitation of the approach of adjusting coefficients to force model agreement with field derived P loads is the possibility of non-unique solutions. That is, there are multiple potential combinations of LUC that could produce the desired L_Y, so it is unclear as to whether these adjusted coefficients are representative of the actual P loading rates. In this study, to examine the uncertainties associated with the LUC and CQ methods of annual L_Y, P loading was estimated, and uncertainty constrained, for three sites along the Charles River. Estimates from the two approaches were compared with each other as well as with corresponding independent estimates. Furthermore, TP analytical uncertainties are briefly addressed by interlaboratory comparison.

A two-year monitoring program was undertaken as part of this study, with sample collection and analysis for P concentration at four mainstem sites and three subwatersheds in the CRW. Due to data quality concerns, which are discussed in more detail in Appendix A, only data from one of these sites is included in the study.

METHODS

Site Descriptions

This study focuses on estimates of L_Y at three sites along the Charles River in eastern Massachusetts – at Watertown Dam (WD), at Waltham (WL), and at Dover (DV), (Figure 2; Table 2). Boston has a warm-summer humid continental climate with average annual precipitation of 111 cm and average annual temperature minimum and maximum are 6.7 ˚C and 14.9˚ C respectively (1971-2010, National Climatic Data Center). The dam at WD separates the middle and upper parts of the CRW from the heavily urbanized lower part. WL and DV are located 4 and 37 river kilometers upstream of WD, respectively, and are locations of active U.S. Geological Survey (USGS) gaging stations. The percentage of forested land cover decreases from 51% in the DV watershed to 43% in the WD watershed, reflecting increasing density of urban environment with proximity to Boston (Figure 2; Table 2). These three sites are locations of historic and contemporary monitoring of Q and TP concentrations undertaken by the Charles River Watershed Association (CRWA), the Massachusetts Water Resources Authority (MWRA), and the USGS (Table 3).

L_Y were estimated for each site using LUC and CQ methods, and Monte Carlo simulations were used to constrain uncertainty. Further, multiple estimates of L_Y were generated using CQ methods with different calibration data sets from distinct time intervals (at DV) and distinct data sources (for WL) as described below. Table 3 indicates the analyses associated with each site, period, and data collection entity.

Sampling

TP and orthophosphate (PO_4^{3-}; hereafter, PO_4) concentrations were monitored weekly from August – December 2016 and monthly from January – May 2017 at WL. In addition to routine samples, water was collected following three rainfall events (event samples, n=20). Samples for PO_4 analysis were filtered with 0.45 μm syringe filters into 125 mL polypropylene bottles. Whole water samples for TP analysis were collected in pre-acidified (H_2SO_4) 250-mL polypropylene bottles, delivered to the National Environmental Laboratory Accreditation Program (NELAP) certified Alpha Analytical Laboratory in Westborough, MA on the day of collection. Samples were analyzed within one week using a SEAL AQ2 Discrete Analyzer according to U.S. Environmental Protection Agency (USEPA) method 119-A, with method reporting limits of 0.01 mg/L and 0.005 mg/L for TP analysis and PO_4, respectively. Field TP replicates (n=7) differed on average by 4.5%. Laboratory TP replicates (n=29) differed on average by 6.5%. An inter-laboratory comparison was conducted between a) Alpha Analytical Laboratory and the MWRA Deer Island Central Laboratory, which have different analytical methods, in March 2017 and b) Alpha Analytical Laboratory and the Nebraska Water Sciences Laboratory in July 2017, which had the same analytical methods (Table B1).

P Loading estimates from Land Use Based P Loading Coefficients

A survey was undertaken to compile existing LUC values estimated from field-based studies in published peer-reviewed and grey literature. Tracking the lineage of the coefficients reported in summary tables in the lower Charles River TMDL report comprised the first mode of literature review, which ultimately led to a document published by the USEPA that compiled nutrient export coefficients (Reckhow et al., 1980). The majority of LUC values reported here are also found in Reckhow et al. (1980). The second approach comprised a general search in scientific databases using the terms phosphorus, loading, and coefficients. Original sources of all of the LUC values were sought after to survey the range of study locations, climate, and project interval; however, not all original sources were found (Table 1). The most common land uses studied are commercial, industrial, low-, med-, and high- density residential (LDR, MDR, HDR, respectively), and forest. The distribution of each of these six land uses within the CRW was delineated in ArcGIS by overlaying the MassGIS Land Use layer with the CRW boundary delineated from USGS StreamStats. The number of land use categories in the GIS data set far exceeded the six land uses targeted in this study, so land uses were binned into six categories (Table 4).

The LUC method relies on the following equation,

$$L_Y = \sum_{i=1}^{n} [E_i \cdot A_i] \text{ (Equation 1)}$$

where L_Y is annual TP load in units of kg/year, E_i is the TP loading coefficient for land use category *i* in units of kg ha^{-1} yr^{-1}, A is the area of land use *i,* and *n* is the total number of land use categories. A Monte Carlo simulation was used to quantify the entire range of possible annual P load outcomes for the DV, WL, and WD watersheds using Equation 1 and the literature reported range of coefficients in each land use category. For each iteration (n=1,000) of the Monte Carlo simulation, the six coefficient values were randomly selected from the log normal distribution of literature reported values for the specified land use, yielding 1,000 values of L_Y for each site.

P Loading estimates from Concentration-Discharge relationships

A range of average L_Y estimates for the DV, WL, and WD sites were estimated using CQ regression models. The form of each regression model was determined using LOADEST, a publicly available tool that uses Adjusted Likelihood Maximum Estimation (AMLE) to estimate constituent loads based on a calibration data set of measured Q and C pairs (Runkel et al., 2004). For this study, model selection was automated, which means 11 possible regression forms were tested in the LOADEST program, and parameterization was optimized based on the lowest Akaike Information Criterion (AIC; Judge et al., 1988). Under these conditions, all analyses resulted in one of the following regression forms:

$$LnC = a_0 + a_1 LnQ + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4\, dtime$$

(Equation 2)

$$LnC = a_0 + a_1 LnQ + a_2 LnQ^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5\, dtime + a_6 dtime^2$$

(Equation 3)

where a_1 through a_6 are regression coefficients, and ln$Q$ is the natural log of daily discharge minus the median of the calibration discharge data. Monte Carlo simulations were used to determine 1,000 estimates of average L_Y at each site based on the random selection of CQ model regression coefficients within the uncertainty bounds reported by LOADEST (i.e., mean ± 2·stdev). The period for which the L_Y were determined was dependent on the data availability at each site (Table 3). Average L_Y were estimated for the DV site for two separate time intervals due to significant differences (p value < 0.05) in TP concentration values between 1970-1984 and 1986-1994 (Figure B1). CQ regressions were developed at WL separately for C data collected in this study for water year (WY) 2017 and C data collected by CRWA from 1996-2016. The WD load estimates from this study were compared with the average L_Y determined for this site over the 1998-2002 interval as reported in the TMDL study (MassDEP, 2007). Table 3 reports the data interval, reporting entity, and relevant analyses for each of the three sites.

For each CQ L_Y estimate, an average annual load for the corresponding period was calculated by multiplying each daily mean Q by the average C value of the whole period. Because the LOADEST models develop a relationship between ln$C$ and ln$Q$, the average C was computed as the average of all ln$C$. This method is referred to as average concentration discharge or ACQ. ACQ has been used to estimate TP loads in routinely monitored sub-basins in Cambridge, MA (Smith, 2017), but is not as common as CQ and LUC. Here, the method is primarily used to understand the CQ relationships at each site.

The L_Y reported in the 2007 Charles River TMDL (MassDEP, 2007) report was also used to test whether multiple unique combinations of LUC values would result in approximately the same L_Y estimate. From the 1,000 results produced by the LUC method Monte Carlo Simulation for WD, L_Y values that fell within 5% of the MassDEP, 2007 load (L_Y = 28,925 kg/yr on average between 1998-2002) were selected to identify the number of LUC combinations that yielded a similar result.

RESULTS

LUCs reported in the literature vary in magnitude but are reported from a relatively limited geographic area (Figures 1 and 3). Of the 66 forested LUC values, 24 are from different watersheds within the same study conducted in southern Ontario, Canada over a 20-month period (Dillon and Kirchner, 1979). The other land use categories are characterized by less than 10 original studies, which were largely published before 1990 (Table 1). Field derived loading coefficients are limited in both space and time (Table 1 and Figure 1). Despite these limits, the available LUC coefficients show that land use affects P loading, with forested land uses consistently having lower P yields than developed land.

The CQ relationship was examined independently of LOADEST; each CQ dataset was plotted and both a linear or power-law least square regression was performed; however, the optimal fit for each dataset is presented here (either the linear or power-law) (Figure 4). WL and WD were fitted with linear regressions which all yielded R^2 values less than 0.01 and p-values greater than 0.05 (Figure 4). Both periods of DV data were fitted with power law regressions which yielded p-values less than 0.01 and R^2 values for DV 1970-1984 and DV 1985-1994 were 0.527 and 0.205, respectively.

Over three orders of magnitude in *Q*, *C* values at each site ranged from 0.03 mg/L to 0.65 mg/L, <0.01 mg/L to 0.45 mg/L, <0.01 mg/L to 0.12 mg/L, <0.01 mg/L to 0.04 mg/L, and 0.03 mg/L to 0.16 mg/L at DV 1970-1984, DV 1985-1994 (n=38), WL (CRWA data, n=79),WL (this study data n=39), and WD (n=247), respectively.

The CQ calibration data for WD are also broken out into seasonal plots (Figure 5); the dataset size at WD (n=247) made it more suitable for this analysis than the other sites. The LOADEST automated regression optimization selected Equation 3 for the DV 1970-1984 CQ data. Equation 2 was the optimized regression form for the remaining four CQ calibration datasets. The observed *C* and LOADEST estimated *C* plotted over time show that the model generally captures the general pattern of *C* changes, but they do not capture the magnitude of these changes (Figures 6 – 10). This behavior is strongly demonstrated in WD data (Figure 10) and the corresponding Nash-Sutcliffe Efficiency Index value (NSE) of 0.439 (Table B5), which is the lowest NSE value of the LOADEST regressions performed in this study. None of the LOADEST regressions resulted in an NSE less than 0 which suggests that in all cases, the data variance was greater than the residuals.

For each study site, the distribution of average annual P loads estimated from the LUC and CQ methods overlap (Figure 11). The MassDEP (2007) TMDL L_Y estimate, which includes all sources in the middle and upper CRW is 21% lower than the median of the LUC method but <1% different than the median CQ L_Y (Figure 11C). The Mass DEP (2007) L_Y and CQ L_Y in this study are based on the same monitoring data. In all cases, the LUC method yielded a wider range of outcomes (Table 6). Figure 12 shows the 97 viable LUCs within each land use category that yielded L_Y within 5% of the MassDEP (2007) value ; these are overlaid with the log normally distributed literature reported LUC.

The annual P loads estimated for WL based on monitoring conducted for this study are non-normally distributed and have a minimum value 75% lower than the minimum L_Y estimated using CRWA data from 1996-2016 (Figure 11B). Average daily Q in 2016-2017 were significantly drier than the period of record (1996-2016, two-tailed t test, p=<0.001; Figure B2A). However, according to the concentration data collected by the CRWA, TP concentrations do not differ between the 1996-2015 (n=78) and 2016 (n=4) periods. Concentration data collected by CRWA between 1996 and 2017 are significantly higher than those reported in this study (two-tailed t tests, p=<0.001 for both) (Figure B2B).

The relative percent differences (RPD) in the interlaboratory comparisons were ≥39% for TP analyses but differed by ≤5% for orthophosphate analyses (Table B1). The differences in TP concentrations between this study and other reporting agencies for the CRW raises concerns about the reproducibility of TP analyses. Of the three labs compared, two use the same analytical methods (Table B1). One of the three labs is NELAP certified (Table B1).

In the case of the DV site analyses, the LUC method generally resulted in greater L_Y estimates for the 1985-1994 period and lower estimates for the 1970-1984 period as compared with the results of the CQ method (Figure 11A). Both the hydrologic conditions and TP concentrations were significantly different between the two periods (two-tailed t test, p = 0.01 and p<0.001) (Figure B3).

DISCUSSION

Concentration-Discharge Relationships

The DV CQ relationship was plotted for 1970-1984 as a demonstration of how the CQ method would be employed in the subsequent analyses. The form of the CQ relationship suggested that higher flows had lower *C*, so a distinguishable CQ relationship was expected at the other sites (Figure 4A). However due to the increased urban drainage area in the WL and WD basins, a different form was anticipated, in part due to higher availability of particulate P during storm events (Song et al., 2017). The form of the TP CQ relationships in the urban watersheds observed had wide variability in the *C* values (0 – 0.5 mg/L) over a limited range of daily flows (Duan et al., 2012). The CQ relationships (Figure 4) at WL and WD demonstrate poor correlation in that there is limited *C* variability over all *Q;* the p-values of the least squares linear regressions for these sites are greater than 0.05 (WL CRWA, WL this study, WD p-values = 0.452, 0.909, 0.619). Both DV periods show significant correlation between *C* and *Q;* power law regressions had p-values less than 0.01 (DV 1970-1985, $y=2.3x^{-0.475}$, n=100, $R^2=0.527$, $p < 0.01$ and DV 1985-1994, $y=0.547x^{-0.403}$, n =38, $R^2=0.205$, $p<0.01$).

Seasonal variations in the CQ relationship are important in determining the overall CQ regression form (Hirsch, 2014). While LOADEST regression forms have seasonal terms (sine and cosine terms in Equations 2 and 3), these terms allow the intercept to vary but keep the regression form the same (Hirsch, 2014). WD seasonal CQ relationships indicate that the seasonal differences are relevant (Figures 5 and 10). Positive correlation between *C* and *Q* in the summer months suggests that high flow events have high TP concentrations (Figure 5). Winter shows a negative correlation between *C* and *Q*, suggesting dilution of TP with higher winter *Q*. The seasonal CQ break down at WD serves as an example of what can complicate the regression fit and why the overall CQ relationship may not always be indicative of the TP dynamics in a system. These seasonal differences should be considered when selecting a load model as they will contribute to uncertainty in the load estimation. The CQ method is best applied in settings and for periods where P sources and dynamics are characterized

DV data are treated in two different periods because of differences in the CQ relationship over time. The differences between *C* and *Q* values during the 1970-1984 and 1985-1994 periods are statistically significant (Figure B1). The transition to lower concentrations and L_Y may be owed to the Clean Water Act and mandated secondary treatment wastewater treatment facilities (WWTF). The DV relationships are also distinct from the other sites and periods because they show that a) *C* and *Q* are correlated and b) *C* generally decreases with higher *Q* (Figure 4A). This correlation suggests that TP loading is dominated by baseflow and not runoff.

Identifying the nature of the CQ relationship is critical to understanding the applicability of load estimation techniques. Another load estimation method called Weighted Regression on Time, Discharge, and Season (WRTDS) may perform better when the CQ relationship is dominated by seasonality (Hirsch, 2014), but this method was not explored here. CQ method is best applied in settings and for periods where a) sufficient data are available to characterize the CQ relationship and b) and P sources and dynamics are characterized

Model Evaluation

Here, the performance of the optimized LOADEST regression models are discussed. While not an exhaustive evaluation of LOADEST performance, these measures are intended to explore the efficacy of the regressions with respect to the observed, raw CQ relationships (Figure 4).

Figures 6-10 demonstrate the regression model performance by comparing observed concentrations to estimated concentrations. Generally, the model captures the *C* trends or patterns through time but not the magnitudes or the extremes. This is most apparent at WD (Figure 10); the estimated and observed *C* have the poorest fit to the 1:1 line, which is reflected in the low NSE value, 0.439. An NSE = 1 is a perfect fit of modeled to observed, E = 0 means the model estimates are equally as appropriate as the observed mean load, and E < 0 indicates the observed mean would be a better estimate of L_Y than the modeled estimate. Even though *C* and *Q* are not correlated at WL and WD (Figure 4), the multi-variate regression forms selected by LOADEST (Equations 1 and 2) are able to capture the general *C* trends because time and seasonality are explanatory variables in the regression equations (Figures 9 and 10). LOADEST regressions performed similarly in studies in the Chesapeake Bay which demonstrate poor CQ correlation but acceptable model performance (Cohn et al., 1992; Duan et al., 2012) (Figure 4 and Figures 6 – 10).

The ACQ method was used as another method to characterize the CQ relationship and evaluate the LOADEST models. This simple method used a single TP concentration, the average for the period of interest, to estimate load instead of estimating TP concentration on non-sampled days. The ACQ and CQ methods for WL 1996-2016 and WD yielded similar results (Figure 11B and C) because these datasets show the least variation in C across all Q values. For WD, the median CQ load estimate and the ACQ estimate are indistinguishable (Figure 11; Table 5). The WL CQ median and ACQ value are similar but the ACQ method is lower by about 9%. In these cases, the LOADEST regression does not seem to provide any advantages over the simplified ACQ method, which is consistent with the poor correlation of lnQ and C.

The DV 1970-1984 ACQ estimate is 43% higher than the median CQ L_Y (Figure 11A). This is also the CQ relationship with the strongest correlation and greatest variation in C (Figure 4A). In this case the LOADEST regression is useful to capture the variance. However, the residuals are still high because the model does not estimate the extremes well (Figure 7).

Interlaboratory Comparison

The TP data collected in this study and analyzed at Alpha Analytical Laboratory (range: <0.01 to 0.043 mg/L, average = 0.017 mg/L) are significantly different than data collected by CRWA and analyzed at MWRA Deer Island laboratory (range: <0.01 mg/L to 0.115 mg/L, average = 0.057 mg/L) (Figure B3). Therefore, load estimations made with the data from this study are suspect. This is evidenced by 1) WL TP data in this study are significantly lower than all other WL data collected by the CRWA (Figure B2B) and 2) interlaboratory comparison results show RPD values between labs are ≥39% (Table B1). The WL L_Y results from the 2017 CQ method are non-normally distributed (Figure 11). Due to the very low C values (Figure B3), the L_Y distribution is compressed near 0 kg/year (Figure 11).

The interlaboratory comparison undertaken in this study was not exhaustive. Additional analyses using standard reference materials would help clarify the results. The similarity between laboratory results for PO_4 analyses (Table B1) suggest that the discrepancies are related to the conversion of particulate and organic P compounds to the dissolved anion form. The laboratories did not all use the same methods; however, other laboratory comparisons for P constituents suggested that difference across methods did not account for dissimilar results (McHale and Chesney, 2007). The WL results analyzed by Alpha Analytical result in low L_Y estimates, which is clearly demonstrated by the ACQ result (Figure 11). These laboratory differences suggest that loading analyses may be significantly impacted by analytical uncertainty.

LUC and CQ Comparisons

The wide ranges of L_Y that result from taking all literature reported LUC values into consideration suggest that the LUC method of estimating annual P loads has high uncertainty that is unaccounted for if a single coefficient from the literature reported range is selected for each land use (Figure 11). The LUC method yielded L_Y that had a total range of 57,662 kg/yr (282% of mean) at DV, 75,246 kg/yr (171% of mean) at WL, and 73,454 kg/yr (196% of mean) at WD with standard deviations (σ) of 36%, 26%, and 32%, respectively (Table 5). Wide L_Y ranges from the LUC method are somewhat expected since this method does not represent a specific period. LUCs derived from field data are representative of the conditions during the study period and at the study location, which is why it is important to consider the origin of these LUC. While applications of the LUC method do recognize that the coefficient method is poorly constrained, the method was used to determine P load reduction scenarios for towns within the CRW (MassDEP, 2007).

The CQ method serves as the underpinning for load modeling of P and other constituents (Huntington and Aiken, 2013; Duan et al., 2012). The 95% confidence interval associated with the error of regression coefficients from LOADEST was used to populate a range of CQ load estimates. In each iteration, the same relationship was used, but the regression coefficients varied normally about the mean. The CQ method yielded a more narrow range of L_Y. L_Y relative percent ranges were 64%, 101%, 33%, and 24% and σ = 8%, 16%, 5%, and 4% at DV 1970-1984, DV 1985-1994, WL, and WD respectively (Table 5). Results from different sites and water years are not intended for comparison to one another. For example, without regard to seasonality WD has the weakest CQ relationship (Figure 4) yet produced the least variable set of L_Y (Figure 11) due to having the greatest number of observations. The purpose of the range calculations is to compare between the CQ and LUC methods at a single site.

The WD CQ median L_Y, ACQ estimate, and WD TMDL reported L_Y are less than 1% different (Figure 11C). With all three methods yielding similar results for WD, it lends confidence to this estimated value but strongly demonstrates that C values at WD are not Q dependent. This suggests that management priorities should not focus solely on stormwater inputs, because at WD this might not be the driving factor in P loading.

The MassDEP (2007) L_Y was also used to test whether multiple combinations of LUC produced similar L_Y. This value was chosen because it was estimated independently of this study. Out of 1,000 different combinations, 97 resulted in a viable L_Y where viable was defined by a value being equal to ± 5% of 28,925 kg/yr (Figure 12). This demonstrates that a LUC combination from the literature may result in a reasonable L_Y for a watershed, but that the LUC values may not reflect the accurate loading scenarios.

The WL median LUC L_Y and the CRWA median CQ L_Y differ by 53% (the L_Y results using data from this study will not be considered here for reasons explained in the Interlaboratory Comparison section). The CQ calibration dataset for WL spanned 20 years, the longest continuous interval (DV's 24 year record was split in two calibration datasets), the length of record may contribute to the large distinction. While the LUC method does not use discharge data, the LUCs were largely derived from one to three year studies. It is more likely that the LUCs do not account for management practices, which may be reducing loads in the Charles River watershed. There is a significant downward trend in L_Y at WL over 1996-2016 (Figure 13) which is not due to increasing Q, annual mean Q over time does not have a significant correlation (y = -4.91x+10206, R^2=0.098, n=20. P=0.178). Installation of best management practices (BMP) and limiting WWTF P discharge, in the WL watershed are also relevant for WD, but the difference between WD L_Y estimates from the CQ and LUC methods are not dramatically different (20%). During and after the TMDL reports of 2007 and 2011 there was likely an increased focus on implementing BMPs; however, an exhaustive list of management practices employed in the CRW was not compiled as part of this work.

CONCLUSION

This study addressed the comparative uncertainties in the CQ and LUC load estimation techniques for several sites in the CRW. The LUC method is advantageous due to its limited data requirements; however, it produces widely variable L_Y as a result of poorly constrained coefficient values (Figures 3 and 11; Table 1). LUC values in the literature are limited in space and time, increasing the number of field derived LUC values may make this method more viable in the future. Applications of the LUC method that rely on values from the literature may produce reasonable L_Y estimates; however, multiple LUC combinations can yield similar L_Y results as demonstrated here (Figure 12).

The CQ method produces a narrower range of results than the LUC method (Figure 11) but may not be applicable depending on the nature of the CQ relationship. The CQ relationship at a single site can vary over time (Figure 4A) and may be seasonally dependent (Figure 5). It can be challenging to understand the relationship without an understanding of P sources and dynamics in the watershed. When C does not vary with Q, the simple ACQ method may serve as a suitable estimation method as demonstrated by the 1998-2002 WD L_Y results (Figure 11C). The CQ method is best applied in settings and for periods where a) sufficient data are available to characterize the CQ relationship and b) and P sources and dynamics are characterized.

The difference in TP concentrations between this study and the other historical and ongoing monitoring programs in the CRW raises concern about the role of analytical uncertainty in load estimates. The L_Y results using data collected in this study are lower than the estimates using CRWA data (Figure 11 and Table 5). The most striking demonstration of this effect is in the WL ACQ results (Figure 11 and Table 5). While laboratory accuracy and precision may be established internally, these results suggest inter-laboratory comparison is required. The comparison undertaken here was limited, the nature of the discrepancies should be characterized with a more comprehensive approach.

The CQ and LUC methods differ in the range of results they produce. The LUC method may serve best as a "back of the envelope" calculation done where little to no C and Q data are available, although uncertainty characterization should accompany these estimates. L_Y estimates are dependent on the method (Figs. 11A, 11B). Full consideration and reporting of L_Y uncertainties are critical to guiding management decisions.

REFERENCES

Aulenbach, B.T., 2006, Annual dissolved nitrite plus nitrate and total phosphorous loads for the Susquehanna, St. Lawrence, Mississippi-Atchafalaya, and Columbia River basins, 1968-2004, U.S. Geological Survey OFR 06-1087,16 p.

Bannerman, R., Konrad, J., Becker, D., Simsiman, G.V., Chesters, G., Goodrich-Mahoney, J., and Abrams, B., 1979, The IJC Menomonee River Watershed Study — Surface Water Monitoring Data: EPA-905/4-79-029. U.S. Environmental Protection Agency, Chicago, IL.

Bicknell, B.R., Imhoff, J.C., Kittle, J.L., Jr., Donigian, A.S., Jr., and Johanson, R.C., 1997, Hydrological Simulation Program--Fortran, User's manual for version 11: U.S. Environmental Protection Agency, National Exposure Research Laboratory, Athens, Ga., EPA/600/R-97/080, 755 p.

Chen, D., Hu, M., Guo, Y., and Dahlgren, R., 2015, Reconstructing historical changes in phosphorus inputs to rivers from point and nonpoint sources in a rapidly developing watershed in eastern China, 1980-2010: The Science of the Total Environment, *v.533*, p.196-204, doi: 10.1016/j.scitotenv.2015.06.079

Cohn, T.A., Caulder, D.L., Gilroy, E.J., Zynjuk, L.D, and Summers, R.M., 1992, The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay, Water Resources Research, v. 28, p. 2353-2363.

CRWA, Charles River Watershed Association, 2011, Total Maximum Daily Load for Nutrients in the Upper/Middle Charles River, Massachusetts: Prepared for the Massachusetts Department of Environmental Protection, CN 272.0, 103 p. plus appendixes, accessed January 29, 2018, at http://www.mass.gov/eea/agencies/massdep/water/watersheds/total-maximum-daily-loads-tmdls.html#9.

Dillon, P.J., and Kirchner, W.B., 1975, The effects of geology and land use on the export of phosphorus from watersheds: Water Research, v. 9, p. 135-148, doi: http://dx.doi.org.proxy.bc.edu/10.1016/0043-1354(75)90002-0.

Duan, S., Kaushal, S.S., Groffman, P.M., Band, L.E., and Belt, K.T., 2012, Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed: Journal of Geophysical Research: Biogeosciences, v. 117, p. n/a-n/a, doi: 10.1029/2011JG001782.

Goonetilleke, A., Thomas, Ev., Ginn, S., and Gilbert, D., 2005, Understanding the role of land use in urban Eventwater quality management: Journal of Env. Management, v. 74, p. 31-42, doi:10.1016/j.jenvman.2004.08.006.

Hamrick J. M., 1992, A three-dimensional environmental fluid dynamics computer code: theoretical and computational aspects. Special Report 317 in Applied Marine Science and Ocean Engineering. The College of William and Mary, Virginia Institute of Marine Science.

Hirsch, R.M., 2014, Large biases in regression-based constituent flux estimates: causes and diagnostic tools: JAWRA, v. 50, doi: 10.1111/jawr.12195.

Huntington, T.G., and Aiken, G.R., 2013, Export of dissolved organic carbon from the Penobscot River basin in north-central Maine: Journal of Hydrology, v. 476, p. 244-256, doi:10.1016/j.jhydrol.2012.10.039

Janke, B., Finlay, J., Hobbie, S., Baker, L., Sterner, R., Nidzgorski, D., and Wilson, B., 2013, Contrasting influences of Eventflow and baseflow pathways on nitrogen and phosphorus export from an urban watershed: Biogeochemistry, v. 121, p. 209-228, doi: 10.1007/s10533-013-9926-1.

Johnes, P.J., 1996, Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: The export coefficient modelling approach: Journal of Hydrology, v. 183, p. 323-349, doi: 10.1016/0022-1694(95)02951-6.

Jones, A., Stevens, D., Horsburgh, J., and Mesner, N., 2011, Surrogate Measures for Providing High Frequency Estimates of Total Suspended Solids and Total Phosphorus Concentrations: Journal of the American Water Resources Association, v. 47, no. 2, p. 239-253.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lutkepohl, H., and Lee, T.C., 1988, Introduction to the theory and practice of econometrics (2d ed.): New York, John Wiley, 1024 p.

Liu, A., Goonetilleke, A., and Egodawatta, P., 2012a, Inadequacy of land use and impervious area fraction for determining urban Eventwater quality: Water Resource Management, v. 26, p. 2259-2265, doi:10.1007/s11269-012-0014-4.

Liu, A., Goonetilleke, A., and Egodawatta, P., 2012b, Inherent errors in pollutant build-up estimation in considering urban land use as a lumped parameter: Journal of Environmental Quality, v. 41, p. 1690-1694, doi:10.2134/jeq2011.0419.

McHale, M.R., and McChesney, D., 2007, Phosphorus concentrations in stream-water and reference samples—An assessment of laboratory comparability: U.S. Geological Survey Open-File Report 2007–1267, 25 p. ONLINE ONLY

Postel, S.L., 1998, Water for food production: will there be enough in 2025?, BioScience, v. 48, no. 8, p. 629-637, doi: 10.2307/1313422.

Reckhow, K.H., Beaulac, M.M., and Simpson, J.T., 1980, Modeling phosphorus loading and lake response under uncertainty: a manual and compilation of export coefficients,https://nepis.epa.gov/Exe/ZyPDF.cgi/00001KWE.PDF?Dockey=00001KWE.PDF (accessed January 2017).

Runkel, R.L., Crawford, C.G., Cohn, T.A., 2004, Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers, U.S. Geological Survey Techniques and Methods Book 4, Chapter A5, 75 p.

Schenk, L.N., Anderson, C.W., Diaz, Paul, and Stewart, M.A., 2016, Evaluating external nutrient and suspended-sediment loads to Upper Klamath Lake, Oregon, using surrogate regressions with real-time turbidity and acoustic backscatter data: U.S. Geological Survey Scientific Investigations Report 2016–5167, 46 p., https://doi.org/10.3133/sir20165167.

Schindler, D.W., 2001, The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium, Can. J. Fish. Aquatic. Sci., v. 58, p. 58-29.

Smith, K.P., 2017, Loads and yields of deicing compounds and total phosphorus in the Cambridge drinking-water source area, Massachusetts, water years 2009–15: U.S. Geological Survey Scientific Investigations Report 2017–5047, 52 p., https://doi.org/10.3133/sir20175047.

Song, K., Winters, C., Xenopooulos, M.A., Marsalek, J., and Frost, P.C., 2017, Phosphorus cycling in urban aquatic ecosystems: connecting biological processes and water chemistry to sediment P fractions in urban stormwater management ponds, Biogeochemistry, v. 132, p. 203-212, doi:10.1007/s10533-017-0293-1.

USEPA, United States Environmental Protection Agency, 1995, QUAL2E Windows Interface User's Manual, Office of Water, EPA-823-B-95-003.

Vollenweider, R.A., 1968, Scientific fundamentals of the eutrophication of lakes and flowing waters with particular reference to nitrogen and phosphorus as factors in eutrophication, OECD Tech. Rep. DAS/CS/68.27, Paris, France.

MassDEP, Massachusetts Department of Environmental Protection and United States Environmental Protection Agency, New England Region, 2007, Total Maximum Daily Load for Nutrients In the Lower Charles River Basin, Massachusetts, CN 301.0, http://www.mass.gov/eea/docs/dep/water/resources/a-thru-m/charlesp.pdf.

Valtanen, M., Sillanpää, N., and Setälä, H., 2014, The Effects of Urbanization on Runoff Pollutant Concentrations, Loadings and Their Seasonal Patterns Under Cold Climate: Water, Air, & Soil Pollution, v. 225, p. 1-16, doi: 10.1007/s11270-014-1977-y.

Wang, Q., Li, S., Jia, P., Qi, C., Ding, F., 2013, A review of surface water quality models: Scientific World Journal, v. 2013, Article ID 231768, 7 pages, doi: 10.1155/2013/231768.

Winter, J.G., and Duthie, H.C., 2000, Export coefficient modeling to assess phosphorus loading in an urban watershed: Journal of the American Water Resources Association, v. 36, n. 5, p. 1053-1061.

TABLES

Table 1. Literature (gray and peer-reviewed) reported LUC sorted by land use and alphabetically by author name. If a copy of the original source was procured, it is indicated by Y, yes. If the original source was not obtained it is indicated by N, no.

Forest	Dillon and Kirchner, 1975	0.077	Y
Forest	Dillon and Kirchner, 1975	0.048	Y
Forest	Dillon and Kirchner, 1975	0.038	Y
Forest	Dillon and Kirchner, 1975	0.027	Y
Forest	Dillon and Kirchner, 1975	0.041	Y
Forest	Dillon and Kirchner, 1975	0.145	Y
Forest	Dillon and Kirchner, 1975	0.092	Y
Forest	Dillon and Kirchner, 1975	0.122	Y
Forest	Dillon and Kirchner, 1975	0.067	Y
Forest	Duan et al. 2015	0.028	Y
Forest	Duffy et al., 1978	0.281	Y
Forest	Duffy et al., 1978	0.306	Y
Forest	Duffy et al., 1978	0.357	Y
Forest	Duffy et al., 1978	0.321	Y
Forest	Duffy et al., 1978	0.226	Y
Forest	Fredriksen, 1972	0.52	N
Forest	Fredriksen, 1979	0.18	N
Forest	Fredriksen, 1979	0.68	N
Forest	Henderson et al., 1977	0.01	N
Forest	Henderson et al., 1977	0.02	N
Forest	Henderson et al., 1977	0.03	N
Forest	Henderson et al., 1977	0.03	N
Forest	Krebs and Golley, 1977	0.275	N
Forest	Likens et al., 1977	0.019	Y
Forest	Nicholson, 1977	0.06	Y
Forest	Nicholson, 1977	0.036	Y

Forest	Schindler and Nighswander, 1970	0.09	Y
Forest	Schindler et al., 1976	0.329	Y
Forest	Schindler et al., 1976	0.435	Y
Forest	Schindler et al., 1976	0.289	Y
Forest	Schindler et al., 1976	0.22	Y
Forest	Singer and Rust, 1975	0.09	Y
Forest	Sylvester, 1960	0.83	Y
Forest	Sylvester, 1960	0.36	Y
Forest	Taylor et al., 1971	0.035	Y
Forest	Taylor et al., 1971	0.072	Y
Forest	Taylor et al., 1971	0.035	Y
Forest	Timmons et al., 1977	0.19	Y
Forest	Timmons et al., 1977	0.38	Y
Forest	Timmons et al., 1977	0.28	Y
Forest	MassDEP, 2007	0.13	Y
Forest	CRWA, 2011	0.17	Y
Forest	Verry, 1979	0.124	Y
Forest	Verry, 1979	0.179	Y
Forest	Verry, 1979	0.157	Y
Forest	White et al., 2015	0.073*	Y
LDR	Bannerman et al., 1979 and 1983	0.04	Y
LDR	Duan et al., 2015	0.032	Y
LDR	Hydrologic Engineering Center, 1977 in Haith and Shoemaker, 1987	0.5	N
LDR	Landon, 1977	0.19	Y
LDR	Landon, 1977	2.7	Y
LDR	MassDEP, 2007	0.045	Y
LDR	CRWA, 2011	0.38	Y
MDR	Bannerman et al., 1979 and 1983	0.58	Y
MDR	Breault et al., 2002	2.28	Y
MDR	Hydrologic Engineering Center, 1977 in Haith and Shoemaker, 1987	0.8	N
MDR	MassDEP, 2007	0.566	Y
MDR	CRWA, 2011	0.62	Y
HDR	Bannerman et al., 1979 and 1983	1.12	Y
HDR	Hydrologic Engineering Center, 1977 in Haith and Shoemaker, 1987	2.7	N
HDR	Landon, 1977	1.1	Y
HDR	Landon, 1977	0.56	Y
HDR	MassDEP, 2007	1.131	Y
HDR	CRWA, 2011	1.11	Y
Commercial	Betson, 1978	4.85	Y

Commercial	EPA 1983 (NURP)	3.4	N
Commercial	Konrad et al., 1978	4.28	Y
Commercial	Landon, 1977	1.7	N
Commercial	Landon, 1977	0.66	Y
Commercial	Much and Kemp, 1978	0.88	N
Commercial	MassDEP, 2007	1.697	Y
Commercial	CRWA, 2011	2.03	Y
Industrial	Bannerman et al., 1979 and 1983	1.49	Y
Industrial	Hydrologic Engineering Center, 1977 in Haith and Shoemaker, 1987	4	N
Industrial	Konrad et al., 1978	1.06	N
Industrial	Much and Kemp, 1978	0.75	N
Industrial	MassDEP, 2007	1.471	Y
Industrial	CRWA, 2011	2.03	Y

* Simulated

Table 2. The three study locations within CRW. The land cover statistics are derived from National Land Cover Database, 2011 in USGS StreamStats v.4.

	USGS Station ID	Site Coordinates	Watershed Area, *A* (km^2)	% Forest	% Urban Developed	% Other
Watertown Dam (WD)	01104615	42˚ 21' 53" 71˚ 11' 25"	707	43	47	10
Waltham (WL)	01104500	42˚ 22' 20" 71˚ 14' 03"	647	46	44	10
Dover (DV)	01103500	42˚ 15' 22" 71˚ 15' 38"	474	51	39	10

Table 3. Summary of discharge (Q) and concentration (C) datasets. The data interval represents only the period used in this report. For example, the discharge records at WL and DV are still active. The analyses indicate which simulations the data were used for; note that none of the data in this table are required for the LUC method. The hydrologic conditions comparisons are reported in the Appendix B and referenced in the discussion section of this text. The two WL sites where C data was collected are ~150 meters apart and are treated as a single location.

Site Name	Data Interval	Reporting Entity	Analyses
Discharge (Q)			
Watertown Dam (WD)	Aug 1999 – Sept 2002	USGS	CQ LOADEST, Monte Carlo Simulation
Waltham (WL)	Aug 1996 – Dec 2016	USGS	CQ LOADEST, Monte Carlo Simulation, Hydrologic Conditions Comparison
Dover (DV)	Nov 1970 – Nov 1994	USGS	CQ LOADEST, Monte Carlo Simulation, Hydrologic Conditions Comparison
TP Concentration (C)			
Watertown Dam (WD)	Aug 1999 – Sept 2002	MWRA	CQ LOADEST, Monte Carlo Simulation
Waltham, Moody St Bridge (WL)	Aug 1996 – Dec 2016	CRWA	CQ LOADEST, Monte Carlo Simulation, Hydrologic Conditions Comparison
Waltham (WL)	Aug 2016 – May 2017	This study	CQ LOADEST, Hydrologic Conditions Comparison
Dover (DV)	Nov 1970 – Nov 1994	USGS	CQ LOADEST, Monte Carlo Simulation

Table 4. Land use binning for LUC method.

Forested	Forested Wetland Brushland/Successional Nursery Orchard Open Land
Low Density Residential	Very Low Density Residential Cemetery Golf Course Transitional Participation Recreation
Medium Density Residential	Multi-Family Residential
High Density Residential	N/A
Commercial	Urban Public Institutional Powerline Transportation
Industrial	Junkyard Waste Disposal Mining

Table 5. Summary statistics of the average annual TP load estimates produced by the CQ and LUC methods and associated Monte Carlo Simulation.

	DV			WL			WD	
	CQ 1970-1984	CQ 1985-1994	LUC	CQ 2017	CQ 1996-2016	LUC	CQ 1998-2002	LUC
Median (L_Y kg/yr)	36873	15236	19095	13819	17235	29666	29155	35443
Mean (L_Y kg/yr)	37078	15483	20474	16015	17264	31291	29180	36851
Standard Deviation (L_Y kg/yr)	3112	2459	7440	9315	908	10472	1116	11210
% Relative Standard Deviation	8	16	36	58	5	33	4	30
Minimum (L_Y kg/yr)	26836	10044	6155	3971	14892	10140	26055	11516
Maximum (L_Y kg/yr)	50382	25657	63818	87122	20621	71694	33188	71431
Range (L_Y kg/yr)	23546	15613	57662	83151	5730	61554	7133	59915
% Relative Range	64	101	282	519	33	197	24	163
ACQ (L_Y kg/yr)	56778	17660	N/A	3310	15787	N/A	29156	N/A

FIGURES

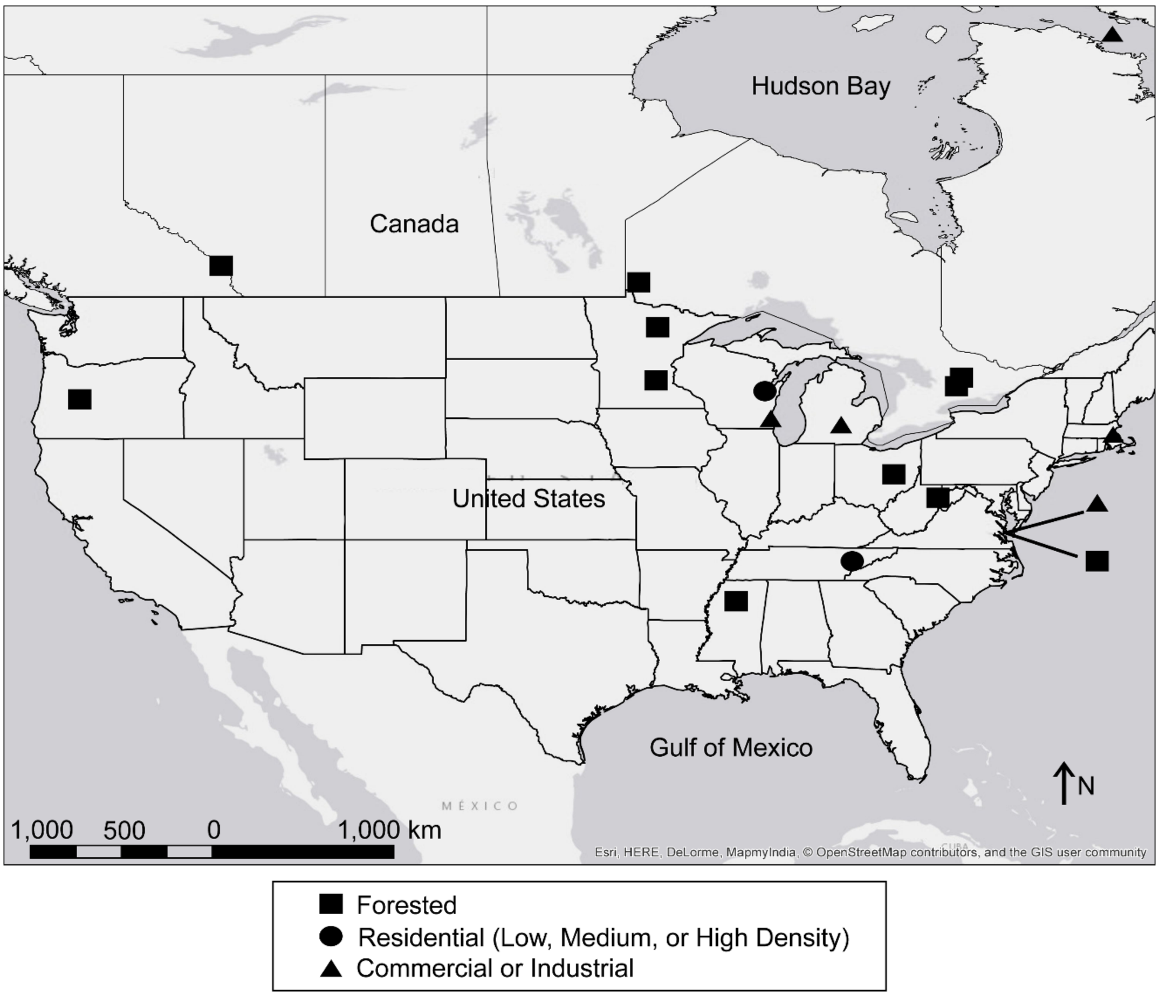


Figure 1. Map of LUC study locations in North America from the literature review. Only those studies for which the original source was obtained are reflected here.

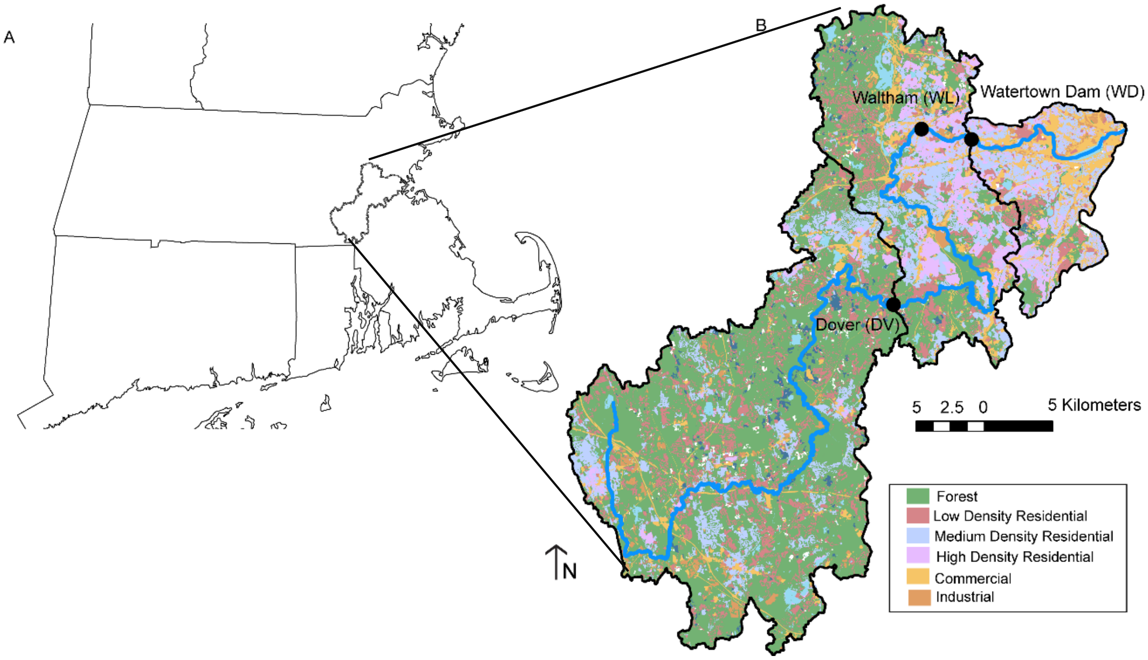


Figure 2. A) Location of CRW within Massachusetts. B) Land cover data for CRW and the three main sites in this study. Black outlines separate the lower (WD to outlet), middle (DV to WD), and upper (upstream of DV) CRW.

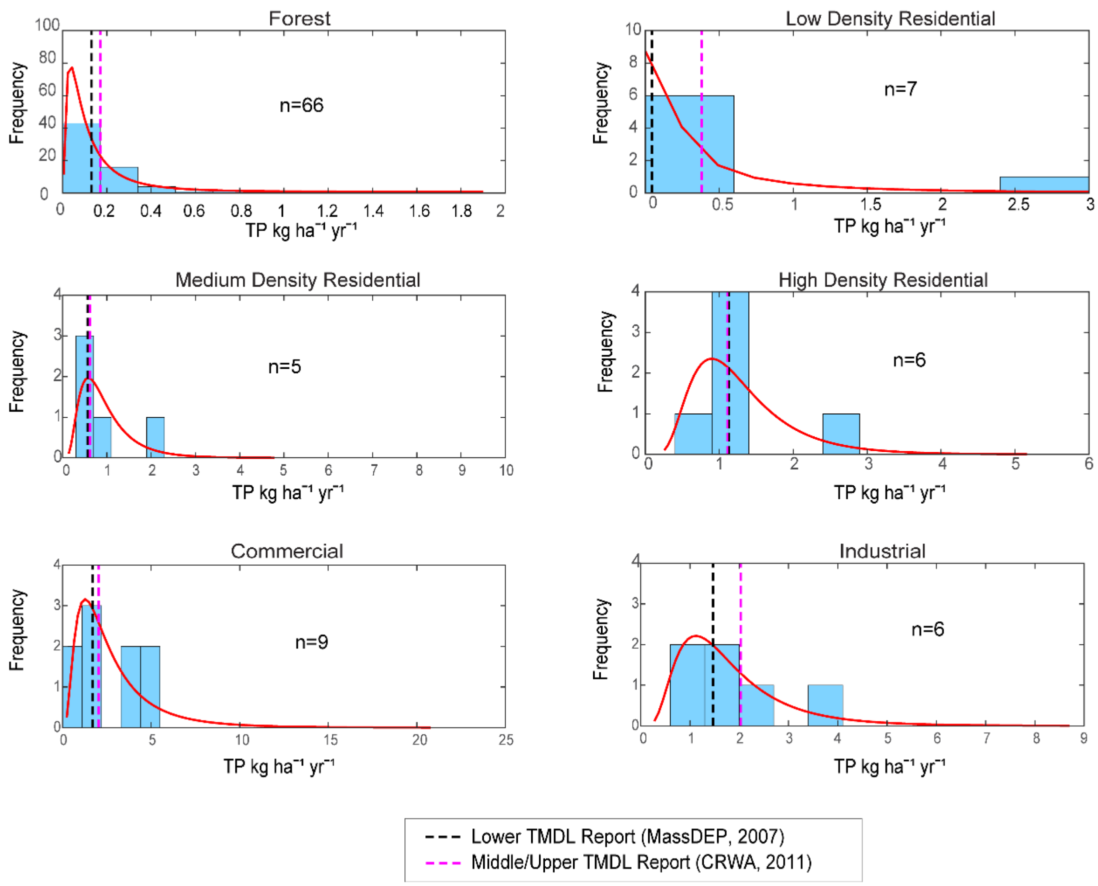


Figure 3. Distribution of reported *LUC* compiled from a review of gray and peer reviewed literature. Red lines are log normal fits and dashed lines are the *LUC* values reported in the two Charles River TMDL Reports.

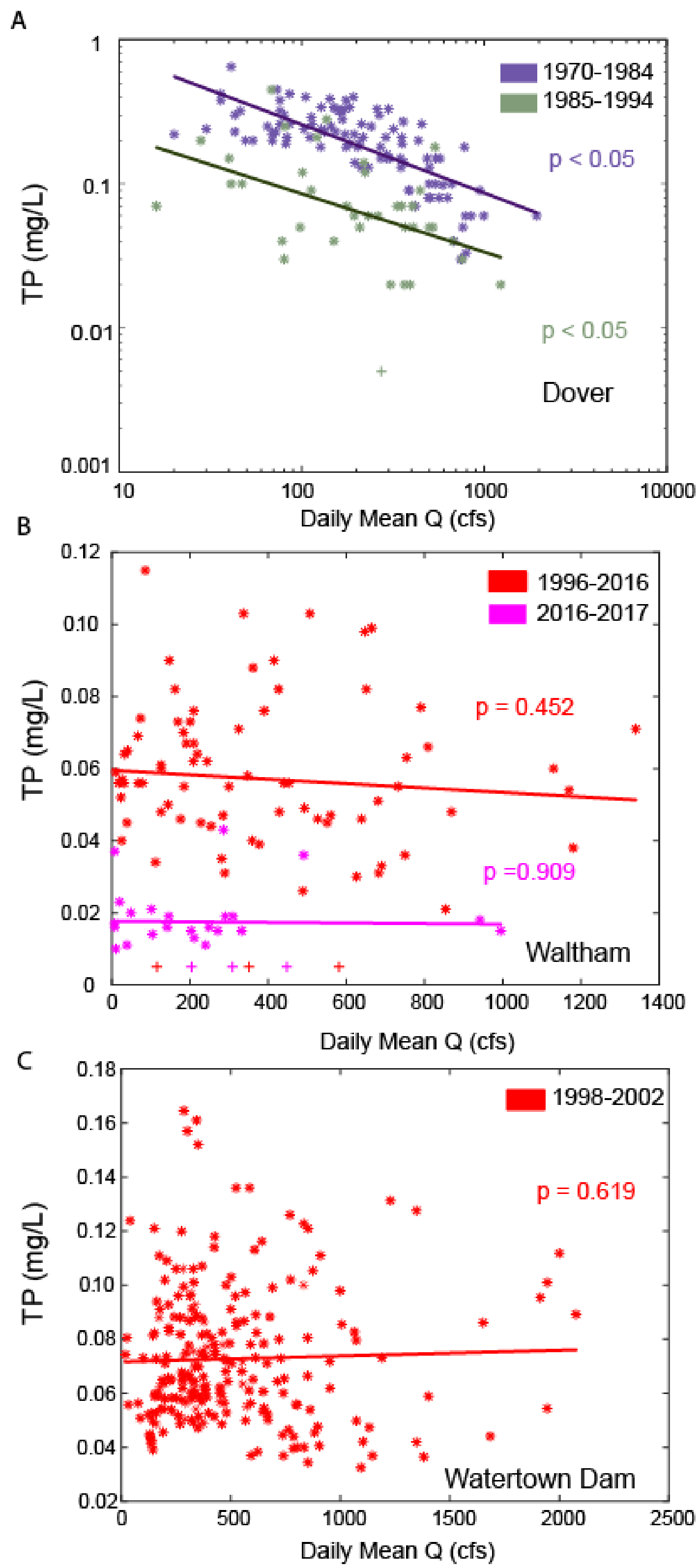


Figure 4. Plots of daily mean discharge (Q) versus instantaneous TP concentration (C), solid lines are least squares regression best-fit lines. A) Least squares power-law regression for Dover

(DV) 1970-1984, $R^2 = 0.527$ and 1985-1994. $R^2 = 0.205$. DV concentration data collected by USGS. B) Least squares linear regression for Waltham (WL), $R^2 < 0.01$. WL concentration data from this study and from CRWA. C) Least squares linear regression for Watertown Dam (WD), $R^2 < 0.01$. WD concentration data collected by MWRA. Data plotted here served as calibration datasets in each respective LOADEST regression. Daily mean *Q* reported by USGS at all three stations. Missing *Q* records in the WD series were estimated using linear regression between WL and WD (Figure B4). The '+' symbol indicates a value below minimum reporting limit, <0.01 mg/L.

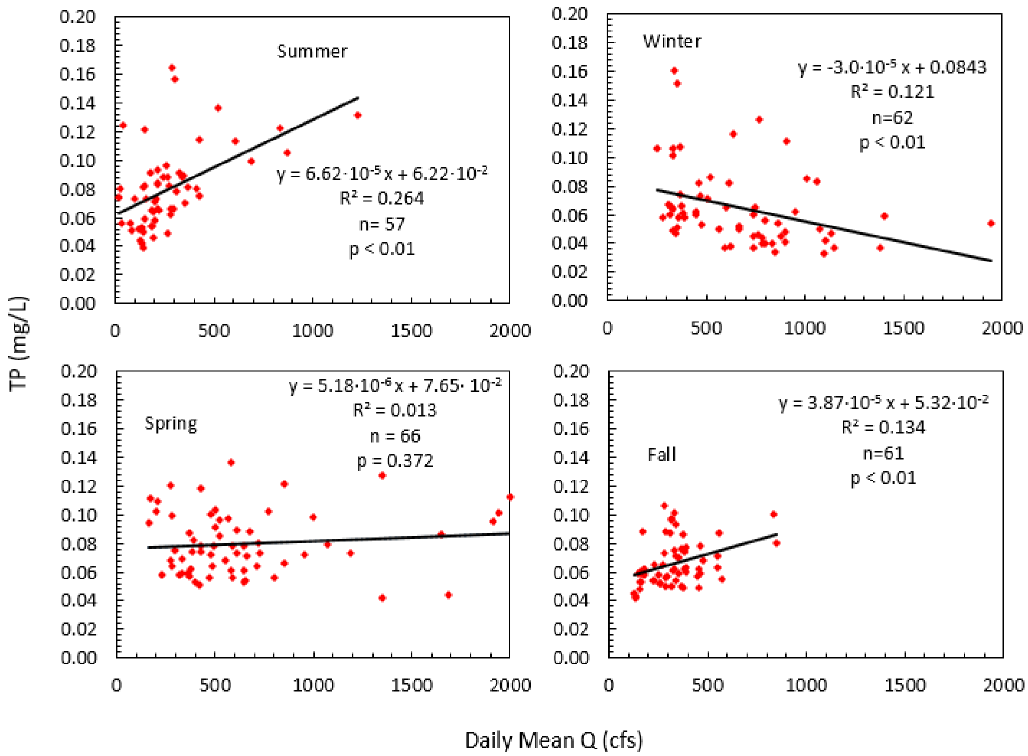


Figure 5. Seasonal plots of WD CQ data from 1998-2002 where spring is April- June, summer is July – September, fall is October- December, and winter is January – February. The solid black lines are the least squares linear regressions.

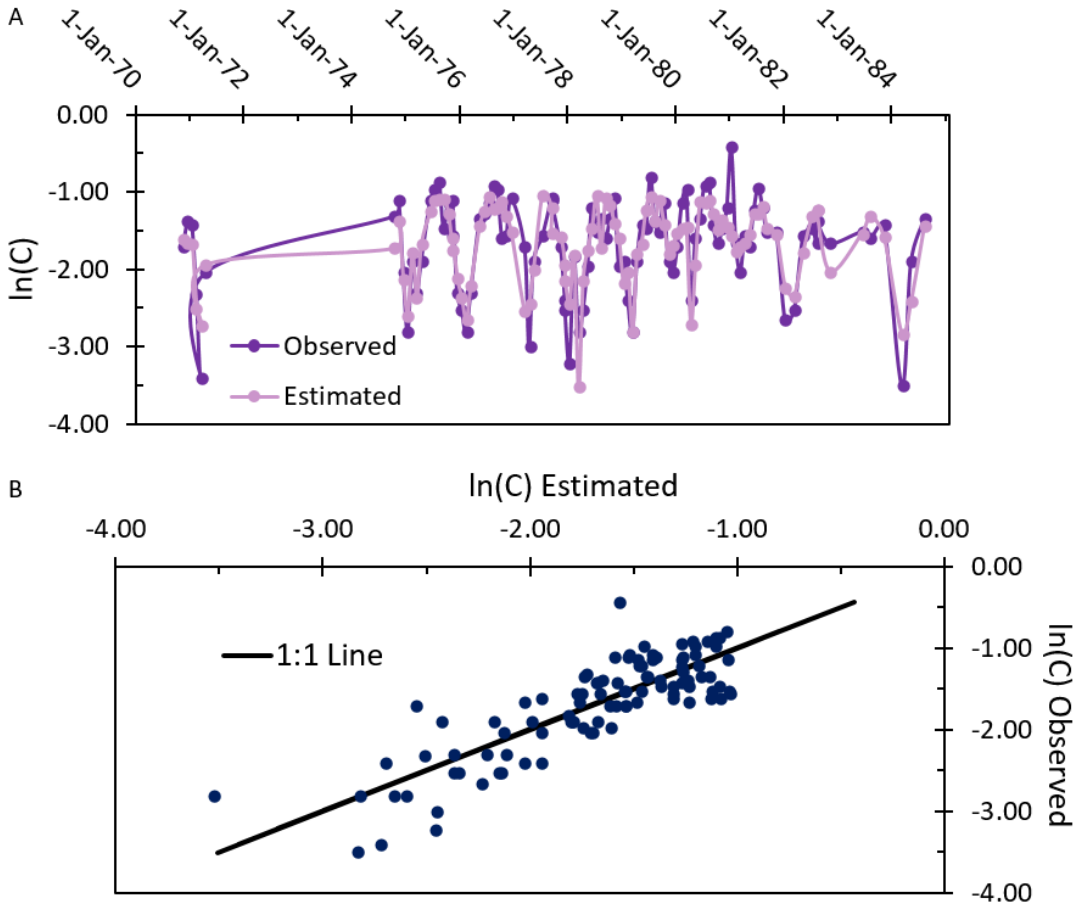


Figure 6. LOADEST regression evaluation for DV 1970-1984. A) Natural log of observed and estimated *C* over time. The points are connected to show trends; however, it should be noted that that only the circle represent data and model estimates. B) The natural log of estimated versus observed *C*. The 1:1 line shows a perfect estimate and observation match for comparison.

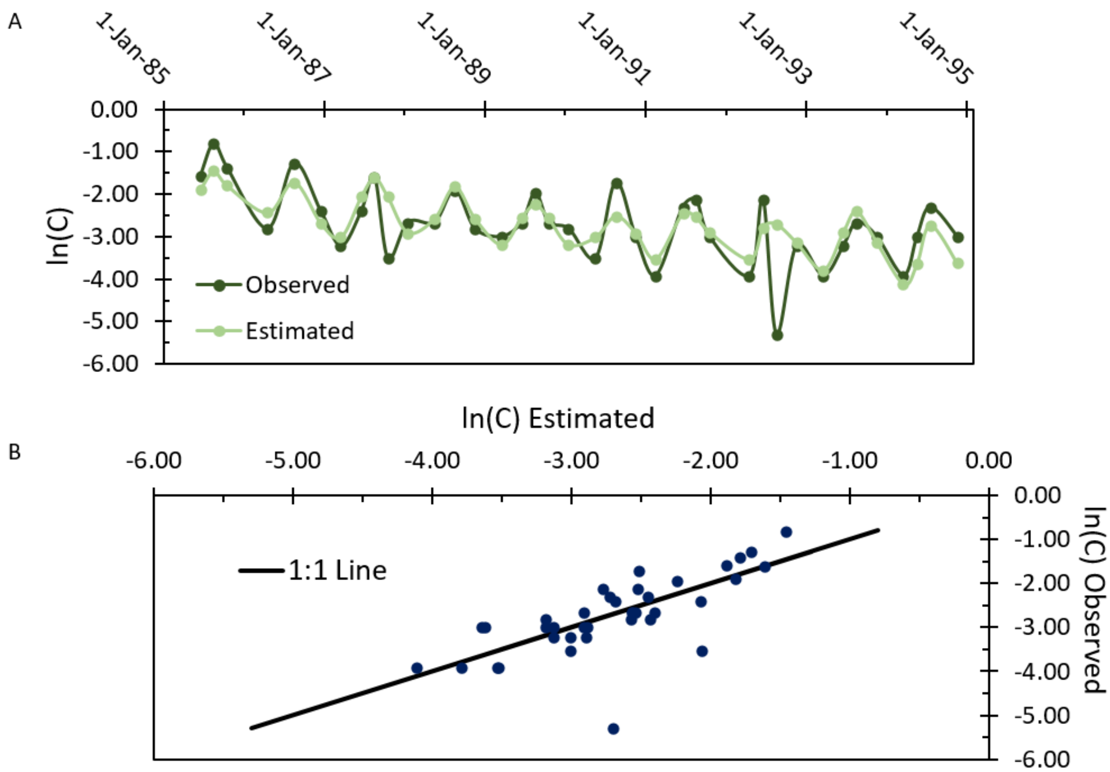


Figure 7. LOADEST regression evaluation for DV 1985-1994. A) Natural log of observed and estimated *C* over time. The points are connected to show trends; however, it should be noted that that only the circle represent data and model estimates. B) The natural log of estimated versus observed *C*. The 1:1 line shows a perfect estimate and observation match for comparison.

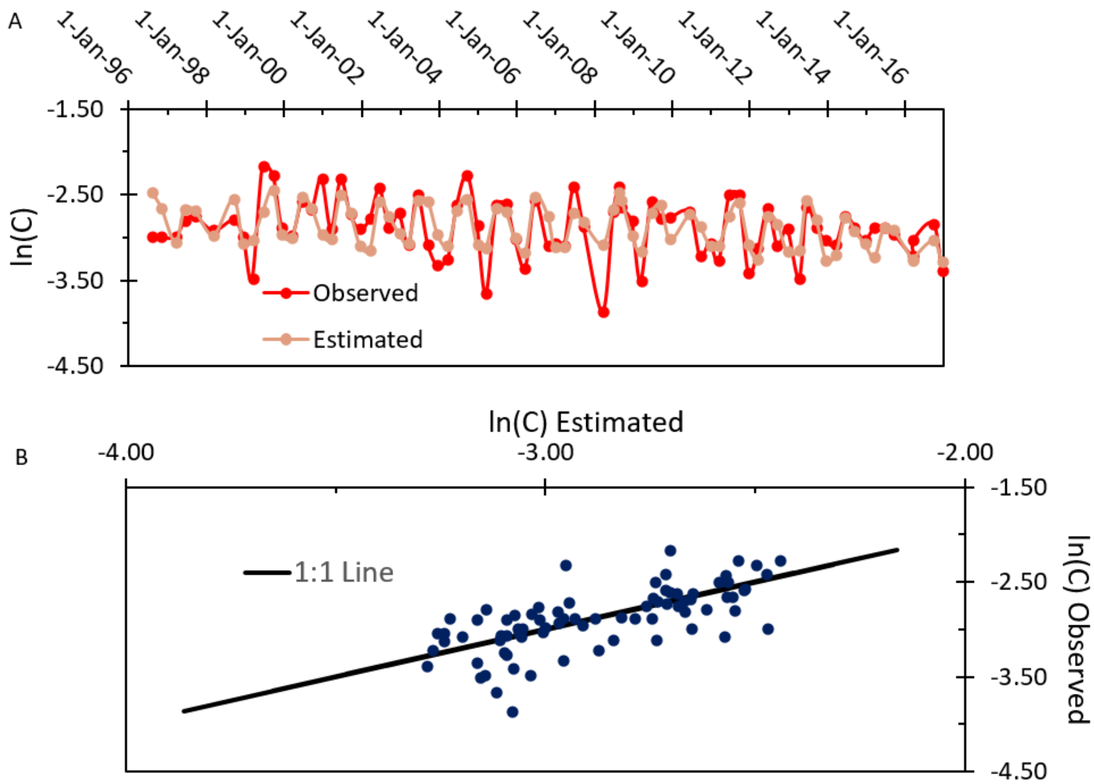
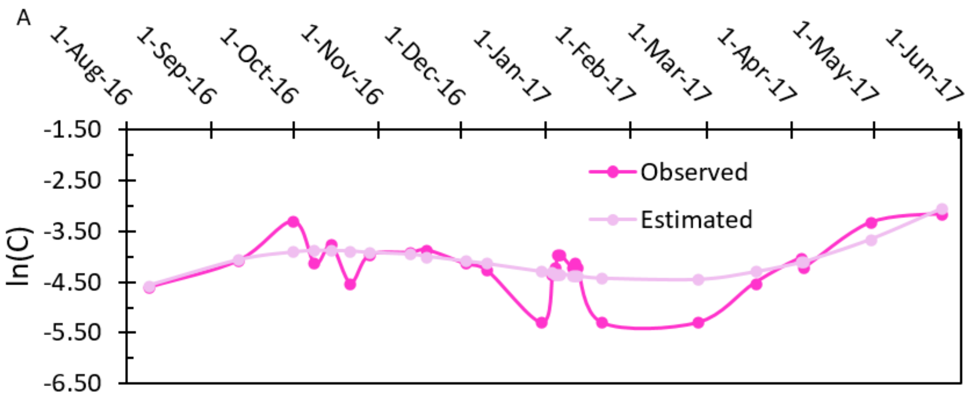


Figure 8. LOADEST regression evaluation for WL 1996-2016. A) Natural log of observed and estimated *C* over time. The points are connected to show trends; however, it should be noted that that only the circle represent data and model estimates. B) The natural log of estimated versus observed *C*. The 1:1 line shows a perfect estimate and observation match for comparison.



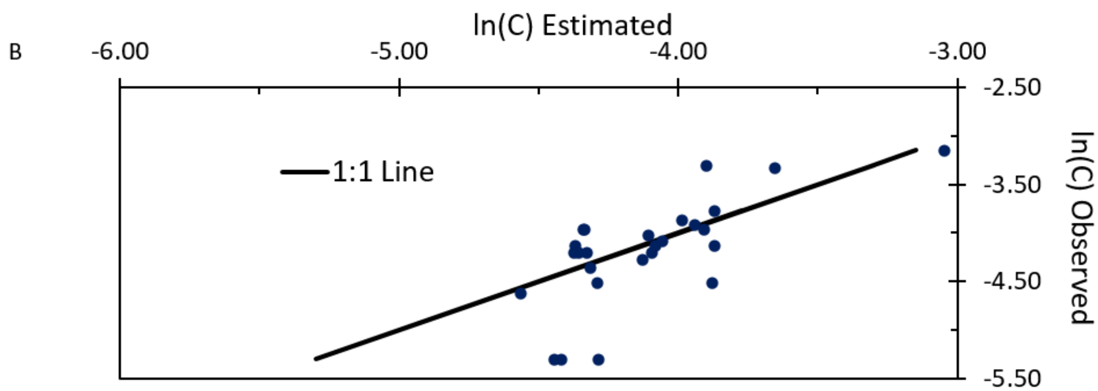


Figure 9. LOADEST regression evaluation for WL 2016-2017. A) Natural log of observed and estimated *C* over time. The points are connected to show trends; however, it should be noted that that only the circle represent data and model estimates. B) The natural log of estimated versus observed *C*. The 1:1 line shows a perfect estimate and observation match for comparison.

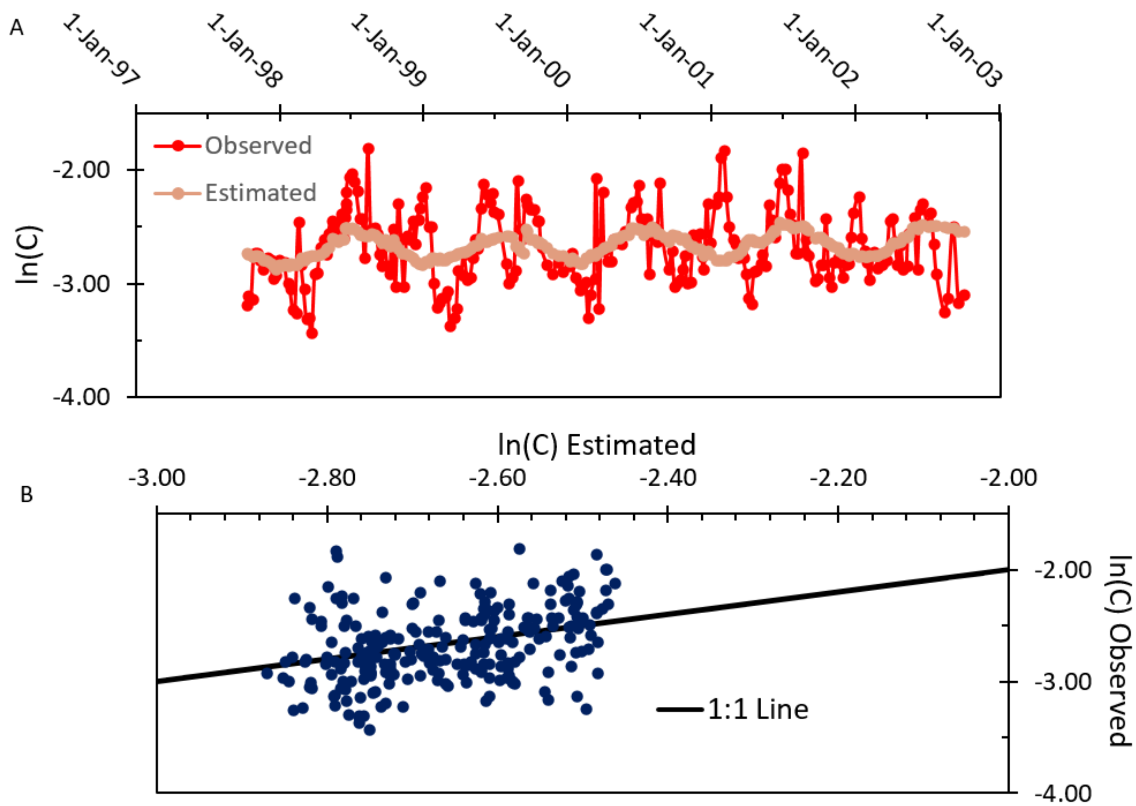


Figure 10. LOADEST regression evaluation for WD 1998-2002. A) Natural log of observed and estimated *C* over time. The points are connected to show trends; however, it should be noted that that only the circle represent data and model estimates. B) The natural log of estimated versus observed *C*. The 1:1 line shows a perfect estimate and observation match for comparison.

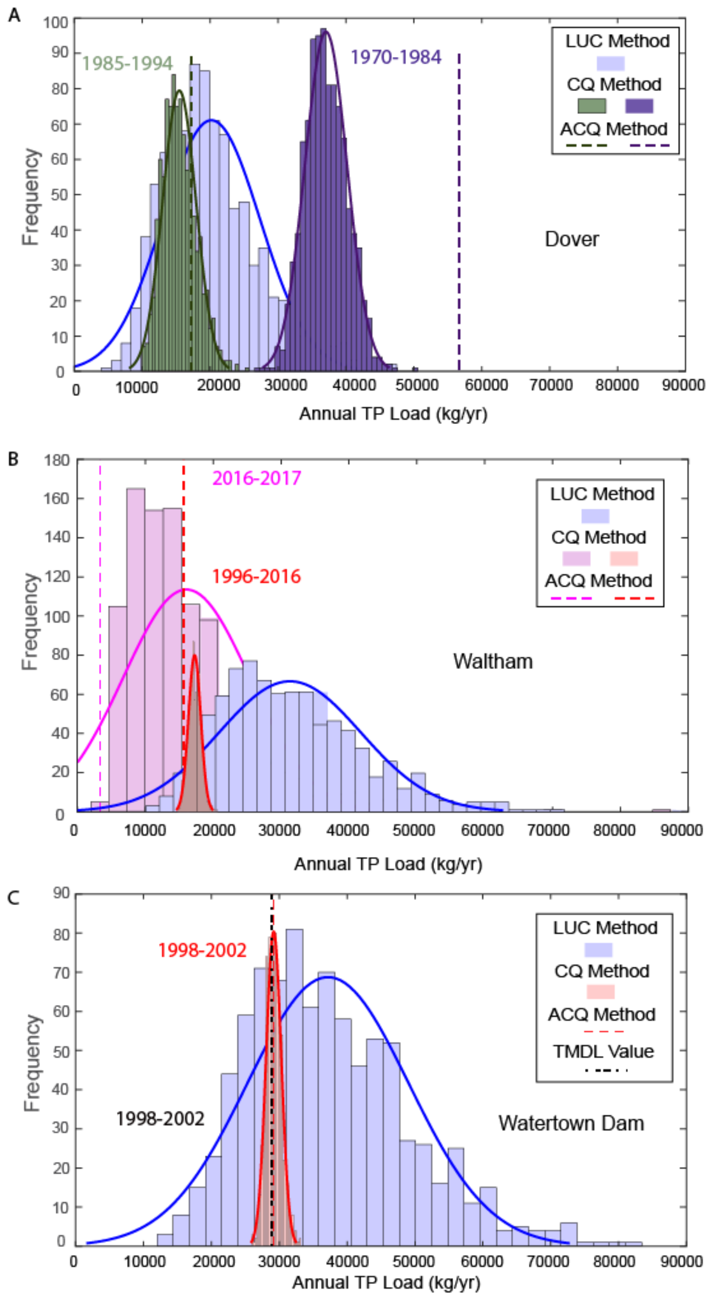


Figure 11. The annual L_Y frequency distributions from LUC, and CQ methods, 1000 iterations each. The dashed lines are the results of the ACQ method. The color-coded years correspond to the period of data used in the simulation for the CQ method. The solid curves correspond to a normal distribution fit of the data. A) For DV, the CQ method was applied to two separate periods due to the distinct nature of the CQ relationship during these periods (Figure 4A). B) For WL, the 2016-2017 period was estimated using data collected in this study and the 1996-2016 estimates use CRWA C data. C) For WD, the black dash-dotted line is the L_Y estimated for 1998-2002 developed during the lower Charles River TMDL (MassDEP, 2007).

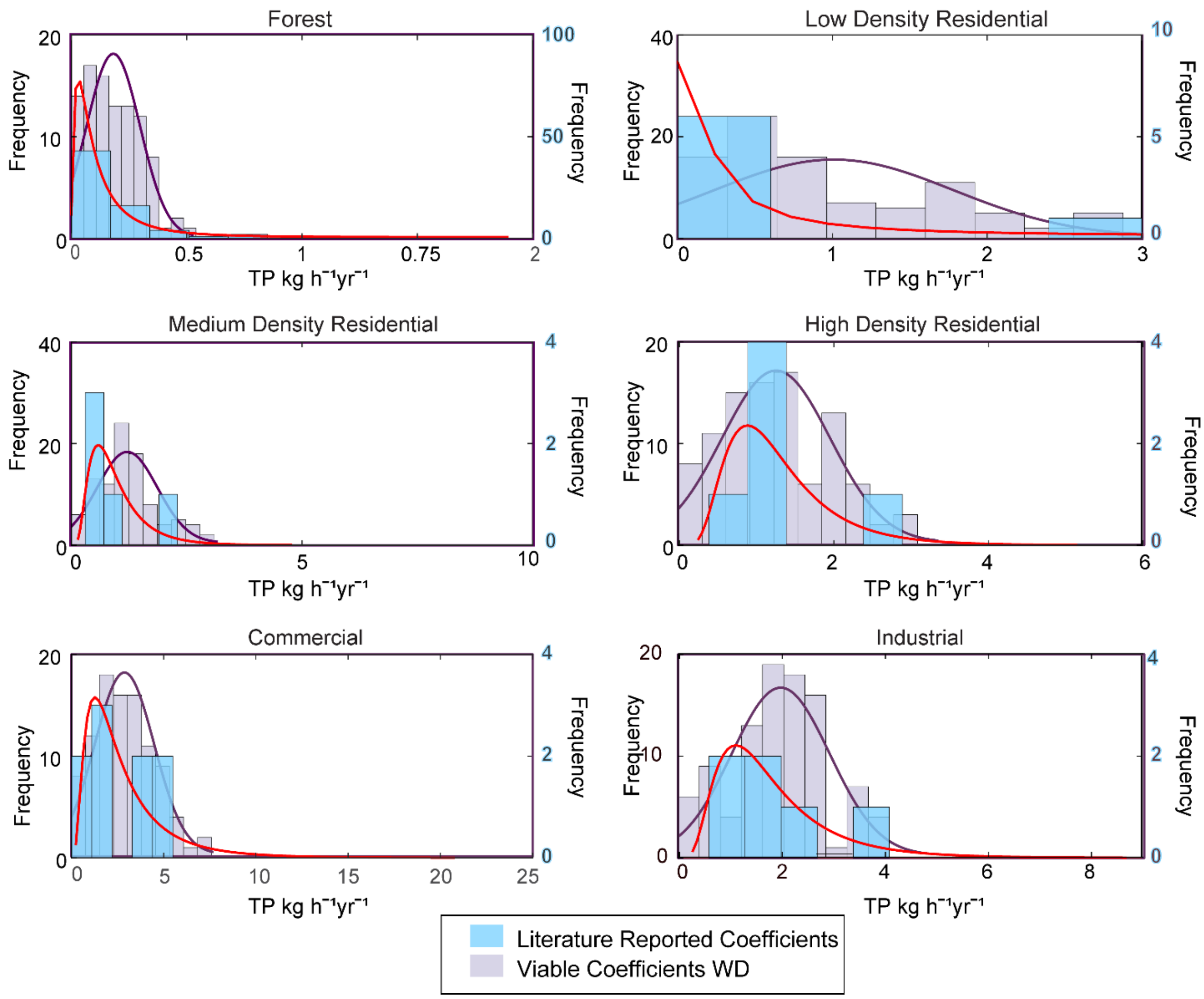


Figure 12. Frequency distribution of literature reported LUC (Figure 3) with log normal fit (red line) and viable coefficients for the WD site with normal fit (purple line). Viable coefficients determined as those *LUC* values which resulted in an L_Y within 5% of the MassDEP, 2007 WD L_Y. Right axis applies to *LUC* values (blue bars, red curve); left axis applies to viable *LUC* (purple bars, purple curve).

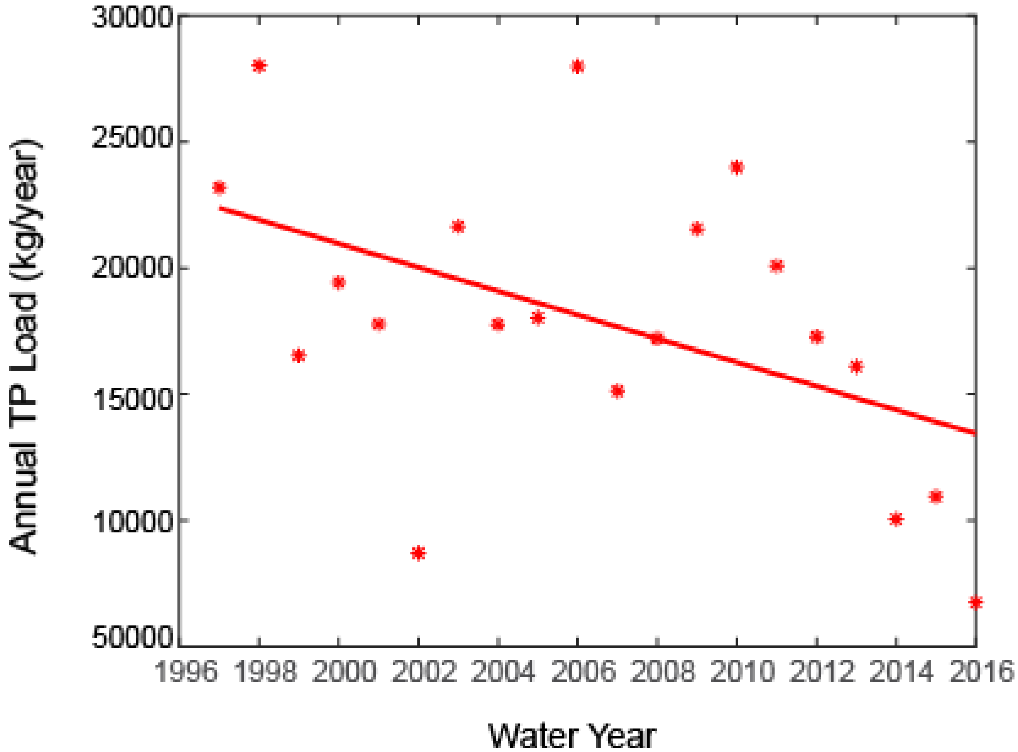


Figure 13. Least squares linear regression for L_Y determined by LOADEST for WL (CRWA data) summed over each water year (Oct 1 – Sept 30) of the calibration period. Significant downward trend where y=-471x+962800, R^2=0.232, n=20, $p<0.05$.

APPENDIX A

Introduction

The following Appendix describes the project details that were not included in the main thesis text, it also outlines challenges associated with these components of the project, but it is not exhaustive. The original intent of this thesis project was to investigate the controls on phosphorus (P) loading in the Charles River watershed (CRW). I hypothesized that P loading was primarily controlled by the dominant land use within a watershed. It is well documented that urbanization alters hydrologic regime and water quality from natural conditions. Pollutant loadings are shown to increase in urbanized watersheds, but these changes occur non-uniformly across space and time (Norvell et al., 1979; Beaulac and Reckhow, 1982; Petrone, 2010; Duan et al., 2012). P loading is critical in shaping the trophic status and ecosystem health of fresh water bodies where P is most commonly limited. It is well known that land use is a critical factor influencing P loading (e.g., Bannerman et al., 1979; Winter and Duthie, 2000; Goonetilleke et al., 2005; Duan et al., 2012; Valtanen et al., 2014) because impervious cover increases runoff, urban drainage systems provide conduits for contaminant transport, and anthropogenic activities can increase P availability (i.e. application of fertilizer and P bearing cleaning agents). The role of land use in altering the hydrologic regime is particularly relevant to phosphorus loading because runoff accounts for 68 to 93% of yearly phosphorus loads in urban watersheds (Janke et al., 2014; Long et al., 2014).

Land use based loading coefficients (LUC), reported in units of kg P ha^{-1} $year^{-1}$, serve as the underpinning in most P loading and water quality models (e.g., Mass DEP, 2007, CRWA, 2011; Liu, 2012b). These coefficients have been applied in environments with different climates, geologies, and management practices resulting in potentially high uncertainties on annual loading estimates (Liu et al., 2012a, 2012b). P loading coefficient values can range considerably even within a single climate and geologic setting (Table 1), evidence which suggests an assessment of the variability in land use dependent P loading coefficients is necessary. Some authors have suggested that a more nuanced consideration of land use may account for nutrient loading coefficient variability within a single region such as considering impervious connectivity, unique management practices, and road cover (Goonetilleke et al., 2005; Liu et al., 2012a). Initially this study endeavored to determine the relative dominance of land use as a control on P loading in the CRW by comparing P loading coefficients based on CRW monitoring data to those reported in the literature and evaluating the consistency of annual P loads estimated from land use-based coefficients and field measurements.

Methods

Seven locations within the CRW were monitored from July 2016 through May 2017. Four sites were located on the mainstem of the Charles River and corresponded to locations of active U.S. Geological Survey (USGS) gaging stations (Figure A1). These locations drain a wide range of land cover categories (Table A1), are located entirely within the middle and upper CRW, which contributes 72% of the TP load to the lower CRW (MassDEP, 2007), and are sites at which discharge is continuously monitored to enable development of CQ relationships and field-based estimation of P loading. The other three sampling sites are located within small sub-watersheds with 70-92% forested or urban land cover in order to derive TP loading coefficients with field data (Table A2; Figure A1). Noanet (NT) and Laundry Brook (LB) are located within the middle CRW, whereas Muddy River (MR) is located in the lower watershed where urbanization is most dense (Figure A1). Existing discharge monitoring was not available at the non-mainstem sites, discharge was determined from velocity profiles collected with a Marsh-McBirney flow meter during routine sampling visits. Pressure transducers (HOBO Water Level U20L Series) installed during summer 2016 monitored water level at 15-minute intervals. Rating curves were prepared using water level and measured instantaneous flow. The goal was to use the rating curve to develop a daily discharge record.

Monthly sampling at all 7 sites occurred to capture seasonal variability of TP concentrations and allow for the calculation of annual TP export. Event samples were taken during Events in January and March 2017 at two sites, Laundry Brook and Waltham, to capture high flow concentrations.

Sampling at each location occurred in the middle of the channel at approximately 15 centimeters below the surface. For sites that could not be waded, a basket sampler was used from a bridge or a pole sampler was used from the bank. Polypropylene sample bottles with a 125-mL capacity were rinsed three times with sample prior to collection. For TP measurement, the sample was unfiltered and acidified to a pH less than 2 with concentrated sulfuric acid. Samples intended for the measurement of PO_4 were filtered in the field with a 0.45µm syringe filter. A small 15 mL polyproplene sample bottle was also filled with unfiltered water and sent to the University of Texas at Austin for major and trace elements analysis. After collection, all sample bottles were covered with

Parafilm and stored on ice. In July and August 2016, I analyzed samples at Boston College. From August 2016 forward samples intended for the analysis of TP and PO_4 were delivered to Alpha Analytical Laboratories in Westborough, MA on the day of collection. Alpha Analytical processed all samples within 1 week of delivery and returned results, including laboratory quality assurance samples via email.

Analyses at Boston College

The environmental laboratory in Devlin Hall at Boston College was not equipped to analyze water samples for TP and PO_4 when this project began. I established laboratory methods based on the Standard Methods for the Examination of Water and Wastewater (APHA, 2005).

Measurement of TP and PO_4 samples occurred in an Evolution 201 spectrophotometer (purchased for this project) equipped with a 50 mm cell at a wavelength of 880 nm. PO_4 samples were processed within 24 hours of collection by treatment with sulfuric acid, potassium antimonyl tartrate, and ammonium molybdate. The primary reaction is the formation of phosphoantimonylmolybdenum blue complex (PAMB), the intensity of the blue color corresponds to the concentration of PO_4 in sample. For TP concentrations, samples first underwent persulfate digestion to convert all organic form of P to the phosphate anion, PO_4^{3-}. After digestion, samples were treated with the same colorimetry methods. Based on data collected by the Charles River Watershed Assocation (CRWA) from 1996-2014, I expected TP samples to be in the range of 21-115 ppb PO_4-P and PO_4 samples to be between 1 and 48 ppb PO_4-P.

Analytical challenges in the Boston College laboratory prompted the use of the commercial lab, Alpha Analytical. Data from samples collected between March and July

2016 were discarded. More than half of the samples for PO_4 were below the 10 ppb detection limit. I attempted a new method to lower the detection limit (Asaoka et al., 2015), but it did not yield better results. August 2016 TP replicate analyses were performed at both Alpha Analytical and Boston College with relative percent differences (RPD) values on all samples <20% (Table A3). Due to the low concentrations and laboratory difficulties, after August 2016 Alpha Analytical was used exclusively. Data from Boston College is not included in the main text of the thesis or elsewhere in this appendix.

Results and Discussion

Concentration

Table A4 summarizes TP and PO_4 concentrations at each site and Table A5 includes all results from the monitoring program. In March 2017, preliminary analyses of the concentration data revealed that they differed significantly from historical data (1996-2015) that was collected by the CRWA and analyzed at the Massachusetts Water Resources Authority (MWRA) Deer Island laboratory (Figures B2B and B3). The MWRA laboratory is not a commercial lab, so I could not submit samples to that lab. However, I accompanied the CRWA sampling volunteers to site 400S in Dover, Massachusetts, located approximately 5.5 river kilometers upstream of the Dover site in this study (Figure A1). The volunteer and I collected concurrent replicates, one of which was delivered to Alpha Analytical. The results of this comparison (Table B2) indicate that the two labs do not produce replicate results for TP concentration within 20% of one another. Additional analyses are required for conclusive results and in the future standard reference materials should be sent to both labs, as opposed to environmental samples.

However, this exercise combined with the dramatically low annual P load estimate produced by direct computation of results from this study was enough to cast doubt on the validity of the entire Alpha Analytical dataset. The two laboratories were not interested in exploring the discrepancy further.

Discharge

At NT, LB, and MR rating curves were developed to estimate daily discharge. However, the instantaneous discharge measurements and water level data did not produce a monotonic relationship at any of the sites. There are several reasons the rating curves had scattered data and poor correlations. The pressure transducers moved during their deployment. Flows at LB and MR, particularly during events, were likely great enough to move the block and attached transducer. Because the transducers were not surveyed at installation, the exact amount of movement was not quantified. Flow at NT was often too low to measure with the flow meter and it was dry during summer months; of 10 attempts to measure flow at NT, only 4 resulted in actual discharge measurements. Too few data points were collected for this site to establish a rating curve. Of 15 attempts at LB to measure discharge, 13 resulted in reasonable values; however, this site experienced the most pressure transducer displacement. At MR, 7 of 10 attempts to measure flow resulted in reasonable values. This site experienced both pressure transducer and staff gage displacement. Overall, the lack of more permanent installations resulted in very poor rating curves. Thus it was determined that discharge could not be estimated at non mainstem sites for the study.

REFERENCES

APHA, American Public Health Association, 2005, Standard Methods for Examination of Water and Wastewater: APHA, Washington, D.C., 21st edition, p. 4-146- 4-160.

Asaoka, S., Kiso, Y., Nagai, M., and Okamura, H., 2015, A membrane extraction method for trace level phosphate analysis: Analytical Methods, v. 7, p. 9268-9273.

Bannerman, R., Konrad, J., Becker, D., Simsiman, G.V., Chesters, G., Goodrich-Mahoney, J., and Abrams, B., 1979, The IJC Menomonee River Watershed Study — Surface Water Monitoring Data: EPA-905/4-79-029. U.S. Environmental Protection Agency, Chicago, IL.

Beaulac, M., and Reckhow, K., 1982, An Examination of Land-use - Nutrient Export Relationships: Water Resources Bulletin, v. 18, p. 1013-1024.

CRWA, Charles River Watershed Association, 2011, Total Maximum Daily Load for Nutrients in the Upper/Middle Charles River, Massachusetts: Prepared for the Massachusetts Department of Environmental Protection, CN 272.0, 103 p. plus appendixes, accessed January 29, 2018, at http://www.mass.gov/eea/agencies/massdep/water/watersheds/total-maximum-daily-loads-tmdls.html#9.

Duan, S., Kaushal, S.S., Groffman, P.M., Band, L.E., and Belt, K.T., 2012 Phosphorus export across an urban to rural gradient in the Chesapeake Bay watershed: Journal of Geophysical Research: Biogeosciences, v. 117, p. n/a-n/a, doi: 10.1029/2011JG001782.

Goonetilleke, A., Thomas, Ev., Ginn, S., and Gilbert, D., 2005, Understanding the role of land use in urban stormwater quality management: Journal of Env. Management, v. 74, p. 31-42, doi:10.1016/j.jenvman.2004.08.006.

Janke, B., Finlay, J., Hobbie, S., Baker, L., Sterner, R., Nidzgorski, D., and Wilson, B., 2014, Contrasting influences of stormflow and baseflow pathways on nitrogen and phosphorus export from an urban watershed: Biogeochemistry, v. 121, p. 209-228, doi: 10.1007/s10533-013-9926-1.

Liu, A., Goonetilleke, A., and Egodawatta, P., 2012a, Inadequacy of land use and impervious area fraction for determining urban stormwater quality: Water Resource Management, v. 26, p. 2259-2265, doi:10.1007/s11269-012-0014-4.

Liu, A., Goonetilleke, A., and Egodawatta, P., 2012b, Inherent errors in pollutant build-up estimation in considering urban land use as a lumped parameter: Journal of Environmental Quality, v. 41, p, 1690-1694, doi:10.2134/jeq2011.0419.

Long, T., Wellen, C., Arhonditsis, G., and Boyd, D., 2014, Evaluation of stormwater and snowmelt inputs, land use and seasonality on nutrient dynamics in the watersheds of Hamilton Harbour, Ontario, Canada: Journal of Great Lakes Research, v. 40, p. 964-979, doi: 10.1016/j.jglr.2014.09.017.

MassDEP, Massachusetts Department of Environmental Protection and United States Environmental Protection Agency, New England Region, 2007, Total Maximum Daily Load for Nutrients In the Lower Charles River Basin, Massachusetts, CN 301.0, http://www.mass.gov/eea/docs/dep/water/resources/a-thru-m/charlesp.pdf.

Norvell, W.A., Frink, C.R., and Hill, D.E., 1979, Phosphorus in Connecticut lakes predicted by land use: Proceedings of the National Academy of Sciences of the United States of America, v. 76, p. 5426-5429.

Petrone, K.C., 2010, Catchment export of carbon, nitrogen, and phosphorus across an agro-urban land use gradient, Swan-Canning River system, southwestern Australia: Journal of Geophysical Research.Biogeosciences, v. 115, p. n/a, doi: http://dx.doi.org.proxy.bc.edu/10.1029/2009JG001051.

Valtanen, M., Sillanpää, N., and Setälä, H., 2014, The Effects of Urbanization on Runoff Pollutant Concentrations, Loadings and Their Seasonal Patterns Under Cold Climate: Water, Air, & Soil Pollution, v. 225, p. 1-16, doi: 10.1007/s11270-014-1977-y.

Winter, J.G., and Duthie, H.C., 2000, Export coefficient modeling to assess phosphorus loading in an urban watershed: Journal of the American Water Resources Association, v. 36, n. 5, p. 1053-1061.

TABLES

Table A1. Land use characterization of the mainstem sub-watersheds. Values represent only that area which lies between one station and the next, not cumulative watershed area and characteristics.

Identification	*Watershed Area (km²)*	*% Forest*	*% Low Density Residential*	*% High Density and Multi-family Residential*	*% Medium Density Residential*	*% Other*
Waltham	104	46	12	6	10	27
Wellesley	73	32	8	20	12	27
Dover	306	53	14	2	8	22
Medway	168	52	12	4	10	22

Table A2. Land use characterization of single land use sub-watershed sampling locations.

Identification	*Watershed Area (km²)*	*% Forest*	*% Developed*	*% Other*
Muddy River	14.5	7.2	88.3	4.5
Laundry Brook	12.0	16.0	70.6	13.4
Noanet	1.6	91.8	0.8	7.3

Table A3. Analytical results for August 8, 2016, BC is Boston College, and AA is Alpha Analytical. ND = not determined due to below minimum reporting level (BRL). The NT samples was not analyzed at BC; therefore, the results from AA are not included here.

	BC	*AA*	*RPD*	*BC*	*AA*	*RPD*
	TP (mg/L)	TP (mg/L)	%	PO4 (mg/L)	PO4 (mg/L)	%
Muddy River	0.144	0.168	4	BRL	0.008	ND
Laundry Brook	0.059	0.075	6	0.041	0.051	5
Waltham	0.019	0.023	5	BRL	0.006	ND
Wellesley	0.020	0.032	11	BRL	0.008	ND
Dover	0.034	0.029	4	0.019	0.013	9
Medway	0.027	0.022	5	BRL	0.017	ND

Table A4. Summary statistics for TP and PO_4 concentrations analyzed at Alpha Analytical between August 2016 and May 2017. BRL means the analytical value was below the minimum reporting level which is 0.01 mg/L for TP and 0.005 mg/L for PO_4. Event samples were collected during a storm event when the hydrograph was either rising, at peak flow, or returning to base level.

	Medway		*Dover*		*Wellesley*		*Waltham*		*Noanet*		*Laundry Brook*		*Muddy River*	
	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)	TP (mg/L)	PO_4 (mg/L)
Average	0.019	0.011	0.017	0.009	0.026	0.007	0.016	0.009	0.028	<0.005	0.101	0.035	0.102	0.011
Median	0.016	0.009	0.017	0.007	0.026	0.007	0.015	0.006	<0.01	<0.005	0.080	0.028	0.096	0.011
Standard Deviation	0.012	0.004	0.011	0.005	0.015	0.004	0.008	0.007	0.043	0.002	0.082	0.021	0.043	0.004
Minimum	<0.01	0.005	<0.01	<0.005	<0.01	<0.005	<0.01	<0.005	<0.01	<0.005	0.017	0.006	0.045	0.005
Maximum	0.044	0.018	0.042	0.021	0.052	0.019	0.043	0.044	0.138	0.008	0.378	0.092	0.177	0.019
Total Number Samples	9	10	11	11	11	11	40	40	8	9	19	19	10	10
Number of Event Samples	0	0	0	0	0	0	22	22	0	0	9	9	0	0
Results BRL	2	0	4	1	2	1	6	7	5	4	0	0	0	0

Table A5. Analytical results for the monitoring program undertaken in this study in CRW August 2016 through May 2017 where BRL means below minimum reporting limit (for TP = 0.01 mg/L and PO_4 0.005 mg/L). The sample regime indicates whether the sample is associated with routine monthly sampling, weekly Waltham sampling, event sampling, or environmental replicate analyses.

Site ID	*Timestamp*	*Total P (mg/L)*	*PO4-P (mg/L)*	*Sample Regime*
Muddy River	8/8/16 9:15	0.168	0.008	Routine
Laundry Brook	8/8/16 10:55	0.075	0.051	Routine
Waltham	8/8/16 12:15	0.01	0.006	Routine
Wellesley	8/8/16 13:30	0.023	0.008	Routine
Noanet	8/8/16 15:00	0.032	0.006	Routine
Dover	8/8/16 17:15	0.022	0.013	Routine
Medway	8/8/16 18:15	0.029	0.017	Routine
Muddy River	9/10/16 8:30	0.177	0.01	Routine
Laundry Brook	9/10/16 9:30	0.08	0.071	Routine
Waltham	9/10/16 10:30	0.017	0.009	Routine
Wellesley	9/10/16 11:30	0.035	0.007	Routine
Noanet	9/10/16 12:50	BRL	0.007	Routine
Dover	9/10/16 13:45	0.019	0.013	Routine
Replicate Dover	9/10/16 13:45	0.02	0.013	Replicate
Medway	9/10/16 14:45	0.016	0.013	Routine
Waltham	9/30/16 9:30	0.037	BRL	Weekly
Waltham Replicate	9/30/16 9:30	0.029	BRL	Replicate
Muddy River	10/8/16 8:00	0.125	0.01	Routine
Muddy River Replicate	10/8/16 8:00	0.125	0.009	Replicate
Laundry Brook	10/8/16 9:15	0.115	0.092	Routine
Waltham	10/8/16 10:00	0.016	0.006	Routine
Wellesley	10/8/16 10:30	0.032	0.005	Routine
Dover	10/8/16 11:30	0.017	0.006	Routine
Medway	10/8/16 13:15	0.02	0.01	Routine
Waltham	10/14/16 10:15	0.023	BRL	Weekly
Waltham	10/28/16 13:15	0.019	BRL	Weekly
Muddy River	11/12/16 8:00	0.111	0.019	Routine
Laundry Brook	11/12/16 9:00	0.077	0.06	Routine
Waltham	11/12/16 9:45	0.02	0.01	Routine
Wellesley	11/12/16 10:30	0.026	0.007	Routine
Dover	11/12/16 11:15	0.019	0.012	Routine
Noanet	11/12/16 12:15	0.029	0.008	Routine
Medway	11/12/16 13:15	0.022	0.014	Routine
Waltham	11/18/16 10:15	0.021	0.008	Weekly
Waltham	12/2/16 10:00	0.016	0.008	Weekly
Muddy River	12/10/16 8:06	0.113	0.016	Routine

Laundry Brook	12/10/16 9:30	0.043	0.027	Routine
Waltham	12/10/16 10:15	0.014	0.005	Routine
Wellesley	12/10/16 11:15	0.014	BRL	Routine
Dover	12/10/16 11:45	0.015	0.005	Routine
Noanet	12/10/16 12:45	0.138	BRL	Routine
Medway	12/10/16 13:45	0.016	0.005	Routine
Waltham	12/30/16 12:15	BRL	0.008	Weekly
Waltham	12/30/16 12:15	BRL	0.008	Replicate
Waltham 01	1/3/17 11:45	0.01	0.005	Event
Waltham 02	1/3/17 16:00	0.012	0.005	Event
Waltham 03	1/3/17 18:00	0.016	0.006	Event
Waltham 04	1/3/17 20:00	0.015	0.006	Event
Waltham 05	1/3/17 22:00	0.013	0.005	Event
Waltham 06	1/4/17 0:00	0.015	0.005	Event
Waltham 07	1/4/17 10:45	BRL	0.005	Event
Waltham 08	1/4/17 13:00	BRL	0.005	Event
Waltham 09	1/5/17 10:30	0.019	0.011	Event
Waltham 10	1/6/17 9:45	0.019	0.007	Event
Waltham Replicate 10	1/6/17 9:45	0.021	ND	Event Replicate
Waltham A	1/10/17 19:45	0.014	0.016	Event
Waltham B	1/10/17 22:45	0.015	0.017	Event
Waltham C	1/10/17 23:45	0.015	0.012	Event
Waltham D	1/11/17 7:45	0.016	0.008	Event
Waltham E	1/11/17 12:30	0.013	0.009	Event
Waltham F	1/11/17 15:45	0.015	0.008	Event
Waltham G	1/12/17 17:00	0.015	0.009	Event
Muddy River	1/21/17 8:45	0.054	0.013	Routine
Laundry Brook	1/21/17 10:00	0.037	0.028	Routine
Waltham	1/21/17 11:00	BRL	0.006	Routine
Wellesley	1/21/17 11:45	0.01	0.005	Routine
Noanet	1/21/17 12:45	BRL	BRL	Routine
Dover	1/21/17 14:30	BRL	0.009	Routine
Medway	1/21/17 15:30	BRL	0.008	Routine
Laundry Brook 01	1/23/17 16:30	0.072	0.044	Event
Laundry Brook 02	1/23/17 22:30	0.139	0.027	Event
Laundry Brook 03	1/24/17 0:30	0.249	0.024	Event
Laundry Brook 04	1/24/17 21:00	0.091	0.04	Event
Laundry Brook 05	1/25/17 14:00	0.083	0.045	Event
Muddy River	2/24/17 8:30	0.045	0.011	Routine
Laundry Brook	2/24/17 9:30	0.106	0.038	Routine
Dover	2/24/17 11:15	BRL	BRL	Routine
Noanet	2/24/17 12:15	BRL	BRL	Routine
Waltham	2/25/17 9:45	BRL	BRL	Routine

Wellesley	2/25/17 10:30	BRL	0.005	Routine
Medway	2/25/17 12:30	0.012	0.006	Routine
Muddy River	3/18/17 8:00	0.079	0.011	Routine
Muddy River Replicate	3/18/17 8:00	0.081	0.009	Replicate
Laundry Brook	3/18/17 9:30	0.017	0.008	Routine
Waltham	3/18/17 10:15	0.011	BRL	Routine
Noanet	3/18/17 12:15	BRL	BRL	Routine
Dover A	3/21/17 6:30	BRL	0.005	Laboratory QC
Wellesley	3/21/17 12:30	BRL	0.006	Routine
Dover B	3/21/17 13:30	BRL	0.005	Laboratory QC
Medway	3/21/17 14:30	BRL	0.006	Routine
Waltham A	4/4/17 8:45	0.023	0.044	Event
Laundry Brook A	4/4/17 9:30	0.119	0.006	Event
Waltham B	4/4/17 10:45	0.026	0.026	Event
Laundry Brook B	4/4/17 11:30	0.099	0.022	Event
Waltham C	4/4/17 12:45	0.024	0.005	Event
Laundry Brook C	4/4/17 17:30	0.056	0.02	Event
Waltham D	4/4/17 18:15	0.018	0.006	Event
Laundry Brook D	4/5/17 12:15	0.378	0.03	Event
Waltham E	4/5/17 13:15	0.015	0.005	Event
Muddy River	4/29/17 8:00	0.064	0.005	Routine
Laundry Brook	4/29/17 9:15	0.042	0.016	Routine
Waltham	4/29/17 10:30	0.036	0.008	Routine
Wellesley	4/29/17 11:15	0.044	0.007	Routine
Wellesley Replicate	4/29/17 11:15	0.038	0.007	Replicate
Dover	4/29/17 13:00	0.031	0.007	Routine
Noanet	4/29/17 14:00	0.005	0.002	Routine
Medway	4/29/17 16:30	ND	0.008	Routine
Muddy River	5/25/17 8:45	0.081	0.011	Routine
Laundry Brook	5/25/17 9:45	0.032	0.018	Routine
Waltham	5/25/17 10:30	0.043	0.018	Routine
Wellesley	5/25/17 11:15	0.052	0.019	Routine
Dover	5/25/17 12:15	0.042	0.021	Routine
Dover Replicate	5/25/17 12:15	0.041	0.021	Replicate
Noanet	5/25/17 13:15	BRL	0.005	Routine
Medway	5/25/17 14:30	0.044	0.018	Routine

FIGURES

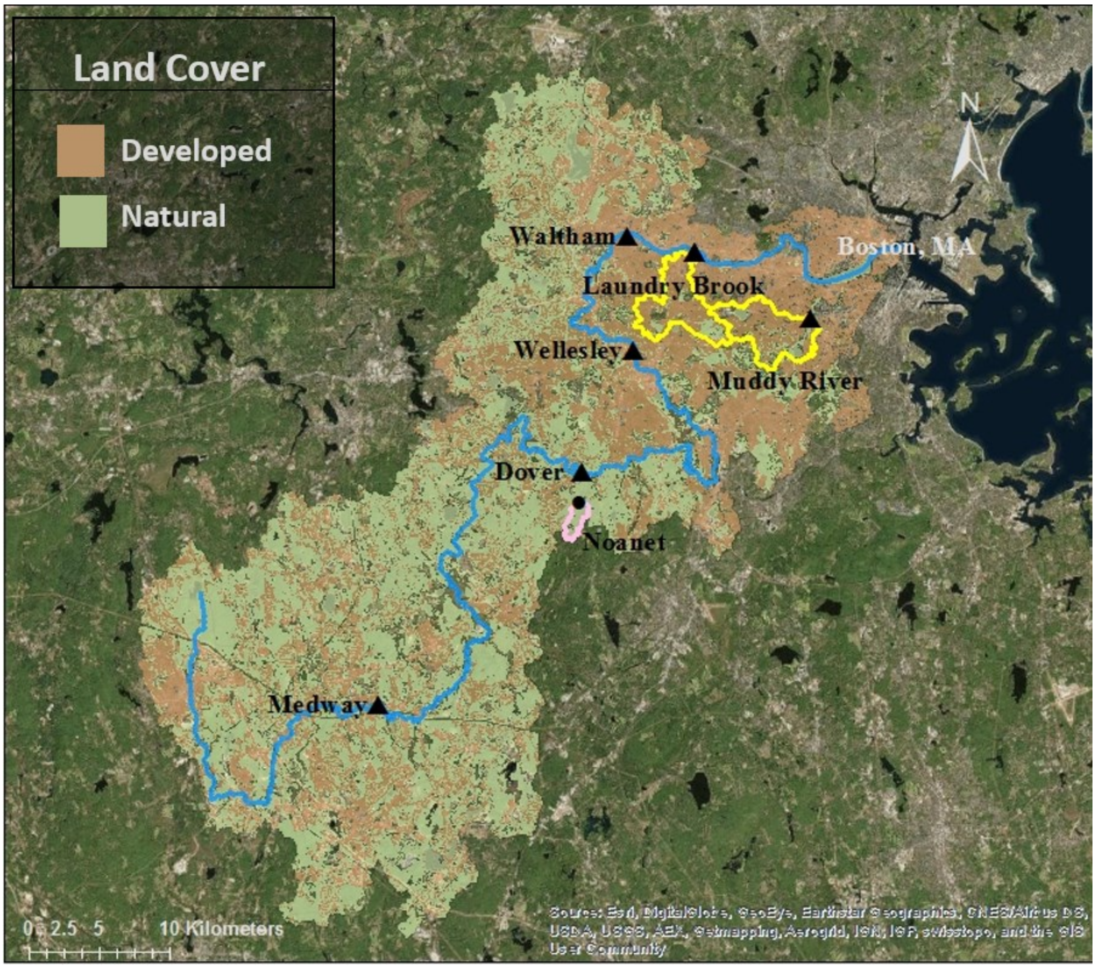


Figure A1. Map of Charles River watershed with mainstem sampling sites indicated by black triangles, single land use sampling locations indicated in yellow and pink outlines. Watertown Dam is the boundary of the lower CRW. Land cover based on bimodal categorization of land use.

APPENDIX B

TABLES

Table B1. Interlaboratory comparison results. Alpha Analytical is the commercial lab used for data collected in this study. MWRA, Deer Island is the laboratory used by the CRWA and by the MWRA. The University of Nebraska lab is used by the authors for other laboratory analyses. Laundry Brook is a tributary to the CR. It was sampled as part of the authors' sampling program beyond the scope of this study.

	TP Method	PO_4 Method	Certifications	Dover (DV), March 2017		Laundry Brook, July 2017	
				TP mg/L	PO_4 mg/L	TP mg/L	PO_4 mg/L
Alpha Analytical	AQ2 method: EPA-119-A Rev. 7, equivalent to EPA 365.1, version 2(1993) SM4500-PB, F(18-20)	Standard Methods for the Examination of Water and Wastewater. APHA-AWWA-WEF. Standard Methods Online.	NELAP	<0.01	0.005	0.049	0.021
MWRA, Deer Island Laboratory	Valderamma J.C., Marine Chemistry, 10, (1981): 109-122. Loder, T. et. al. unpublished data and communications, University of New Hampshire, 12/18/95.	EPA 365.1 Revision 2.0 (1993)	Massachusetts Department of Environmental Protection	0.020	0.005	N/A	N/A
University of Nebraska, Water Sciences Laboratory	Total P in water EPA 365.1, Seal Analytical Phosphorus-P, total, in Surface and Saline Waters and Domestic and Industrial	Soluble Phosphate AQ2 EPA 365.1 1. Seal Analytical EPA-118-A. 2. (1993) EPA 365.1 Determination of Phosphorus by Semi-Automated Colorimetry	Unknown	N/A	N/A	0.033	0.022
			Relative Percent Difference	>66	0	39	5

Tables B2-B6. LOADEST fit results for DV, WL, and WD. These analyses did not include data collected during this study. Censored calibration pairs include a TP concentration that was reported below the minimum reporting limit.

B2.

Dover 1970-1984							
Number of CQ calibration pairs			100				
Number of censored calibration pairs			0				
AIC			0.788				
Model Selected (9)			$LnC = a_0 + a_1 LnQ + a_2 LnQ^2 + a_3 \sin(2\pi dtime) + a_4 \cos(2\pi dtime) + a_5\, dtime + a_6 dtime^2$				
	a_0	a_1	a_2	a_3	a_4	a_5	a_6
Coefficient	4.7137	0.6015	-0.1948	-0.2001	-0.0393	-0.0142	-0.0061
Standard Deviation	0.0550	0.0498	0.0344	0.0646	0.0511	0.0116	0.0026
AMLE Regression Statistics							
R- Squared (%)			72.76				
Residual Variance			0.1200				
Nash Sutcliffe Efficiency Index (NSE)			0.587				

B3.

Dover 1985-1996					
Number of CQ calibration pairs			38		
Number of censored calibration pairs			1		
AIC			1.925		
Model Selected (7)			$LnC = a_0 + a_1 LnQ + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4\, dtime$		
	a_0	a_1	a_2	a_3	a_4
Coefficient	3.1401	0.8432	0.1793	-0.4675	-0.01466
Standard Deviation	0.1084	0.01204	0.1373	0.1688	0.0339
AMLE Regression Statistics					
R- Squared (%)			66.89		
Residual Variance			0.3334		
Nash Sutcliffe Efficiency Index (NSE)			0.599		

B4.

Waltham (CRWA data)					
Number of CQ calibration pairs	78				
Number of censored calibration pairs	3				
AIC	0.099				
Model Selected (7)	$LnC = a_0 + a_1 LnQ + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4\, dtime$				
	a_0	a_1	a_2	a_3	a_4
Coefficient	3.0	1.1037	-0.3812	-0.0191	-0.0110
Standard Deviation	0.0333	0.344	0.0513	0.0421	0.0052
AMLE Regression Statistics					
R- Squared (%)	95.41				
Residual Variance	0.0598				
Nash Sutcliffe Efficiency Index (NSE)	0.890				

B5.

Waltham (This Study data)					
Number of CQ calibration pairs	26				
Number of censored calibration pairs	3				
AIC	0.859				
Model Selected (7)	$LnC = a_0 + a_1 LnQ + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4\, dtime$				
	a_0	a_1	a_2	a_3	a_4
Coefficient	1.1456	0.9843	-0.9267	-0.2967	3.4040
Standard Deviation	0.0873	0.1246	0.2492	0.1556	0.9513
AMLE Regression Statistics					
R- Squared (%)	96.41				
Residual Variance	0.0927				
Nash Sutcliffe Efficiency Index (NSE)	0.918				

B6.

Watertown Dam					
Number of CQ calibration pairs			247		
Number of censored calibration pairs			0		
AIC			0.917		
Model Selected (7)			$LnC = a_0 + a_1 LnQ + a_2 \sin(2\pi dtime) + a_3 \cos(2\pi dtime) + a_4\, dtime$		
	a_0	a_1	a_2	a_3	a_4
Coefficient	3.9894	1.0581	0.1002	-0.1122	0.0293
Standard Deviation	0.0253	0.0443	0.0347	0.0439	0.0179
AMLE Regression Statistics					
R- Squared (%)			78.99		
Residual Variance			0.1435		
Nash Sutcliffe Efficiency Index (NSE)			0.439		

FIGURES

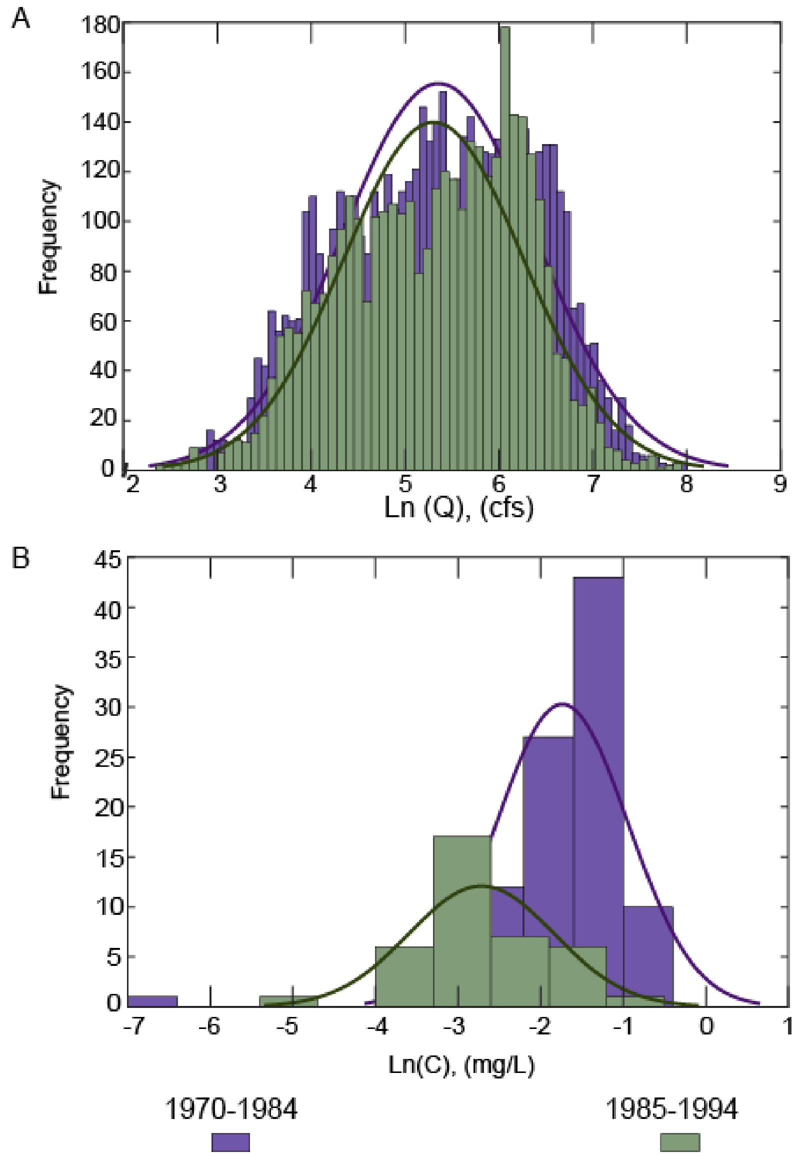


Figure B1. Log transformed frequency distributions of discharge (A) and concentration (B) at DV from 1970-1994, solid lines are normal fit curves. Two tail Student's T-Test indicates that the 1970-1984 period is significantly different from 1985-1994 with respect to both C and Q ($p < 0.05$ and $p < 0.01$ respectively).

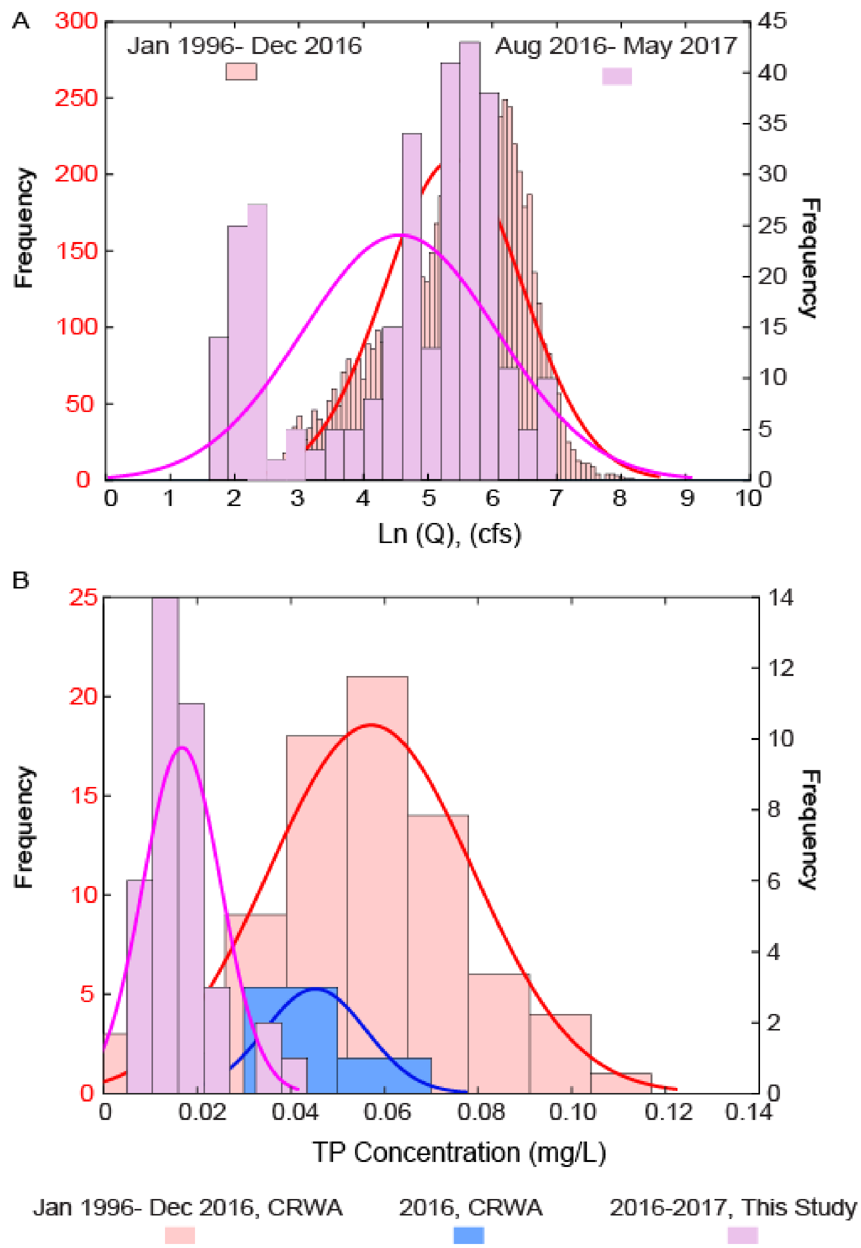


Figure B2. A) Log transformed discharge at WL from 1996-2016 compared with the period of sample collection in this study, August 2016 –May 2017, which are significantly different ($p < 0.001$). B) TP concentration data collected by the CRWA quarterly from 1996-2016 and TP concentrations collected during this study at WL. 2016 CRWA and full record CRWA are not significantly different ($p > 0.05$) but both periods are significantly different than data from this study ($p < 0.001$).

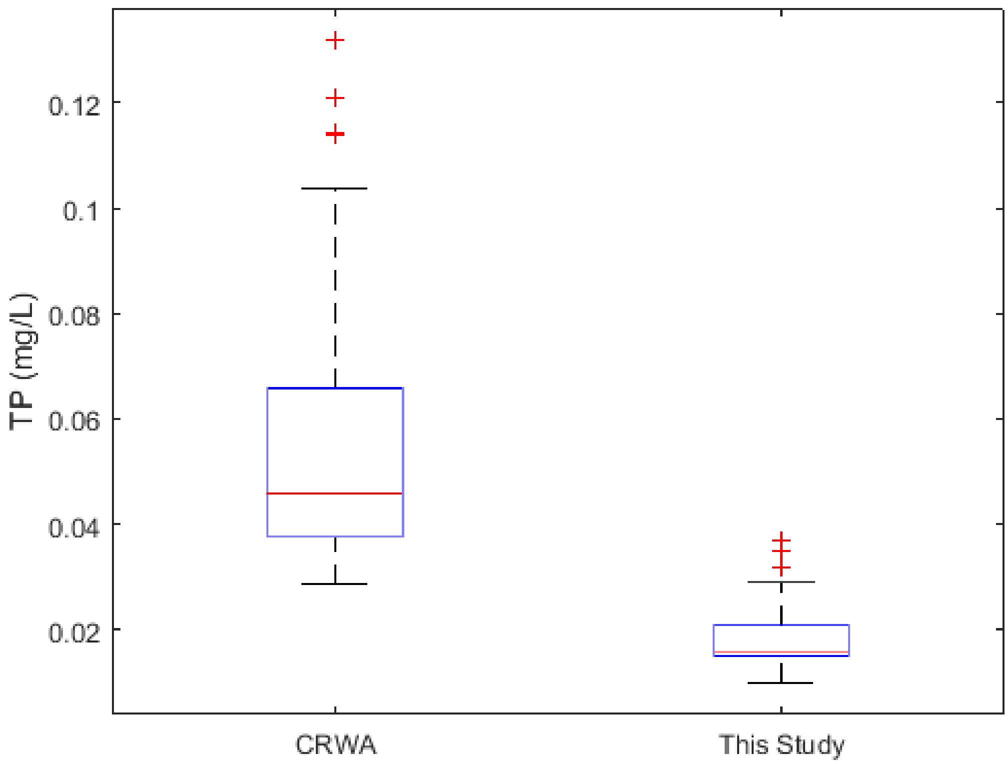


Figure B3. Box-and-whisker plot showing all TP data collected by CRWA and in this study on the mainstem of the Charles River between 2016 and 2017, CRWA n=56, this study n=52.

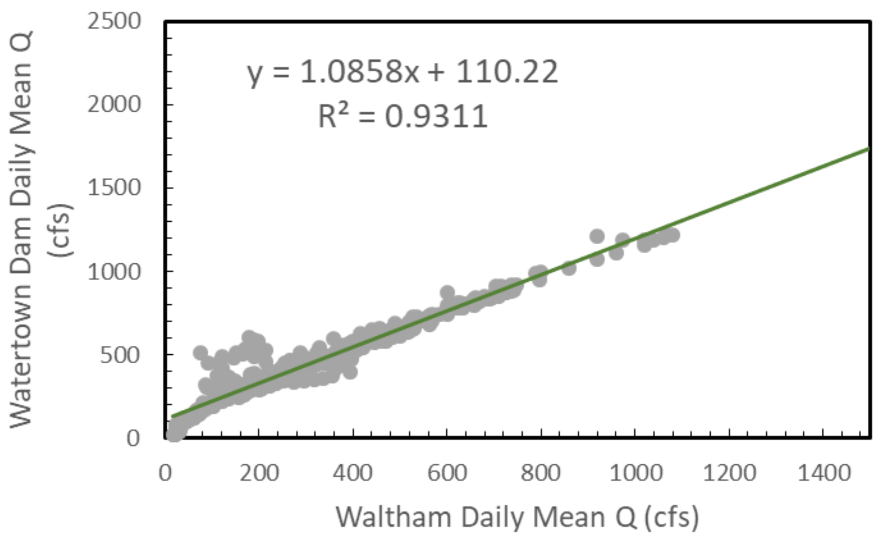


Figure B4. Linear regression of monitored daily mean *Q* at WL and WD for the overlapping period of record, 1999-2001. The relationship was used to estimate the missing WD *Q* data since the WL record includes 1998-2002.