Draft submitted to Journal of the American Water Resources Association Sept 2020
Not for distribution

Coupling PCSWMM and WASP to evaluate green stormwater infrastructure impacts to storm sediment loads in an urban watershed

Sarah Beganskas, Robert J. Ryan, Evelyn Walters, Manahel Soro,
Elizabeth Cushman, and Laura Toran

Department of Earth and Environmental Science (Beganskas, Toran, Cushman), Department of Civil and Environmental Engineering (Ryan, Walters), and Ambler Campus (Soro), Temple University, Philadelphia, Pennsylvania, USA
(Correspondence to Beganskas: sarah.beganskas@temple.edu)

ABSTRACT: We coupled rainfall–runoff and in-stream water quality models to evaluate total suspended solids (TSS) in Wissahickon Creek, a mid-sized urban stream near Philadelphia, PA. Using stormwater runoff and in-stream field data, we calibrated the model at a sub-daily scale and focused on storm responses. We demonstrate that treating event mean concentrations as a calibration parameter rather than a fixed input can substantially improve model performance. Urban stormwater TSS concentrations vary widely in time and space and are difficult to represent simply. Suspended and deposited sediment pose independent stressors to stream biota and model results suggest that both currently impair stream health in Wissahickon Creek. Retrofitting existing detention basins to prioritize infiltration reduced in-stream TSS loads by 20%, suggesting that infiltration mitigates sediment more effectively than detention. Infiltrating stormwater from 30% of the watershed reduced in-stream TSS loads by 47% and cut the frequency of TSS exceeding 100 mg/L by half. Settled loads and the frequency of high TSS values were reduced by a smaller fraction than suspended loads and duration at high TSS values. A widely-distributed network of infiltration-focused projects is an effective stormwater management strategy to mitigate sediment stress. Coupling rainfall–runoff and water quality models is an important way to integrate watershed-wide impacts and evaluate how management directly affects urban stream health.

RESEARCH IMPACT STATEMENT: This study models urban runoff and in-stream sediment loads, evaluates how well potential stormwater control measures ameliorate sediment stress, and considers implications for stream health.

KEYWORDS: green stormwater infrastructure; urban areas; total suspended solids (TSS) loading; sediment transport; rainfall–runoff modeling; in-stream water quality modeling; stormwater management; Wissahickon Creek; WASP; PCSWMM

INTRODUCTION

Urbanization results in many stressors to stream ecosystems, increasing flood intensity, flood frequency, sediment loads, bank scouring, and stream turbidity (Blaszczak *et al.*, 2019; Walsh *et al.*, 2005). Urban streams have greater sediment loads than forested streams due to expanded sources of mobile material (e.g., construction activities) and increased erosion during higher flows (Taylor and Owens, 2009). High total suspended solids (TSS) levels and settling rates both severely impact stream ecosystems. Elevated TSS concentrations are associated with reduced primary productivity (Bilotta and Brazier, 2008), decreased fish growth and survival rates (Kemp *et al.*, 2011), and low diversity in benthic macroinvertebrates (Buendia *et al.*, 2013). Excess settling (burial) reduces habitat availability and dissolved oxygen exchange within the streambed (Burdon *et al.*, 2013; Izagirre *et al.*, 2009). Furthermore, higher TSS concentrations are often associated with higher levels of metals and other pollutants (Berndtsson, 2014).

Storms play a significant role in degrading urban stream health and delivering sediment to streams (Lawler *et al.*, 2006). For example, Clinton et al. (2006) observed a stream that flowed through an urban area and then a forest; at baseflow, the forested segment showed improved water quality compared to the urban area, but during storms the forested segment was equally as impaired as the urban area. Storm size is a significant explanatory variable for predicting TSS in urban watersheds (Fisher *et al.*, 2016).

Green stormwater infrastructure (GSI), also commonly referred to as low-impact development (LID), stormwater control measures (SCMs) or best management practices (BMPs) (Fletcher *et al.*, 2015), refers to decentralized engineering approaches to reduce stormwater runoff generation during storms that incorporate soils and vegetation; these “green” approaches may be more cost-effective than traditional grey infrastructure approaches (Vineyard *et al.*,

2015). Detention-focused GSI, including wetlands and detention basins, focus on temporarily storing stormwater (Fletcher *et al.*, 2013). Infiltration-focused GSI aims to infiltrate as much stormwater as possible through swales, infiltration basins, bioretention systems, or permeable pavements (Eckart *et al.*, 2017), while harvest-focused GSI collects stormwater and re-uses it for other purposes, such as irrigation or toilet-flushing (Askarizadeh *et al.*, 2015; Shuster *et al.*, 2013). Both infiltration- and harvest-focused GSI ultimately re-route stormwater and reduce the total volume that ends up in streams (Jefferson *et al.*, 2017), while detention-focused GSI may be less effective at reducing flooding during storms (Emerson *et al.*, 2005; Fletcher *et al.*, 2013).

At a project scale, GSI tends to be effective at reducing stormwater sediment loads, with removal rates commonly in the 50–100% range (Ahiablame *et al.*, 2012); removal has been reported consistently for large and small storms (Wadzuk *et al.*, 2010). Variability in performance can be attributed to the specific project type, size, location, design, and maintenance. Modeling can help optimize GSI project design (Binns *et al.*, 2019; Yang and Chui, 2018) and configuration (Bahaya *et al.*, 2019; Sparkman *et al.*, 2017) to maximize pollutant load reduction.

There is a need for larger-scale studies to address integrated watershed-wide GSI impacts and incorporate a broader range of GSI (Prudencio and Null, 2018). GSI is decentralized by design and most GSI models focus on a small number of projects and only a single type of GSI. In a recent review of 100 watershed-scale GSI studies, a third focused on watersheds ≤ 1 km^2 and only 10 incorporated larger watersheds ≥ 100 km^2 (Jefferson *et al.*, 2017).

There is also a need for watershed-scale GSI studies to incorporate stream health impacts more directly. The overwhelming majority of studies focus on the impact of GSI on flow and pollutant loading to streams without considering in-stream processing. Some GSI studies couple

rainfall–runoff models with groundwater models to evaluate impacts on recharge rates (Barron *et al.*, 2013; Locatelli *et al.*, 2017) and some couple rainfall–runoff models with hydraulic stream channel models to evaluate downstream flooding risks (Giacomoni *et al.*, 2014; Schubert *et al.*, 2017). However, comparatively few couple rainfall–runoff models with in-stream water quality models, which is necessary to quantify impacts on in-stream processes, stream biota, and downstream pollutant loads.

In this study, we evaluate TSS in a mid-sized urban watershed (165 km^2) by coupling a rainfall–runoff model (PCSWMM) with an in-stream water quality model (WASP). We address three primary questions: (1) How effective are land cover–based EMCs at representing stormwater TSS dynamics in a coupled watershed-scale model? (2) What are the current sediment dynamics in Wissahickon Creek and how do they impact stream health? (3) What is the integrated impact of expanding GSI throughout the watershed on TSS loading to the stream, in-stream suspended sediment concentrations, and settled loads within the stream?

METHODS

Study Area: Wissahickon Creek

Wissahickon Creek is an urban stream that discharges into the Schuylkill River in Philadelphia, Pennsylvania, USA. The creek runs for 43.5 km through Montgomery and Philadelphia Counties (*Figure 1*); including its tributaries, the watershed has a stream length of 185 km and drains 165 km^2 (PWD, 2007). 52% of land area in the watershed is developed, including structures, roads, and lawns (*Figure 1*). Wissahickon Creek and its tributaries are classified as impaired (PADEP, 2018) and are characterized by low species diversity and tolerant

benthic macroinvertebrate and fish species (PWD, 2007). Four wastewater treatment plants (WWTPs) contribute a considerable fraction of baseflow discharge.

Stormwater Collection and Analysis

Stormwater runoff samples were collected from six locations: two locations in the headwaters of Sandy Run, the largest tributary to Wissahickon Creek, and two locations in each of two nearby headwater streams in adjacent watersheds (*Figure 1*). The adjacent headwaters have similar land cover distribution to the Wissahickon watershed. Based on accessibility, each location had 1–3 sites where overland stormwater runoff was directly sampled using Thermo Scientific NALGENE® Storm Water Sampler bottles. The high-density polyethylene (HDPE), 1-L bottles are designed with a debris-shedding dome and floating ball valve to maintain sample integrity. The day before an anticipated storm, we placed bottles in storm grates or buried them underground in a secure housing for passive sample collection. Following a storm, all sample bottles were collected within 48 hr and inspected for physical irregularities. Across 13 storms from May to October 2018, 65 samples were collected (*Table S1*).

TSS filtration was completed for 400–850 mL of each sample. The samples were filtered through dried and pre-weighed 0.45-µm filter paper under vacuum pressure. The filtered sediment samples were dried in an oven for 48 hr at 30°C. After drying, the filtered samples were placed in a desiccator for 20 min to cool and then weighed. TSS concentration in mg/L was calculated by dividing the difference in pre- and post-sample weight by the volume of filtered sample.

A catchment area delineation for each sample site was created in ArcMap using flow accumulation rasters. To ensure accuracy, in some cases the catchments were adjusted manually

based on field observations of surface flow during storms. The catchment area for sample sites varied widely, from 75 m^2 to 0.32 km^2, with most falling between 1000 and 8000 m^2. 1-m land cover data (University of Vermont Spatial Analysis Laboratory, 2016) was used to calculate land cover distribution within each catchment with the following categories: barren, tree canopy, low vegetation (grass), impervious roads, other impervious surfaces, and structures.

Stream Data Collection: TSS–Turbidity Relationships

Turbidity data were obtained from the United States Geological Survey (USGS) National Water Information System (NWIS) database at two stations: Fort Washington (#01473900) and Mouth (#01474000) (*Figure 1*). During two storm events on March 2, 2018 and April 16, 2018, ISCOs were deployed at these stations to automatically sample stream water. ISCOs began collecting discrete samples when water level rose to a certain point on the rising limb. During the storm on March 2, the Fort Washington ISCO was overturned, but we collected three grab samples. TSS concentrations were measured in each sample using the laboratory procedure outlined above with 900–950 mL samples.

Linear relationships were developed between turbidity and TSS concentrations and used to convert 30-min turbidity data into 30-min TSS data (*Figure 2*). TSS–turbidity relationships are commonly applied to generate high-frequency TSS records, but these relationships can vary with location and over time (Esteves *et al.*, 2019; Gippel, 1995). A distinct relationship was developed for each location using data from two storms. While data from the March 2 storm fell into the lower range of observed values, there was enough overlap between the two storms to provide confidence in the TSS–turbidity relationship. The maximum observed linear TSS concentration at the Mouth was 195 mg/L and at Fort Washington was 416 mg/L. Calculated TSS values were

included in the final record only if they fell within the range of observed TSS values; the linear TSS–turbidity relationship can break down at high TSS concentrations (Ziegler *et al.*, 2014).

Rainfall–Runoff Model: PCSWMM

PCSWMM Version 7.1.2480 was used to model rainfall and runoff processes throughout the Wissahickon watershed and requires a number of input datasets. 15-min precipitation data were obtained at four locations (*Figure 3*): three weather stations operated by the Philadelphia Water Department (PWD) and one on Temple University's Ambler campus; the Ambler station also provided air temperature, which is required to calculate evapotranspiration. Delaware Valley Regional Planning Commission (DVRPC) data were obtained for 2015 land use (DVRPC, 2017) and impervious cover (DVRPC, 2018). Soil type data were obtained from the USDA's Web Soil Survey (USDA, 2019) and used to estimate infiltration characteristics. In developed areas, soil types "Made Land" and "Urban Soil" are common and have unknown properties; in these cases, we assumed a type 'C' soil (*Table S2*). A 0.6-m digital elevation model (DEM) for the region was developed in coordination with BAE Systems (Mt Laurel Township, NJ). The DEM was used to divide the Wissahickon watershed into 137 subcatchments and calculate the slope within each. The subcatchments ranged in size from 0.15 to 5.47 km^2, with a median area of 1.2 km^2. Flow was routed through the subcatchments using a network of 963 nodes.

Stormwater basins. As of 2014, there were 227 stormwater detention basins in the Wissahickon watershed with a total storage capacity of 4.7×10^5 m^3 (Fromuth, 2014). Of these basins, 69 were explicitly included in the model (*Figure 3*), including all 52 basins with a storage volume >2500 m^3 (basin depths range 1.5–7.62 m, with an average of 2.6 m) and 17 smaller basins with readily available location, storage, and outlet information. The other 158 smaller

basins were implicitly included in the model through the calibration process in adjustments to subcatchment curve numbers, surface roughness, routing, and/or impervious cover.

For the explicitly-modeled basins, site visits were conducted for 53 basins to confirm outlet configuration and dimensions. The other 16 basins were represented using design information from engineer's reports. The DEM was used to develop a stage–surface area relationship and calculate the drainage area of each basin. These 69 basins collect runoff from 4.3% of the watershed.

Infiltration and runoff. The Curve Number (CN) method (USDA, 1986) was selected to model infiltration and runoff generation; this method is widely applied in stormwater modeling (Shuster and Pappas, 2011). Curve numbers are defined based on land use and soil type (*Table S2*).

Event mean concentrations (EMCs) were defined to characterize TSS concentrations in runoff from each land use type in the model. EMCs from the Pennsylvania Department of Environmental Protection (PADEP) were used (PADEP, 2005); however, the PADEP land cover categories did not coincide with DVRPC land use types, which were used in PCSWMM. To rectify this, DVRPC land use types were considered as weighted combinations of PADEP land cover categories. For example, single-family detached home was represented as 60% lawn, 20% rooftop, 10% low traffic residential street, and 10% residential driveway (*Table S3*).

Flow calibration. The PCSWMM model was calibrated for flow at five locations (*Figure 3*) for March–November 2017; December, January, and February were excluded to avoid the impacts of snow, which is not recorded at local weather stations. Two calibration points coincide with USGS flow gages (Fort Washington and Mouth). At the other three locations, stage–discharge relationships were developed based on 15-min depth measurements using Onset

HOBO pressure transducers and periodic flow measurements using SONTEC River Surveyor or SONTEC ADV instruments.

Calibration parameters for each subcatchment included overland flow width, percent of flow routed from impervious to pervious land surfaces, groundwater characteristics, depression storage, Manning's roughness coefficient, and curve number (*Table S4*). The primary calibration metric was the Nash–Sutcliffe Efficiency (NSE), which can range from $-\infty$ to 1. A NSE value of 0.5 is typically considered suitable for hydrologic modeling (Moriasi *et al.*, 2007; Rosa *et al.*, 2015). NSE values for flow reported at 15-minute time steps were 0.56–0.72 (*Table 1*), indicating a well-calibrated flow model.

Scenario construction. The calibrated model represented a Baseline scenario. The Baseline scenario expanded the one year of observed data (2017) to a three-year period (2016–2018) with a wider range of conditions. Two alternative scenarios were developed to represent watershed-scale GSI expansion. The Detention to Infiltration scenario modeled the impact of retrofitting all existing detention basins to prioritize infiltration rather than storage (e.g., by raising outlet structures, diverting flow, and/or incorporating plantings). Explicitly-modeled stormwater detention basins were adjusted to infiltrate stormwater runoff according to the Green and Ampt equation. Several additional parameters needed to be defined. Saturated hydraulic conductivity was set to 0.41 in/hr, a representative value for C soils (USDA, 2019). Suction head was set to 5.56 in, porosity to 0.477, and field capacity to 0.238 (Rossman and Huber, 2016).

The 30% Infiltration scenario represented extensive expansion of GSI to collectively infiltrate runoff from 30% of the watershed. Rather than explicitly modeling specific additional GSI projects, we focused on integrating the watershed-scale impact of many projects by modeling the hydrologic outcomes of additional GSI (enhanced infiltration and reduced

stormwater runoff generation). We adjusted subcatchment curve numbers such that runoff generation during a 2-year design storm was reduced by 60% in the area treated by infiltration-based GSI (see *Table S5* for details). This approach assumed that the area treated by GSI was distributed evenly among land use and soil types present in the watershed. We also reduced the effective impervious area in each subcatchment by 30%. Other modeling studies have adjusted curve numbers to represent individual GSI projects (Damodaram *et al.*, 2010; Liu *et al.*, 2015; Wu *et al.*, 2015) or watershed-scale GSI implementation (Giacomoni *et al.*, 2014; Gu *et al.*, 2019). A benefit to applying this broader approach is that one model scenario may represent many different GSI configurations that have a similar hydrologic impact.

Stream Model: WASP

Model construction. WASP Version 8.32 (Ambrose and Wool, 2017) was used to model in-stream processes. WASP is commonly used to model a variety of water quality parameters in streams, including suspended sediment (Borah *et al.*, 2006; Caruso, 2005; Ma *et al.*, 2018), and has been coupled with rainfall–runoff models, including SWMM (Burian *et al.*, 2002; Ekdal *et al.*, 2011). Wissahickon Creek and its tributaries were divided into 144 stream segments using the U.S. Environmental Protection Agency's (EPA) BASINS Version 4.5 model (EPA, 2019). The length, average width, and slope of each segment was characterized using field-surveyed cross sections from PWD. Each segment has a corresponding streambed onto which suspended sediment can settle and from which sediment can become resuspended. Streamflow was modeled using the kinematic wave method. Four old mill dams were incorporated along the lower main stem of Wissahickon Creek.

15-min stormwater volume and TSS concentration data from associated PCSWMM nodes and subcatchments were used as input to each WASP segment. PCSWMM results cannot be directly imported into WASP. A C++ tool was used to extract PCSWMM results (PWD, 2019) and Python was used to combine the output of multiple PCSWMM subcatchments and/or nodes to a single WASP segment and format the data for WASP input. Discharge and TSS concentration records from each of the four WWTPs (*Figure 1*) were obtained from electronic Monthly Discharge Monitoring Reports submitted by each facility to PADEP. Where WWTP TSS records were not available, a constant 30 mg/L was used, which is EPA's average monthly limitation for wastewater effluent (EPA, 2010).

In-stream TSS calibration. 30-min in-stream TSS concentration records (generated using TSS–turbidity relationships) were combined with 30-min USGS flow records to calculate daily TSS concentrations and loads for calibration at Fort Washington and the Mouth (*Figure 1*). Using flow at the Mouth, storms were identified using a Python code; the start of the rising limb and end of the falling limb were noted for each storm. For each of 30 storms identified during the calibration period (4/1/17–11/10/17), the total in-stream TSS load, average in-stream TSS concentration, peak in-stream TSS concentration, time above 50 mg/L TSS, time above 100 mg/L TSS, peak streamflow, and total streamflow volume were calculated. NSE values were calculated for daily and storm TSS records. One outlier storm with TSS concentrations that exceeded the maximum observed value for a protracted time was removed from NSE calculations. Qualitatively, the shape of hourly TSS responses during storms was also used to evaluate each model run's accuracy; however, WASP output not expected to closely reproduce hourly TSS patterns. The model time scale (with 15-min PCSWMM output and 30-min observed data used in WASP) is appropriate for daily evaluation, but finer-scale data would be required to

develop an accurate hourly model. Additionally, it is possible that sub-daily timing errors were propagated between PCSWMM and WASP.

Sediment settling velocity (in WASP), sediment resuspension velocity (in WASP), and TSS EMCs (in PCSWMM) were varied for calibrating in-stream TSS within WASP. In-stream settling and resuspension velocities can be set individually for each segment in WASP. Segments were divided into three groups: Lower Main Stem (below confluence with Sandy Run), Upper Main Stem (above confluence with Sandy Run), and Tributaries. Settling and resuspension velocities were defined separately for each group to incorporate heterogeneity (*Table S6*). In-stream sediment settling velocity was varied between 0.25 and 60 m/d. Stoke's Law, a relatively simple tool for estimating a particle's settling velocity (Williams *et al.*, 2008), guided the range of values tested in calibration. According to Stoke's Law, the range in expected settling velocity for silt particles (0.002–0.05 mm diameter) is 0.3–194 m/d. Sediment resuspension velocity in ponds is on the order of 0.00005 m/d (EPA, 2017); accordingly, we varied in-stream resuspension velocity between 0.00005 and 1 m/d.

EMCs were varied in PCSWMM by multiplying the TSS EMC values for each land use type by a constant factor (*Table S3*). Four WASP models were created, each receiving PCSWMM input with a different TSS EMC. These four WASP models were then calibrated independently, adjusting sediment settling and resuspension velocities to achieve the best possible fit with that EMC. In total, over 700 WASP model runs were evaluated, with at least 100 runs for each EMC.

This TSS calibration procedure incorporated many metrics: NSE values for daily average concentration, daily total load, storm average concentration, storm peak concentration, and storm total load at Fort Washington and the Mouth. To evaluate the performance of each model run, we

qualitatively weighed these metrics according to our best judgment. Priority was given to results at the Mouth, which integrate responses from the entire watershed. TSS response during storms was also considered particularly important, as storms are critical for generating TSS (Clinton and Vose, 2006; Fisher *et al.*, 2016).

Scenarios and post-processing. The Baseline, Detention to Infiltration, and 30% Infiltration scenarios were run for March–November, 2016–2018, at a dynamic sub-hourly time step. The total stormwater volume (PCSWMM), streamflow volume (WASP), stormwater TSS load (PCSWMM), in-stream TSS load (WASP), and in-stream TSS settled load (WASP) during each calendar year were calculated for each scenario. Because the models were calibrated using observed in-stream data, we focus on streamflow and in-stream TSS loads. Settled loads are presented but are less certain.

Rain events were defined using 15-min precipitation data from four weather stations (*Figure 3*). Each event comprised at least 6.4 mm (0.25 in) rainfall and had gaps with no rainfall of <24 hr. Rain events were characterized in terms of total rainfall, rainfall duration, average rainfall intensity, and maximum rainfall intensity.

Using Python, stormflows were identified by finding peaks in modeled flow data at the Mouth (as described above); storm start and end times were defined using Baseline scenario results and the stormflow events were linked to rain events. In some cases, multiple consecutive rain events corresponded to a single multi-peak stormflow event. Modeled flow and TSS concentration data at the Mouth were used to calculate the total streamflow, total TSS load, average TSS concentration, peak TSS concentration, and duration during which TSS concentrations exceeded 50 and 100 mg/L during each storm for each scenario. *p*-values from

pairwise t-tests were used to compare results for each storm between the Baseline and GSI scenarios.

The recurrence interval of each stormflow event was characterized using a partial duration series analysis of a 29-year flow record from the Mouth. Python was used to identify all peak discharges >1.5 m^3/s. To avoid the impact of multi-peak storms, only the largest peak within any sliding 24-hour window was selected. A relationship between peak discharge and recurrence interval (inverse of exceedance frequency) was generated (see *Figure S1* for more details). Storms were divided into three categories for analysis. Large storms have peak flow >75 m^3/s and recurrence interval >100 days; medium storms have peak flow between 10 and 75 m^3/s with recurrence interval between 13 and 100 days; small storms have peak flow <10 m^3/s and recurrence interval <13 days.

STORMWATER TSS RESULTS AND EMC CALIBRATION

TSS in Stormwater Runoff

TSS concentrations in stormwater samples ranged from 11 to 8600 mg/L, with a median of 750 mg/L and an arithmetic mean of 1720 mg/L. TSS concentrations at most sites varied across different storms, although some sites had consistently higher TSS (*Figure 4*). Stormwater TSS values fall at the higher end of values reported in other urban studies (Bach *et al.*, 2010; Berndtsson, 2014; Hathaway and Hunt, 2011; Kayhanian *et al.*, 2003; Toor *et al.*, 2017).

Overall, the spatial variability in TSS concentrations did not correlate strongly with catchment land cover or size (*Figure S2*). Percent tree canopy had a weak negative relationship (r^2 = 0.4) with average stormwater TSS, such that stormwater from less-developed areas had lower TSS, which is not surprising (Taylor and Owens, 2009; Walsh *et al.*, 2005). The impact of

specific types of development (e.g., lawns vs. roads vs. buildings) on stormwater TSS concentrations was impossible to determine (Cushman, 2019); all other land cover variables had correlation coefficients close to 0.0 with average stormwater TSS. Similarly, Goonetilleke et al. (2005) found that runoff from urban catchments with similar impervious coverage had very different TSS signals and Bahaya et al. (2019) found that urban "hot spots" associated with large sediment loads had variable land cover. The lack of a clear land cover–TSS relationship could be due to varying land management methods in urban settings, varying configurations of developed surfaces (Ferreira *et al.*, 2019), and/or transient sediment sources (e.g., construction). The variability in TSS concentrations between storms at each site (*Figure 4*) demonstrates that factors unrelated to land cover also affect stormwater TSS.

Incorporating EMCs into Stream Model Calibration

EMCs are a critical model parameter, directly controlling storm sediment loading; measuring local stormwater TSS concentrations was intended to guide the selection of EMC values for stormwater modeling. Stormwater TSS (mean 1720 mg/L) was approximately an order of magnitude greater than PADEP EMCs defined for urban settings (21–195 mg/L). However, it is important to note that measured TSS values do not represent EMCs. It is common for stormwater early in a storm event to have particularly high TSS concentrations due to a first flush effect (Bach *et al.*, 2010); though, particularly in urban settings, the first flush effect for TSS is inconsistent and not always observed (Deletic, 1998; Deletic and Maksimovic, 1998; Lawler *et al.*, 2006; Todeschini *et al.*, 2019). Peak stormwater TSS concentration is generally not more than 3 times the EMC for that storm (Bannerman *et al.*, 1993; Charbeneau and Barrett, 1998; Cho and Lee, 2017; Göbel *et al.*, 2007; Peng *et al.*, 2016); thus, even if measured

stormwater TSS values are conservatively considered to be the peak TSS concentration occurring during each storm, these values still exceed the expected range defined by PADEP EMCs by a considerable margin (*Figure 4*). As a result, during calibration we chose to raise EMCs to 1, 2, 3, or 5 times the PADEP values (*Table S3*).

As expected based on field TSS data, the best model run with PADEP values (1X EMC) did not reproduce TSS well during storms: the model had NSEs of 0.41 and 0.22 for storm average TSS concentration at the Mouth and Fort Washington, respectively, and even lower NSEs (0.13 and –0.16) for storm peak TSS concentration (*Table 2*). Regardless of how WASP sedimentation parameters were tuned (varying over several orders of magnitude), the 1X EMC model simply did not generate enough sediment during storms to match observed TSS values (*Figure 5A*). Furthermore, qualitative analysis of TSS concentrations during storms showed that the 1X EMC model matched temporal patterns poorly (*Figure 5C*). This performance was deemed unacceptable, given the importance of storms for sediment transport (Clinton and Vose, 2006; Fisher *et al.*, 2016).

Increasing EMCs to double (2X EMC) or triple (3X EMC) the PADEP values substantially improved accuracy for storm and daily TSS concentrations at both locations (*Table 2*; *Figures 5B*, *S3*) and more closely matched field stormwater TSS data (*Figure 4*). The 3X EMC model modestly outperformed the 2X EMC model, particularly for storms at the Mouth. Qualitatively, the 3X EMC model better matched observed TSS patterns during storms (*Figure 5D*), though no WASP model closely matched hourly observed TSS patterns, as expected; higher-resolution input data would be required to reproduce hourly patterns in a coupled model. Further increasing EMCs to five times (5X EMC) the PADEP values resulted in poorer performance in most metrics, notably daily TSS concentrations and loads at both locations

(*Table 2*). Some storm metrics were improved in the 5X EMC model compared to the 3X EMC model, but sediment peaks were overestimated by a factor of 1.4–2.4 during smaller storm events in the 5X EMC model, indicating that too much sediment was produced. The 3X EMC model was thus selected as the most accurate.

The final calibrated model represented streamflow and TSS at the Mouth well during the calibration period on a daily time step and during storms (*Figures 6*, *S3*). All NSE values exceeded 0.5 except at Fort Washington during storms (*Tables 2*, *3*), likely due to Fort Washington having a poorer fit to stormflow, which influences load calculations. Storms at the Mouth were well-represented; this location was considered a higher priority as it integrates signals from the complete watershed and is the focus of model output analyses.

Applications for Hydrologic Modeling

This study demonstrated that varying EMCs during calibration can improve model performance (*Figures 5*, *S3; Table 2*). Incorporating generic EMCs and treating them as a static input parameter is a common modeling practice (Alamdari *et al.*, 2017; Fisher *et al.*, 2016; Steinman *et al.*, 2015). In this study, applying the PADEP EMCs would have resulted in significant underestimation of TSS loads in stormwater and incorrect representation of settling and resuspension within the stream. Collecting field data guided the range of EMCs explored in this study; field data also emphasized the variability of stormwater TSS concentrations in urban settings (*Figure 4*) and the difficulty in tying this variability to static factors such as land cover (*Figure S2*). It is challenging to develop EMCs that apply broadly. However, EMCs can be a helpful starting point for calibration, even when local field data are not available.

Calibration Limitations

The limits of observed TSS posed a constraint on model calibration, particularly for large storms. The TSS–turbidity relationship requires simultaneous high-quality TSS and turbidity data, both of which are logistically challenging to collect during intense storms. For TSS, it is often unsafe to collect grab samples during storms and ISCO auto-samplers placed near the stream may be at risk of flooding or tipping over. In-stream turbidity loggers may be blown out or clogged during high flow events. Furthermore, Wissahickon Creek has been observed to top its banks during large storms, which contributes a more complex, nonlinear relationship between TSS and turbidity.

As TSS–turbidity relationships are likely to become nonlinear at high TSS concentrations (Ziegler *et al.*, 2014), it is not reasonable to extrapolate the linear relationship beyond observed TSS values. Thus, it is impossible to calibrate the model for storms during which TSS concentrations exceed observed values for an extended time because observed TSS is essentially unknown. One storm (out of 30 identified for calibration) was not included in any NSE calculations for this reason. Furthermore, nine storms were excluded from peak TSS NSE calculations, as peak TSS concentrations exceeded the threshold. Thus, model output for the largest storms will be less accurate than for smaller storms. Modeled TSS concentrations did not exceed the maximum observed values at the Mouth or Fort Washington. While we are not as confident in specific TSS values at elevated levels, we are confident that at these times, TSS concentrations are high and potentially damaging. Elevated TSS levels below the threshold may be harmful to stream biota (Bilotta and Brazier, 2008).

SEDIMENT LOADING IN WISSAHICKON CREEK UNDER CURRENT CONDITIONS

In-Stream Results: Baseline Scenario

Calendar year 2018 brought more than 155 cm (61 in) of precipitation (rainfall + melted snow) to the Philadelphia region; this was the second-highest annual total since recordkeeping began in the early 1870s (Franklin Institute, 2017). 2016 and 2017 were relatively drier, with 2017 close to the mean annual precipitation and 2016 a little drier than average. In this study, we modeled the snow-free months of each year, March–November. As total rainfall increased from 2016 to 2018, so did modeled total streamflow volume and in-stream TSS load (*Table 3*). In 2018, there was 1.7 times more rainfall than in 2016, but the total in-stream TSS load during storms was more than 5 times greater. This demonstrates a non-linear relationship between annual rainfall amount and in-stream sediment load.

Overall, 70–88% of annual streamflow volume occurred during storms, but 91–99% of annual in-stream TSS load occurred during storms. This highlights the importance of storms for generating and transporting sediment loads. During March–November of 2016–2018, 113 stormflow events were identified in the model. The largest storm occurred during 2018 with a peak flow rate of 274 m^3/s; baseflow at the Mouth is typically <1 m^3/s. This storm has an estimated recurrence interval of 5 years, while all other storms had recurrence intervals <1.5 years. Five additional storms from 2017 and 2018 met the criteria defined for a "large" storm, with peak flow >75 m^3/s and recurrence interval >100 days.

Storms with higher peak flows tended to have higher peak TSS concentrations; the two metrics had a Spearman rank correlation coefficient of 0.86. The highest peak TSS concentration during a small storm was 105 mg/L, though 75% of small storms had peak TSS <48 mg/L (*Figure 7A*). Medium storms had the most variability in peak TSS, ranging between 41 and 180

mg/L. Overall, TSS concentrations peaked above 50 mg/L 61 times, on average >2 times per month over the modeled period, and exceeded 100 mg/L 24 times, about once a month.

TSS loads during large storms were more than an order of magnitude greater than TSS loads during small storms (*Figure 7B*). Altogether, the six large storms resulted in 39% of the TSS load that occurred over the three-year period. Medium storms brought 57% of the TSS load and small storms just 3.5%. The median large storm delivered 11.6 times as much sediment as the median medium storm, but medium storms occur 10–20 times more frequently.

Stream Health: Baseline Scenario

In-steam suspended sediment at the peak levels observed and modeled in this study (50–200 mg/L) directly harm periphyton by reducing light penetration and thus photosynthetic efficiency (Izagirre *et al.*, 2009), macroinvertebrates by interfering with drift (Béjar *et al.*, 2017), and fish by abrading gills (Kemp *et al.*, 2011), among many other impacts (Bilotta and Brazier, 2008). While stream biota may be resilient to infrequent sediment loading events during some large storms (Monk *et al.*, 2008; Ryan, 1991), the increased frequency, intensity, and duration of such events in anthropogenically-modified streams like Wissahickon Creek poses a serious threat (Bilotta and Brazier, 2008; Walsh *et al.*, 2005).

As exposure duration is an important factor determining how high TSS levels affect stream health, we used model results to estimate time above 50 mg/L TSS and time above 100 mg/L TSS at the Mouth during each storm. A validation analysis with observed storms in 2017 (*Figure S4*) shows favorable NSEs (0.59 and 0.58) for these two metrics. While TSS concentrations at the Mouth did not often exceed 50 mg/L during small storms, almost every medium and large storm resulted in TSS concentrations >50 mg/L for several hours (*Figure 7C*).

The longest durations with TSS concentration >50 mg/L did not occur during the largest storms, as was expected, but during sustained, multi-peak, medium-sized storms. Four medium-sized storms each caused TSS concentration to be >50 mg/L for longer than 28 hours, a greater duration than any large storm. TSS concentrations exceeded 100 mg/L for 4–13 hours during every large storm.

Fine-grained sediment deposition is another process that disrupts habitat for many stream biota and limits oxygen exchange for benthic macroinvertebrates and fish embryos (Conroy *et al.*, 2018; Kemp *et al.*, 2011; Larsen *et al.*, 2011). WASP reports settled sediment loads in addition to in-stream TSS loads; though we do not have data to calibrate settled loads, settling rate is a critical calibration parameter. Model results indicated that 2.4×10^7 kg of sediment settled in the stream channel over three years, while 6.1×10^6 kg flowed out of the watershed into the Schuylkill River (*Table 3*). Even with some uncertainty, the large settled sediment loads under current conditions indicate stress to stream biota. Overall, it is likely that both suspended and deposited sediment contributes to impaired stream health in Wissahickon Creek.

Implications of Baseline Scenario Results for Wissahickon Creek

Overall results from the Baseline scenario indicate a stream whose biota is likely heavily impacted by high suspended and settled sediment loads, despite implementation of a sediment Total Maximum Daily Load (TMDL) in 2003 (EPA, 2003) and hundreds of stormwater detention basins around the watershed (Fromuth, 2014). A recent report found that the watershed is impaired in terms of water quality and macroinvertebrate assemblages (EPA, 2015), which is consistent with the sediment conditions observed and modeled in this study. Model results show that the sediment load leaving the watershed was well below the sediment TMDL (3.3×10^6

kg/yr at the Mouth) in 2016 and 2017. In 2018, the TMDL was exceeded, as in-stream TSS load at the Mouth was 3.8×10^6 kg over March–November. As estimated storm recurrence intervals in this study were ≤5 years, the conditions modeled were not outside the ordinary.

In-stream TSS loading from Wissahickon Creek was five times greater in 2018 (a wet year) than in 2016 (a drier-than-average year), which has implications for future stormwater and sediment management under climate change. Despite the differences in TSS loads at the Mouth over the three years modeled, peak storm TSS concentrations remained consistently high. On average, TSS concentrations at the Mouth exceeded 100 mg/L once a month and exceeded 50 mg/L twice a month. Large storms, with substantial in-stream TSS loads, occurred more than once per year on average. In the northeastern United States, precipitation accumulation and intensity are increasing (and are expected to continue increasing) more than any other part of the country (Martinez-Villalobos and Neelin, 2018). It is likely that TSS loading will increase as precipitation accumulation and intensity increase, which will make meeting existing TMDLs more difficult. GSI may also be less effective at reducing peak and total stormflows during more intense storms (Sohn *et al.*, 2019).

SEDIMENT LOADING IN WISSAHICKON CREEK WITH ADDITIONAL GSI

In-Stream Results: GSI Scenarios

After retrofitting all detention basins to prioritize infiltration (Detention to Infiltration scenario), total in-stream TSS load at the Mouth was reduced by 20% or 5.3×10^5 kg/yr (*Table S7*). Increasing the area treated by infiltration-focused GSI from 4.3% to 30% (30% Infiltration scenario) was, as expected, more effective and reduced the total in-stream TSS load at the Mouth by 47% or 1.3×10^6 kg/yr.

The same 113 modeled storm events across 2016–2018 were evaluated for the GSI scenarios. TSS load reductions relative to the Baseline scenario varied from storm to storm (*Figures 8, S5, S6*). During the Detention to Infiltration scenario, storm TSS loads at the Mouth were 0–47% lower than the Baseline scenario, with up to 1.4×10^5 kg less sediment passing through the Mouth in a single storm. During the 30% Infiltration scenario, storm TSS loads at the Mouth were 22–83% lower than the Baseline scenario, with up to 2.3×10^5 kg less sediment in a single storm.

As GSI treatment area increased, so did the percentage of in-stream TSS load occurring during the largest six storms (*Table S8*), which had peak flow ≥75 m^3/s during the Baseline scenario. At Baseline, these large storms accounted for 39% of in-stream TSS load; these storms accounted for 43% during the 30% Infiltration scenario. This finding illustrates that the GSI reduced sediment loads more during smaller storms and is consistent with other studies (Sohn *et al.*, 2019).

Notably, peak storm TSS concentrations were not reduced much relative to the Baseline Scenario, especially considering substantial decreases in sediment loads (*Figure 8*). During the Detention to Infiltration scenario, the average storm had an 18% reduction in TSS load and 7% reduction in peak TSS concentration; 72 of 113 storms showed changes in peak TSS concentration of <5%. During the 30% Infiltration Scenario, the average storm had a 60% reduction in TSS load but just a 35% reduction in peak TSS concentration. During the largest storm on June 10, 2018, TSS concentration peaked at ~200 mg/L in all three scenarios.

Stream Health: GSI Scenarios

Model results showed that GSI can reduce both the frequency and duration of high-TSS events, though it may be more effective at reducing high-TSS duration. Time above 50 mg/L was reduced by 10% or 52 hr and 41% or 208 hr during the Detention to Infiltration and 30% Infiltration scenarios, respectively; the time spent above 100 mg/L was reduced by 21 hr (21%) and 54 hr (55%) (*Table S7*). While the high-TSS-concentration durations were reduced, high-TSS-concentration frequencies were similar during the Detention to Infiltration and Baseline scenarios: TSS concentrations exceeded 50 mg/L at the Mouth 61 times during both scenarios and exceeded 100 mg/L 24 times (Baseline) and 21 times (Detention to Infiltration). During the 30% Infiltration scenario, these frequencies were substantially reduced. TSS concentration exceeded 50 mg/L at the Mouth 39 times, about 1/3 less often. TSS concentration exceeded 100 mg/L at the Mouth 13 times, half as often as during the Baseline scenario (an average of once every two months rather than once every month).

GSI may more effectively reduce suspended rather than settled loads. During the Detention to Infiltration scenario, in-stream storm settled loads were 0–19% lower than the Baseline scenario; storm settled loads were 17–46% lower during the 30% Infiltration Scenario than the Baseline scenario. For both GSI scenarios, in-stream suspended TSS load was reduced by a larger fraction than was settled load. As a result, the fraction of sediment settling on the streambed (rather than remaining in suspension) increased with additional GSI from the Baseline scenario to the 30% Infiltration scenario.

Implications for GSI Management and Modeling

Many field and modeling studies have documented the impact of project design and configuration on GSI performance (Zhang and Chui, 2018). We applied a novel approach in this

study, indirectly modeling the impact of hundreds of GSI projects spread throughout a 160-km^2 watershed without linking the model to specific types, sizes, or locations of GSI projects. This approach addresses a need in the literature for more studies to evaluate the watershed-scale impact of a variety of GSI projects (Prudencio and Null, 2018; Zhang and Chui, 2019).

Results demonstrate that retrofitting 69 existing detention basins to prioritize infiltration had a measurable benefit in Wissahickon Creek, reducing in-stream TSS loads by 20% and suggesting that infiltration-focused GSI is more effective at reducing TSS loads than detention-focused GSI. A primary mechanism for pollutant removal by GSI projects is the physical reduction of stormwater volumes (Jefferson *et al.*, 2017); detention-focused GSI is effective at slowing stormwater down, but generally not at reducing total stormwater volume (Fletcher *et al.*, 2013).

While extensive investment and space would be required to build enough GSI to infiltrate stormwater runoff from 30% of the watershed, model results show that doing so in the Wissahickon watershed would reduce in-stream TSS loads by 47% and cut the frequency of TSS concentrations exceeding 100 mg/L at the Mouth by half. These results demonstrate the potential effectiveness of building a distributed network of infiltration-focused GSI throughout a watershed. In practice, the long-term effectiveness of GSI is largely determined by how well projects are maintained over time (Roy-Poirier *et al.*, 2010), though it is beyond the scope of this study to evaluate this effect quantitatively. Many types of GSI can be designed to promote infiltration, including permeable pavement, rain gardens, trenches, and bioinfiltration basins (Eckart *et al.*, 2017). Infiltration-focused GSI is a space-efficient choice, particularly in urban environments where limited area is available for GSI (Zhang and Chui, 2018).

Not all TSS metrics were lowered equally in the model as a result of GSI installation. Model results suggest that GSI more effectively reduces suspended loads and the time spent above 50 mg/L than settled loads, peak TSS concentrations, or the frequency of TSS exceeding 50 mg/L. The impact of suspended sediment on stream biota is related to how high TSS concentrations get, for how long, how often, and during what season (Bilotta and Brazier, 2008; Mathers *et al.*, 2017); furthermore, different species have widely varying tolerance to sediment (McKenzie *et al.*, 2020). These complexities make it difficult to characterize an "ideal" TSS regime.

The modeling techniques applied in this study allowed for in-depth evaluation of the current state of sediment delivery to and fate within Wissahickon Creek as well as the impacts of watershed-scale GSI. It was crucial to model both stormwater inputs throughout the watershed (PCSWMM) and in-stream settling and transport (WASP) together to distinguish suspended and settled loads, which pose unique threats to stream biota. High-temporal-resolution model output (15-min for PCSWMM and 45-min for WASP) required bigger files and longer run times, but ultimately allowed for detailed analysis of individual storm events, including peak TSS concentrations and the duration of high TSS levels.

SUMMARY AND CONCLUSIONS

Coupling rainfall–runoff and in-stream water quality models allowed insight into EMC calibration, impacts of sediment on stream health, and the potential of GSI to ameliorate sediment stress. The model was calibrated to evaluate daily TSS loads and sub-daily TSS concentrations during storms in Wissahickon Creek, a mid-sized urban watershed in Philadelphia, Pennsylvania, USA.

Treating EMCs as a calibration parameter rather than a known input substantially improved model performance. Urban stormwater TSS concentrations vary widely in time and space and are difficult to represent simply. Field stormwater TSS data indicated that recommended EMCs from PADEP would likely underestimate TSS loading to Wissahickon Creek; as expected, regardless of how other model parameters were tuned, in-stream TSS during storms was poorly simulated using PADEP EMCs. Field data were used to inform a range of EMCs to test in the model, and performance was best at triple the PADEP EMC values. The final calibrated model accurately represented (NSE >0.5) average TSS concentration, peak TSS concentrations, and TSS load during storms at the Mouth.

Model results suggest that both suspended and deposited sediment contribute to impaired stream health in Wissahickon Creek. On average, TSS concentrations at the Mouth exceeded 100 mg/L once a month and exceeded 50 mg/L twice a month. A recurrence interval analysis demonstrated that the modeled flow conditions were not atypical; one storm had an estimated recurrence interval of 5 years while all others were <1.5 years. While settled loads are uncalibrated, the model indicated that more sediment was deposited on the streambed than remained in suspension, indicating a potential for high burial rates. These threats may be exacerbated by climate change, which will likely increase TSS loads alongside increasing precipitation accumulation and intensity. The wettest year in this study (2018) had 1.5 times as much rainfall and 5 times the in-stream TSS load as the driest year (2016).

A distributed network of infiltration-focused GSI throughout a watershed can be an effective stormwater management strategy to mitigate sediment stress in urban streams. Retrofitting all detention basins to prioritize infiltration had a measurable benefit in Wissahickon Creek, reducing in-stream TSS and settled loads by 20% and 8%, respectively. This result

suggests that GSI promoting infiltration, rather than focusing on temporary stormwater detention, is more effective at reducing sediment loads. Building enough GSI to infiltrate stormwater runoff from 30% of the watershed would reduce in-stream TSS (47%) and settled (27%) loads and cut the frequency of high TSS concentrations at the Mouth by half. Modeling the impact of extensive, decentralized GSI without linking the model to specific projects allowed us to focus on integrating the watershed-wide impact of successful GSI rather than the efficacy of individual project designs.

The modeling techniques applied in this study allowed for in-depth evaluation of the current state of sediment in Wissahickon Creek and the potential benefits of watershed-scale GSI. While analyses focused on results from the Mouth, calibrating TSS concentrations and loads at two locations helped develop a model that could distinguish suspended and deposited loads. High-temporal-resolution model output allowed for detailed analysis of individual storm events, including peak TSS concentration and the duration of high TSS concentrations. Model results showed that not all TSS metrics were lowered equally as a result of GSI installation; for instance, in-stream TSS loads were reduced much more than settled load or peak TSS concentration. Coupling rainfall–runoff and in-stream water quality models allowed us to assess TSS loads delivered to the stream as well as the fate of that sediment in-stream and is an important way to evaluate how stormwater management directly impacts stream health.

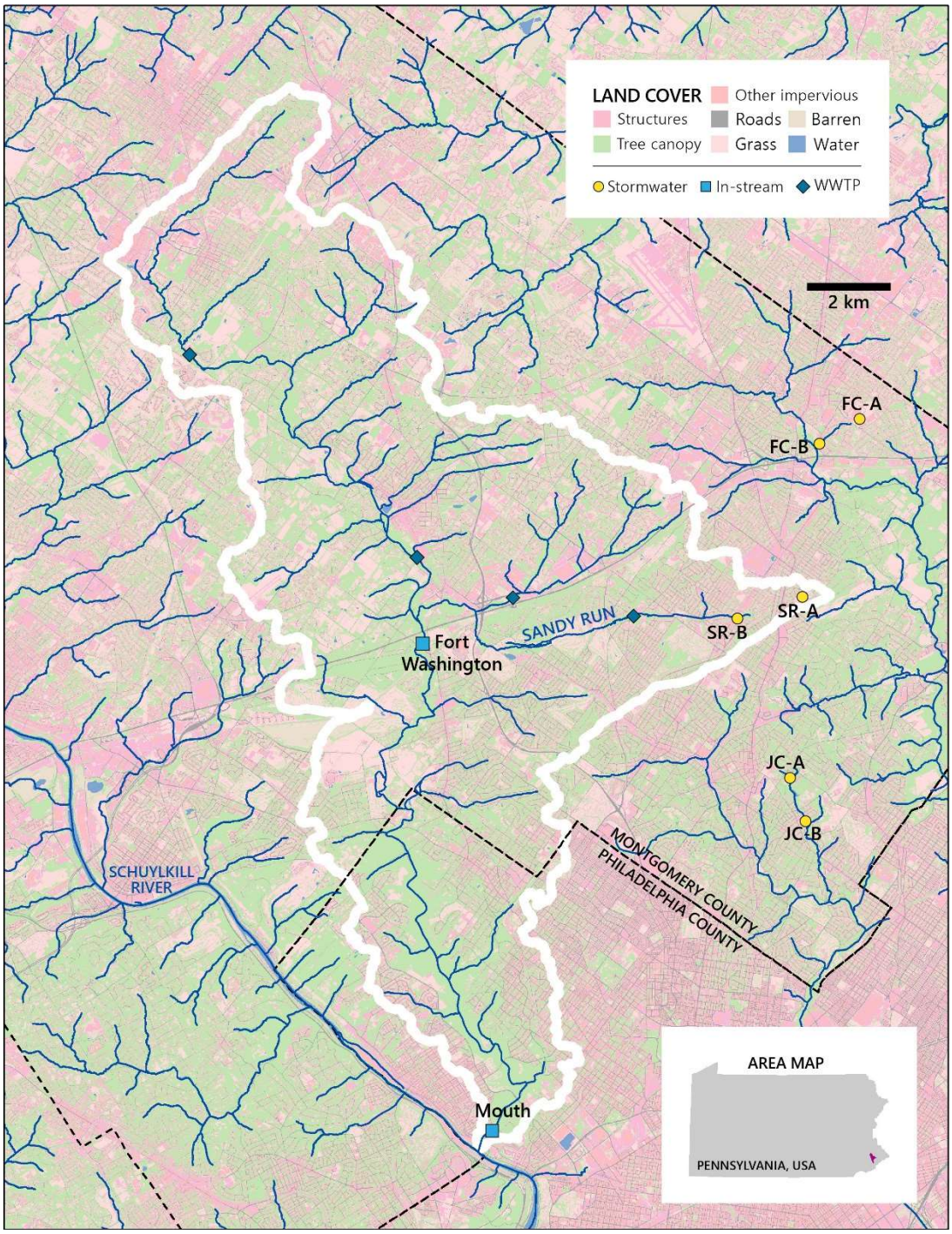

Figure 1. Wissahickon Creek is an impaired urban watershed in Philadelphia. The Wissahickon watershed (*white*) contains four wastewater treatment plants (WWTPs; *navy diamonds*). Stream (*blue squares*) and stormwater (*yellow dots*) sampling locations are shown.

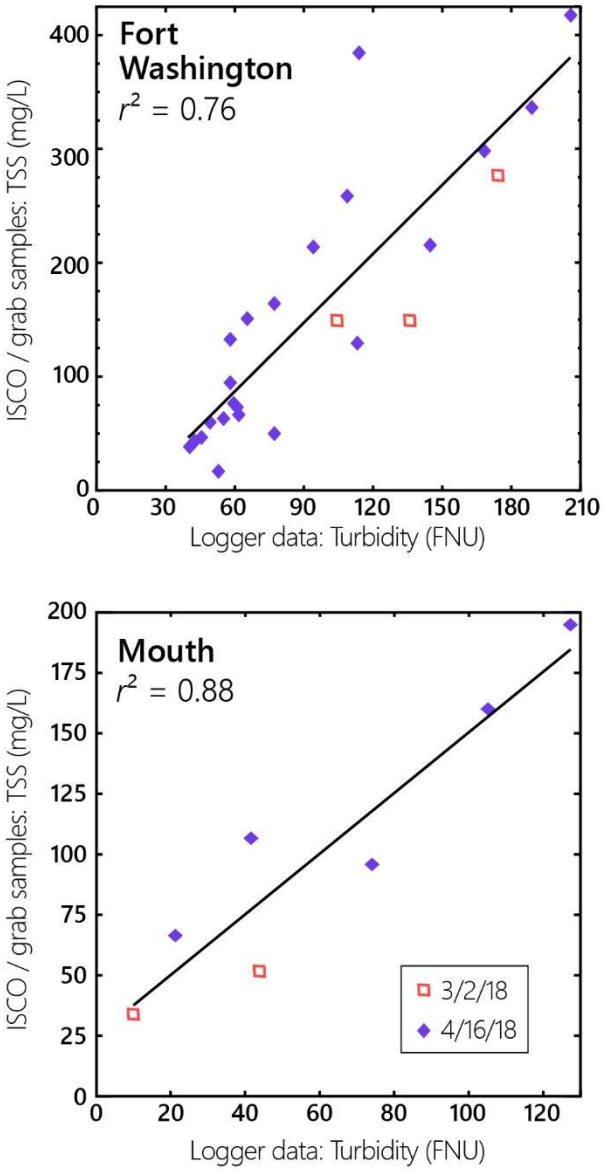

Figure 2. Linear relationships between turbidity and total suspended solids (TSS) were developed using USGS turbidity data and TSS concentrations in stream samples during two storms. TSS–turbidity relationships from Fort Washington (*top*) and Mouth (*bottom*) are shown.

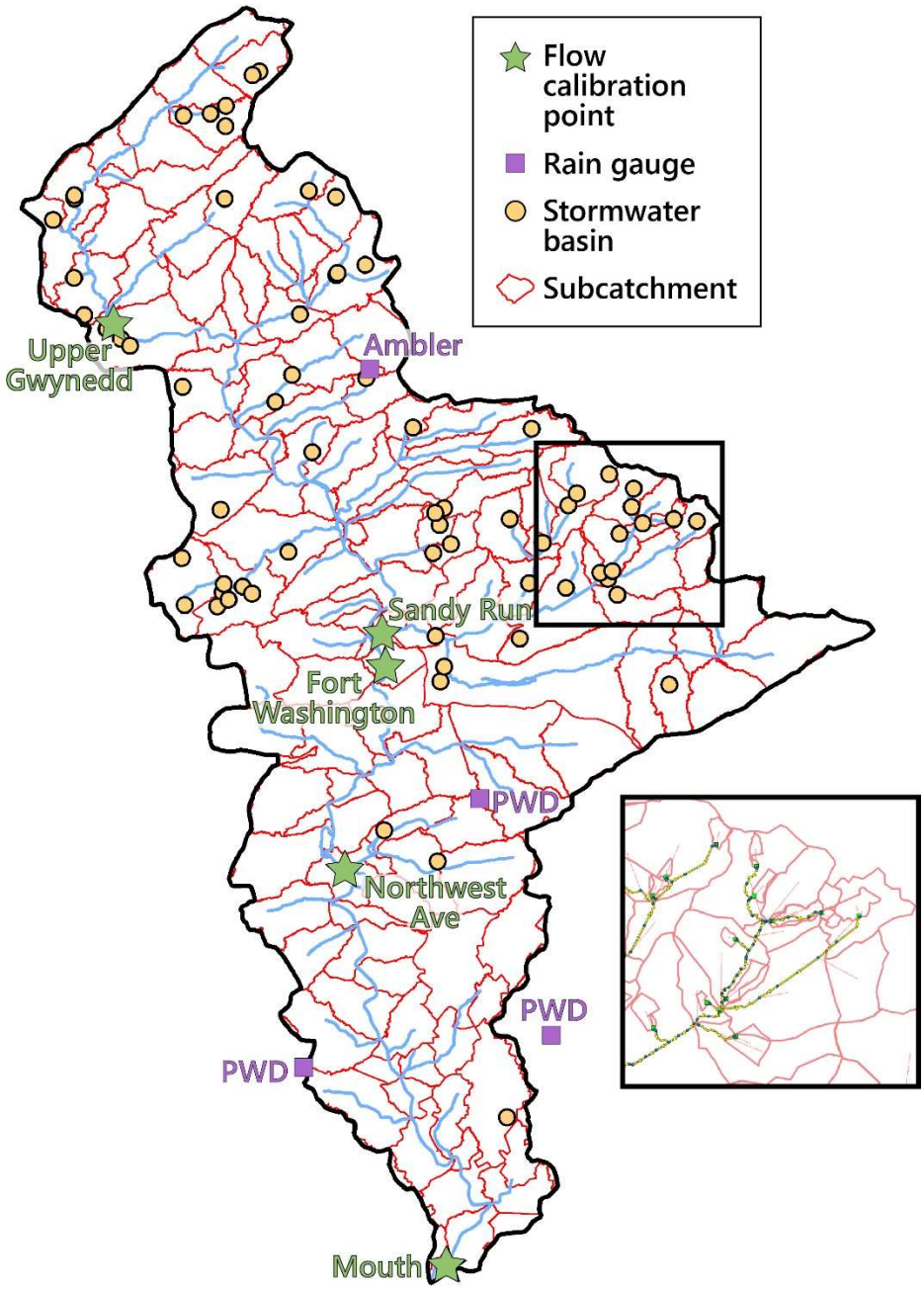

Figure 3. An extensive PCSWMM model was developed for the 165-km^2 Wissahickon watershed. Subcatchments (*red*), flow calibration points (*green stars*), stormwater basins (*yellow circles*), and rain gauges (*purple squares*) are shown. *Inset* shows more detailed stormwater routing within and between subcatchments via stormwater basins (*green squares*), links (*yellow arrows*), and nodes (*blue circles*).

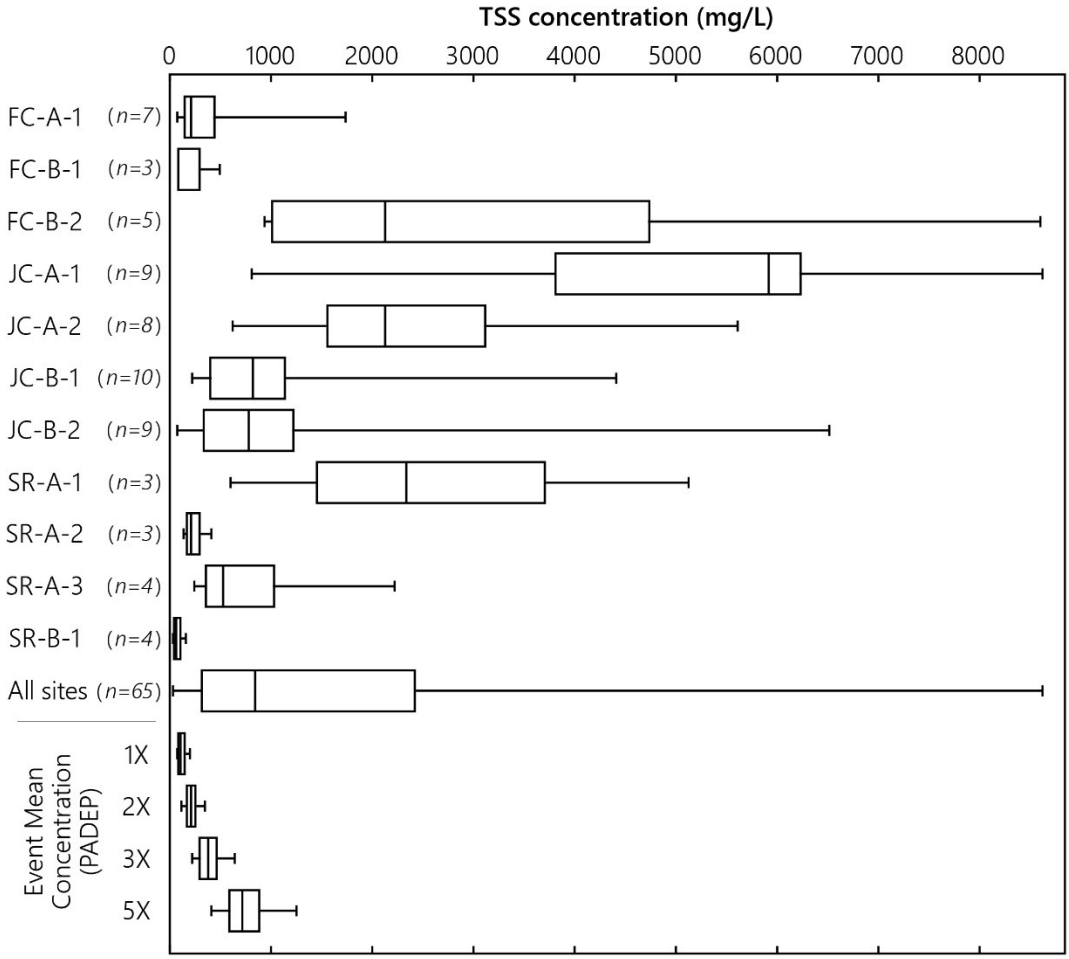

Figure 4. Stormwater TSS concentrations varied between sites and on average exceeded event mean concentrations (EMCs) defined by the Pennsylvania Department of Environmental Protection (PADEP) by an order of magnitude. The distributions of stormwater TSS concentrations from each sampling site are shown. Below that, the ranges of EMC values used in calibration are shown; values are 1, 2, 3 or 5 times the PADEP values. See *Figure S2* for TSS–land cover analysis.

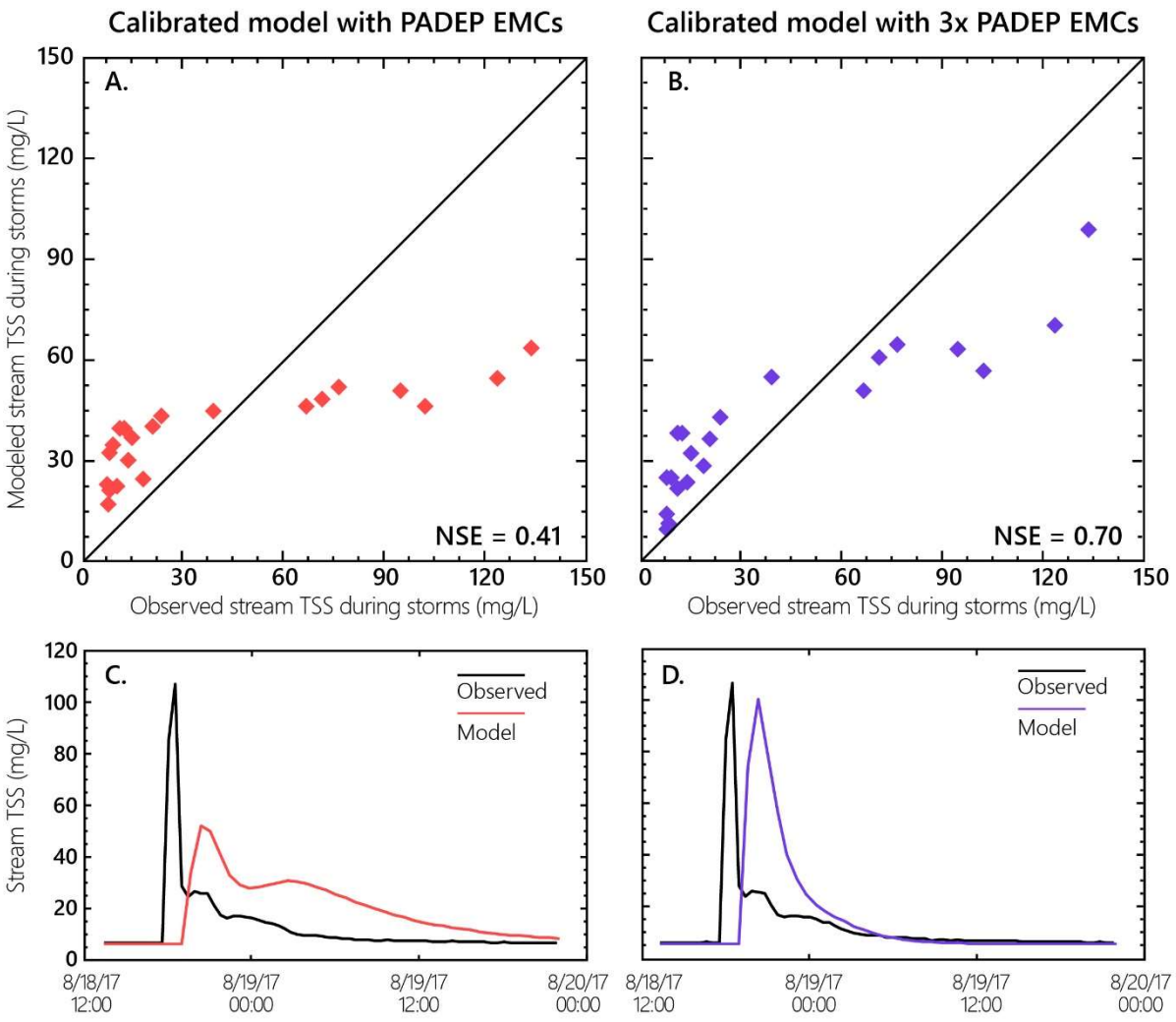

Figure 5. Tripling TSS EMCs substantially improved model performance, especially during storms. Modeled vs. observed TSS concentration at the Mouth during storms is shown for calibrated WASP models using PADEP EMC values (*A*) and triple these values (*B*). Time series of TSS concentration at the Mouth during a storm in August 2017 with calibrated WASP models using PADEP EMC values (*C*) and triple these values (*D*). See *Figure S3* for daily time series results for 1X EMC and 3X EMC models.

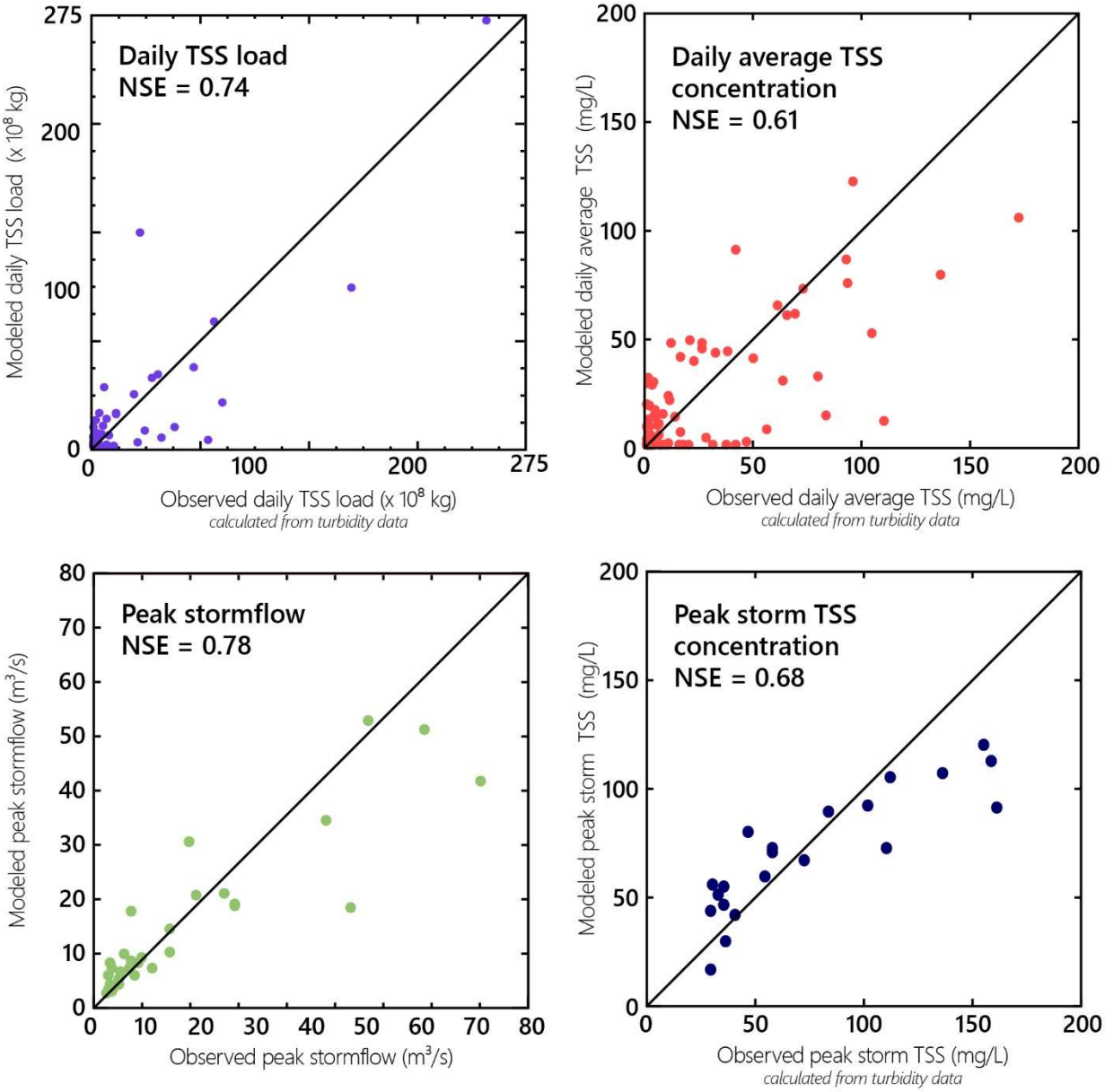

Figure 6. The calibrated PCSWMM–WASP model represents TSS concentration and flow at the Mouth well on a daily scale and during storms. Scatter plots for daily TSS load, daily average TSS concentration, peak stormflow, and peak storm TSS concentration at the Mouth. Observed TSS concentrations are calculated from turbidity data, and values that exceeded the maximum observed value (195 mg/L) were not considered in NSE calculations due to uncertainty. See *Figure S4* for scatter plots for time >50 mg/L and time >100 mg/L.

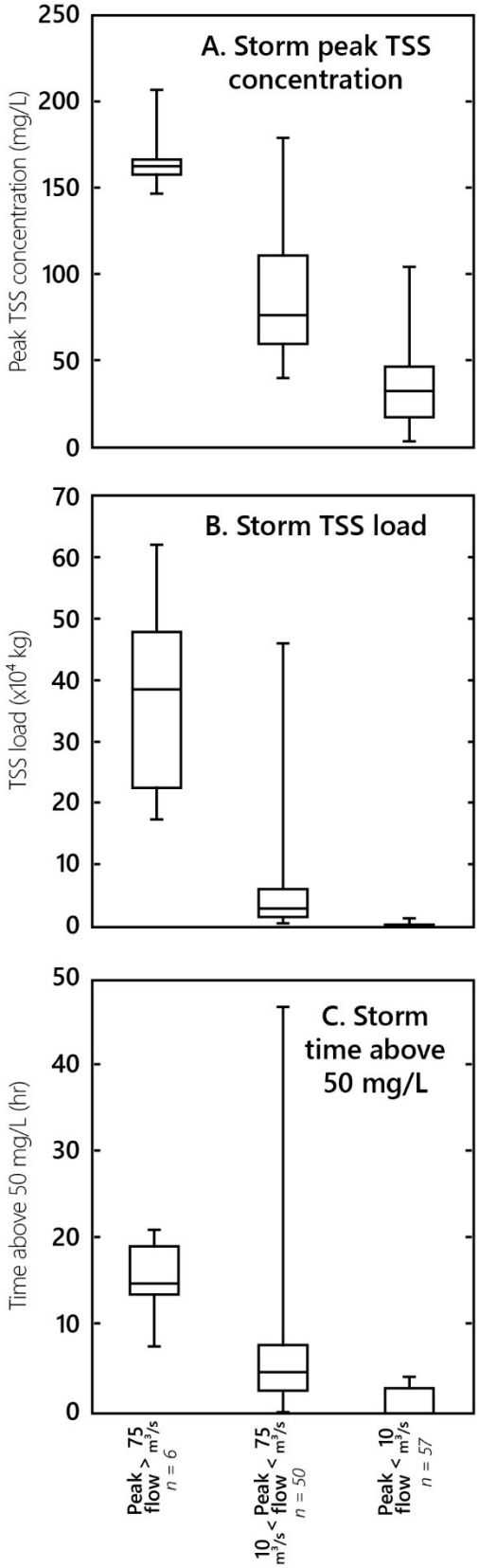

Figure 7. Large storms (with peak flow >75 m^3/s) deliver by far the most sediment to Wissahickon Creek; on average storms of this size currently occur more than once per year. Distribution of peak TSS concentration (*top*), total TSS load (*middle*), and duration for which in-stream TSS concentration exceeded 100 mg/L (*bottom*) at the Mouth during small, medium, and large storms. Storm size is defined based on peak flow: small storms have peak flow <10 m^3/s (*n=57*, recurrence interval <13 days), medium storms between 10 and 75 m^3/s (*n=50*, recurrence interval 13–100 days), and large storms >75 m^3/s (*n=6*, recurrence interval >100 days).

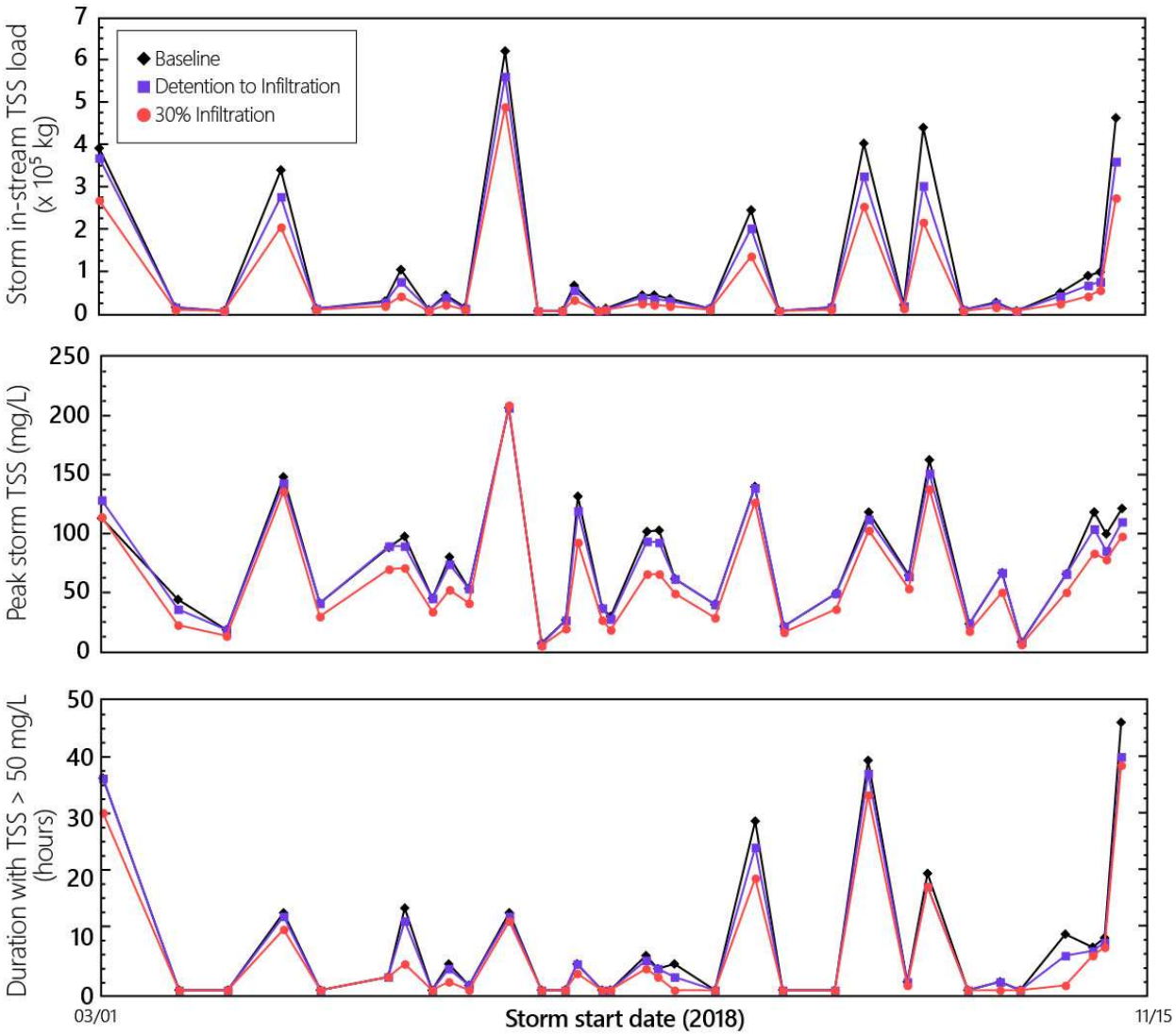

Figure 8. Increasing GSI is associated with reduced TSS load during storms, though peak TSS concentrations are not reduced as much. For each storm identified in 2018, the total TSS load exiting the Mouth during that storm is shown (*top*) as is the peak TSS concentration occurring during that storm (*middle*) and duration for which TSS concentration at the Mouth exceeded 50 mg/L for each storm (*bottom*). See *Figures S5 & S6* for analogous storm data from 2016 and 2017.

TABLES

Table 1. PCSWMM calibration: NSE values for flow metrics. Bold values are ≥0.5, which is a common threshold for acceptable NSE in hydrologic modeling. Flow calibration locations are shown on *Figure 3*.

	Mouth	Northwest Avenue	Fort Washington	Upper Gwynedd	Sandy Run
15-min Flow	**0.68**	**0.56**	**0.62**	**0.63**	**0.72**
Storm Total Flow	**0.50**	**0.83**	0.38	0.40	0.47
Storm Peak Flow	**0.78**	**0.71**	0.13	**0.73**	**0.76**

Table 2. WASP TSS calibration: NSE values for best calibration model runs with 1X, 2X, 3X, and 5X EMC. Bold values are ≥0.5, which is a common threshold for acceptable NSE in hydrologic modeling. The highlighted row indicates the model that was selected.

	MOUTH					FORT WASHINGTON				
EMC	Storm Average Conc.	Storm Peak Conc.	Storm Load	Daily Conc.	Daily Load	Storm Average Conc.	Storm Peak Conc.	Storm Load	Daily Conc.	Daily Load
1X	0.41	0.13	0.45	**0.54**	**0.74**	0.22	–0.16	0.25	0.40	0.40
2X	**0.62**	**0.58**	**0.55**	**0.61**	**0.78**	0.31	0.10	0.24	**0.50**	0.48
3X	**0.70**	**0.68**	**0.52**	**0.61**	**0.74**	0.33	0.20	0.18	**0.56**	**0.50**
5X	**0.80**	**0.62**	0.44	**0.58**	**0.55**	0.21	0.32	–0.07	**0.52**	0.43

Table 3. Model results for Baseline scenario. All values in this table represent the period between March and November each year.

	Total storm rainfall (in) [A]	Streamflow volume ($\times 10^6$ m^3)	% Flow occurring during storms	In-stream TSS load ($\times 10^6$ kg)	% TSS load occurring during storms	In-stream settled load ($\times 10^6$ kg) [B]	TSS above 50 mg/L (hr) [C]	TSS above 100 mg/L (hr) [C]
2016	30.7	31.1	69.9	0.6	91.5	3.8	83.25	6.75
2017	32.7	43.3	83.7	1.7	99.8	6.6	150.75	33
2018	47.8	73.3	88.2	3.8	92.2	13.5	269.25	48

[A] Average rainfall recorded at four stations (*Figure 3*).
[B] In-stream settled loads are not calibrated, so values are uncertain.
[C] Validation for time above 50 mg/L and 100 mg/L is shown in *Figure S4*.

SUPPORTING INFORMATION

Additional supporting information may be found online under the Supporting Information tab for this article: Tables with stormwater TSS data, curve number and EMC definition, calibration parameters, and scenario model results. Figures with peak flow–recurrence interval relationship, land cover–TSS relationships, additional calibration/validation, and storm-specific scenario data for 2016 and 2017.

ACKNOWLEDGMENTS

The William Penn Foundation (Grant # 39-18) provided funding for this study. Computational Hydraulics International (CHI) provided the license for PCSWMM software (University Grant #5840). Chelsea Kanaley, Ashlynn Young, and Dr. Sarah Ledford collected and analyzed data for TSS calibration, Dr. Sarah Ledford collected flow calibration data, and Tyler Wong provided some of the Python code used in this study.

LITERATURE CITED

Ahiablame, L.M., B.A. Engel, and I. Chaubey, 2012. Effectiveness of Low Impact Development Practices: Literature Review and Suggestions for Future Research. Water, Air, & Soil Pollution 223:4253–4273.

Alamdari, N., D. Sample, P. Steinberg, A. Ross, and Z. Easton, 2017. Assessing the Effects of Climate Change on Water Quantity and Quality in an Urban Watershed Using a Calibrated Stormwater Model. Water 9:464.

Ambrose, R.B. and T.A. Wool, 2017. WASP8 Stream Transport - Model Theory and User's Guide. U.S. Environmental Protection Agency, Atlanta, GA. https://www.epa.gov/ceam/wasp-model-documentation.

Askarizadeh, A., M.A. Rippy, T.D. Fletcher, D.L. Feldman, J. Peng, P. Bowler, A.S. Mehring, B.K. Winfrey, J.A. Vrugt, A. AghaKouchak, S.C. Jiang, B.F. Sanders, L.A. Levin, S. Taylor, and S.B. Grant, 2015. From Rain Tanks to Catchments: Use of Low-Impact Development To Address Hydrologic Symptoms of the Urban Stream Syndrome. Environmental Science & Technology 49:11264–11280.

Bach, P.M., D.T. McCarthy, and A. Deletic, 2010. Redefining the Stormwater First Flush Phenomenon. Water Research 44:2487–2498.

Bahaya, B., M. Al-Quraishi, and C. Gruden, 2019. Utilizing SWMM and GIS to Identify Total Suspended Solids Hotspots to Implement Green Infrastructure in Lucas County, OH. Environmental Progress & Sustainable Energy 38. doi:10.1002/ep.13240.

Bannerman, R.T., D.W. Owens, R.B. Dodds, and N.J. Hornewer, 1993. Sources of Pollutants in Wisconsin Stormwater. Water Science and Technology 28:241–259.

Barron, O.V., A.D. Barr, and M.J. Donn, 2013. Effect of Urbanisation on the Water Balance of a Catchment with Shallow Groundwater. Journal of Hydrology 485:162–176.

Béjar, M., C.N. Gibbins, D. Vericat, and R.J. Batalla, 2017. Effects of Suspended Sediment Transport on Invertebrate Drift: Drift and Sediment Transport. River Research and Applications 33:1655–1666.

Berndtsson, J.C., 2014. Storm Water Quality of First Flush Urban Runoff in Relation to Different Traffic Characteristics. Urban Water Journal 11:284–296.

Bilotta, G.S. and R.E. Brazier, 2008. Understanding the Influence of Suspended Solids on Water Quality and Aquatic Biota. Water Research 42:2849–2861.

Binns, A.D., A. Fata, A.M. Ferreira da Silva, H. Bonakdari, and B. Gharabaghi, 2019. Modeling Performance of Sediment Control Wet Ponds at Two Construction Sites in Ontario, Canada. Journal of Hydraulic Engineering 145:05019001.

Blaszczak, J.R., J.M. Delesantro, D.L. Urban, M.W. Doyle, and E.S. Bernhardt, 2019. Scoured or Suffocated: Urban Stream Ecosystems Oscillate between Hydrologic and Dissolved Oxygen Extremes. Limnology and Oceanography 64:877–894.

Borah, D.K., G. Yagow, A. Saleh, P. L. Barnes, W. Rosenthal, E. C. Krug, and L. M. Hauck, 2006. Sediment and Nutrient Modeling for TMDL Development and Implementation. Transactions of the ASABE 49:967–986.

Buendia, C., C.N. Gibbins, D. Vericat, R.J. Batalla, and A. Douglas, 2013. Detecting the Structural and Functional Impacts of Fine Sediment on Stream Invertebrates. Ecological Indicators 25:184–196.

Burdon, F.J., A.R. McIntosh, and J.S. Harding, 2013. Habitat Loss Drives Threshold Response of Benthic Invertebrate Communities to Deposited Sediment in Agricultural Streams. Ecological Applications 23:1036–1047.

Burian, S.J., T.N. McPherson, M.J. Brown, G.E. Streit, and H.J. Turin, 2002. Modeling the Effects of Air Quality Policy Changes on Water Quality in Urban Areas. :12.

Caruso, B.S., 2005. Simulation of Metals Total Maximum Daily Loads and Remediation in a Mining-Impacted Stream. Journal of Environmental Engineering 131:777–789.

Charbeneau, R.J. and M.E. Barrett, 1998. Evaluation of Methods for Estimating Stormwater Pollutant Loads. Water Environment Research 70:1295–1302.

Cho, J. and J. Lee, 2017. Stormwater Runoff Characteristics and Effective Management of Nonpoint Source Pollutants from a Highland Agricultural Region in the Lake Soyang Watershed. Water 9:784.

Clinton, B.D. and J.M. Vose, 2006. Variation in Stream Water Quality in an Urban Headwater Stream in the Southern Appalachians. Water, Air, and Soil Pollution 169:331–353.

Conroy, E., J.N. Turner, A. Rymszewicz, M. Bruen, J.J. O'Sullivan, D.M. Lawler, S. Stafford, and M. Kelly-Quinn, 2018. Further Insights into the Responses of Macroinvertebrate Species to Burial by Sediment. Hydrobiologia 805:399–411.

Cushman, E., 2019. Concentration of Suspended Solids and Nutrietnts in Overland Flow in Suburban Philadelphia Watersheds. M.S. Thesis, Temple University, Philadelphia, PA.

Damodaram, C., M.H. Giacomoni, C. Prakash Khedun, H. Holmes, A. Ryan, W. Saour, and E.M. Zechman, 2010. Simulation of Combined Best Management Practices and Low Impact Development for Sustainable Stormwater Management. JAWRA Journal of the American Water Resources Association 46:907–918.

Deletic, A., 1998. The First Flush Load of Urban Surface Runoff. Water Research 32:2462–2470.

Deletic, A.B. and C.T. Maksimovic, 1998. Evaluation of Water Quality Factors in Storm Runoff from Paved Areas. Journal of Environmental Engineering 124:869–879.

DVRPC, 2017. 2015 Land Use Update, Phase 1. Delaware Valley Regional Planning Commission. https://dvrpc-dvrpcgis.opendata.arcgis.com/datasets/dvrpc-2015-land-use-enhanced-feature-service-new-.

DVRPC, 2018. 2015 Impervious Surface Montgomery County, PA. Delaware Valley Regional Planning Commission. https://dvrpc-dvrpcgis.opendata.arcgis.com/datasets/dvrpc-2015-impervious-surface-montgomery-county-pa.

Eckart, K., Z. McPhee, and T. Bolisetti, 2017. Performance and Implementation of Low Impact Development – A Review. Science of The Total Environment 607–608:413–432.

Ekdal, A., M. Gürel, C. Guzel, A. Erturk, A. Tanik, and I.E. Gonenc, 2011. Application of WASP and SWAT Models for a Mediterranean Coastal Lagoon with Limited Seawater Exchange. Journal of Coastal Research:6.

Emerson, C.H., C. Welty, and R.G. Traver, 2005. Watershed-Scale Evaluation of a System of Storm Water Detention Basins. Journal of Hydrologic Engineering 10:237–242.

EPA, 2003. Nutrient and Siltation TMDL Development for Wissahickon Creek, Pennsylvania. Environmental Protection Agency, Philadelphia, PA.

EPA, 2010. National Pollutant Discharge Elimination System (NPDES) Permit Writer's Manual. U.S. Environmental Protection Agency, Washington, D.C.

EPA, 2015. Total Phosphorous TMDL for the Wissahickon Creek Watershed, Pennsylvania. Environmental Protection Agency, Philadelphia, PA.

EPA, 2017. Sediment Transport. WASP7 Course, U.S. Environmental Protection Agency.

EPA, 2019. Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) Version 4.5 User's Manual. U.S. Environmental Protection Agency.

Esteves, M., C. Legout, O. Navratil, and O. Evrard, 2019. Medium Term High Frequency Observation of Discharges and Suspended Sediment in a Mediterranean Mountainous Catchment. Journal of Hydrology 568:562–574.

Ferreira, C.S.S., R. Moruzzi, J.M.G.P. Isidoro, M. Tudor, M. Vargas, A.J.D. Ferreira, and J.L.M.P. de Lima, 2019. Impacts of Distinct Spatial Arrangements of Impervious Surfaces on Runoff and Sediment Fluxes from Laboratory Experiments. Anthropocene 28:100219.

Fisher, J.R., B.I. Dvorak, and D.M. Admiraal, 2016. Pollutant Load Estimates Using Regression Models with In-Stream Measurements. Journal of Environmental Engineering 142:04015081.

Fletcher, T.D., H. Andrieu, and P. Hamel, 2013. Understanding, Management and Modelling of Urban Hydrology and Its Consequences for Receiving Waters: A State of the Art. Advances in Water Resources 51:261–279.

Fletcher, T.D., W. Shuster, W.F. Hunt, R. Ashley, D. Butler, S. Arthur, S. Trowsdale, S. Barraud, A. Semadeni-Davies, J.-L. Bertrand-Krajewski, P.S. Mikkelsen, G. Rivard, M. Uhl, D. Dagenais, and M. Viklander, 2015. SUDS, LID, BMPs, WSUD and More – The Evolution and Application of Terminology Surrounding Urban Drainage. Urban Water Journal 12:525–542.

Franklin Institute, 2017. Historical Weather Data for Philadelphia. Franklin Institute, Philadelphia, PA. https://www.fi.edu/history-resources/philadelphia-weather-data.

Fromuth, R. (ed), 2014. Wissahickon Creek Act 167 Plan. Center for Sustainable Communities and Newell, Tereska, & Mackay Engineering.

Giacomoni, M.H., R. Gomez, and E.Z. Berglund, 2014. Hydrologic Impact Assessment of Land Cover Change and Stormwater Management Using the Hydrologic Footprint Residence. JAWRA Journal of the American Water Resources Association 50:1242–1256.

Gippel, C.J., 1995. Potential of Turbidity Monitoring for Measuring the Transport of Suspended Solids in Streams. Hydrological Processes 9:83–97.

Göbel, P., C. Dierkes, and W.G. Coldewey, 2007. Storm Water Runoff Concentration Matrix for Urban Areas. Journal of Contaminant Hydrology 91:26–42.

Goonetilleke, A., E. Thomas, S. Ginn, and D. Gilbert, 2005. Understanding the Role of Land Use in Urban Stormwater Quality Management. Journal of Environmental Management 74:31–42.

Gu, C., K. Cockerill, W.P. Anderson, F. Shepherd, P.A. Groothuis, T.M. Mohr, J.C. Whitehead, A.A. Russo, and C. Zhang, 2019. Modeling Effects of Low Impact Development on Road Salt Transport at Watershed Scale. Journal of Hydrology 574:1164–1175.

Hathaway, J.M. and W.F. Hunt, 2011. Evaluation of First Flush for Indicator Bacteria and Total Suspended Solids in Urban Stormwater Runoff. Water, Air, & Soil Pollution 217:135–147.

Izagirre, O., A. Serra, H. Guasch, and A. Elosegi, 2009. Effects of Sediment Deposition on Periphytic Biomass, Photosynthetic Activity and Algal Community Structure. Science of The Total Environment 407:5694–5700.

Jefferson, A.J., A.S. Bhaskar, K.G. Hopkins, R. Fanelli, P.M. Avellaneda, and S.K. McMillan, 2017. Stormwater Management Network Effectiveness and Implications for Urban Watershed Function: A Critical Review. Hydrological Processes 31:4056–4080.

Kayhanian, M., A. Singh, C. Suverkropp, and S. Borroum, 2003. Impact of Annual Average Daily Traffic on Highway Runoff Pollutant Concentrations. Journal of Environmental Engineering 129:975–990.

Kemp, P., D. Sear, A. Collins, P. Naden, and I. Jones, 2011. The Impacts of Fine Sediment on Riverine Fish. Hydrological Processes 25:1800–1821.

Larsen, S., G. Pace, and S.J. Ormerod, 2011. Experimental Effects of Sediment Deposition on the Structure and Function of Macroinvertebrate Assemblages in Temperate Streams. River Research and Applications 27:257–267.

Lawler, D.M., G.E. Petts, I.D.L. Foster, and S. Harper, 2006. Turbidity Dynamics during Spring Storm Events in an Urban Headwater River System: The Upper Tame, West Midlands, UK. Science of The Total Environment 360:109–126.

Liu, Y., V.F. Bralts, and B.A. Engel, 2015. Evaluating the Effectiveness of Management Practices on Hydrology and Water Quality at Watershed Scale with a Rainfall-Runoff Model. Science of The Total Environment 511:298–308.

Locatelli, L., O. Mark, P.S. Mikkelsen, K. Arnbjerg-Nielsen, A. Deletic, M. Roldin, and P.J. Binning, 2017. Hydrologic Impact of Urbanization with Extensive Stormwater Infiltration. Journal of Hydrology 544:524–537.

Ma, C.-Y., Y.-T. Sheu, K.-F. Hsia, C.-D. Dong, C.-W. Chen, Y.-C. Huang, and C.-M. Kao, 2018. Development of Water and Sediment Quality Management Strategies for an Urban River Basin: A Case Study in Taiwan. Journal of Water Supply: Research and Technology-Aqua 67:810–823.

Martinez-Villalobos, C. and J.D. Neelin, 2018. Shifts in Precipitation Accumulation Extremes during the Warm Season over the United States. Geophysical Research Letters 45:8586–8595.

Mathers, K.L., S.P. Rice, and P.J. Wood, 2017. Temporal Effects of Enhanced Fine Sediment Loading on Macroinvertebrate Community Structure and Functional Traits. Science of The Total Environment 599–600:513–522.

McKenzie, M., K.L. Mathers, P.J. Wood, J. England, I. Foster, D. Lawler, and M. Wilkes, 2020. Potential Physical Effects of Suspended Fine Sediment on Lotic Macroinvertebrates. Hydrobiologia 847:697–711.

Monk, W.A., P.J. Wood, D.M. Hannah, and D.A. Wilson, 2008. Macroinvertebrate Community Response to Inter-Annual and Regional River Flow Regime Dynamics. River Research and Applications 24:988–1001.

Moriasi, D.N., J.G. Arnold, M.W. van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE 50:885–900.

PADEP, 2005. Pennsylvania Stormwater Best Management Practices Manual Draft. Pennsylvania Department of Environmental Protection. http://www.dep.state.pa.us/dep/subject/advcoun/stormwater/Manual_DraftJan05/Section09-jan-rev.pdf.

PADEP, 2018. Pennsylvania Integrated Water Quality Monitoring and Assessment Report. Pennsylvania Department of Environmental Protection, Harrisburg, PA. https://www.depgis.state.pa.us/2018_integrated_report/index.html.

Peng, H.-Q., Y. Liu, H.-W. Wang, X.-L. Gao, and L.-M. Ma, 2016. Event Mean Concentration and First Flush Effect from Different Drainage Systems and Functional Areas during Storms. Environmental Science and Pollution Research 23:5390–5398.

Prudencio, L. and S.E. Null, 2018. Stormwater Management and Ecosystem Services: A Review. Environmental Research Letters 13:033002.

PWD, 2007. Wissahickon Creek Watershed Comprehensive Characterization Report. Philadelphia Water Department, Philadelphia, PA.

PWD, 2019. SWMMout2csv 0.8.7. Philadelphia Water Department. https://github.com/OOW/SWMMout2csv/tree/master/final_files.

Rosa, D.J., J.C. Clausen, and M.E. Dietz, 2015. Calibration and Verification of SWMM for Low Impact Development. JAWRA Journal of the American Water Resources Association 51:746–757.

Rossman, L.A. and W.C. Huber, 2016. Storm Water Management Model Reference Manual Volume I - Hydrology. National Risk Management Laboratory, Cincinnati, OH.

Roy-Poirier, A., P. Champagne, and Y. Filion, 2010. Review of Bioretention System Research and Design: Past, Present, and Future. Journal of Environmental Engineering 136:878–889.

Ryan, P.A., 1991. Environmental Effects of Sediment on New Zealand Streams: A Review. New Zealand Journal of Marine and Freshwater Research 25:207–221.

Schubert, J.E., M.J. Burns, T.D. Fletcher, and B.F. Sanders, 2017. A Framework for the Case-Specific Assessment of Green Infrastructure in Mitigating Urban Flood Hazards. Advances in Water Resources 108:55–68.

Shuster, W.D., D. Lye, A. De La Cruz, L.K. Rhea, K. O'Connell, and A. Kelty, 2013. Assessment of Residential Rain Barrel Water Quality and Use in Cincinnati, Ohio. JAWRA Journal of the American Water Resources Association 49:753–765.

Shuster, W.D. and E. Pappas, 2011. Laboratory Simulation of Urban Runoff and Estimation of Runoff Hydrographs with Experimental Curve Numbers Implemented in USEPA SWMM. Journal of Irrigation and Drainage Engineering 137:343–351.

Sohn, W., J.-H. Kim, M.-H. Li, and R. Brown, 2019. The Influence of Climate on the Effectiveness of Low Impact Development: A Systematic Review. Journal of Environmental Management 236:365–379.

Sparkman, S.A., D.M. Hogan, K.G. Hopkins, and J.V. Loperfido, 2017. Modeling Watershed-Scale Impacts of Stormwater Management with Traditional versus Low Impact Development Design. JAWRA Journal of the American Water Resources Association 53:1081–1094.

Steinman, A.D., E.S. Isely, and K. Thompson, 2015. Stormwater Runoff to an Impaired Lake: Impacts and Solutions. Environmental Monitoring and Assessment 187:549.

Taylor, K.G. and P.N. Owens, 2009. Sediments in Urban River Basins: A Review of Sediment–Contaminant Dynamics in an Environmental System Conditioned by Human Activities. Journal of Soils and Sediments 9:281–303.

Todeschini, S., S. Manenti, and E. Creaco, 2019. Testing an Innovative First Flush Identification Methodology against Field Data from an Italian Catchment. Journal of Environmental Management 246:418–425.

Toor, G.S., M.L. Occhipinti, Y.-Y. Yang, T. Majcherek, D. Haver, and L. Oki, 2017. Managing Urban Runoff in Residential Neighborhoods: Nitrogen and Phosphorus in Lawn Irrigation Driven Runoff Y. F. Tsang (Editor). PLOS ONE 12:e0179151.

University of Vermont Spatial Analysis Laboratory, 2016. High-Resolution Land Cover, Commonwealth of Pennsylvania, Chesapeake Bay Watershed and Delaware River Basin. University of Vermont. https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=3193.

USDA, 1986. Urban Hydrology for Small Watersheds. United States Department of Agriculture, National Resources Conservation Service, Washington, D.C.

USDA, 2019. Web Soil Survey. United States Department of Agriculture, National Resources Conservation Service, Washington, D.C. https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm.

Vineyard, D., W.W. Ingwersen, T.R. Hawkins, X. Xue, B. Demeke, and W. Shuster, 2015. Comparing Green and Grey Infrastructure Using Life Cycle Cost and Environmental Impact: A Rain Garden Case Study in Cincinnati, OH. JAWRA Journal of the American Water Resources Association 51:1342–1360.

Wadzuk, B.M., M. Rea, G. Woodruff, K. Flynn, and R.G. Traver, 2010. Water-Quality Performance of a Constructed Stormwater Wetland for All Flow Conditions. JAWRA Journal of the American Water Resources Association 46:385–394.

Walsh, C.J., A.H. Roy, J.W. Feminella, P.D. Cottingham, P.M. Groffman, and R.P. Morgan, 2005. The Urban Stream Syndrome: Current Knowledge and the Search for a Cure. Journal of the North American Benthological Society 24:706–723.

Williams, N.D., D.E. Walling, and G.J.L. Leeks, 2008. An Analysis of the Factors Contributing to the Settling Potential of Fine Fluvial Sediment. Hydrological Processes 22:4153–4162.

Wu, H., J.P. Bolte, D. Hulse, and B.R. Johnson, 2015. A Scenario-Based Approach to Integrating Flow-Ecology Research with Watershed Development Planning. Landscape and Urban Planning 144:74–89.

Yang, Y. and T.F.M. Chui, 2018. Rapid Assessment of Hydrologic Performance of Low Impact Development Practices under Design Storms. JAWRA Journal of the American Water Resources Association 54:613–630.

Zhang, K. and T.F.M. Chui, 2018. A Comprehensive Review of Spatial Allocation of LID-BMP-GI Practices: Strategies and Optimization Tools. Science of The Total Environment 621:915–929.

Zhang, K. and T.F.M. Chui, 2019. Linking Hydrological and Bioecological Benefits of Green Infrastructures across Spatial Scales – A Literature Review. Science of The Total Environment 646:1219–1231.

Ziegler, A.D., S.G. Benner, C. Tantasirin, S.H. Wood, R.A. Sutherland, R.C. Sidle, N. Jachowski, M.A. Nullet, L.X. Xi, A. Snidvongs, T.W. Giambelluca, and J.M. Fox, 2014. Turbidity-Based Sediment Monitoring in Northern Thailand: Hysteresis, Variability, and Uncertainty. Journal of Hydrology 519:2020–2039.

SUPPORTING INFORMATION

Coupling PCSWMM and WASP to evaluate green stormwater infrastructure impacts to storm sediment loads in an urban watershed

Sarah Beganskas, Robert J. Ryan, Evelyn Walters, Manahel Soro,
Elizabeth Cushman, and Laura Toran

Table S1. Total suspended solids (TSS) concentration in stormwater samples. TSS concentration (in mg/L) for each stormwater sample collected. Bottles were installed in advance of 13 storms, but all 11 sites were not sampled during every storm.

Site	5/12/18	5/19/18	6/10/18	6/24/18	6/28/18	7/6/18	7/27/18	8/4/18	8/11/18	8/31/18	9/18/18	10/11/18	10/26/18
FC-A-1	–	–	–	170	1,700	540	**DNF**	180	310	**DNF**	48	–	65
FC-B-1	–	–	–	**DNF**	**DNF**	**DNF**	470	–	**DNF**	**DNF**	48	–	63
FC-B-2	–	–	–	**DNF**	8,600	**DNF**	4,700	–	2,100	**DNF**	980	–	920
JC-A-1	770	6,400	5,900	6,100	4,600	**DNF**	**DNF**	6,200	8,600	3,800	–	–	3,200
JC-A-2	4,000	1,000	5,600	**DNF**	2,800	**DNF**	**DNF**	1,700	2,400	1,800	–	–	600
JC-B-1	4,400	400	1,100	820	770	–	NM	–	250	1,100	370	1,300	290
JC-B-2	–	750	1,600	6,500	310	**DNF**	120	–	1,200	370	1,100	–	52
SR-A-1	–	–	–	–	–	–	**DNF**	–	2,300	**DNF**	5,100	–	590
SR-A-2	–	–	–	–	–	–	130	–	**DNF**	**DNF**	380	–	170
SR-A-3	–	–	–	–	–	–	2,200	–	**DNF**	230	630	–	380
SR-B-1	–	120	**DNF**	**DNF**	**DNF**	**DNF**	**DNF**	**DNF**	11	–	29	–	55

– = no bottle installed
DNF = bottle did not adequately fill
NM = bottle filled but TSS concentration was not measured

Table S2. Curve numbers based on Hydrologic Soil Group and land use category.

	Hydrologic Soil Group (HSG)	A	B	C	D
	Infiltration Rate (in/hr)	>0.3	0.15–0.3	0.05–0.15	<0.05
Land use category	Agriculture	49	69	79	84
	Commercial	89	92	94	95
	Community Services	81	88	91	93
	Manufacturing: Light Industrial	81	88	91	93
	Military	63	77	85	88
	Parking	98	98	98	98
	Recreation	49	69	79	84
	Residential: Mobile Home	77	85	90	92
	Residential: Multi-Family	77	85	90	92
	Residential: Row Home	77	85	90	92
	Residential: Single-Family	57	72	81	86
	Transportation	83	89	92	93
	Utility	89	92	94	95
	Vacant	77	85	90	92
	Water	100	100	100	100
	Wooded	36	60	73	79

Table S3. EMCs classified based on land cover were adjusted to land use categories for characterization within PCSWMM. For each Delaware Valley Regional Planning Commission (DVRPC) land use category, the percentage contribution of each land cover type is shown, along with the calculated TSS EMCs.

DVRPC Land Use	PADEP Land Cover: Forest	Meadow	Fertilized planting area	Lawn	Athletic field	Rooftop	Medium traffic street	Low traffic residential street	Driveway	High traffic parking	Low traffic parking	Calculated TSS EMC (mg/L)	2X EMC (mg/L)	3X EMC (mg/L)	5X EMC (mg/L)
Residential: Single-Family				60%		20%		10%	10%			**127**	**254**	**381**	**635**
Residential: Row Home				35%		50%		15%				**86**	**172**	**258**	**430**
Residential: Mobile Home				35%		50%		15%				**86**	**172**	**258**	**430**
Residential: Multi-Family				35%		50%		15%				**86**	**172**	**258**	**430**
Manufacturing				28%		40%				32%		**77**	**154**	**231**	**385**
Transportation		15%					85%					**103**	**206**	**309**	**515**
Agriculture		20%	70%								10%	**54**	**108**	**162**	**270**
Commercial				15%		15%					70%	**71**	**142**	**213**	**355**
Recreation		15%		50%	25%						10%	**153**	**306**	**459**	**765**
Wooded	100%											**39**	**78**	**117**	**195**
Vacant		90%									10%	**48**	**96**	**144**	**240**
Parking										100%		**120**	**240**	**360**	**600**
Utility		25%									75%	**55**	**110**	**165**	**275**
Community Services				15%		15%					70%	**71**	**142**	**213**	**355**
Military				15%		15%					70%	**71**	**142**	**213**	**355**
PADEP TSS EMC (mg/L)	**39**	**47**	**55**	**180**	**200**	**21**	**113**	**86**	**60**	**120**	**58**				

Table S4. Calibration parameter ranges for flow within PCSWMM. Parameter values were varied independently for each subcatchment. Ranges with a % indicate that individual values for each subcatchment were varied as a percentage of their initial value.

Parameter	Range tested
Width of overland flow	10% – 160%
Manning's N for impervious area	0.01 – 0.03
Manning's N for pervious area	0.05 – 0.6
Depth of depression storage on impervious	10% – 150%
Depth of depression storage on pervious	10% – 150%
Percent of runoff routed between subareas	10% – 160%
SCS runoff curve number	96% – 104%
Groundwater flow coefficient	0% – 105%
Groundwater flow exponent	0 – 2

Table S5. Adjusted curve numbers (CNs) for the 30% Infiltration scenario (New CN) compared to Baseline scenario curve numbers (Original CN). Subcatchment CNs were adjusted so that runoff from the treated area was reduced by 60% during a 2-year, 24-hr design storm (3.27 in). This 60% reduction metric was based on the modeled performance of infiltration-based GSI projects in the Wissahickon watershed.

The following equation was used: $New\ CN = (1 - x) * Original\ CN\ +\ x * Adjusted\ CN$

Adjusted CN = the CN that reduces subcatchment runoff by 60% during the design storm

x = the fraction of the watershed treated by infiltration-based GSI (in this case, 30%)

Original CN	New CN	Original CN	New CN	Original CN	New CN	Original CN	New CN
60.7	58.3	70.0	66.5	74.6	70.6	82.0	77.2
63.2	60.4	70.3	66.8	75.2	71.1	82.1	77.3
64.4	61.5	70.5	67.0	75.6	71.5	82.7	77.8
64.6	61.8	70.6	67.0	75.8	71.7	82.8	77.9
65.3	62.3	71.0	67.4	75.9	71.8	83.0	78.1
65.4	62.4	71.2	67.6	77.1	72.9	83.7	78.7
65.8	62.8	71.3	67.7	77.2	73.0	83.7	78.8
66.7	63.6	71.4	67.8	77.3	73.1	84.0	79.0
66.8	63.7	71.6	68.0	77.6	73.3	84.7	79.6
66.9	63.8	71.8	68.2	77.7	73.4	85.0	79.9
67.1	63.9	71.9	68.2	78.1	73.8	85.7	80.5
67.4	64.2	72.1	68.4	78.2	73.9	86.0	80.8
67.8	64.6	72.3	68.6	78.4	74.0	86.7	81.4
67.9	64.7	72.4	68.6	78.6	74.2	87.0	81.7
68.0	64.8	72.5	68.7	78.7	74.3	88.0	82.6
68.1	64.8	72.9	69.1	79.0	74.5	88.3	82.8
68.2	64.9	73.1	69.3	79.1	74.7	88.8	83.2
68.4	65.1	73.4	69.6	79.6	75.1	89.0	83.5
68.5	65.2	73.5	69.7	79.7	75.2	89.8	84.1
68.7	65.4	73.6	69.8	80.0	75.4	90.0	84.3
68.8	65.4	73.7	69.8	80.1	75.5	91.0	85.2
68.9	65.5	73.8	69.9	80.6	76.0	92.0	86.1
69.0	65.6	74.0	70.1	80.7	76.0	93.0	87.0
69.1	65.7	74.1	70.2	81.0	76.3	93.2	87.2
69.5	66.1	74.2	70.3	81.3	76.6	94.0	87.9
69.6	66.2	74.3	70.3	81.4	76.7	95.0	88.8
69.7	66.2	74.4	70.4	81.7	76.9	96.0	89.7
69.8	66.3	74.5	70.5	81.8	77.0	98.0	91.5

Table S6. Calibration parameter ranges for TSS concentrations within WASP.

Parameter	Region	Range tested (m/d)	Calibrated value (m/d)
Sediment settling velocity	Tributaries	0.25 – 30	10
	Upper main stem	0.25 – 45	11
	Lower main stem	0.25 – 60	12
Sediment resuspension velocity	All	0.00005 – 1	0.0001

Table S7. Model results for GSI scenarios compared to Baseline scenario. *p*-values represent results from paired t-tests comparing values for each storm between the Baseline scenario and the GSI scenarios.

		IN-STREAM SEDIMENT LOAD			IN-STREAM SETTLING LOAD [A]		
	Number of distinct storms	**Baseline total (× 10^6 kg)**	**Detention to Infiltration total (× 10^6 kg)**	**30% Infiltration total (× 10^6 kg)**	**Baseline total (× 10^6 kg)**	**Detention to Infiltration total (× 10^6 kg)**	**30% Infiltration total (× 10^6 kg)**
2016	41	0.6	0.5 *p=2E-03*	0.3 *p=1E-05*	3.8	3.6 *p=9E-04*	2.6 *p=2E-09*
2017	39	1.7	1.3 *p=3E-03*	0.8 *p=1E-04*	6.6	6.2 *p=6E-04*	4.8 *p=1E-07*
2018	33	3.8	3.0 *p=1E-03*	2.2 *p=6E-05*	13.5	12.2 *p=8E-04*	10.0 *p=2E-07*
Total	**113**	**6.1**	**4.8** ***p=6E-06***	**3.3** ***p=5E-08***	**23.9**	**22** ***p=9E-06***	**17.4** ***p=2E-12***

TIME ABOVE 50 MG/L [B]			TIME ABOVE 100 MG/L [B]		
Baseline total (hr)	**Detention to Infiltration total (hr)**	**30% Infiltration total (hr)**	**Baseline total (hr)**	**Detention to Infiltration total (hr)**	**30% Infiltration total (hr)**
83.3	73.5 *p=4E-03*	30.0 *p=2E-05*	6.8	6.0 *p=2E-01*	4.5 *p=4E-02*
150.8	139.5 *p=3E-04*	69.0 *p=3E-06*	33.0	25.5 *p=4E-02*	11.3 *p=5E-03*
269.3	238.5 *p=1E-03*	196.5 *p=2E-04*	58.5	45.8 *p=1E-01*	28.5 *p=8E-03*
503.4	**451.5** ***p=2E-06***	**295.5** ***p=2E-12***	**98.3**	**77.3** ***p=3E-02***	**44.3** ***p=3E-04***

[A] In-stream settling loads are not calibrated, so values are uncertain.
[B] Time above 50 mg/L and 100 mg/L are not calibrated, but validation is shown in Figure S4.

Table S8. Percentage of streamflow and in-stream TSS load occurring during small, medium, and large storms. Storm size was determined based on peak streamflow during the Baseline scenario (see Figure 7).

Scenario	Percent in-stream TSS load occurring during		
	Small storms [A] *n* = 57	Medium storms [B] *n* = 50	Large storms [C] *n* = 6
Baseline	3.5	57.1	39.4
Detention to Infiltration	3.6	57.1	39.3
30% Infiltration	2.6	53.9	43.5

[A] Small storms had peak flow <10 m^3/s during the Baseline scenario.
[B] Medium storms had peak flow between 10 and 75 m^3/s during the Baseline scenario.
[C] Large storms had peak flow >75 m^3/s during the Baseline scenario.

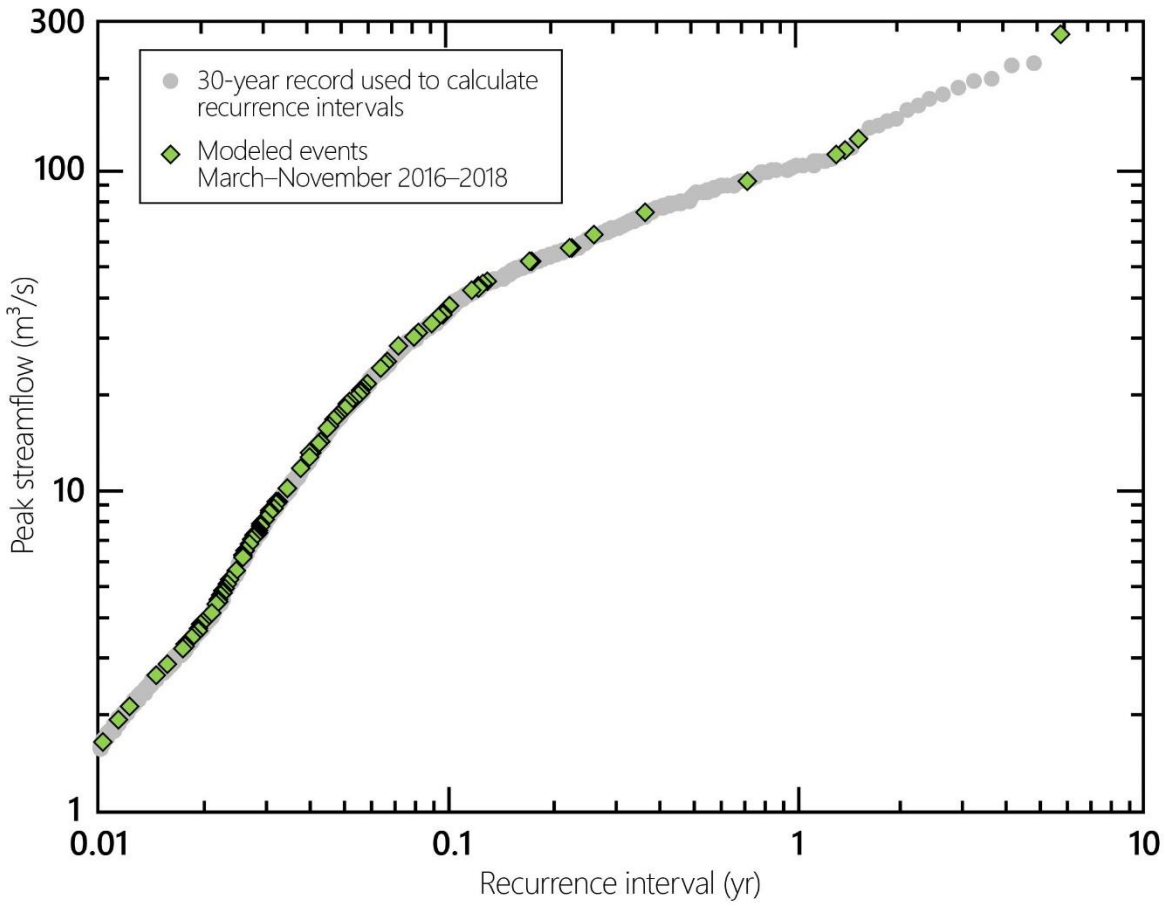

Figure S1. All but one storm modeled this study had a calculated recurrence interval ≤1.5 years. A 29-year partial duration series of all peaks >1.5 m^3/s (baseflow = ~1 m^3/s) was generated from 30-min flow data at the Mouth of Wissahickon Creek. Peaks were then declustered to avoid the impact of multi-peak storms, such that only one peak was selected within any sliding 24-hour window.

To calculate the recurrence interval (in yr) for a given flow x (in m^3/s), we applied these formulas:

$$RI_x = \frac{1}{freq_x} \qquad freq_x = \frac{\#\ of\ occurences \geq x}{30\ yr}$$

where RI_x = the recurrence interval of flow x
$freq_x$ = the average frequency of occurrence of flow x

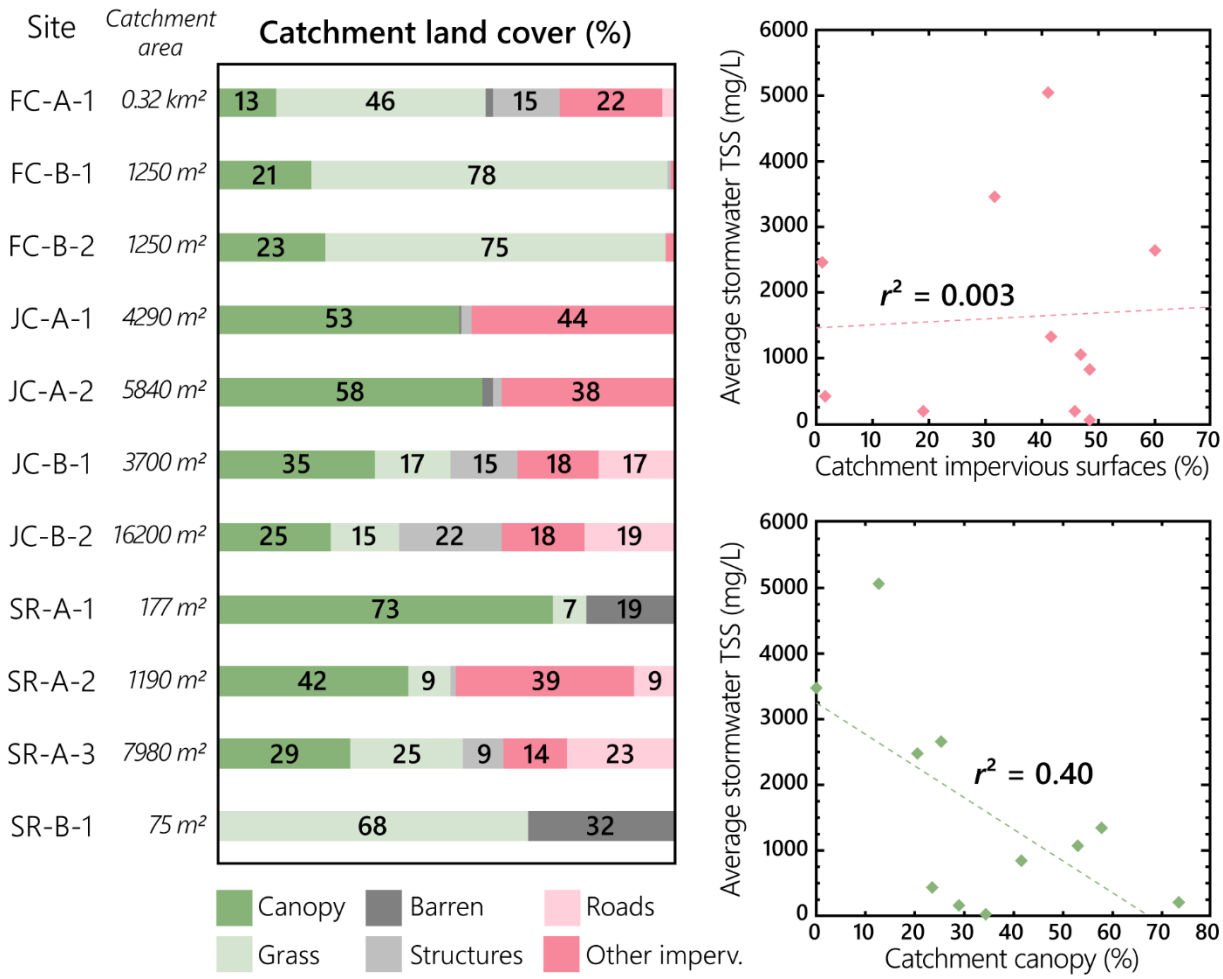

Figure S2. Land cover variability in the catchments draining to each stormwater sample site does not fully explain the variation in observed TSS concentrations. *Left*. Land cover in the catchment for each stormwater sampling site. *Right*. Average stormwater TSS concentration at each site plotted against percent impervious surfaces (sum of barren, structures, roads, and other impervious) and percent canopy in the catchment. While catchment canopy has a weak negative correlation ($r^2 = 0.4$) with average stormwater TSS concentration, no other land cover variable has a significant correlation.

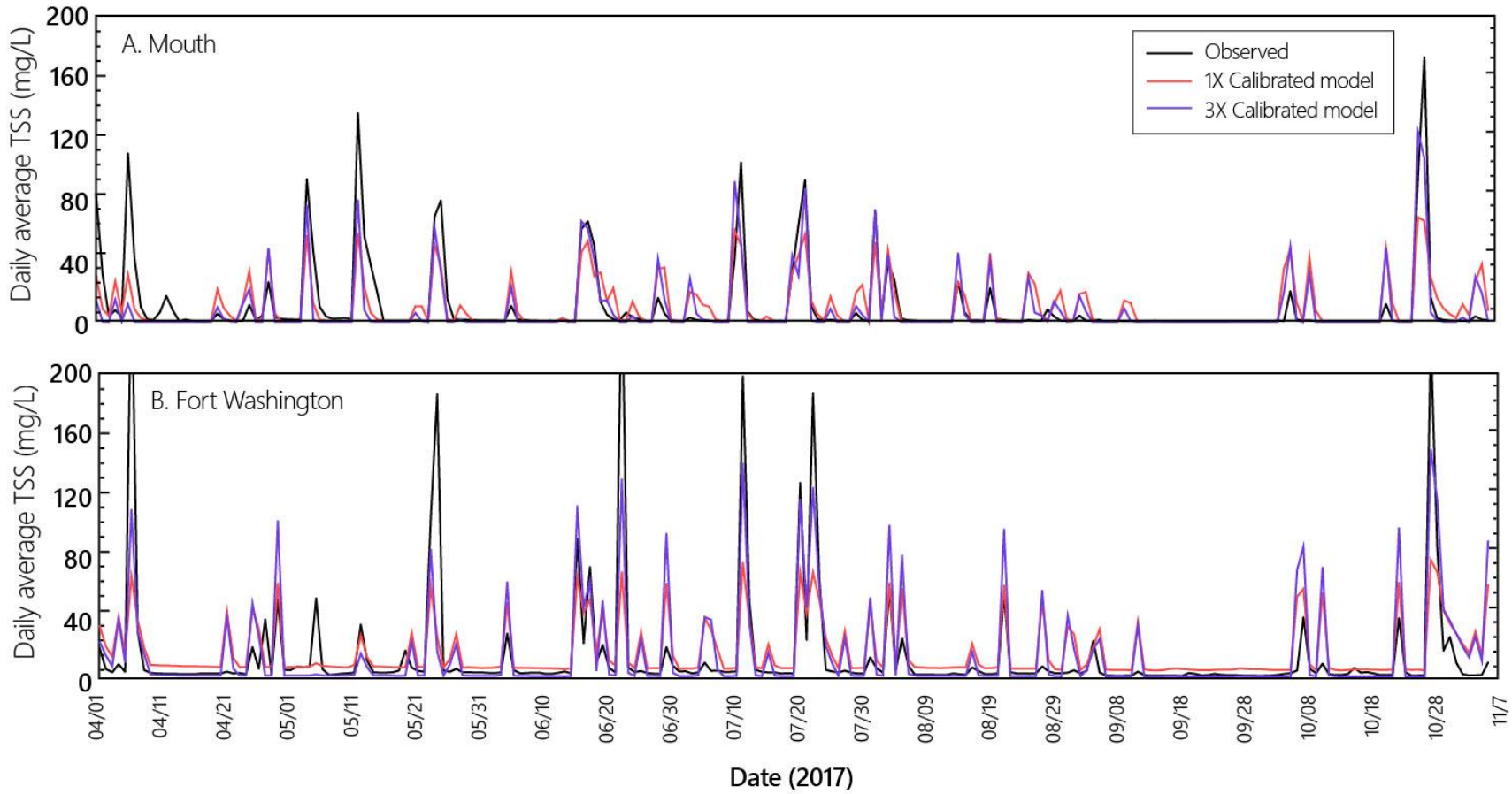

Figure S3. Increasing EMC to three times PADEP values improved TSS calibration. Time-series of daily average TSS concentrations at the Mouth (*A*) and Fort Washington (*B*) for observed records (*black*), calibrated 1X EMC model (*red*) and calibrated 3X EMC model (*purple*).

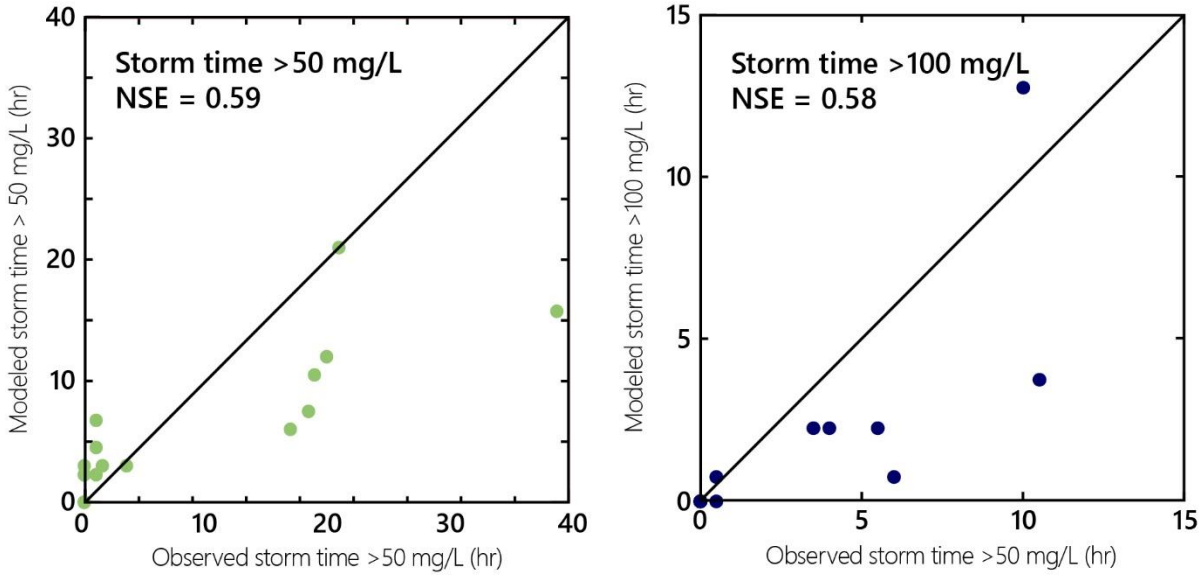

Figure S4. The model represented time above 50 mg/L and time above 100 mg/L during storms reasonably well. Modeled vs. observed time above 50 mg/L (*left*) and time above 100 mg/L (*right*) at the Mouth during storms is shown for the calibrated WASP model.

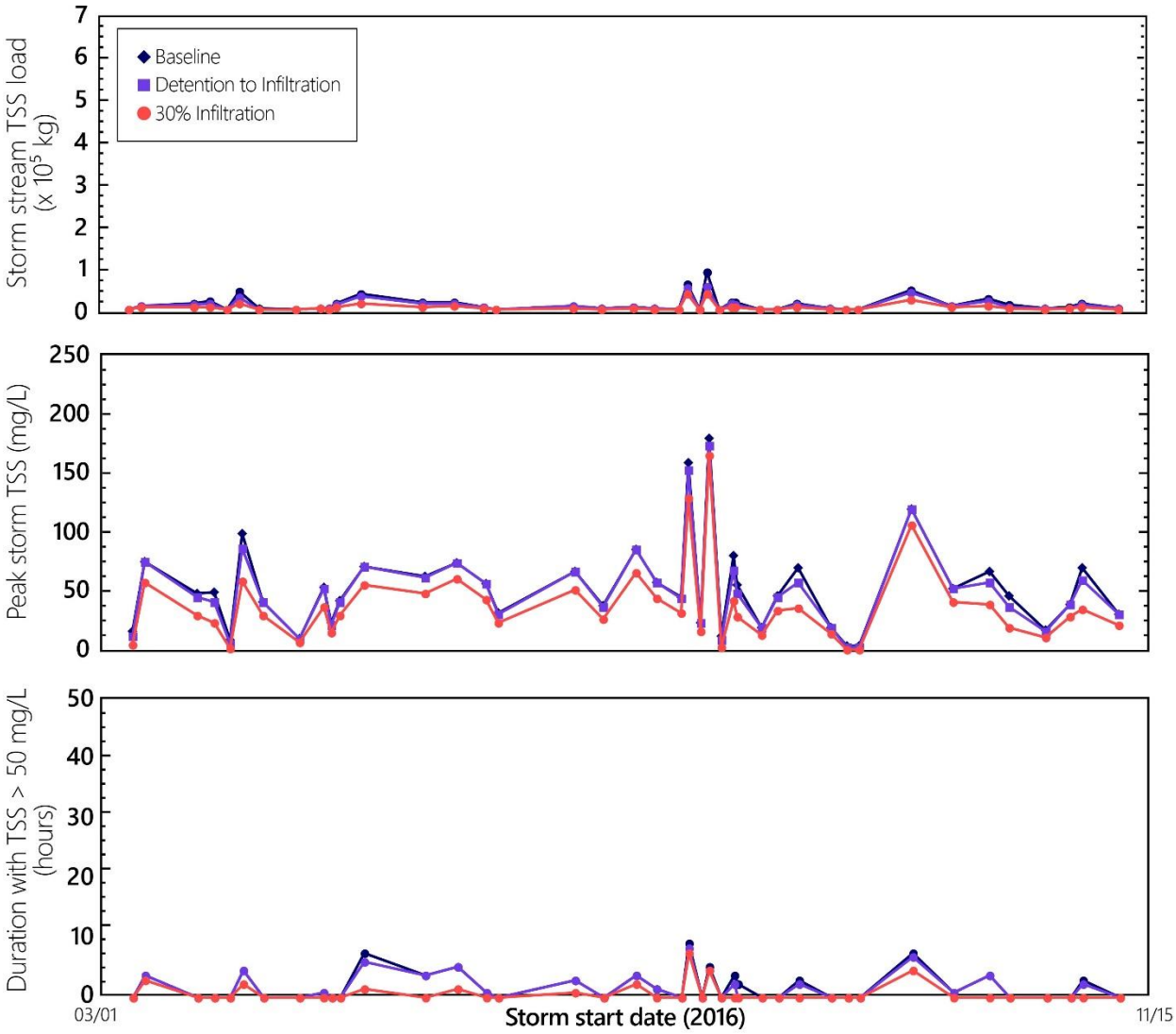

Figure S5. For each storm identified in 2016, the total TSS load exiting the Mouth during that storm is shown (*top*) as is the peak TSS concentration occurring during that storm (*middle*) and duration for which TSS concentration at the Mouth exceeded 50 mg/L for each storm (*bottom*).

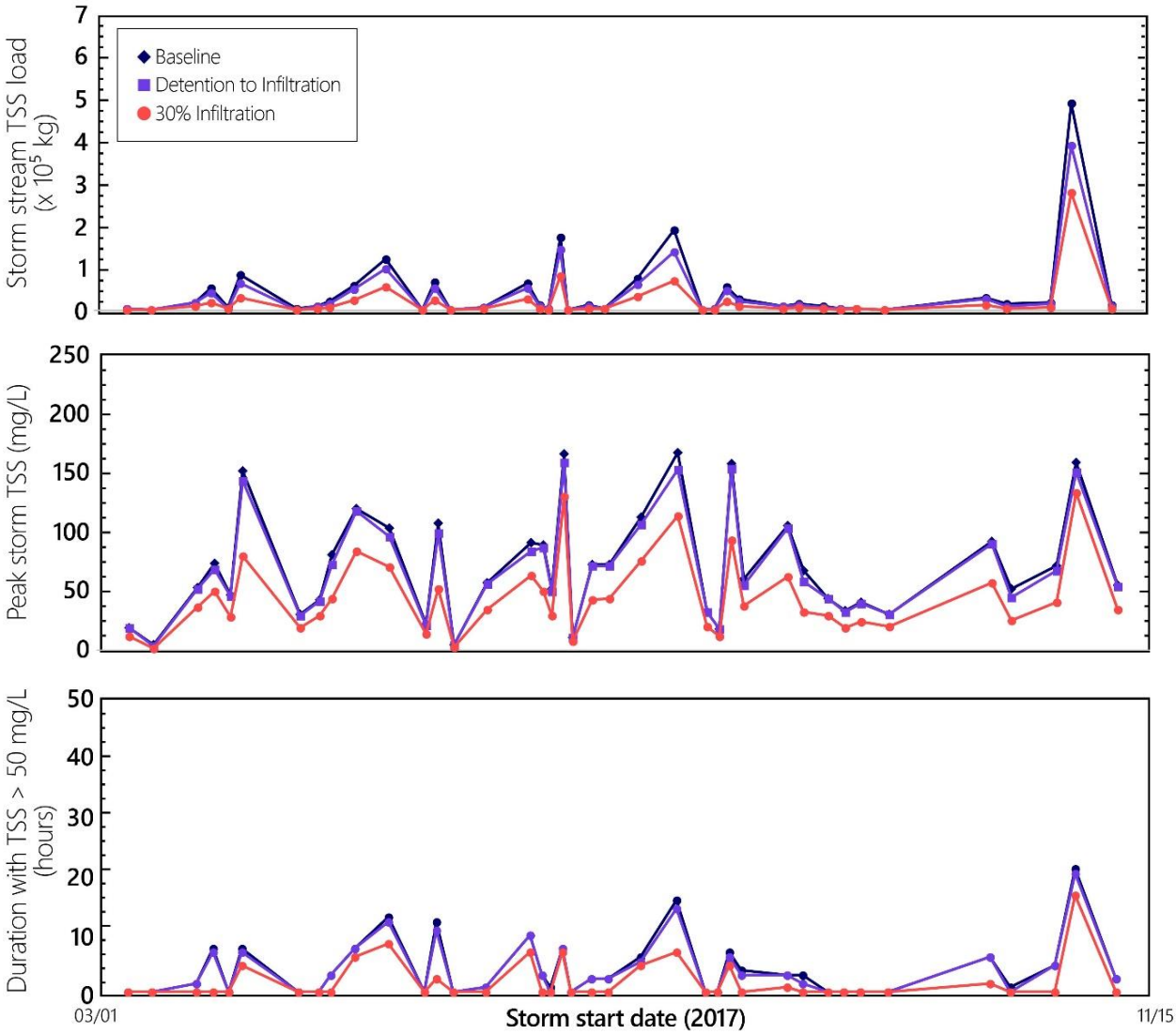

Figure S6. For each storm identified in 2017, the total TSS load exiting the Mouth during that storm is shown (*top*) as is the peak TSS concentration occurring during that storm (*middle*) and duration for which TSS concentration at the Mouth exceeded 50 mg/L for each storm (*bottom*).