

Effects of Watershed Land Use Data on HSPF Water Quality in the Upper Opequon Watershed in northern Virginia, USA

Isaac A. Alukwe^{1,*}, Theo Dillaha²

¹Energy and Environmental Engineering, Mount Kenya University, Thika, Kenya

²Biological Systems Engineering, Virginia Tech, Blacksburg, USA

*Corresponding author: isaaca6@vt.edu

Received June 12, 2014; Revised July 09, 2014; Accepted July 13, 2014

Abstract Land use data source can contribute to errors in watershed modeling. This paper evaluated the effects of using site-specific versus county-level aggregated land use data on Hydrologic Simulation Program-Fortran (HSPF) simulated contaminant losses. Site-specific land use was derived from the local watershed land use inventory while aggregated land use was derived from county-level data (percentage of county land in various land use categories and sub-categories). County level data are useful when modeling large watersheds such as the Chesapeake Bay Watershed when collection and use of site-specific data may be cost prohibitive. The study site was the 14,941 ha predominately rural Upper Opequon Watershed in northern Virginia, USA. Percentage relative errors in model output were calculated and compared using the two land use data sources. Results showed that use of aggregated land use data resulted in 13, 3 and 4 percent higher simulated sediment, and total nitrogen and phosphorus losses, respectively due to overestimated cropland area. The higher contaminant losses would suggest the need for more management measures to meet water quality goals. This study suggests that while the use of county-level aggregated land use data may be appropriate for developing basin scale pollutant reduction goals such as those in total maximum daily load (TMDL) plans, it should be used with extreme caution for watershed planning and implementation activities on smaller watersheds that may mandate site-specific changes in land management and costs for landowners. For smaller watersheds, TMDLs and their watershed implementation plans should utilize local site-specific spatial data that accurately reflects watershed conditions. This will help target resources where they are most needed and maintain credibility with local stakeholders while improving the accuracy of the developed pollution reduction plans.

Keywords: HSPF, water quality modeling, land use, TMDLs, watershed management

Cite This Article: Isaac A. Alukwe, and Theo Dillaha, "Effects of Watershed Land Use Data on HSPF Water Quality in the Upper Opequon Watershed in northern Virginia, USA." *American Journal of Water Resources*, vol. 2, no. 3 (2014): 54-62. doi: 10.12691/ajwr-2-3-1.

1. Introduction

Water quality models are commonly used in predicting watershed nonpoint source (NPS) pollution. In the US, detailed land use, agricultural production, and livestock population data collected by United States Department of Agriculture–National Agricultural Statistics Service (USDA-NASS) are often only available at county and state scales. The data are collected at the farm and/or field-scale, but the data are then aggregated to the county level to meet congressional landowner privacy mandates. These data are extensively used in large watershed-scale models such as the Chesapeake Bay Watershed Model (CBWM). While appropriate at the large watershed-scale, use of aggregated county level data for smaller watersheds, particularly those requiring site-specific TMDLs, may not be appropriate because such data may poorly represent local site-specific conditions. Studies [1,2,3] have

addressed the impacts of geographical information system (GIS) spatial data resolution of site-specific data on model output uncertainty. However, there is considerably less information on uncertainty due to the disaggregation of county-level data to local watersheds requiring watershed management efforts such as TMDLs. This is particularly important for the CBWM, which relies on the use of county-level aggregated land use and BMP data and which some have suggested could be used for the development of local TMDLs. Studies that assess the impacts of using the county-level aggregated spatial data on local watersheds are largely missing. In the Chesapeake Bay Program (CBP), water quality modeling is accomplished at the basin or major tributary scale for planning and modeling, which yields insufficient detail for establishing specific loads or for determining specific, desired management changes at the small or subwatershed scale [4]. For instance, serious concerns have been raised by different stakeholders about local, subwatershed reduction

goals derived from the CBWM-TMDL [5]. The concerns relate to the manner in which the TMDL assumes a level of precision beyond its capability and without regard for the economic consequences to stakeholders. Thus, the objective of this study was to evaluate the effects of different spatial land use data sources on pollutant losses simulated by the HSPF model. Differences in HSPF simulated flow, sediment, total nitrogen (TN) and phosphorus (TP) losses using county-level aggregated land use data which were disaggregated to watershed and sub-basin levels were compared to those from local site-specific land use.

2. Methodology

2.1. Study Area

Upper Opequon Watershed located in northern Virginia was selected as the study watershed due to availability of historical flow, water quality, and land use data. The watershed is 14,941 ha in size and is south of the City of Winchester in Frederick County, Virginia. The Upper Opequon (39°06'30" N; 78°10'30" W) is part of the Opequon Creek Watershed (Figure 1).



Figure 1. Study area –Upper Opequon Watershed, Virginia, USA

Upper Opequon Creek runs for approximately 40 km from its headwaters to its confluence with Abrams Creek, which flows into the Lower Opequon and then into the Potomac River, which empties into the Chesapeake Bay [6]. The Upper Opequon watershed was delineated into 16 subwatersheds (Figure 2).

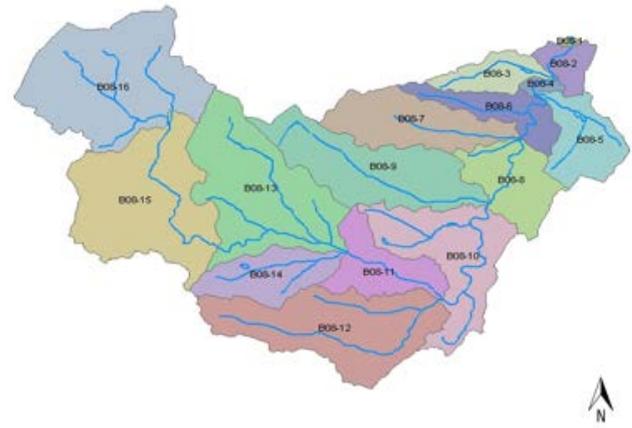


Figure 2. Subwatershed delineation in Upper Opequon Watershed

2.2. Description of HSPF Model

The HSPF model is the core model of the CBWM and is supported by U.S. Environmental Protection Agency (USEPA) and U.S. Geological Survey (USGS) [7]. The HSPF model is a comprehensive watershed model and is commonly used for simulating point and NPS pollution and for developing TMDL plans for impaired water bodies [8]. The HSPF model simulates hydrology and water quality constituents, including sediment, bacteria, pesticides, nutrients (nitrogen, N and phosphorus, P) [9].

Cropland can be simulated with a detailed NPS simulation module known as agricultural chemicals (AGCHEM) that includes application of fertilizer and manures, atmospheric deposition, crop uptake, soil binding, denitrification, and surface export [10,11]. The AGCHEM sub-modules are process-based and include PEST (pesticides), NITR (nitrogen), and PHOS (phosphorus) and MSTLAY (soil moisture). The sub-modules provide detailed simulation of complex nutrient balances and transport in the soil profile [12,13]. Comprehensive nutrient balances are important in identifying potential watershed areas for improvements as well as assessing the impact on water quality resulting from installed BMPs and land use changes [10,14]. Extensive inputs and parameters are required during AGCHEM application to represent physical, chemical and biological processes in the soil profile [15]. A simplified pollutant tracking module, PQUAL, which uses a simple regression relationship, can also be used to simulate pollutant losses from pervious areas but it does not consider nutrient cycling and balances. In this paper, three sub-modules of AGCHEM (MSTLAY, NITR and PHOS) were utilized in accordance with HSPF Users' Manual guidance [12].

Table 1. Model input weather data stations

Type of Data	Location	Source	Frequency of Recording	Period of Record	Latitude Longitude
Percentage of Possible Sun, Dew Point, Wind Speed	Washington Reagan National Airport	NCDC (USW00013743)	1 Day	1945 – present	35°52' N 77°02' W
Hourly Rainfall	Star Tannery	NCDC (USC00448046)	1 Hour	1982 - present	39°4' N 78°25' W
Daily Rainfall, Min and Max Air Temp	Winchester	NCDC (USC00449181)	1 Day	1982 – present	39°11' N 78°07' W

2.3. Input Weather Data

Weather input data were obtained from the National Climatic Data Center (NCDC) weather stations (Table 1) [16]. The data required for creating the HSPF weather data files included hourly precipitation, average daily temperatures (maximum, minimum and dew point), average daily wind speed, and percentage sun.

2.4. Observed Flow and Water Quality Data

The available observed monthly water quality data included total suspended solids (TSS), nitrate (NO₃-N) and phosphorus (PO₄-P) concentrations as shown in Table 2.

Table 2. Flow and Water Quality Observed Data

Data type	Source	Time-step	Data availability
Stream Flow	USGS	Daily	Jan. 1980- present
TSS	VADEQ	Monthly	Aug. 1991-Apr. 2007
NO ₃ -N	VADEQ	Monthly	Aug. 1991-Dec. 2007
PO ₄ -P	VADEQ	Monthly	Aug.1991-Dec. 2007

Courtesy of VADEQ (Sieber, 2011 (personal communication))

2.5. Representation of Nutrients and Crops in HSPF-AGCHEM Model

Crop rotation in the watershed’s cropland is typically a six-year cycle consisting of corn, soybeans, and four years of hay. Ninety percent of the corn/soybean acreage was simulated as receiving manure following discussions with the Lord Fairfax Soil and Water Conservation District (LFSWCD) extension specialist (Rau 2012, personal communication). Model parameters for planting, harvesting, fertilizer and manure application rates and dates, plant uptake of nutrients and erosion-related process were developed and included in the model user control input (UCI) file.

The Upper Opequon Watershed is predominantly a beef producing watershed, it was assumed that no commercial fertilizer was applied on pasture. Fertilizer was only applied to field crops (corn, soybeans, winter cover crops and rotational hay) using application rates from the [17]

and USDA’s Agricultural Census data (1992-2007) for Frederick County, VA. For cropland, N and P were applied in April and October. Urban and residential lawn fertilizer application rates were obtained from the Phase 4.3 CBWM [18]. Atmospheric deposition nutrient inputs were supplied to the model via the HSPF-WDMUtil (Weather Data Management Support Tool) over the simulation period. Both the dry and wet deposition time series data for Frederick County, VA were available from the CBWM [10,19] for the period 1984 -2005 and these were extrapolated for the simulation period using the BASINS WDM utility tool.

The annual cattle confinement schedule for the study watershed was obtained from VADCR and VADEQ (2003a). Manure application rates on cropland, application rates were estimated using reported values from the bacterial TMDL report for the Upper Opequon Watershed [6].

Solid manure accumulated during confinement was applied to cropland January through May (prior to planting) and October through November after harvest. There were no dairies in the watershed. During spring, manure from cattle and poultry litter were surface applied to hay and no-till corn, and incorporated into the soil for corn in conventional tillage. Fall manure application was incorporated into the soil, and surface-applied to cropland under rotational hay. According to [6], there were 1,060 cow/calf pairs on pasture in the Upper Opequon Watershed.

The actual number of beef cattle varied throughout the year due to the presence and absence of calves. Cattle were distributed in subwatersheds 6, 9, 10, 15 and 16 as 20, 100, 100, 446 and 394 cow/calf pairs, respectively. These were figures for 2003 and coincide with the 2002/03 land use data used in the study. Manure loading on pasture was estimated by multiplying the total number of cow/calf pairs on pasture by the amount of manure produced, 81 kg/ha-day [20]. During confinement, solid manure produced by the cattle was collected for land application. Manure produced in each subwatershed was estimated based on the populations of beef cattle in the subwatershed and their confinement schedules, and land applied to cropland and pasture (Table 3).

Table 3. Estimated monthly manure N and P inputs to cropland and pasture

Month	Cropland				Pasture			
	NH ₄ -N (kg/ha)	Org-N (kg/ha)	PO ₄ -P (kg/ha)	Org-P (kg/ha)	NH ₄ -N (kg/ha)	Org-N (kg/ha)	PO ₄ -P (kg/ha)	Org-P (kg/ha)
J	0.00	0.00	0.00	0.00	0.85-5.02	1.17-6.81	0.03-1.63	0.03-1.63
F	18.46	24.43	5.83	5.83	0.76-4.53	1.08-6.19	0.05-2.51	0.05-2.51
M	36.92	48.87	14.55	14.55	0.11-0.61	0.58-3.41	0.03-1.52	0.03-1.52
A	36.92	48.87	14.55	14.55	0.11-0.61	0.13-0.82	0.01-0.51	0.01-0.51
M	14.77	19.55	11.63	11.63	0.11-0.62	0.15-0.84	0.01-0.51	0.01-0.51
J	3.70	4.89	2.91	2.91	0.11-0.58	0.13-0.80	0.02-0.76	0.02-0.76
J	0.00	0.00	0.00	0.00	0.43-2.40	0.58-3.30	0.01-0.51	0.01-0.51
A	0.00	0.00	0.00	0.00	0.43-2.40	0.58-3.30	0.02-1.01	0.02-1.01
S	0.00	0.00	0.00	0.00	0.43-2.40	0.56-3.27	0.00-0.13	0.00-0.13
O	5.62	7.44	8.85	8.85	0.21-1.24	0.29-1.69	0.01-0.51	0.01-0.51
N	5.62	7.44	8.85	8.85	0.21-1.21	0.28-1.65	0.08-4.04	0.08-4.04
D	0.00	0.00	0.00	0.00	0.21-1.26	0.29-1.70	0.03-1.63	0.03-1.63

2.6. Modeling Watershed Spatial Scales of Land Use Data

In 2003, Virginia Tech [6,21] verified the Upper Opequon site-specific land use based on the Virginia Department of Conservation and Recreation (VADCR) 1995-1997 aerial photographs and NLCD (2002) data. The VADCR identified 24 site-specific land uses in the Upper Opequon Watershed. This paper consolidated the 24 land uses into 6 categories (cropland, pasture, loafing lot, forest, residential, urban) for modeling purposes. These data provided the site-specific land use dataset for the Upper Opequon Watershed.

The Upper Opequon Watershed county-level aggregated land use data were developed for 2002 and were obtained from the CBP Phase 5.32 model database [22]. The county-level aggregated data were then disaggregated to subwatersheds based on the subwatershed to watershed area ratios. The distribution to the watershed was necessary because the county scale of CBWM output limits the development of targeted management actions at a finer spatial scale [4]. The county-wide aggregated land uses in the CBWM database had 25 land use categories. The CBWM combines GIS spatial data and Classification and Regression Tree (CART) analysis to generate the Bay-wide land use data. The data sources used are National Land Cover Database (NLCD) imagery, USDA-Agricultural Census data by county, NOAA Coastal Change Analysis Program (CCAP), and Landsat Raster Reflectance [10]. Many of the CBWM agricultural categories were not strictly land uses. Some land uses also reflected BMPs, and these did not strictly equate to spatially-derived land use categories. The 25 CBWM land uses were reclassified and consolidated into 6 broader agricultural and urban/residential land use categories to match the site-specific land uses.

2.7. Statistical Analysis

The model performance criteria was evaluated using the following equations for percent bias (PBIAS), the coefficient of determination (R^2), and the Nash-Sutcliffe Efficiency (NSE):

$$PBIAS = \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim}) * 100}{\sum_{i=1}^n Y_i^{obs}} \right] \quad (1)$$

where Y_i^{obs} is the i th observation for the constituent being evaluated, Y_i^{sim} is the i th simulated value for the constituent being evaluated, and n is the total number of observations.

$$R^2 = \left[\frac{\sum_{i=1}^n [(Y_i^{obs} - Y_{mean}^{obs})(Y_i^{sim} - Y_{mean}^{sim})]}{\left[\sum_{i=1}^n (Y_i^{obs} - Y_{mean}^{obs})^2 \right]^{1/2} \left[\sum_{i=1}^n (Y_i^{sim} - Y_{mean}^{sim})^2 \right]^{1/2}} \right] \quad (2)$$

where Y_{mean}^{obs} and Y_{mean}^{sim} are the means of the observed and simulated data, respectively, for the constituent being evaluated.

$$NSE = \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - Y_{mean}^{obs})^2} \right] \quad (3)$$

The Percent bias (PBIAS) (measures the tendency of the simulated data to deviate from the observed values [23], coefficient of determination (R^2) which measures the level of co-linearity between simulated and observed data [24] and Nash-Sutcliffe Efficiency (NSE) (determines the relative magnitude of the residual variance compared to the variance of observed data [25]. [26] recommended a general satisfactory model simulation if $NSE > 0.50$ and $PBIAS \pm 25\%$ for daily streamflow, $PBIAS \pm 55\%$ sediment, and $PBIAS \pm 70\%$ for nitrogen and phosphorus, respectively at monthly time-steps. Donigian (2000) suggested that PBIAS be considered satisfactory if $\pm 25\%$ for streamflow, $\pm 45\%$ for sediment and $\pm 35\%$ for nutrients, respectively at monthly time-steps for the HSPF model. Positive PBIAS values suggest model under-prediction while negative values represent systematic model over-prediction of the concerned parameter[26]. [27,28,29] suggested that R^2 value of at least 0.50 is acceptable for satisfactory model performance.

Model predictions were compared based on outputs from the site-specific and disaggregated land use data for the period 1988 to 2009. Differences in modeled constituents were quantified as percentage relative errors (RE) using the following equation:

$$RE = \frac{LU_d - LU_{ss}}{LU_{ss}} * 100\% \quad (4)$$

where LU_d = modeled output from disaggregated land use and LU_{ss} = the modeled output from site-specific land use. Bench-mark conditions were based on the site-specific land use as it represented the local conditions in the watershed. Statistical significance tests were performed using the Wilcoxon signed-rank test at 95% significance level due to its wide application in watershed studies [30-35].

3. Results and Discussions

3.1. Model Calibration and Validation

The Upper Opequon HSPF model hydrology was calibrated for the period 1987 to 1992 and validated from 1993-1997 in the 2003 Opequon Creek bacteria TMDLs [6,21] at a daily time-step. Thus, hydrology recalibration was not necessary and the existing 2003 bacterial TMDLs calibrated hydrology parameter values were judged adequate based on 1998-2002 hydrology validation run results. For sediment, TN and TP sensitivity analysis was used to identify candidate parameters for calibration (1988-1997) and validation (1998-2002). A one year data period, January to December 1987, prior to the calibration period was used utilized as a dynamic model initialization/spin-up period to allow for stabilization of model state variables. Simulated data from this period was discarded and not used for model calibration or validation. Sensitive parameters were identified via calculating relative sensitive indices and ranking them [36,37]. Sensitivity analysis characterizes the impact of changes in model inputs on the model outputs [38].

Table 4 provides a summary of model evaluation results for flow, TSS, NO₃-N and PO₄-P outputs.

Table 4. Model evaluation results for daily flow and monthly water quality constituents

	Daily flow statistics					
	Calibration			Validation (1998 – 2002)		
	PBIAS	R ²	NSE	PBIAS	R ²	NSE
Remarks	-	-	-	7.34	0.60	0.65
	-	-	-	Very good	Acceptable	Good
Monthly statistics for water quality constituents						
Constituent	Calibration (1988 – 1997)			Validation (1998 – 2002)		
	PBIAS	R ²	NSE	PBIAS	R ²	NSE
TSS	19	0.63	-0.74	-18	0.67	-0.97
Remarks	Good	Acceptable	Not Acceptable	Good	Acceptable	Not Acceptable
NO ₃ -N	6	0.50	-0.84	-4	0.50	-1.05
Remarks	Very good	Acceptable	Not Acceptable	Very good	Acceptable	Not Acceptable
PO ₄ -P	-37	0.89	-1.19	46	0.86	-0.56
Remarks	Good	Acceptable	Not Acceptable	Satisfactory	Acceptable	Not Acceptable

Generally, based on the criteria defined in section 2.7 and the land use distribution over the subwatersheds for the two land use categories, the model performance was satisfactory except for the NSE statistics for water quality constituents calibrated and validated. The negative NSE values <0.0 indicate that the mean observed value is a better predictor than the simulated value, which indicates unacceptable performance. The NSE statistic is sensitive to extreme values, which were observed in the monitored data for all the simulated water quality constituents. For instance, high TSS values were reported even on days with no or very little precipitation. The corresponding generated flow on such occasions was barely available to transport sediment downstream. Such data points might have resulted from some upstream disturbance during sampling, such as livestock in the stream and and/or extremely high algae concentrations during low flow periods. High extreme values (7 out of 56 observed data points) not matching rainfall occurrence were treated as outliers and removed from the observed data time-series but still 10 extreme data points of the 56 observed values were kept to maintain a sufficient sample size.

The poor NSE statistic performance can be attributed to its sensitivity to high extreme values. [39] reported that NSE is very sensitive to extreme values due to their squared differences. A review by [26] reported NSE values of -3.35 and -2.46 for flow and sediment, respectively at a monthly time-step. A study by [40] reported NSE value of 0.05 for a stream flow simulation at a monthly time step. In all the cases the NSE statistic was attributed to sensitivity associated with input data deficiencies. Better results can be obtained through the use of more detailed, complete and more accurate data [40]. Additionally, the HSPF model was observed to over predict nutrient concentrations during extreme dry weather flows of less than 12 mm. A similar observation was also reported in bacteria HSPF simulations under low-flow conditions [41,42,43].

In the case of bacteria simulations, the HSPF model is modified by introducing a “cut-off stage” which simply stops the model from simulating bacteria loads from direct deposit sources[41,42]. This approach could not work for nutrient simulations and therefore this study concluded that over-prediction during low flow conditions is a limitation in HSPF that requires further research. However, this model limitation did not affect the expected modeled total nutrient loads, because the model calculates mass as

a product of concentration and discharge (volume) at a given point in time, and therefore the high concentration at low flow conditions did not affect the pollutant load predictions.

3.2. Annual Modeled Flow and Water Quality Responses

The comparison of land use distribution across the subwatersheds of the study area showed mixed results with both major and minor differences observed in land use areas for both data sources (site-specific and disaggregated). The differences varied across the subwatersheds. At the watershed level, these differences in area were minor except for cropland. The cropland area for the entire watershed differed by 56% based on the site-specific and disaggregated land use areas of 805 and 1,258 ha, respectively. Loafing lots differed by 1,381% though their areas were 1 and 16 ha for site-specific and disaggregated land use categories, respectively. Large differences in cropland and loafing lots areas can result in significant variations in modeled loadings as they are major sources of nutrient loads per unit area. For the watershed as a whole, the overall difference in total area was 43.5 ha (based on the total site-specific land use area of 14,941 ha and disaggregated land use area of 14,898 ha) for the two land use data categories. This difference represented about 0.30% of the total watershed area which is negligible when compared to the total site-specific watershed area of 14,941 ha. The minor difference in the total area resulted from the different approaches used in deriving the land use categories (section 2.6). These results imply that with increasing watershed scale, disaggregated land use data are appropriate for basin-scale modeling but may not be accurate at a local scale. The effects of land use area differences on average annual simulated flow and loadings from the two land use assessments are given in Table 5.

Simulated sediment loads at the watershed outlet from cropland and loafing lot had the largest variation with RE of 105 and 225%, respectively (Table 5). The large RE in sediment loss from cropland would be expected as the land use had a difference of 453 ha representing 56% of the 805 ha and 1,258 ha of the site-specific and disaggregated land use data, respectively. The rest of the land uses showed the site-specific predictions being either slightly more or less than outputs from disaggregated land

use data. Results in Table 5, show Total N and TP REs are similarly higher for cropland and loafing lots. The RE of loss of TN from cropland was 17% while that of TP is 33%, respectively. Similarly, RE of total N and TP losses from loafing lots are 111% and 63%, respectively. However, the differences in simulated pollutant losses from the loafing lots at the watershed scale are of little significance because little land (site-specific (1 ha) and disaggregated (16 ha)) is involved. Compared to the other simulated constituents, sediment has the largest RE of

13% (Table 5) at the outlet of the watershed with most of the difference due to at the cropland losses. Total N and TP had REs of 3 and 4%, respectively at the watershed level. Flow has RE of 5% at the outlet of the watershed. Results in Table 5 show that modeled predictions from both land use datasets balance out at the watershed scale. The variations in simulated constituents were observed to correspond to differences in land use areas, primarily due to differences in loadings from cropland.

Table 5. Effects of land use on average annual simulated flow, sediment, TN and TP loadings at watershed outlet

Land Use	Flow (mm)			Sediment (Mg/yr)			TN (kg/yr)			TP (kg/yr)		
	site-spec ¹	Disagg ²	RE*	site-spec	disagg	RE	site-spec	disagg	RE	site-spec	disagg	RE
Cropland	51	82	60.8	2,879	5,906	105.1	69,198	80,932	17.0	6,816	9,028	32.5
Pasture	94	93	(1.1)	6,370	6,227	(2.2)	85,955	85,873	(0.1)	10,850	10,754	(0.9)
Loafing Lot	30	41	36.7	4	13	225.0	10,156	21,415	110.9	1,719	2,799	62.8
Forest	97	92	(5.2)	8,241	7,831	(5.0)	71,630	67,446	(5.8)	9,612	8,514	(11.4)
Urban	72	63	(12.5)	148	116	(21.6)	67,560	68,711	1.7	8,952	8,897	(0.6)
Residential	72	67	(6.9)	708	567	(19.9)	79,244	70,424	(11.1)	8,134	8,081	(0.7)
Total	416	438	5.3	18,350	20,660	12.6	383,743	394,801	2.9	46,083	48,073	4.3

¹site-specific land use; ²disaggregated land use; *% relative error between sit-specific and disaggregated land uses, and quantities in brackets indicate simulated constituents from site-specific land use are higher than disaggregated land use.

Table 6 presents a summary of RE in annual simulated flow and loads and their corresponding tests of significance differences at the subwatershed level.

For stream flow, most subwatersheds generally had differences that were less than 10%, except cropland (61%), loafing lots (37%) and urban land uses (13%). The differences in flow due to land use source were significantly different in all subwatersheds except subwatershed 6 ($p = 0.095$). County-level disaggregated land use data produced higher loads than site-specific land use data. Sediment has maximum and minimum differences of 20% in subwatershed 3 and 0.2% in subwatershed 15, respectively. Test of significance showed sediment not significantly different in

subwatersheds 8 ($p = 0.6640$) and 9 ($p = 0.2340$). Total N has its highest difference of 19% in subwatershed 16 and minimum RE of -0.3% in subwatershed 12, and the load differences were significantly different in all subwatersheds ($p < 0.0001$). In subwatersheds 12 and 14, modeled TN from site-specific land use data exceeded the disaggregated land use output by -0.3 and -3.2%, respectively. Total P had its maximum and minimum differences of 22% and 2% in subwatersheds 16 and 14, respectively. Total P loads were significantly different except in subwatersheds 9 ($p = 0.485$) and 14 ($p = 1.000$). In general, differences in flow and pollutant losses were significantly different in almost every subwatershed with the two land use data sources.

Table 6. Effects of site-specific and disaggregated land use on flow, sediment, TN and TP at subwatershed level ($\alpha = 0.05$)

Subwatershed	Flow		Sediment		Total N		Total P	
	difference (mm)	RE	difference (Mg-sed)	RE	difference (kg-TN)	RE	difference (kg-TP)	RE
1^a	22	5.3	2,310	12.6	11,058	2.9	1,990	4.3
2	(6.4)	3.0	2,222	12.3	13,946	(3.7)	1,651	4.2
3	3.5	3.3	112	19.9	202	4.5	65	10.3
4	5.5	2.4	1,953	11.0	9,589	7.6	1,685	9.3
5	1.0	7.3	115	12.3	659	9.3	93	9.4
6	(0.7)	(0.2)*	1,835	0.5	8,792	7.2	1,571	9.1
7	(2.2)	(8.8)	432	2.2	1,422	14.2	256	19.8
8	1.0	0.3	1,212	0.4*	5,972	5.3	1,107	6.8
9	(3.4)	(9.2)	241	0.6*	655	3.5	148	2.1*
10	3.5	1.4	866	0.3	4,210	4.6	774	5.7
11	2.3	1.2	781	0.4	3,761	5.9	835	8.9
12	(0.5)	(1.3)	311	0.6	-63	(0.3)	878	9.4
13	3.3	2.5	366	0.2	4,860	6.6	2.6	7.6
14	(2.1)	(11.5)	263	1.7	-303	(3.2)	25	2.0*
15	5.3	6.0	193	0.2	4,278	7.9	731	10.9
16	2.9	6.2	432	0.7	4,869	18.7	726	21.5

^a watershed outlet, *indicates output not significantly different, RE = percentage relative error from equation (1) and quantities in brackets indicate simulated constituents from site-specific land use are higher than disaggregated land use.

3.3. Implications of Load Variations on TMDL Implementation Strategies

In general, TMDL plans are selected and implemented based on the long-term average annual loadings. Most TMDLs are developed with a 10% margin of safety (MOS) [6,44,45,46,47]. In this study, the difference of approximately 13% in sediment loads at the watershed outlet exceeds the commonly used 10% MOS and suggests that differences due to land use sources may need to be considered. For TN and TP, the relative errors at the watershed outlet were 3 and 4%, respectively, which are generally less than the 10% MOS limit. The results suggest that the planning process based on simulated loads obtained from disaggregated land use data sources may result in the developing of more management plans over the actual practices required. Additionally, the higher loads with disaggregated land use when compared to site-specific land use have direct implications on watershed pollution control management cost estimates for the study watershed. The costs emanate from the number of additional BMP units required to meet target loads. For instance, estimated loading produced using disaggregated data would require more area with nutrient management plans (NMPs) and cover crop BMPs and more streamside fencing to achieve the target loadings. There are also extra unit cost estimates for design and installation, and hiring personnel for educational outreach. The cost estimates would be less in the present study with site-specific land use data, which resulted in lower loadings and BMP needs.

To illustrate this point, NMPs are used to show the effect on probable costs that might be incurred if watershed implementation plans (WIPs) are developed on the basis of modeled disaggregated land use outputs. The average annual loads for TN from each land use distribution are 383,743 and 394,801 kg-TN/yr for site-specific and disaggregated land use, respectively (Table 5). The corresponding TP loads are 46,083 and 48,083 kg-TP/yr, respectively. Estimated costs for implementing NMPs on cropland include the cost of equipment and labor for soil testing and possibly manure storage, hiring a consultant to design the plan, and the costs of applying nutrients in a different manner. There is also the possibility that NMPs will reduce farming costs due to reduced nutrient application costs but analysis focuses on the additional costs associated with cost-shared NMPs. Assuming a 3-year useful life for a plan once it is developed, and including the costs of soil testing, implementation and in some cases cost savings and yield increases, net cost estimates range from -\$74.10/ha-yr (that is, a net cost savings due to cost savings from manure transportation out of the farm) to \$34.58/ha-yr in 2001 dollars [48,49].

In this paper, a cost estimate of \$34.58/ha-yr is used, the worst case possible. This cost estimate does not specifically refer to the practice in the study watershed, and thus is only used for comparison purposes. Assuming that NMPs are implemented on cropland, Table 4 provides 56% (453 ha) more cropland area as a result of using disaggregated land use. Additional costs of implementing NMPs on cropland would be \$15,665/yr. The additional cost to meet target TN loads for cropland is \$1.42 /kg-TN/yr reduced. Thus, errors in assessing the correct land

use distribution in a watershed can affect pollution abatement estimates during planning and can have other implications in terms of modeled loadings. The CBP has considered using county-level aggregated land use data in modeling local watersheds to develop local TMDLs. The choice of using county-level aggregated land use data should be determined by comparing the costs associated with acquiring and utilizing site-specific land use data, which will provide more accurate NPS pollution control assessment and targeting of limited pollution control resources with the potentially inflated control costs with disaggregated land use data. Generally, this assessment implies that using disaggregated data results in more uncertainty during TMDL development that must be addressed during the development of the implementation plan. Saving costs during TMDL development may result in less confidence in required TMDL loads. This would result in more effort (higher WIPs costs) during implementation planning to better characterize the uncertainty and develop attainable WIPs. In the study watershed, the use of disaggregated land use data (compared to the site-specific land use) overestimated cropland and loafing lot areas and pollutant losses. However, this may not be the case for other watersheds. For example, watersheds with high proportions of cropland might have less cropland using county aggregated data and thus end up with lower estimated losses.

4. Conclusions and Recommendations

From this research, the modeled responses showed that sediment, TN and TP were all affected by the differences in the land use data sources. Disaggregated county-level land use data provided higher cropland and loafing lot estimated areas than site-specific land use for the study watershed, which contributed to higher simulated loadings using disaggregated data. It is concluded that for smaller watersheds, depending on the accuracy required for modeled responses, use of the disaggregated data is less desirable than the use of site-specific data. If good resolution and quality local satellite or other land use imagery is available, combining it with the aggregated data may provide a better option than using disaggregated data alone and future research should investigate how this can be done in an effective manner. Online tools such as google earth, google maps and bing search tools are likely to provide useful supplementary land use data information for a given location. Additionally, local land use data in broad categories of urban, forest, and agricultural are almost always available from remote sensed data sources. These data might be combined with disaggregated county data on agricultural land uses to better represent different agricultural land use subcategories (cropland, loafing lots, pastureland, hay land, etc.) in the watershed. Generally, county-level land use data disaggregated to sub-county watersheds based solely on watershed area do adequately represent site-specific watershed conditions for quality planning purposes. The study suggests that site-specific land use data are preferred during TMDL implementation planning to maintain credibility with local stakeholders and improve the accuracy of the developed WIPs.

Acknowledgements

This work was supported by Biological Systems Engineering Department at Virginia Tech, Blacksburg, VA. We appreciate the support and help of Karen Kline.

List of Abbreviations

AGCHEM	AGCHEM is the HSPF module that allows users to model nutrients as part of the processes that take place in pervious segments of watersheds
BMPs	Best Management Practices which refer to reasonable and cost-effective means to reduce the likelihood of pollutants entering a water body
BASINS	Better Assessment Science Integrating Point and Nonpoint Sources, an environmental modeling package that integrates a geographical information system (GIS) and nation-wide watershed data, used for environmental assessment and modeling
CBWM	Chesapeake Bay Watershed Model or simply Bay Model
CBP	Chesapeake Bay Program
HSPF	Hydrologic Simulation Program-FORTRAN is a water quality model
IQUAL	HSPF modules that allow users to model the water quality processes occurring on impervious areas
N	Nitrogen, a chemical element
NCDC	U.S. National Climatic Data Center for weather data in the United States
NOAA	U.S. National Oceanic and Atmospheric Administration
NMP	Nutrient Management Plan, a documented farm plan
NPS	Non-Point Source, diffuse and unregulated pollution sources
NSE	Nash-Sutcliffe Efficiency, a statistical data analysis statistic
P	Phosphorus, a chemical element
PBIAS	Percent bias is a statistical analysis metric for data analysis
PQUAL	The simpler empirical HSPF modules that allow users to model the water quality processes occurring in pervious areas
RE	Relative Error, error approximation in some data representing a discrepancy between an exact value and some approximation to it, percentage.
SWCD	Soil and Water Conservation Districts
TMDL	Total Maximum Daily Loads, is a program mandated by the Clean Water Act, and is a watershed management process that integrates watershed planning with water quality assessment and protection. It is the maximum amount of pollution that a waterbody can receive and still meet the applicable water quality standards

TN	Total N - sum of nitrate-nitrogen (NO ₃ -N), nitrite-nitrogen (NO ₂ -N), ammonia-nitrogen (NH ₃ -N) and organically bonded nitrogen
TP	Total P - includes the amount of P in solution (reactive) and in particulate form (adsorbed and organic)
USDA- NASS	United States Department of Agriculture – National Agricultural Statistics Service (Agricultural Census)
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
VADCR	Virginia Department of Conservation and Recreation
VADEQ	Virginia Department of Environmental Quality

References

- [1] Chaubey, I., Sensitivity Analysis, Calibration, and Validation for a Multi-variable SWAT Model. *Journal of the American Water Resources Association*, 41(5): 1077-1089. 2005.
- [2] Eckhardt, K., L. Beuer, and H.G. Frede, Parameter uncertainty and the significance of simulated land use change effects. *Journal of Hydrology*, 273: 164-176. 2003.
- [3] Patil, A., Z. Q. Deng, and R.F. Malone, Input data resolution-induced uncertainty in watershed modeling. *Hydrologic Processes*, 25:2302-2312. 2011.
- [4] Yagow, G., B. Benham, K. Kline, B. Zeckoski, and C. Wallace, A Method for Disaggregating Existing Model Pollutant Loads for Subwatersheds. *Watershed Science Bulletin. Journal of the Association of Watershed & Stormwater Professionals*, 3(1): 49-61. 2012.
- [5] VADCR. Virginia Bay Total Maximum Daily Load, 2012. Available: <http://www.dcr.virginia.gov/vabaytmdl/index.shtml>.
- [6] VADCR and VADEQ. Bacteria TMDLs for Abrams Creek and Upper and Lower Opequon Creek Located in Frederick and Clarke County, Virginia, 2003. Available: <http://www.deq.state.va.us/tmdl/apptmdls/shenrvr/abropebc.pdf>.
- [7] USEPA. Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) v. 3.0 User's Manual, EPA-823-B01-001, Washington, D.C., Office of Water, USEPA, 2001. Available: <http://water.epa.gov/scitech/datait/models/basins/index.cfm>.
- [8] USEPA. Chesapeake Bay Total Maximum Daily Load, 2010. Available: http://www.epa.gov/reg3wapd/pdf/pdf_chesbay/FinalBayTMDL/CBayFinalTMDLExecSumSection1through3_final.pdf.
- [9] Migliaccio, W.K. and P. Srivastava. Hydrologic Components of Watershed-Scale Models. *ASABE 100th Anniversary Soil and Water Invited Review Series*, 1695-1703]. Aug. 2007. Available: <http://trec.ifas.ufl.edu/kwm/documents/Migliaccio-Srivastava.pdf>.
- [10] CBP. Chesapeake Bay Phase 5.3 Community Watershed Model, 2010. Available: http://www.chesapeakebay.net/model_phase5.aspx?menuitem=26169.
- [11] Hunsaker, C.T., C.T. Garten, and P.J. Mulholland. Modeling Nitrogen Cycling in Forested Watersheds of Chesapeake Bay, 1995. Available: www.ntis.gov/search/product.aspx?ABBR=DE95008845.
- [12] Bicknell, B. R., J. C. Imhoff, J. John, L. Kittle, T. H. Jobes, and A. S. Donigan, *Hydrological Simulation Program - Fortran (HSPF Ver.12.2) User's Manual*. Athens, GA: USEPA, 2005. Available: <http://www.epa.gov/ceampubl/swater/hspf/>.
- [13] Im, S., K. Brannan, S. Mostaghimi, and J. Cho, A Comparison of SWAT and HSPF Models for Simulating Hydrologic and Water Quality Responses from an Urbanizing Watershed. *ASABE Annual International Meeting*, 2003. Available: <http://swatmodel.tamu.edu/media/1303/asae032175.pdf>.
- [14] Deliman, P. N., C. E. Ruiz, A. S. Donigan, T. H. Jobes, E. J. Nelson, and C. T. Manwaring, Application of HSPF-AGCHEM Module Within the WMS for the LeSueur Basin. 2001. Available:

- <http://el.erdc.usace.army.mil/wots/pubs.cfm?Topic=TechNote&Code=watqual>.
- [15] Liu, Y., Application and Sensitivity Analysis of a Watershed Model Application using HSPF with the Nutrient Algorithm AGCHEM in Upper Broad Run Watershed, Virginia, in Civil and Environmental Engineering, Virginia Tech [ETDS - PhD Dissertation], 2011.
- [16] NCDC. National Climatic Data Center, 2012. Available: <http://www.ncdc.noaa.gov/oa/ncdc.html>.
- [17] MAWP. The Mid-Atlantic Nutrient Management Handbook, 2006. Available: http://www.mawaterquality.org/capacity_building/mid-atlantic%20nutrient%20management%20handbook/MANMH_complete.pdf.
- [18] CBP, Chesapeake Bay Watershed Model Application and Calculation of Nutrient and Sediment Loadings, 1997, A Report of the Chesapeake Bay Program Modeling Subcommittee: Annapolis, MD.
- [19] CBP. Chesapeake Bay model Tributary Tools, Loads and Landuse Acreage, 2010. Available: <http://www.chesapeakebay.net/tribtools.htm>. [Accessed Jan. 2010].
- [20] MWPS, Livestock Waste Facilities Handbook (II ed.). Ames, Iowa: MidWest Plan Service, Iowa State Univ., 1993.
- [21] VADCR and VADEQ. Opequon Watershed TMDLs for Benthic Impairments: Abrams Creek and Lower Opequon Creek, Frederick and Clarke Counties, Virginia, 2003. Available: <http://www.deq.state.va.us/tmdl/apptmdls/shenrvr/abropebc.pdf>.
- [22] CBP, Chesapeake Bay Phase 5.3 Community Watershed Model Land Use Data., 2012. Available: ftp://ftp.chesapeakebay.net/Modeling/phase5/Phase532/land_use/.
- [23] Gupta, H.V., S. Sorooshian, and P.O. Yapo, Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2): 135-143. 1999.
- [24] Van Liew, M.W., Veith, T. L., Bosch, D. D. and Arnold J. G., Suitability of SWAT for the conservation effects assessment project: A comparison on USDA-ARS experimental watersheds. *Hydrologic Engineering*, 12(2): 173-189. 2007.
- [25] Nash, J.E. and J.V. Sutcliffe, River flow forecasting through conceptual models: Part 1. A discussion of principle. *Hydrology*, 10(3): 282-290. 1970.
- [26] Moriasi, D.N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L., Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3): 885-900. 2007.
- [27] Ramanarayanan, T.S., R. Srinivasan, and J.G. Arnold, Modeling Wister Lake Watershed Using a GIS-Linked Basin Scale Hydrologic/Water Quality Model, in *Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling, National Center for Geographic Information and Analysis*, 1996. Santa Fe, New Mexico.
- [28] Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R. and Hauck, L. M., Validation of the SWAT Model on A large River Basin with Point and Nonpoint Sources. *Journal of the American Water Resources Association*, 37(5): 1169-1188. 2001.
- [29] Singh, J., H.V. Knapp, and M. Demissie. Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT. 2008. Available: <http://www.isws.illinois.edu/pubdoc/CR/ISWSCR2004-08.pdf>.
- [30] Cosgrove, D.M. and J. Taylor, Preliminary Assessment of Hydrogeology and Water Quality in Ground Water in Canyon, County, Idaho, Idaho Water Resources Research Institute, University of Idaho for the Idaho Department Environmental Quality, 2007.
- [31] Dibike, Y.B., Gachon, P., St-Hilaire, A., Ouarda, T. B., Nguyen M. J. and Van T. V., Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada. *Theoretical and Applied Climatology*, 91(1): 149-170. 2008.
- [32] Meentemeyer, R.K., Rank, N. E., Anacker, B. L., Rizzo, D. M. and Cushman, J. H., Influence of Land-Cover Change on the Spread of an Invasive Forest Pathogen. *Ecological Applications*, 18(1): 159-171. 2008.
- [33] Rupert, M.G., Decadal-scale changes of nitrate in ground water of the United States, 1988-2004. *Journal of Environmental Quality*, 37: 240. 2008.
- [34] Thompson, J.A., J.C. Bell, and C.A. Butler, Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma*, 100(1-2): 67-89. 2001.
- [35] Yang, G., Bowling, L. C., Cherkauer, K. A., Pijanowski, B. C. and Niyogi, D., Hydroclimatic Response of Watersheds to Urban Intensity: An Observational and Modeling-Based Analysis for the White River Basin, Indiana. *Journal of Hydrometeorology*, 11(1): 122-138. 2010.
- [36] Fontaine, T.A. and V.M.F. Jacomino, Sensitivity analysis of simulated contaminated sediment transport. *Journal American Water Resources Association*, 33(2): 313-326. 1997.
- [37] Storm, D.E., Dillaha, T. A., Mostaghimi, S. and Shanholtz, V. O., Modeling Phosphorus Transport in Surface Runoff. *Transactions of ASAE*, 31(1): 117-0127. 1988.
- [38] Tang, Y., Reed, P., Wagener, T. and van Werkhoven, K., Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and earth systems science*, 11(2): 793-817. 2007.
- [39] Legates, D.R. and G.J. McCabe, Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1): 233-241. 1999.
- [40] Obiero, J.P.O., M.A. Hassan, and L.O.M. Gumbe, Modelling of Streamflow of a Catchment in Kenya. *Journal of Water Resource and Protection*, 3: 667-677. 2011.
- [41] Hall, K., Comparing Alternative Methods of Simulating Bacteria Concentrations With HSPF under Low-Flow Conditions, in *Biological Systems Engineering*, MSc. Thesis, Virginia Tech, 2007.
- [42] Hall, K.M., Brian L. B., Brannan, K. M. and Zeckoski R. W. Assessing Alternative Fecal Coliform Direct Deposit Modeling Approaches, in *21st Century Watershed Technology: Improving Water Quality and Environment, ASABE Proceedings of the 29 March - 3 April 2008 Conference*. Concepcion, Chile.
- [43] Zeckoski, R. Rebecca Zeckoski's Engineering Blog. Available: <http://rzeckoskiengineering.blogspot.com/2011/10/low-flow-conundrum-part-1.html>. [Accessed May 12, 2013].
- [44] MDDEQ, Total Maximum Daily Load of Phosphorus in the Antietam Creek Watershed, Washington County, Maryland, 2012, Maryland Department of Environmental Quality (MDDEQ): 1800 Washington Boulevard, Suite 540, Baltimore, Maryland 21230-1718.
- [45] Walker, L., Calleguas Creek Nutrient TMDLs, 2001, Larry Walker Associates. Prepared for Los Angeles Regional Water Quality Control Board: Santa Monica, CA.
- [46] Yagow, G. Using GWLF for development of "reference watershed approach" TMDLs, in *2004 Proceedings of the American Society of Agricultural and Biological Engineers and the Canadian Society for Bioengineering annual international meeting, paper no. 042262*. St. Joseph, MI: American Society of Agricultural and Biological Engineers.
- [47] Yagow, G., Mostaghimi, S., Dillaha, T., Brannan, K., Wynn, J., Zeckoski, R. and Brian, B. Linville Creek TMDL for A Benthic Impairment, in *Proceedings of the 8-12 November 2003 Conference*. Albuquerque, New Mexico USA: Transactions of ASAE.
- [48] USEPA, National Management Measures for the Control of Nonpoint Pollution from Agriculture. USEPA-Office of Water, U.S. Environmental Protection Agency (U.S. EPA): Washington D.C. Jul., 2003.
- [49] USEPA, Economic Analyses of Nutrient and Sediment Reduction Actions to Restore Chesapeake Bay Water Quality, in U.S. EPA Region III- Chesapeake Bay Program Office, U.S. Environmental Protection Agency (U.S. EPA): Annapolis, MD, 2003.