Science of the Total Environment xxx (xxxx) xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Simulating internal watershed processes using multiple SWAT models

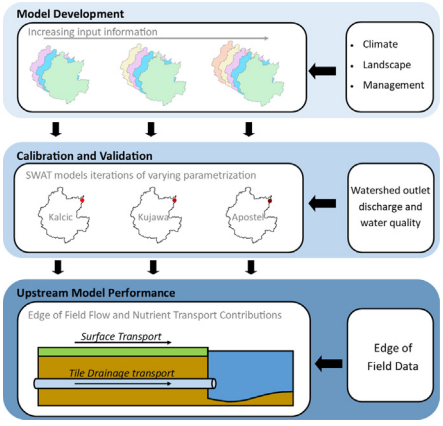
Anna Apostel [a,*], Margaret Kalcic [a,b], Awoke Dagnew [c,d], Grey Evenson [a], Jeffrey Kast [a], Kevin King [e], Jay Martin [a,f], Rebecca Logsdon Muenich [g], Donald Scavia [c]

[a] Food, Agricultural and Biological Engineering, The Ohio State University, Columbus, OH, USA
[b] Ohio State University Translational Data Analytics Institute, Columbus, OH, USA
[c] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA
[d] Environmental Consulting and Technology, Inc., Ann Arbor, MI, USA
[e] USDA-ARS Soil Drainage Research Unit, Columbus, OH, USA
[f] Ohio State University Sustainability Institute, Columbus, OH, USA
[g] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

HIGHLIGHTS

- Internal watershed processes depend on model inputs, parameters, and structure
- These model variations were captured through use of multiple models
- Edge of field data was used to evaluate internal watershed process simulation
- Calibration at watershed outlet did not ensure accurate simulation upstream
- Model structure limited accurate internal partitioning of P transport

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 August 2020
Received in revised form 15 October 2020
Accepted 29 October 2020
Available online xxxx

Editor: Ouyang Wei

Keywords:
Soil and Water Assessment Tool (SWAT)
Ensemble modeling
Calibration/validation
Nutrient loading
Field scale
Subsurface drainage

ABSTRACT

The need for effective water quality models to help guide management and policy, and extend monitoring information, is at the forefront of recent discussions related to watershed management. These models are often calibrated and validated at the basin outlet, which ensures that models are capable of evaluating basin scale hydrology and water quality. However, there is a need to understand where these models succeed or fail with respect to internal process representation, as these watershed-scale models are used to inform management practices and mitigation strategies upstream. We evaluated an ensemble of models—each calibrated to in-stream observations at the basin outlet—against discharge and nutrient observations at the farm field scale to determine the extent to which these models capture field-scale dynamics. While all models performed well at the watershed outlet, upstream performance varied. Models tended to over-predict discharge through surface runoff and subsurface drainage, while under-predicting phosphorus loading through subsurface drainage and nitrogen loading through surface runoff. Our study suggests that while models may be applied to predict impacts of management at the basin scale, care should be taken in applying the models to evaluate field-scale management and processes in the absence of data that can be incorporated at that scale, even with the use of multiple models.

* Corresponding author at: 590 Woody Hayes Dr., Columbus, OH 43210, Department of Food, Agricultural, and Biological Engineering, the Ohio State University, USA.
E-mail address: apostel.4@osu.edu (A. Apostel).

https://doi.org/10.1016/j.scitotenv.2020.143920
0048-9697/

1. Introduction

Shifts in land use and management have transformed many landscapes, altering water resources and raising environmental concerns among managers and other stakeholders (Guswa et al., 2014). Understanding and predicting the impacts of these changes is key to developing effective mitigation strategies that balance development and environmental goals. Policy- and decision-making often rely on computer model simulations, raising concerns around appropriate model use and prediction confidence. Significant progress in model error analysis and uncertainty estimation has been made; however, models considered to be accurate can have inaccurate representations of dominant processes in a system (Hrachowitz and Clark, 2017). Models varying in structure and parameterization can effectively reproduce similar system behavior—the equifinality concept—allowing for multiple possible internal representations of a system (Beven, 2006). Capturing this process-level information results in more robust hydrologic and land management models and is key for effective decision-making.

Distributed parameter models have become a common tool for hydrologic analysis and watershed management planning. With greater access to spatially variable and fine resolution datasets, the representation of natural systems within these models can be vastly improved (Vieth et al., 2008; Daggupati et al., 2015). The Soil and Water Assessment Tool (SWAT) is a highly parameterized process-based model developed for watershed scale assessment of climate, land management, and land-use change (Neitsch et al., 2011). While originally intended for use on ungauged basins, modelers generally improve SWAT model performance through calibration and validation to streamflow, and potentially water quality, at a single downstream outlet location. To the extent that models are only assessed at the watershed outlet, they can fail to take into account key intra-watershed processes—internal processes which govern global responses—and can thereby lead to successful models that lack realistic system representation (Yen et al., 2014; Daggupati et al., 2015). Yet capturing these processes is important as most management implementation and improvement strategies in agricultural regions occur at the field scale (Diebel et al., 2008; Muenich et al., 2017).

Data capable of validating these internal processes at the necessary scales are rarely available, especially in large watersheds. This use of smaller scale data is important as studies have shown that, while sometimes difficult to incorporate and calibrate to, the use of direct field-scale data can significantly impact field-scale model performance (Merriman et al., 2018; Muenich et al., 2017; Kalcic et al., 2015; Daggupati et al., 2015). Two important sources of field-scale data that have become more available and have been used in upstream validation are remote-sensed data and edge of field (EOF) monitoring. The use of remotely sensed data, capable of effectively describing the internal distribution of key watershed characteristics, such as soil moisture (Rajib et al., 2016) and vegetation (Ma et al., 2019; Rajib et al., 2020), has greatly improved SWAT calibration and performance. EOF monitoring provides valuable site-specific information on management practices, as well as key nutrient transport budgets (Pease et al., 2017; Hanrahan et al., 2019), which can inform soft validation for upland performance. As their availability grows, these data present an opportunity to add to the understanding and refinement of upstream representation and prediction accuracy of hydrologic models.

A lack of data availability, compounded with model framework uncertainties and computational constraints, may prohibit thorough uncertainty characterization of deterministic models such as SWAT. The use of multiple models that vary in model inputs, parameterization, and structure can capture some of the variability that is lost using a single deterministic model, and therefore play a role in uncertainty analysis (Hrachowitz and Clark, 2017). Such ensemble modeling has been used in a variety of SWAT modeling efforts to evaluate watershed water quality and policy-relevant management (Evenson et al., 2020; Martin et al., 2019; Scavia et al., 2017). Model ensembles can help capture uncertainties in model structure, parameterization, input data, and assumptions, thus enabling the exploration of the impacts of these facets on model performance.

The aim of this study was to assess the extent to which watershed hydrologic models, calibrated and validated at the watershed outlet, can capture field-scale dynamics upstream. This was done by developing the latest version of the Maumee River Watershed (MRW) SWAT model, Apostel, and comparing it against EOF monitoring data and two previous MRW SWAT model iterations, Kalcic et al. (2016) and Kujawa et al. (2020). While the Kalcic and Kujawa models varied with respect to management inputs, structure, and parameterization, both achieved acceptable Moriasi et al. (2007, 2015) performance metrics at the watershed outlet (Kalcic et al., 2016; Kujawa et al., 2020). Our objective was to assess the ability of the model ensemble to capture upstream field level discharges and nutrient loadings. The results will be discussed in the context of critical differences between models in terms of structure and inputs.

2. Methods

2.1. Study area

The Maumee River Watershed (MRW) is the largest watershed draining to Lake Erie. Excess inputs of nutrients from the MRW, especially phosphorus, have increased harmful algal blooms in the Lake's western basin and hypoxia (low oxygen concentrations) in its central basin that threaten both ecosystems and human health (Scavia et al., 2014, 2016; Michalak et al., 2013; Ohio EPA, 2010, 2013). The MRW's impact on the health of regional resources has made it a key target for management and nutrient mitigation research (GLWQA, 2015).

Located in northwest Ohio, northeast Indiana and southwest Michigan (Fig. 1), the MRW is over 17,000 km^2 in area and its land use is dominated by over 70% row crop agriculture (corn, soybean, and wheat). It consists of 253 HUC-12 units across 27 counties in the tristate area. Formerly containing the Great Black Swamp, a majority of the watershed's soils are poorly drained, requiring systematic, subsurface ("tile") drainage to facilitate agricultural crop production. The MRW is the largest contributor of total phosphorus into Lake Erie, accounting for 48% of the total phosphorus (TP) load to the Western Lake Erie Basin (Maccoux et al., 2016).

2.2. Ensemble of watershed models

The ensemble of watershed models used in this study were developed consecutively over a span of six years using the Soil and Water Assessment Tool (SWAT). SWAT is a watershed-scale model, developed by the U.S. Department of Agriculture's Agricultural Research Service to assess the impacts of land management on water quality in large ungauged basins (Neitsch et al., 2011). SWAT is commonly used in agricultural settings because it permits the implementation of detailed agricultural management practices and schedules. This model has been used worldwide to assess agricultural impacts on discharge, nutrient loads, and crop growth (Francesconi et al., 2016).

Two earlier versions of the MRW SWAT model used in this ensemble are described in Kalcic et al. (2016) (Kalcic) and the University of Michigan SWAT model within the Kujawa et al. (2020) ensemble (Kujawa). The latest version (Apostel) has not been previously described and so is detailed below. All three models were developed using similar baseline inputs of 30-m resolution elevation data (USGS-NED, 2016), county-based soils data (SSURGO, 2016), 30-m resolution gridded land use data (NASS-CDL, 2012), regional land based climate station data (Menne et al., 2012), and medium resolution stream shapefiles (USGS-NHD, 2016). Each model maintains the same watershed area (17,305 km^2) with 358 subbasins. However, the key difference in this newest iteration is the discretization of HRUs. By default, HRUs in SWAT are classified as regions of unique slope, soil, and land cover

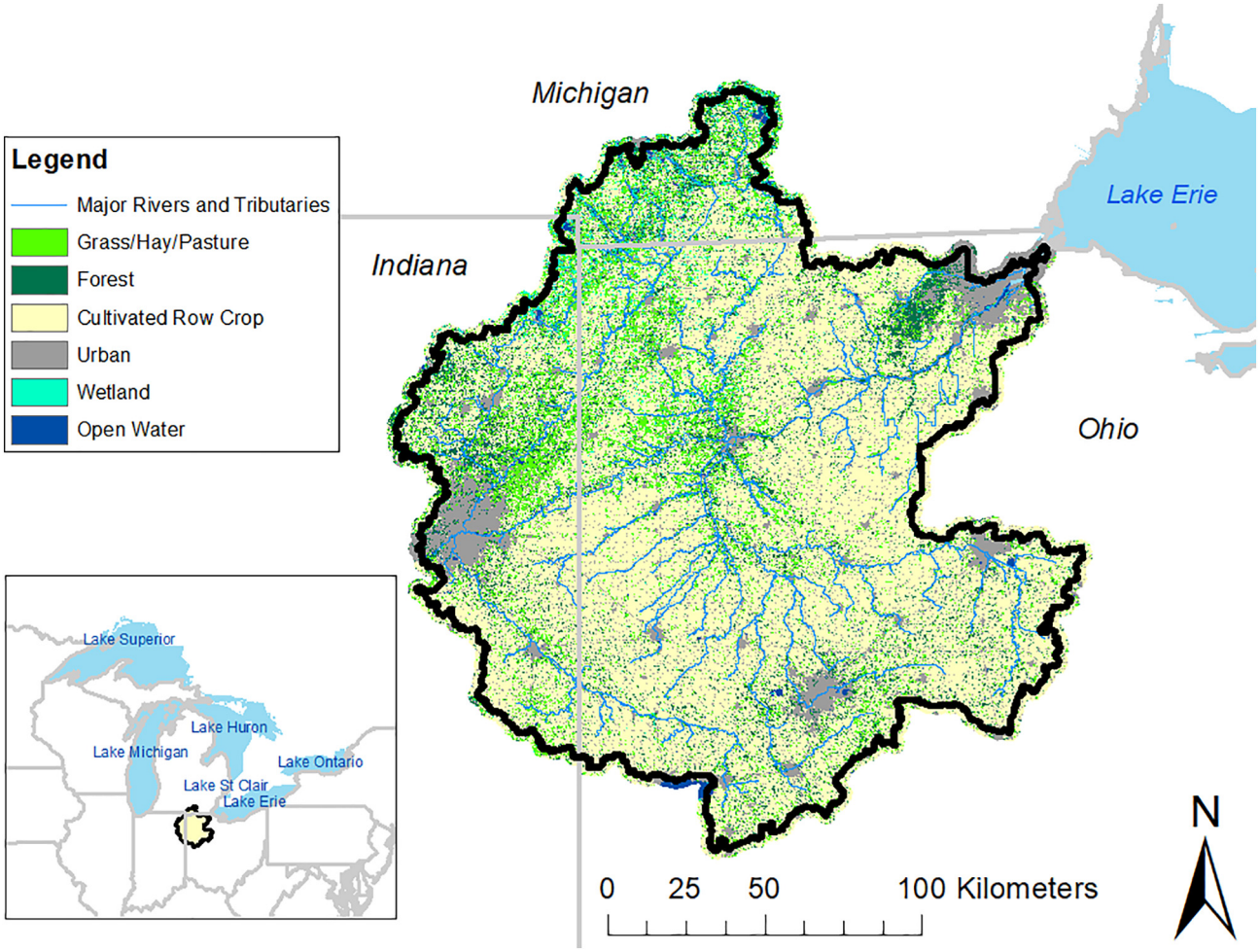


Fig. 1. Map of Maumee River Watershed and regional land use.

combinations, which can result in a single HRU spanning multiple, unattached areas. The Apostel model does not allow for this spatial discontinuity in the HRU definition, instead setting HRU boundaries based on approximate field boundaries, resulting in 24,256 HRUs (Fig. 2), compared to 10,266 spatially discontinuous HRUs in Kalcic and Kujawa. The Apostel model has 18,018 agricultural HRUs at the near field scale, 12,676 of which contain subsurface (tile) drainage, with an average size of 74 ha. This is compared to 870 agricultural HRUs, 645 of which contain tile drains, with an average size of 1385 ha, in the Kalcic and Kujawa models.

All models were run using the SWAT 2012 Revision 635 source code with a modification to properly move soluble phosphorus through subsurface tile drains (Kalcic et al., 2016). All three models used the same model subroutines, with the exception of preferential flow (through SWAT's soil cracking routine), which was enabled only in the Kujawa and Apostel models. The models were run for the 2005–2015 time period and model performance was assessed using Moriasi et al. (2015) updated performance criteria for 'satisfactory' model fit.

Additional information on model parameterization can be found in Table S-25 of the Supplemental Information (SI). Key changes in the Apostel model are further described below and referenced in Table 1.

2.3. Apostel SWAT model development

The Apostel model was developed with the intent of being able to incorporate direct field-scale management information. It is not standard practice to use models delineated at the field scale, especially in watersheds of this size, though it allows for the incorporation of more realistic management spatial heterogeneity (Karki et al., 2020). To overcome spatially discontinuous lumped units, the Apostel model was delineated at the near field scale and populated with the best available regional data. Management operations were estimated from several sources (described below), such that field-scale HRUs were representative of current (2005–2015) production practices, and yet data on specific management operations were not available at the field scale.

2.3.1. HRU delineation

HRU delineation techniques developed by Teshager et al. (2016) were used for creating near field-scale HRU boundaries for the model. This approach used the USDA National Agricultural Statistics Service Cropland Data Layer (NASS CDL) to develop land use polygons that were then further subdivided by roadways and stream networks. These predefined polygons were intersected with the model's subbasins and a single dominant soil type was assigned to each individual HRU polygon. This process allows the model to have spatially continuous HRUs at the near field scale, enabling assignment of all management parameter values and some field-level physical parameter values for individual fields. This permits greater resolution compared to other SWAT models in the region (Kujawa et al., 2020), as well as use of survey data to better bundle field-specific management practices.

2.3.2. Management practices

Management practices were incorporated according to regional and county level information, similar to the previous model iterations, Kalcic and Kujawa. These practices included cropping systems, fertilizer

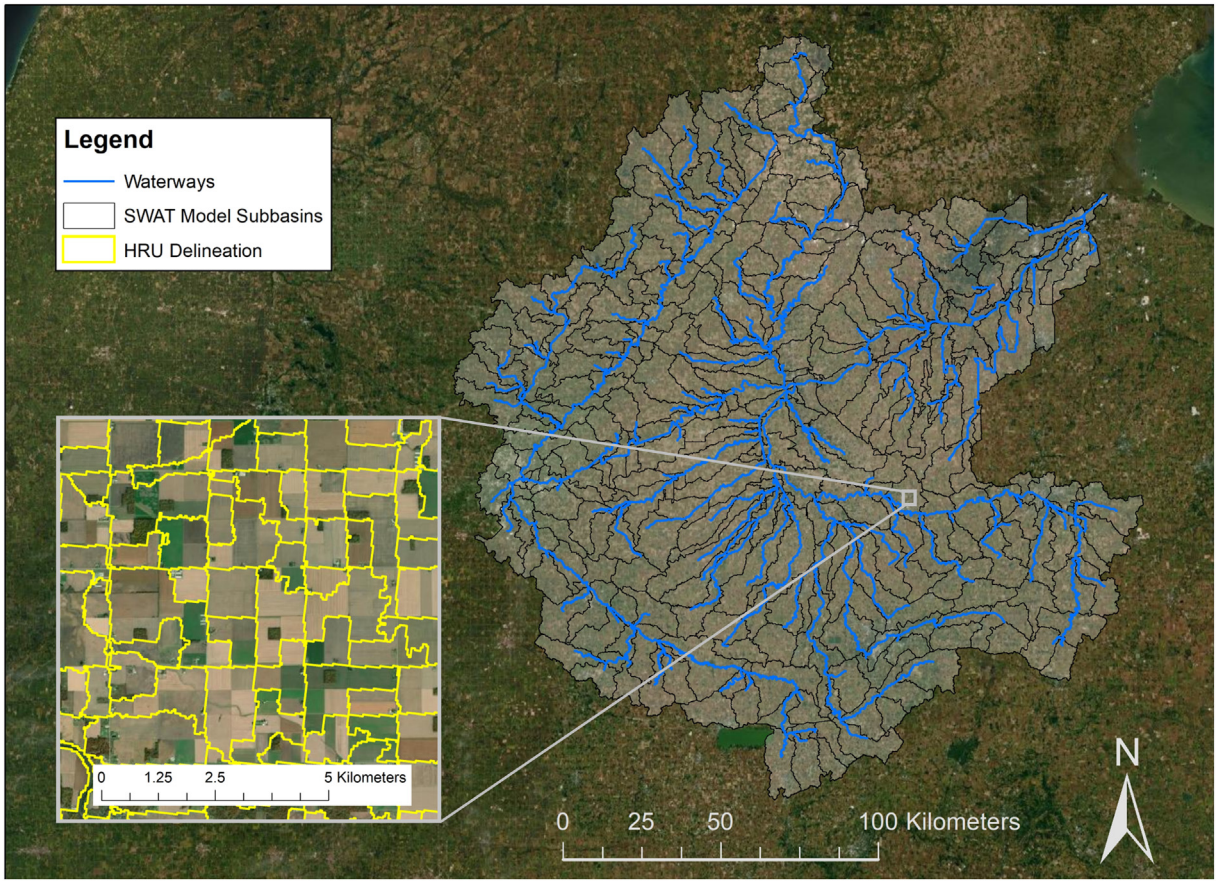


Fig. 2. HRU delineation in the Apostel MRW SWAT. Teshager et al. (2016) field delineation methodology led to near field-scale boundary matches that have an average size of 79 ha.

applications, tillage, best management practices, and tile drainage (Table 1). Regional surveys were used to capture common linked practices to guide management patterns and practices throughout the watershed, similar to the Kujawa model.

Unique combinations of crop rotations were developed by assessing annual crop cover from the USDA-NASS CDL for 2000–2010 and monitoring crop rotation patterns over time. These data were condensed to identify the most common 29 agricultural rotations based on soybean, corn, and wheat, as well as continuous alfalfa and pastureland, capturing dominant rotations while achieving watershed-wide percent coverage (Tables S2; S3). Kalcic and Kujawa had only twelve rotations, and within each rotation all management operations were identical (Table 1). In the Apostel model, the 29 crop rotations were applied across HRUs prior to setting management operations, thus increasing the heterogeneity of land management across the watershed.

Organic fertilizer application rates were based on manure application practices in the tri-state area (Kast et al., 2019), using both mass-balance and spatial-balance approaches and regional Confined Animal Feeding Operations (CAFOs). Individual rates were determined for swine, cattle, and poultry manure applied on cropland, as well as for swine and cattle manure applied on pastureland (SI - Section 5.4). Previous models used county level livestock counts alone to determine manure application rates and did not apply manure based on a spatial balance determined by proximity to CAFO (Kast et al., 2020) (Table 1).

Inorganic fertilizer application rates were based on county-level data from the Nutrient Use Geographic Information System (NuGIS) from the International Plant Nutrition Institute (IPNI, 2011). County application rates for corn, soybean, and wheat were distributed to crop rotations in proportion to the tristate application standard recommendations for maintenance-level phosphorus and maximum crop yields (Vitosh et al., 1995) to ensure mass balances for each county were maintained and crops received the appropriate nutrients. This scaled fertilizer distribution was then adjusted in locations having manure application to account for the nutrients contributed by inorganic sources. The end result is that all identical crop rotations with identical manure application patterns within a county would have the same nutrient application. Previous models used county level sales to determine inorganic fertilization rates. The Kujawa model additionally scaled sales data according to individual crop needs using the tri-state standards, similar to the Apostel model scaling approach (Table 1).

Regional survey data from multiple sources were used to establish land management in the Apostel model. Region-wide percentages for practices such as tillage, best management practices, and tile drainage were obtained as target level implementations for the watershed. The use of producer survey data was first incorporated into the Kujawa model to establish further refined percentages of practices throughout the watershed to supplement previous stakeholder engagement in the Kalcic model and refine the extent of existing agricultural management and conservation practices. In the Apostel model, survey data were enhanced to link practices, tillage, fertilizer application methods, and other conservation practices through disaggregating regional data to individual fields.

Tillage within the Apostel model was classified as continuous tillage (5%), no tillage (25%) and rotational tillage (70%) (CTIC, 2013; NRCS-CEAP, 2016). Continuous tillage was defined as annual use of a tillage implement, whereas no tillage referred to no soil disturbance throughout a rotation. Rotational tillage was defined as tillage associated with certain crops in the rotation, and in this case was applied prior to planting corn. Over the last several years, this watershed has experienced a significant shift from continuous tillage to rotational and minimum tillage (Smith et al., 2015a). To maintain the MRW tillage distributions, crop rotations that included wheat were identified as no-tillage fields.

Table 1
Multi-model comparison of model structure and inputs.

		Kalcic	Kujawa	Apostel
General model inputs and characteristics	Model version	Rev. 635 – Modified to fix a bug where soluble P was not properly moving through subsurface drains	Rev. 635 – Modified to fix a bug where soluble P was not properly moving through subsurface drains	Rev. 635 – Modified to fix a bug where soluble P was not properly moving through subsurface drains
	Elevation	National Elevation Dataset (NED) 30 m	NED 30 m	NED 30 m
	Climate	NOAA NCDC – precipitation and temperature (Menne et al., 2012)	NOAA NCDC – precipitation and temperature (Menne et al., 2012)	NOAA NCDC – precipitation and temperature (Menne et al., 2012)
	Point sources	National Pollution Discharge Elimination System (NPDES) permits	Measured data from EPA Discharge Monitoring Reports (DMR); aggregated to average monthly	Measured data from EPA DMR; aggregated to average monthly
	Landuse	National Agricultural Statistics Service Cropland Data Layer (NASS CDL) (NASS-CDL, 2012)	National Land Cover Database (NLCD) 2006; NASS CDL 2007–2012 data layers (NASS-CDL, 2012)	NASS CDL 2000–2010 data layers (NASS-CDL, 2012)
	Calibration time period	2001–2005	2005–2014	2005–2015
	Watershed size (km^2)	17,305	17,305	17,305
	No. Subbasins	358	358	358
	No. HRUs	10,266	10,266	24,256
	Spatially discrete HRUs?	No	No	Yes
	No. Agricultural HRUs	870	870	18,018
Soils	Soil inputs	Soil Survey Geographic (SSURGO) Database (SSURGO, 2016)	SSURGO Database (SSURGO, 2016)	SSURGO Database (SSURGO, 2016)
	Soil stratification	Default	Default	Top soil layer stratified
	Soil phosphorus	Default	Default	Soil phosphorus initialization adjusted according to regional soil test phosphorus values
Agricultural management	Inorganic fertilizer	Estimated from county fertilizer sales data from 2002	Estimated from county fertilizer sales data from 2002; Scaled for individual plants according to tristate recommendations for maintenance application (Vitosh et al., 1995)	Estimated from NUGIS IPNI county level data (2007–2014); Scaled for individual plants according to tristate recommendations (Vitosh et al., 1995)
	Organic fertilizer (manure)	Estimated from NASS Agricultural Census Yield and Fertilizer Use data 1990–2010	Estimated from NASS Agricultural Census Yield and Fertilizer Use data 1990–2010; Estimated from county-level livestock count	Mass and spatial balances based on CAFOs located within the watershed and regional management guides (Kast et al., 2020); Manure distributed based on distance from CAFO site (Kast et al., 2020)
	Crop rotations	12 Corn, soy, wheat rotations; Continuous pastureland and alfalfa field	12 Corn, soy, wheat rotations; Continuous pastureland and alfalfa field	29 corn, soy, wheat rotations; Continuous pastureland and alfalfa field; Additional rotation information Tables S2 and S3
	Tillage	Estimated from Conservation Tillage Information Center (CTIC) data (CTIC, 2013)	Estimated from CTIC (CTIC, 2013); Estimated from CEAP report (NRCS-CEAP, 2016)	Estimated from CTIC (CTIC, 2013); Estimated from CEAP report (NRCS-CEAP, 2016); Estimated according to crop planted and incorporation
	Best management practices	Grassed waterways; Buffer strips	Grassed waterways; Buffer strips	Grassed waterways; Buffer strips
	No. tile drained HRUs	645	645	12,676
	Tile implementation	70% of agricultural fields targeting poorly and very poorly drained soils	70% of agricultural fields targeting poorly and very poorly drained soils	70% of agricultural fields targeting poorly and very poorly drained soils; Drainage intensity (spacing between drains) adjusted to increase intensity of very poorly drained fields

However, corn-soybean-wheat rotations only accounted for 10% of rotations. Therefore, the remaining 15% of no tillage was applied to corn-soybean rotations. Continuous tillage was first assigned to those fields where phosphorus fertilizer was to be incorporated. Once the 5% threshold was met, the remaining phosphorus incorporation was assigned to rotational tilled fields scheduled for corn.

For tillage practices, certain parameters known to vary based on tillage intensity were defined prior to calibration. These parameters were: biological mixing efficiency (BIOMIX), Manning's N for overland flow (OV_N), and the Universal Soil Loss Equation support practice (USLE_P). Precise values for BIOMIX according to tillage practice were unknown; however, it is established that no-tillage systems have greater biological mixing than systems having soil disturbance (Kladivko et al., 1997). The surface roughness (OV_N) was adjusted based on tillage practice and the crop present using values from Table 19–1 from the SWAT Input/Output documentation, with rotational tillage being given the mean value between continuous and no-tillage practices (Arnold et al., 2012). The USLE_P factor defines the ratio of soil loss from a specific management practice to that of the up-and-down slope culture (Neitsch et al., 2011). Typically, tillage management is a component of the C (cover management) factor in the Universal Soil Loss Equation. However, in SWAT, this factor is solely based on crop planted. In order to incorporate tillage management, the management factor was moved to USLE_P. Therefore, continuous-tillage,

no-tillage, and rotational-tillage practices were used to adjust USLE_P according to the increased likelihood of soil loss with increased tillage intensity (Arabi et al., 2008).

Fertilizer placement was added to the management schedule and tied to tillage practice. Phosphorus fertilizer was incorporated through tillage in 60% of all row-cropped fields, with 10% having subsurface placement of phosphorus fertilizer, and 30% having broadcast application followed by incorporation. The 40% of fields lacking phosphorus fertilizer incorporation had either broadcast fertilizer application or manure application. These percentage values and pairings were adopted based on values from Burnett et al. (2015), Prokup et al. (2017), Wilson et al. (2013), and the NRCS Conservation Effects Assessment Program (NRCS-CEAP, 2016), and adjusted to fit MRW land uses and acreage percentages.

Best management practices, including grassed waterways, buffer strips, and non-winter wheat cover crops, were incorporated into the model, similar to the Kujawa model. Buffer strips intercepting runoff from 35% of row crop acreage and grassed waterways intercepting surface runoff from 21% of row crop acreage were added based on regional CEAP surveys (NRCS-CEAP, 2016). Cover crops were added to 10% of agricultural row crops (NRCS-CEAP, 2016), selecting from fields that had no wheat in the rotation.

2.3.3. Tile drainage

Farmer surveys from the western Lake Erie basin suggest that greater than approximately 70% of the watershed cropland benefits from artificial drainage (NRCS-CEAP, 2016). Limited information exists on the specific locations of tile drainage in the MRW. To account for tile drainage in the model, tile drains were implemented on 72% of agricultural acreage, with greater tile intensity inversely proportional to drainage classification (i.e., drainage intensity was greater for very poorly drained soils than for somewhat poorly drained soils) (Fig. S8). Variability in drainage intensity was achieved through adjustments to drainage parameters within the model (SI Section 7). The daily drainage coefficient (DRAIN_CO) and the distance between drains (SDRAIN) were set to simulate tile density (Neitsch et al., 2011). These value ranges were determined from personal communication with local drainage experts. The allocation of tile drain placement based on soil characteristics is similar to the Kalcic and Kujawa models, however tile intensity refinement was a new addition (Table 1).

2.3.4. Soil phosphorus stratification and phosphorus concentration

Soil phosphorus stratification and initialization according to regional soil test phosphorus values were unique to the Apostel model. Stratification of soil phosphorus in the presence of reduced tillage has become increasingly prevalent in this region (Baker et al., 2017). The model's default soil structure has up to 10 soil horizons of variable depth, but the number simulated is based on the underlying incorporated SSURGO soils data. To incorporate soil stratification, all soil horizons were divided at 5 and 20 cm based on studies reporting concentration differentials around this depth (Baker et al., 2017).

Regional soil test phosphorus values were used to determine an initial labile phosphorus concentration for the stratified layers. A soil test with a phosphorus value of 39.6 mg/kg Mehlich-3P for the region (Williams et al., 2015) was converted to Bray phosphorus (Culman et al., 2019) and then converted to model labile phosphorus (Sharpley et al., 1984) for a final value of 21.5 mg/kg soil. This value of 21.5 ppm was achieved for the mean concentration in the top 20 cm of the soil profile. The vertical concentration stratification was then adjusted to follow regional stratification patterns based on Baker et al. (2017), who reported that 68% of the phosphorus found in a soil test core was in the top 5 cm. This resulted in a core mean labile phosphorus concentration of 31 ppm in the top 5 cm and 18 ppm concentration in the 6–20 cm layer. The initial soluble phosphorus concentration in the soil layer (SOL_SOLP) was set as described above to implement field-level soil test phosphorus values and phosphorus stratification.

2.3.5. Snow parameters set as inputs

Snow parameters were determined in the calibration process for the Kalcic and Kujawa models. For the Apostel model, these values were determined prior to model calibration from a regional climate assessment using data obtained from NOAA's Global Historical Climatology Network, including a daily record of snowfall and snow depth (Menne et al., 2012). Temperature data were used as input to SWAT and the value of SFTMP (snowfall temperature) was changed until SWAT predictions largely agreed with the observed data on which days had precipitation in the form of snow.

A focused calibration was conducted to determine the remaining snowfall parameters (SMTMP, SMFMX, SMFMN, TIMP, SNOCOVMX, and SNO50COV). Resulting snowfall and snow melt values were extracted from modeled SWAT outputs. We calculated daily snow accumulation (depth) using these values for each day. SWAT output values, which are given in mm of water, were converted into mm of snow using a formula from Qi et al. (2016):

$$ds = SNO/s \quad (1)$$

where ds is the depth of snow (cm), SNO is the snow water equivalent (mm) and s is the snow density (g cm^{-3}). The optimum combination of parameters with highest value of Nash-Sutcliff Efficiency (NSE) and smallest absolute value of percent bias (PBIAS) was chosen by comparing these converted values to daily observed snow depth values from the NOAA GHCN data set. All snowfall parameters were singular basin-wide values.

2.3.6. Apostel model calibration and validation

Using an approach similar to Kalcic and Kujawa models, detailed manual calibration was performed for discharge, TP, dissolved reactive phosphorus (DRP), total nitrogen (TN), nitrate, and sediment near the outlet of the Maumee River at the Waterville gauge station at daily and monthly timescales for 2005–2015, with validation for 2000–2004. Key parameters and parameter changes used in the calibration process are identified in Table S16 of the SI. Calibration and validation were evaluated with the coefficient of determination (R^2), Nash-Sutcliffe Efficiency (NSE), and percent bias (PBIAS) following Moriasi et al. (2007) and Moriasi et al. (2015) standards to achieve 'satisfactory' or better performance.

2.4. Upstream edge of field comparisons

Observed EOF data were used to assess simulated internal watershed processes. This EOF hydrology and water quality data included surface and subsurface discharge and nitrogen and phosphorus loading from eight Northwest Ohio fields that are part of the USDA-ARS edge-of-field network (Williams et al., 2016). HRUs for comparison against EOF data were chosen for each model based on the presence of cultivated row crop land management and the presence of tile drainage, conditions characteristic of all EOF sites. Only observed data within the 2005–2015 period was used for comparison, and HRU outputs were compiled for the periods overlapping with observations. Output from each model was compared to EOF data for 2013–2015, the period when model output and EOF data overlap. This resulted in 1935 (Kalcic), 1932 (Kujawa) and 38,025 (Apostel) individual yearly data points from simulated HRUs compared to 13 individual yearly data points from the observed network. Simulated annual surface and tile discharge and nutrient loading were summarized and compared to measured data. Surface and tile discharge, along with nutrient loading via these two transport pathways, were compared along with the relative contributions from each pathway using a surface to tile discharge ratio. A non-parametric Wilcoxon Rank Sum Test was used to test for significance between the simulated and measured data due to the large variance in n values between samples.

Table 2
Multi-model outlet validation results. Aims are based on Moriasi et al. (2007) and Moriasi et al. (2015) performance criteria for 'Satisfactory' performing models. Italicized entries indicate values that did not meet the minimum criteria for 'Satisfactory' performances.

			Kalcic (2005–2015)		Kujawa (2005–2015)		Apostel (2005–2015)		Apostel Validation (2000–2004)	
	Statistic	Aim	Daily	Monthly	Daily	Monthly	Daily	Monthly	Daily	Monthly
Flow	R^2	>0.6	0.72	0.91	0.76	0.95	0.87	0.95	0.83	0.88
	NSE	>0.5	0.72	0.89	0.76	0.95	0.87	0.95	0.82	0.86
	PBIAS	<±15	9.60	9.62	2.04	2.08	−0.83	−0.88	−10.03	−10.11
TP	R^2	>0.4*	0.56	0.69	0.53	0.59	0.6	0.61	0.47	0.51
	NSE	>0.35*	0.52	0.61	0.46	0.46	0.58	0.52	0.46	0.44
	PBIAS	<±30	18.61	19.69	3.40	4.79	−3.76	−3.23	−18.53	−18.35
DRP	R^2	>0.4*	*0.33*	0.61	*0.34*	0.68	0.63	0.68	0.63	0.74
	NSE	>0.35*	*0.32*	0.56	*0.18*	0.67	0.62	0.67	0.63	0.73
	PBIAS	<±30	−8.85	−9.34	9.98	9.13	2.03	1.51	−9.89	−10.22
TN	R^2	>0.3*	0.67	0.85	0.61	0.81	0.63	0.78	0.75	0.82
	NSE	>0.35*	0.51	0.64	0.43	0.75	0.55	0.69	0.68	0.71
	PBIAS	<±30	*34.14*	*34.34*	2.79	2.16	−0.4	−1.24	−6.44	−6.73
Sediment	R^2	>0.4*	0.60	0.78	0.62	0.81	0.68	0.8	0.59	0.75
	NSE	>0.45*	0.56	0.66	0.62	0.80	0.65	0.75	0.58	0.70
	PBIAS	<±20	*26.77*	*27.83*	−3.27	−1.88	1.62	2.06	*−27.21*	*−26.09*

* Recommended statistic for monthly temporal scale only.

3. Results

3.1. Calibration and validation

Final calibration of the Apostel model involved adjustments in 40 parameters (Table S26) and resulted in 'satisfactory' or better performance standards (Table 2; Moriasi et al., 2007, 2015). Validation for 2000–2004 also achieved 'satisfactory' model performance, although sediment PBIAS fell just outside the target range. Primary parameterization differences among the Kalcic, Kujawa, and Apostel models were parameters for tile drainage, soil nutrient initialization and transport, snow characteristics, and activation of the soil cracking (ICRK) subroutine.

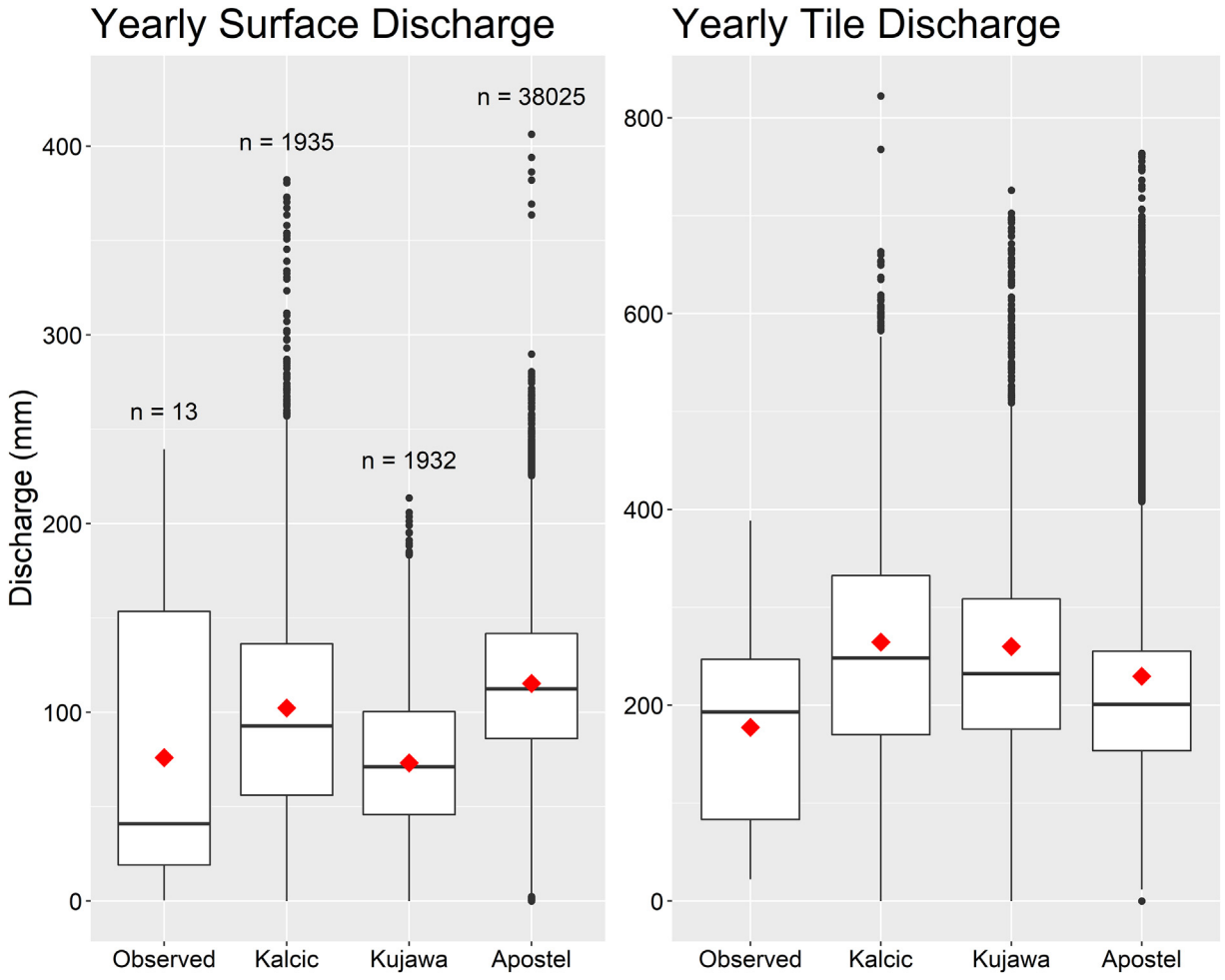


Fig. 3. Box and whisker plots of average annual surface and tile flows across modeled HRUs and observed sites from 2013 to 2015. Diamonds represent mean values. Boxes are bounded by 25th and 75th percentiles with whiskers representing 10th and 90th percentiles.

Table 3
Means ± standard deviations of modeled HRU and observed field level annual discharge and loads.

		Observed	Kalcic	Kujawa	Apostel
Discharge *(mm)*	Surface	76.00 ± 76.89	102.29 ± 66.44	73.23 ± 41.15	115.32 ± 46.67
	Tile	177.40 ± 109.75	264.51 ± 122.56	260.01 ± 121.37	229.56 ± 119.11
TP *(kg/ha)*	Surface	0.37 ± 0.45	1.85 ± 1.99	2.33 ± 2.10	2.68 ± 2.59
	Tile	0.34 ± 0.29	0.14 ± 0.08	0.33 ± 0.34	0.04 ± 0.03
DRP *(kg/ha)*	Surface	0.15 ± 0.23	0.13 ± 0.17	0.07 ± 0.07	0.29 ± 0.24
	Tile	0.12 ± 0.12	0.14 ± 0.08	0.33 ± 0.34	0.04 ± 0.03
TN *(kg/ha)*	Surface	6.36 ± 9.39	6.12 ± 5.10	12.67 ± 12.61	7.10 ± 5.57
	Tile	26.38 ± 25.16	36.44 ± 30.06	32.76 ± 29.02	25.31 ± 27.04
NO3 *(kg/ha)*	Surface	4.48 ± 7.57	1.01 ± 0.57	3.19 ± 8.89	2.82 ± 3.27
	Tile	23.22 ± 23.55	36.44 ± 30.06	32.76 ± 29.02	25.31 ± 27.04

To facilitate comparisons, the two previously calibrated SWAT models (Kalcic, calibrated 2001–2005; Kujawa, calibrated 2005–2014) were run and performance was assessed for the Apostel calibration period (2005–2015) at the outlet. Performance statistics for these two models were 'satisfactory' for all measures, with the exception of PBIAS for TN and sediment in the Kalcic model according to Moriasi et al., 2015 (Table 2). While daily DRP validation for the Kalcic and Kujawa displayed poor performance, its performance was satisfactory at a monthly time step.

3.2. Edge of field comparisons

Simulated mean surface runoff and tile discharge were within the range of measured data across all models. EOF measurements exhibited greater tile discharge than surface runoff, indicating subsurface dominance (Fig. 3; Table 3). With the exception of Kujawa, models tended to overestimate surface discharge (Table 3). Similarly, simulated tile discharge tended to be greater than EOF measurements. Kalcic and Kujawa showed significant differences from the observed ($p < 0.05$), while Apostel did not ($p = 0.09$). However, simulated surface to tile discharge ratios (Kalcic: 0.39; Kujawa: 0.28; Apostel: 0.50) were similar to the measured ratio of 0.43.

Observed annual TP load showed no dominant transport pathway. Measured mean annual surface TP load was similar to tile TP load with a surface to tile ratio of 1.1 (Table 3). In contrast, all models favored surface-dominated transport (Fig. 4). Simulated annual surface TP loads were significantly greater than observed (Wilcoxon p-values were $p < 0.05$ for each paired-group comparison), while simulated annual tile TP loads were significantly less than observed ($P < 0.05$) with the exception of Kujawa ($p = 0.8$). The overestimated surface contributions led to surface to tile ratios (Kalcic: 16.6; Kujawa: 5.6; and Apostel: 72.4) considerably greater than the measured ratio (1.1).

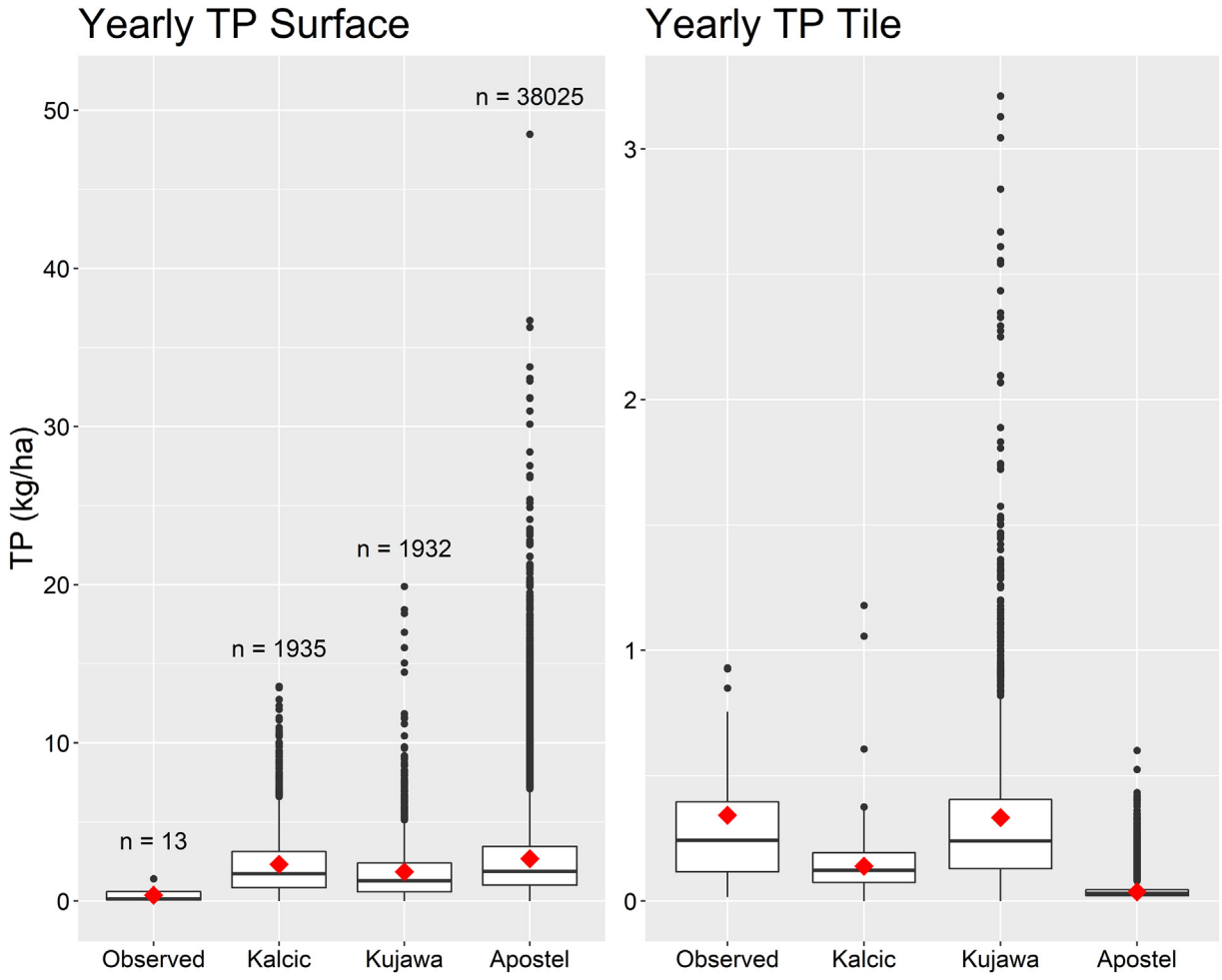


Fig. 4. Box and whisker plots of average annual surface and tile TP loads across models HRUs and observed sites from 2013 to 2015. Diamonds represent mean values. Boxes are bounded by 25th and 75th percentiles with whiskers representing 10th and 90th percentiles.

Observed annual DRP load showed similar trends to TP (Fig. 5). Specifically, measured mean annual surface DRP load was 0.15 ± 0.23 kg/ha while mean annual DRP load from tiles was 0.12 ± 0.12 kg/ha, resulting in a surface to tile ratio of 1.2. Trends in simulated loads varied across models. Surface DRP loads from Kalcic and Kujawa were less than measured loads while annual surface DRP loads from Apostel were greater (Table 3). In contrast, simulated annual tile DRP loads from Kalcic and Kujawa were greater than the measured load, while the simulated load from Apostel was less than measured load (Table 3). The simulated loads resulted in tile-dominated DRP partitioning ratios for Kalcic (0.50) and Kujawa (0.39), and a strong surface-dominated DRP load for Apostel (7.8).

Observed mean annual surface TN load was 6.36 ± 9.39 kg/ha compared to 26.38 ± 25.16 kg/ha for tile drainage, resulting in a surface to tile ratio of 0.24. Simulated tile TN loads were comparable to measured loads ($p = 0.08$, Kalcic; $p = 0.22$, Kujawa; $p = 0.68$, Apostel). Surface TN loads varied significantly ($p < 0.05$) except for Kujawa ($p = 0.11$) (Fig. 6; Table 3). The resultant surface to tile ratios of 0.35 (Kalcic), 0.18 (Kujawa), and 0.28 (Apostel) were similar to the observed.

Surface nitrate loads were significantly underpredicted by all three models ($p < 0.05$) (Fig. 7; Table 3). The models tended to over predict tile nitrate loads, though only Kalcic did so significantly ($p < 0.05$, Kalcic; $p = 0.05$, Kujawa; $p = 0.71$, Apostel). Similar to TN, nitrate loads were tile-dominated, having surface to tile ratios considerably less than 1 (0.19, observed; 0.09, Kalcic; 0.03, Kujawa; and 0.11, Apostel).

4. Discussion

4.1. Newly developed field-scale MRW SWAT model (Apostel)

The Apostel MRW SWAT model provides a platform for incorporating field-specific management-level decisions. While direct field-scale HRU data were not available for this watershed, a more realistic representation of management practices and schedules was implemented near the farm field scale, which is not typical in SWAT models (Karki et al., 2020). The Apostel model performance at the outlet was similar to that of previous iterations of the model, though model run time tripled due to greater numbers of agricultural HRUs. While an increase in computational time is a common issue when increasing model complexity (Her and Chaubey, 2015), this spatial refinement allows for more realistic implementations of management strategies and spatial heterogeneity in future work.

The Apostel model performed similarly to its two predecessor models with regards to discharge and nutrient loads at the watershed outlet. Due to a lack of data on management practices at the HRU scale, model improvements for the Apostel model were difficult to assess in the context of field level model improvements. Overpredictions of TP and DRP movement in general, and particularly with surface runoff, indicated that the Apostel model was unable to accurately capture field-scale TP and DRP loading. Developing a well performing model that is not representative of the system is possible for highly parameterized models such as SWAT, but these issues should be reduced by the

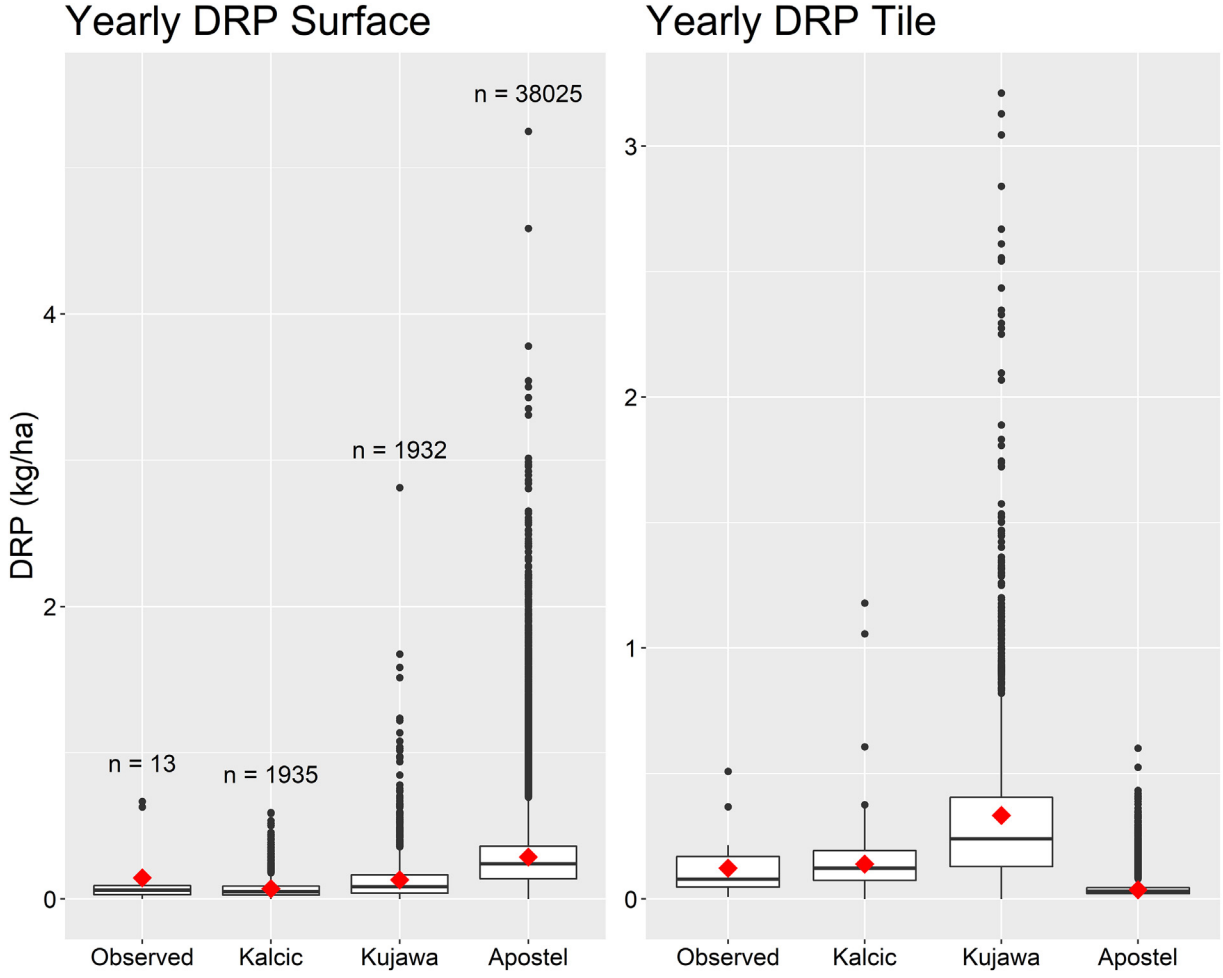


Fig. 5. Box and whisker plots of average annual surface and tile DRP loads across models HRUs and observed sites from 2013 to 2015. Diamonds represent mean values. Boxes are bounded by 25th and 75th percentiles with whiskers representing 10th and 90th percentiles.

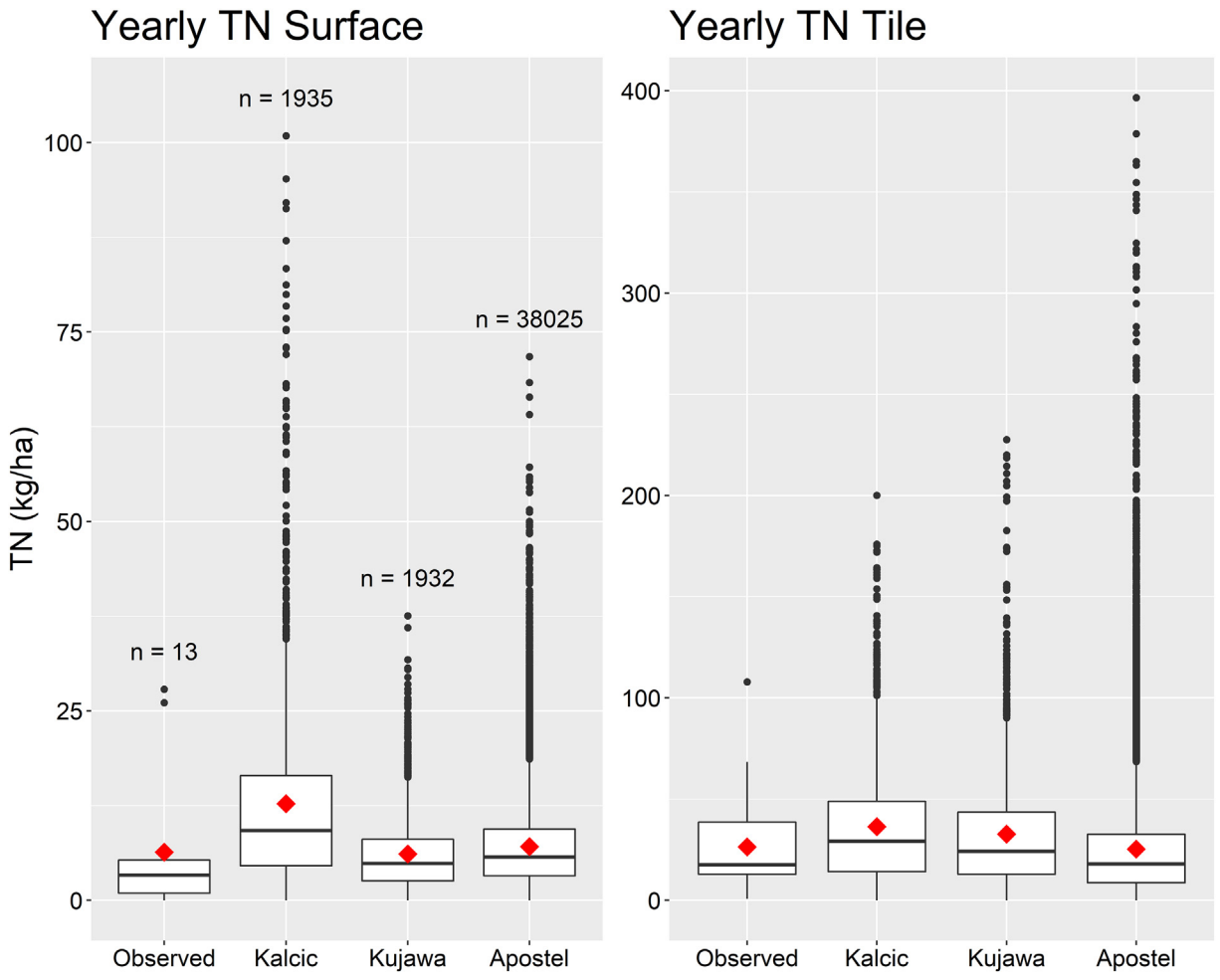


Fig. 6. Box and whisker plots of average annual surface and tile TN loads across models HRUs and observed sites from 2013 to 2015. Diamonds represent mean values. Boxes are bounded by 25th and 75th percentiles with whiskers representing 10th and 90th percentiles.

use of more detailed intra-watershed information and knowledge (Daggupati et al., 2015; Yen et al., 2014; Cibin et al., 2010). However, our study does not find clear support for improvements in the representation of these systems through regional information improvements alone. In the following sections we explore several factors that explain why model improvements of inputs and parameterization may have altered model performance.

4.2. Management data impacts

While all models had good model performance at the outlet, at the field level, Kujawa showed an increase in tile TP and decrease in the surface N compared to Kalcic. Fertilizer application varied between models, as more robust approaches were used to determine distribution and placement of both organic and inorganic forms of P. However, general nutrient application amounts remained similar among models. While a redistribution of fertilizer may have impacted these outlet level contributions, Evenson et al. (2020) found that fertilizer application rate distribution was only sometimes correlated with upstream phosphorus contributions when determining critical source areas, suggesting these input changes alone had limited impact.

Initializing soil phosphorus values based on observed soil test phosphorus measurements was one of the most significant input data enhancements in the Apostel model. This is the primary change to the phosphorus mass balance upstream of the outlet because the amount of fertilizer and manure applied was similar in all three models. At the field scale, we used median regional soil test phosphorus levels (Williams et al., 2015) across all agricultural fields to better represent legacy phosphorus, an important contributor of phosphorus in this system (Duncan et al., 2017; Muenich et al., 2016; NRCS-CEAP, 2016). As SWAT simulates the phosphorus transported with surface runoff and macropore flow based on its concentration in the top soil layer, greater edge of field phosphorus losses would be expected for greater soil test phosphorus values (King et al., 2017). Labile phosphorus in agricultural fields was increased by 30 times in the Apostel model compared to Kalcic and Kujawa. The impact was amplified because our stratification scheme further increased concentrations in the top 5 cm of the soil layer. To better match observed outlet phosphorus loads during calibration, tile phosphorus loads were reduced through a substantial decrease in the soil cracking parameter (SOL_CRK) to compensate for considerable increases in surface phosphorus losses. These changes meant that improving DRP loading at the outlet worsened the field level surface and subsurface partitioning of DRP. This partitioning is important in subsurface-dominated watersheds because tile drains have been shown to be a key contributor of DRP, resulting in a shift in focus to drainage management strategies (Smith et al., 2015b). While the Apostel model is an improved representation of the high legacy phosphorus values within the watershed, a blanket soil phosphorus initialization may not have been sufficient to capture the range of values seen in the monitored fields. Finer scale data may be needed to more accurately represent the magnitudes and pathways of DRP losses.

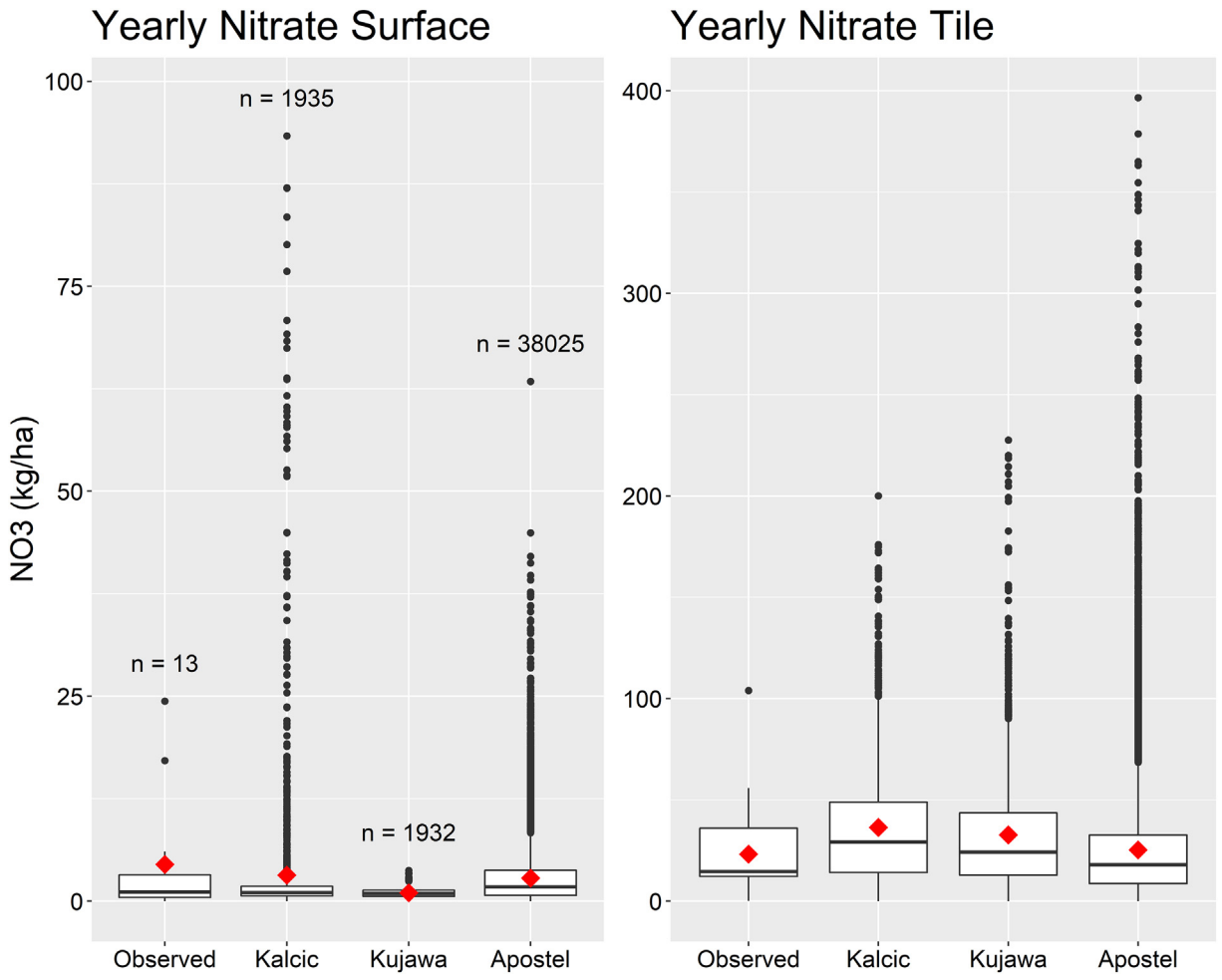


Fig. 7. Box and whisker plots of average annual surface and tile NO3 loads across models HRUs and observed sites from 2013 to 2015. Diamonds represent mean values. Boxes are bounded by 25th and 75th percentiles with whiskers representing 10th and 90th percentiles.

4.3. Parameter constraints on calibration

Two key differences in the parameterization of the Apostel model compared to the other two models were the refinement of certain model parameters based on regional data, and the use of a different model subroutine. These refinements at the HRU and basin scales are based on improved intra-watershed process knowledge and were expected to produce more realistic processes at the field scale. However, in doing so, they reduce the number of sensitive parameters available for calibration. Therefore, we used the ensemble to investigate how the parameterization informed by regional data influenced the upstream model performance. Three key refinements made through the adjustment of parameter values in the Apostel model included: (1) the intensity and contribution of tile drainage and preferential flow, (2) initializing soil phosphorus content based on regional data, and (3) setting fixed snow parameters based on regional climate.

In significantly tile-drained watersheds, fine tuning calibration is heavily focused on tile drainage parameters (Guo et al., 2018; Daggupati et al., 2015). Several parameters associated with tile drainage were refined and set to fixed values for all three models, including SDRAIN (tile spacing distance) and DRAIN_CO (daily drainage coefficient). SDRAIN and DRAIN_CO adjust the rate of flow traveling to and through the subsurface tiled system (Moriasi et al., 2012, 2013). LATKSATF was a fixed HRU-level input in Kalcic and Kujawa but used for parameter calibration in Apostel as a larger value was needed to simulate a larger tile discharge in this system. In the Apostel model, we adjusted values based on soil drainage characteristics to mimic tile density patterns in the region, where very poorly drained fields were more likely to have closer drain spacing (Smith et al., 2015a). Overall, the three models were similarly constrained by setting tile drainage parameters as fixed inputs.

Soil cracking and preferential flow are important pathways of flow and phosphorus to tile drains. Preferential flow has been shown to be prevalent in the heavy clay soils of the MRW (Smith et al., 2015a) and the ICRK routine is a SWAT sub-model that allows flow through cracks in the soil (Arnold et al., 2005). The soil cracking routine (ICRK) differed among the models, with this subroutine turned on in Kujawa and Apostel to simulate preferential flow, but turned off in the Kalcic model. The parameters LATKSATF and SOL_CRK impact soil hydraulic conductivity and crack volume in the soil profile, both of which determine contribution to tile drains. The presence of this flow pathway, a major structural difference between the Kalcic and Kujawa model iterations, resulted in a decrease in the surface flow and increase in tile flow that was more representative of the system's greater tile contributions, as well as a new pathway for DRP to enter tile drains.

A limitation of the current SWAT tile drainage simulation is that it does not incorporate the transport of particulate-bound phosphorus through tile drains. Studies have shown that this can contribute as much as 90% of TP through tile drains (Merriman et al., 2018; Christianson et al., 2016) while in our EOF sites it contributed nearly 50%. Inclusion of only DRP through tiles in SWAT resulted in a strong underestimation of TP, skewing the surface to tile ratios further in our modeled results. Even with available tile drainage data at the field scale, achieving correct tile distributions of phosphorus is difficult

under these model structural deficiencies (Merriman et al., 2018; Guo et al., 2018).

The introduction of soil test phosphorus initialization, discussed in the section above, was also a source of limited calibration flexibility introduced in the Apostel model. The HRU-level soil phosphorus concentration (SOL_SOLP) was incorporated as a fixed parameter input instead of being used in the calibration process, which had played a key part in previous model calibrations. While the soil phosphorus values initialized in the model are more representative of the watershed, we were unable to reach desirable performance at the watershed outlet while still meeting a soft calibration target of approximately 40% of DRP through tiles that was used in Kalcic and Kujawa calibrations (Kalcic et al., 2016; Kujawa et al., 2020). While Kalcic and Kujawa were able to achieve good performance at the watershed outlet with considerable DRP loading through tile drains, they were not based on data-driven initial soil phosphorus concentrations. This could indicate an underlying structural issue concerning phosphorus initialization and transport in the model. Potential issues of phosphorus transport through leaching and surface runoff may explain the large increase of phosphorus at the outlet that was seen in the Apostel model when initial soil phosphorus values were increased and before recalibration took place (Lu et al., 2016; Bauwe et al., 2019).

Using predetermined snow parameters in the Apostel calibration further constrained parameter selection. While this provided a better representation of snowfall for the region, it hindered our ability to calibrate other key model components that influence both hydrology and nutrient transport. Merriman et al. (2018) found air temperature, snowfall, and snowmelt temperature threshold, as well as parameters impacting surface/subsurface separation, were important in the accuracy of tile flow simulation. Our use of local information to identify the parameter values prior to calibration limited calibration flexibility that could have been used to improve flow and nutrient dynamics.

While the three MRW SWAT models had several key differences in their use of subroutines and fixed parameter values, these changes were incorporated based on regional information that did not capture the spatial heterogeneity expected among farm fields. Refinement of parameter estimates to smaller scales is needed to model these complex systems.

4.4. Future work

The use of multiple models gives us a window into the capabilities of the underlying framework of SWAT to simulate internal watershed processes. By allowing for alternative internal representations of a system, which all produce acceptable watershed level results, we can explore a range of potential internal simulations highlighting critical limitations in model structure, parameterization, and data availability.

More fine-scale data is needed to calibrate and validate models as complexity increases and spatial scales are further refined (Karki et al., 2020). This is evident in this work as a spatial discrepancy persists between HRUs and management data, limiting true field-scale implementation. While the framework exists within the Apostel model to use field-level information, limited data exist for such fine-grained information across large regions. A growing EOF network can provide increased information to implement field-level management schemes. However, with cost constraints limiting implementation at scales as large as the Maumee, further expansion of EOF monitoring could still provide greater confidence and range in nutrient transport budgets for use in model calibration. Technological advancements in remote sensing can also increase spatially-explicit data for model inputs and calibration. The use of remotely sensed data for refined DEMs through LiDAR, land cover through satellite-based LAI (Ma et al., 2019), and tile drainage detection (Gökkaya et al., 2017), are promising sources of field-scale information. As more data become available for developing and calibrating models, there is a need for advanced methods of computer learning techniques and data assimilation to effectively use the increased levels of information (Sarkar and Mukunda, 2018; Jiang and Wang, 2019).

Growing model complexity increases the potential for misrepresenting a system through model structure or over-parameterization. As we have shown, a model can be successfully calibrated to produce acceptable end point outcomes, yet not properly simulate internal watershed processes. Narrowing the parameter space, through the inclusion of increased input and validation data, can increase accuracy, but can also appear unrealistic if the model structure does not support the additional restrictions. Continuing to add high-resolution data for inputs and parameterization can make calibration of the model more difficult, as shown by field-scale studies that include additional model performance constraints (Guo et al., 2018; Merriman et al., 2018). This is especially important for parameters in the SWAT model for which a single value is used across the watershed, such as the snow parameters discussed in this work. However, the increased inflow of data can allow for more thorough uncertainty characterization and the identification of structural improvements needed in the model.

The ability of a model to properly capture certain system processes dictates the potential uses of a model. If a model is unable to capture key watershed transport processes, then we may not be able to reliably use the model to assess dynamics and management scenarios related to those processes. In this study we find that the simulation of field level dynamics of phosphorus transport through tile drains and surface runoff were fairly uncertain, and so processes and scenarios based on phosphorus transport pathway may also be uncertain. Future work in this area should focus on potential structural issues around surface runoff generation and surface phosphorus loading to more accurately represent these processes at the field scale, as well as phosphorus transport in tiles. In short, the identification of model weaknesses and knowledge gaps are key for model improvements required for effective model use.

5. Conclusions

Ensemble modeling has been used to gauge uncertainty and reduce potential biases introduced through decisions made during model development. We used three models, calibrated and validated at the watershed outlet, to assess performance at the farm field scale for capturing field level hydrology. This was done without incorporating true field-scale management data, so that we could test whether baseline model performance was representative across tile-drained agricultural fields. The ensemble models were developed iteratively within the same research group, and so there are similarities among the models, as well as consistent improvement of input data and assumptions over time. We found that:

- Performance was similar among MRW SWAT models for discharge and nutrient loading at the watershed outlet, as well as at the field scale, apart from field level phosphorus loading.
- While the latest model development, the Apostel model, provides finer resolution in land management and parameterization, it showed minimal improvement in field-scale model performance. This was at least partly caused by limited access to field-scale management data to match the model's spatial refinement.
- Incorporating representative soil phosphorus initialization based on a median regional soil test phosphorus concentration is a key enhancement of the Apostel model. However, the sensitive nature of SWAT's runoff processes to surface soil phosphorus concentration resulted in unrealistic partitioning of surface and subsurface phosphorus loading.
- Refinement of inputs and parameters based on intra-watershed information restricted the number of sensitive parameters available for adjustment during calibration, hindering successful field level process representation.
- While increased data availability can limit model calibration flexibility, it can highlight the need for structural changes or key data improvements necessary for realistic model performance.
- Therefore, as new data and management techniques are introduced,

modelers should evaluate performance across a range of model structures and parameterizations.

CRediT authorship contribution statement

Anna Apostel: Writing – Original Draft, Methodology, Conceptualization, Data Curation, Formal Analysis, Visualization, Software, Validation

Margaret Kalcic: Funding Acquisition, Project Administration, Conceptualization, Methodology, Writing – Review & Editing, Supervision, Software, Validation

Awoke Dagnew: Methodology, Software

Grey Evenson: Methodology, Software, Writing – Review and Editing

Jeffrey Kast: Methodology, Writing – Review and Editing

Kevin King: Resources, Data Curation, Writing – Review and Editing

Jay Martin: Writing – Review and Editing, Supersvision

Rebecca Logsdon Muenich: Methodology, Software, Writing – Review and Editing, Supervision

Donald Scavia: Funding Acquisition, Project Administration, Conceptualization, Writing – Review and Editing, Supervision

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project was funded by the National Science Foundation Coastal SEES grant OCE-1600012, with additional support from a Harmful Algal Bloom Research Initiative grant from the Ohio Department of Higher Education, and the U.S. Department of Agriculture, Agricultural Research Service Soil Drainage Unit.

Appendix A. Supplementary data

SI contains detailed information on the Apostel MRW SWAT model development and set up including assumptions, as well as parameter adjustment and additional model calibration and validation details for all three models. Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.143920.

References

Arabi, M., Frankenberger, J., Engel, B.A., Arnold, J.G., 2008. Representation of agricultural conservation practices with SWAT. Hydrol. Process. 22, 3042–3055. https://doi.org/10.1002/hyp.6890.

Arnold, J.G., Potter, K.N., King, K.W., Allen, P.M., 2005. Estimation of soil cracking and the effect on surface runoff in a Texas Blackland Prairie watershed. Hydrol. Process. 19, 589–603. https://doi.org/10.1002/hyp.5609.

Arnold, J.G., Kiniry, J.R., Srinivasan, R., Williams, J.R., Haney, E.B., Neitsch, S.L., 2012. Input/output documentation, version 2012. Texas Water Resources Institute TR-439.

Baker, D.B., Johnson, L.T., Confesor, R.B., Crumrine, J.P., 2017. Vertical stratification of soil phosphorus as a concern for dissolved phosphorus runoff in the Lake Erie Basin. J. Environ. Qual. 46, 1287–1295. https://doi.org/10.2134/jeq2016.09.0337.

Bauwe, A., Eckhardt, K.-U., Lennarts, B., 2019. Predicting dissolved reactive phosphorus in tile-drained catchements using modified SWAT model. Ecohydrol. Hydrobiol. 19, 198–209. https://doi.org/10.1016/j.ecohyd.2019.03.003.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Burnett, E.A., Wilson, R.S., Roe, B., Howard, G., Irwin, E., Zhang, W., Martin, J., 2015. Farmers, Phosphorus and Water Quality: Part II. A Descriptive Report of Beliefs, Attitudes and Best Management Practices in the Maumee Watershed of the Western Lake Erie Basin. Columbus, OH: The Ohio State University, School of Environment & Natural Resources.

Christianson, L.E., Harmel, R.D., Smith, D., Williams, M.R., King, K., 2016. Assessment and synthesis of 50 years of published drainage phosphorus losses. J. Environ. Qual. 45, 1467–1477. https://doi.org/10.2134/jeq2015.12.0593.

Cibin, R., Sudheer, K.P., Chaubey, I., 2010. Sensitivity and identifiability of stream flow generation parameters of the SWAT model. Hydrol. Process. 24, 1133–1148. https://doi.org/10.1002/hyp.7568.

CTIC, Conservation Technology Innovation Center, Regional summary of county tillage data for the Western Lake Erie Basin. Date obtained: January 17, 2013.

Culman, S., Mann, M., Sharma, S., Saeed, M., Fulford, A., Lindsey, L., Brooker, A., Dayton, E., Eugene, B., Warden, R., Steinke, K., Camberato, J., Joern, B., 2019. Converting between Mehlich-3, Bray P and ammonium acetate soil test values. The Ohio State University, College of Food, Agricultural, and Environmental Sciences. Fact Sheet: ANR-75.

Daggupati, P., Yen, H., White, M.J., Srinivasan, R., Arnold, J.G., Keitzer, C.S., Sowa, S.P., 2015. Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin. Hydrol. Process. https://doi.org/10.1002/hyp.10536.

Diebel, M.W., Maxted, J.T., Nowak, P.J., Zanden, M.J.V., 2008. Landscape planning for agricultural nonpoint source pollution reduction I: a geographical allocation framework. Environ. Manag. 42, 789–802. https://doi.org/10.1007/s00267-008-9186-3.

Duncan, E.W., King, K.W., Williams, M.R., LaBarge, G., Pease, L.A., Smith, D.R., Fausey, N.R., 2017. Linking soil phosphorus to dissolved phosphorus losses in the Midwest. Agricultural and Environmental Letters. 2, 170004. https://doi.org/10.2134/ael2017.02.0004.

Evenson, G., Kalcic, M., Wang, Y.-C., Robertson, D., Scavia, D., Martin, J., Apostel, A., Aloysuis, N., Kast, J., Kujawa, K., Boles, C., Redder, T., Confesor, R., Guo, T., Muenich, R., Dagnew, A.T., Murumkar, A., Brooker, M., 2020. Uncertainty in critical sources area predictions from watershed-scale hydrologic models. J. Environ. Manag. https://doi.org/10.1016/j.jenvman.2020.111506 In press.

Francesconi, W., Srinivasan, R., Pérez-Miñana, E., Willcock, S.P., Quintero, M., 2016. Using the soil and water assessment tool (SWAT) to model ecosystem services: a systematic review. J. Hydrol. 535, 625–636. https://doi.org/10.1016/j.jhydrol.2016.01.034.

GLWQA, (Great Lakes Water Quality Agreement) 2015. Recommended phosphorus loading targets for Lake Erie: annex 4 objectives and targets task team final report to the nutrients annex subcommittee.

Gökkaya, K., Budhathoki, M., Christopher, S.F., Hanrahan, B.R., Tank, J.L., 2017. Subsurface tile drained area detection using GIS and remote sensing in an agricultural watershed. Ecol. Eng. 108B, 370–379. https://doi.org/10.1016/j.ecoleng.2017.06.048.

Guo, T., Gitau, M., Merwade, V., Arnold, J., Srinivasan, R., Hirschi, M., Engel, B., 2018. Comparison of performance of tile drainage routines in SWAT 2009 and 2012 in an extensively tile-drained watershed in the Midwest. Hydrol. Earth Syst. Sci. 22, 89–110. https://doi.org/10.5194/hess-22-89-2018.

Guswa, A.J., Brauman, K.A., Brown, C., Hamel, P., Keeler, B.L., Sayre, S.S., 2014. Hydrologic modeling to support decision making. Water Resour. Res. 50, 1–10. https://doi.org/10.1002/2014WR015497.

Hanrahan, B.R., King, K.W., Williams, M.R., Duncan, E.W., Pease, L.A., LaBarge, G.A., 2019. Nutrient balanaces influence hydrologic losses of nitrogen and phosphorus across agricultural fields in northwestern Ohio. Nutr. Cycl. Agroecosyst. 113, 231–245. https://doi.org/10.1007/s10705-019-09981-4.

Her, Y., Chaubey, I., 2015. Impact of the numbers of observations and calibration parameter on equifinality, model performance, and output and parameter uncertainty. Hydrol. Process. 29, 4220–4237. https://doi.org/10.1002/hyp.10487.

Hrachowitz, M., Clark, M.P., 2017. HESS opinions: the complementary merits of competing modelling philosophies in hydrology. Hydrol. Earth Syst. Sci. 21, 3953–3973. https://doi.org/10.5194/hess-21-3953-2017.

IPNI. 2011. A Nutrient Use Information System (NuGIS) for the U.S. Norcross, GA. November 1, 2011. Available online: www.ipni.net/nugis. Date Accessed: September, 2016.

Jiang, D., Wang, K., 2019. The role of satellite-based remote sensing in improving simulated streamflow: a review. Water. 11 (8), 1615. https://doi.org/10.3390/w11081615.

Kalcic, M.M., Chaubey, I., Frankenberger, J., 2015. Defining soil and water assessment tool (SWAT) hydrologic response units (HRUs) by field boundaries. International Journal of Agricultural and Biological Engineering. 8, 1–12. https://doi.org/10.3965/j.ijabe.20150803.951.

Kalcic, M.M., Kirchhoff, C., Bosch, N., Muenich, R.L., Murray, M., Gardner, J.G., Scavia, D., 2016. Engaging stakeholder to define feasible and desirable agricultural conservation in Western Lake Erie watersheds. Environ. Sci. Technol. 50, 8135–8145. https://doi.org/10.1021/acs.est.6b01420.

Karki, R., Srivastava, P., Veith, T.L., 2020. Application of the soil and water assessment tool (SWAT) at the field-scale: categorizing methods and review of applications. Transactions of the ASABE. 63; doi: 10.13031/trans.13545.

Kast, J., Long, C.M., Muenich, R.L., Martin, J., Kalcic, M.M., 2019. Manure management at Ohio confined animal feeding facilities in the Maumee River Watershed. J. Great Lakes Res. 45, 1162–1170. https://doi.org/10.1016/j.jglr.2019.09.015.

Kast, J., Apostel, A.M., Kalcic, M.M., Muenich, R.L., Dagnew, A., Long, C.M., Evenson, G., Martin, J., 2020. Source contribution to phosphorus loads from the Maumee River watershed to Lake Erie. J. Environ. Manag. (Accepted).

King, K.W., Williams, M.R., Johnson, L.T., Smith, D.R., LaBarge, G.A., Fausey, N.R., 2017. Phosphorus availability in Western Lake Erie Basin drainage waters: legacy evidence across spatial scales. J. Environ. Qual. 46, 466–469. https://doi.org/10.2134/jeq2016.11.0434.

Kladivko, E., Akhouri, N.M., Weesies, G., 1997. Earthworm populations and species dsitributions under no-till and conventional tillage in Indiana and Illinois. Soil Biol. Biochem. 29, 613–615. https://doi.org/10.1016/S0038-0717(96)00187-3.

Kujawa, H., Kalcic, M., Martin, J., Aloysius, N., Apostel, A., Kast, J., Murumkar, A., Evenson, G., Becker, R., Boles, C., Confesor, R., Dagnew, A., Guo, T., Muenich, R.L., Redder, T., Scavia, D., Wang, Y.-C., 2020. The hydrologic model as a source of nutrient loading uncertainty in a future climate. Sci. Total Environ. 724, 138004.

Lu, S., Andersen, H.E., Thodsen, H., Rubæk, G.H., Trolle, D., 2016. Extended SWAT model for dissolved reactive phosphorus transport in tile-drained fields and catchments. Agric. Water Manag. 175, 78–90. https://doi.org/10.1016/j.agwat.2015.12.008.

Ma, T., Duan, Z., Li, R., Song, X., 2019. Enhancing SWAT with remotely sensed LAI for improved modelling of ecohydrolgical process in subtropics. J. Hydrol. 570, 802–815. https://doi.org/10.1016/j.jhydrol.2019.01.024.

Maccoux, M.J., Dove, A., Backus, S.M., Dolan, D.M., 2016. Total and soluble reactive phosphorus loadings to Lake Erie a detailed accounting by year, basin, country, and tributary. J. Great Lakes Res. 42, 1151–1165. https://doi.org/10.1016/j.jglr.2016.08.005.

Martin, J.F., Kalcic, M.M., Aloysius, N., Apostel, A.M., Brooker, M.R., Evenson, G., Kast, J.B., Kujawa, H., Murumkar, A., Becker, R., Boles, C., Redder, T., Confesor, R., Guo, T., Dagnew, A., Long, C.M., Muenich, R., Scavia, D., Wang, Y., Robertson, D., 2019. Evaluating Management Options to Reduce Lake Erie Algal Blooms with Models of the Maumee River Watershed - Final Project Report.

Menne, M.J., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. Global historical climatology network - daily (GHCN-daily), version 3. NOAA National Climatic Data Center https://doi.org/10.7289/V5D21VHZ.

Merriman, K.R., Daggupati, P., Srinivasan, R., Toussant, C., Russell, A.M., Hayhurst, B., 2018. Assessing the impact of site-specific BMPs using spatially explicit, field-scale SWAT model with edge-of-filed and tile hydrology and water-quality data in Eagle Creek watershed, Ohio. Water. 10, 1299. https://doi.org/10.3390/w10101299.

Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloglu, I., Depinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., Laporte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., Zagorski, M.A., 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Natl. Acad. Sci. U. S. A. 110, 6448–6452. https://doi.org/10.1073/pnas.1216006110.

Moriasi, D., Rossi, C., Arnold, J., Tomer, M., 2012. Evaluating hydrology of the soil and water assessment tool (SWAT) with new tile drain equations. J. Soil Water Conserv. 67, 513–524. https://doi.org/10.2489/jswc.67.6.513.

Moriasi, D.N., Arnold, J.G., Liew, M.W. Van, Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*. 50, 885–900; doi: 10.13031/2013.23153.

Moriasi, D.N., Gowda, P.H., Arnold, J.G., Mulla, D.J., Ale, S., Steiner, J.L., Tomer, M.D., 2013. Evaluation of the Hooghoudt and Kirkham tile drain equations in the soil and water assessment tool to simulate tile flow and nitrate-nitrogen. J. Environ. Qual. 42, 1699–1710. https://doi.org/10.2134/jeq2013.01.0018.

Moriasi, D.N., Gitau, M.W., Daggupati, P. 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Transactions of the ASABE. 58, 1763-1785; doi: 10.13031/trans.58.10715.

Muenich, R.L., Kalcic, M., Scavia, D., 2016. Evaluating the impact of legacy P and agricultural conservation practices on nutrient loads from the Maumee River watershed. Environ. Sci. Technol. 50, 8146–8154. https://doi.org/10.1021/acs.est.6b01421.

Muenich, R.L., Kalcic, M.M., Winsten, J., Fisher, K., Day, M., O'Neil, G., Wang, Y.-C., Scavia, D. 2017. Pay-for-performance convervation using SWAT highlight need for field-level agricultural conservation. *Transactions of the ASABE*. 60, 1925-1937; doi:10.13031/trans.12379.

NASS-CDL, 2012. USDA National Agricultural Statistics Services. Cropland Data Layer. Available online: https://nassgeodata.gmu.edu/CropScape/ Date Accessed: July 2016.

Neitsch, S., Arnold, J., Kiniry, J., Williams, J., 2011. Soil & Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute TR-406.

NRCS-CEAP, Natural Resources Conservation Service Conservation Effects Assessment Project, U.S. Department of Agriculture. 2016. Effects of conservation practice adoption on cultivated cropland acres in Western Lake Erie Basin, 2003–06 and 2012. (120 pp).

Ohio EPA, 2010. Ohio Lake Erie Phosphorus Task Force (Final Report). Ohio Environmental Protection Agency, Division of Surface Water, Columbus, Ohio.

Ohio EPA, 2013. Ohio Lake Erie Phosphorus Task Force II (Final Report). Ohio Environmental Protection Agency, Division of Surface Water, Columbus, Ohio.

Pease, L.A., King, K.W., Williams, M.R., LaBarge, G.A., Duncan, E.W., Fausey, N.R., 2017. Phosphorus export from artificially drained fields across the eastern Corn Belt. J. Great Lakes Res. https://doi.org/10.1016/j.jglr.2017.11.009.

Prokup, A., Wilson, R., Zubko, C., Heeren, A., Roe, B., 2017. 4R Nutrient Stewardship in the Western Lake Erie Basin. The Ohio State University, School of Environment & Natural Resources, Columbus, OH.

Qi, J., Li, S., Li, Q., Xing, Z., Bourque, C.P.A., Meng, F.R., 2016. A new soil-temperature module for SWAT application in regions with seasonal snow cover. J. Hydrol. 538, 863–877. https://doi.org/10.1016/j.jhydrol.2016.05.003.

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrological model using spatially distributed remotely sensed/in-situ soil moisture. J. Hydrol. 536, 192–207. https://doi.org/10.1016/j.jhydrol.2016.02.037.

Rajib, M.A., Kim, I.L., Golden, H.E., Lane, C.R., Kumar, S.V., Yu, Z., Jeyalakshmi, S., 2020. Watershed modeling with remotely sensed big data: MODIS leaf area index improves hydrology and water quality predictions. Remote Sens. 12 (13), 2148. https://doi.org/10.3390/rs12132148.

Sarkar, T., Mukunda, M., 2018. Soil erosion susceptibility mapping with the application of logistic regression and artificial neural network. Journal of Geovisualization and Spatial Analysis 2 (1). https://doi.org/10.1007/s41651-018-0015-9.

Scavia, D., Allan, J.D., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, T.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Richards, R.P., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T.M., Zhang, H., Zhou, Y., 2014. Assessing and addressing the re-eutrophication of Lake Erie: central basin hypoxia. J. Great Lakes Res. 40, 226–246. https://doi.org/10.1016/j.jglr.2014.02.004.

Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. J. Great Lakes Res. 42, 1139–1150. https://doi.org/10.1016/j.jglr.2016.09.007.

Scavia, D., Kalcic, M., Logsdon Muenich, R., Aloysius, N., Bertani, I., Boles, C., Confesor, R., DePinto, J., Gildow, M., Martin, J., Read, J., Redder, T., Robertson, D., Sowa, S., Wang, Y., Yen, H., 2017. Multiple models guide strategies for agricultural nutrient reductions. Front. Ecol. Environ. 15, 126–132. https://doi.org/10.1002/fee.1472.

Sharpley, A.N., Jones, C.A., Gray, C., Cole, C.V., 1984. A simplified soil and plant phosphorus model: II. Prediction of labile, organic, and Sorbed phosphorus. Soil Sci. Soc. Am. J. 48, 805–809. https://doi.org/10.2136/sssaj1984.03615995004800040021x.

Smith, D.R., King, K.W., Johnson, L.T., Francesconi, W., Richards, P., Baker, D., Sharpley, A.N., 2015a. Surface and tile drainage transport in the midwestern United States. J. Environ. Qual. 44, 495. https://doi.org/10.2134/jeq2014.04.0176.

Smith, D.R., King, K.W., Williams, M.R., 2015b. What is causing the harmful algal blooms in Lake Erie? Soil and Water Conservation. 70, 27A–29A. https://doi.org/10.2489/jswc.70.2.27A.

SSURGO, Soil Survey Geographic (SSURGO) Database. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Available online at https://sdmdataaccess.sc.egov.usda.gov. Date Accessed: July 2016.

Teshager, A.D., Gassman, P.W., Secchi, S., Schoof, J.T., Misgna, G., 2016. Modeling agricultural watersheds with the soil and water assessment tool (SWAT): calibration and validation with a novel procedure for spatially explicit HRUs. Environ. Manag. 57, 894–911. https://doi.org/10.1007/s00267-015-0636-4.

USGS-NED, 2016. U.S. Geological Survey, National Elevation Dataset, 1/3 Arc-Second. Available online: https://nationalmap.gov/elevation.html Date Accessed: July 2016.

USGS-NHD, 2016. U.S. Geological Survey, National Hydrography Dataset, Medium Resolution Streams. Available online: https://www.usgs.gov/core-science-systems/ngp/national-hydrography/access-national-hydrography-products Date Accessed: July 2016.

Vieth, T.L., Sharpley, A.N., Arnold, J.G. 2008. Modeling a small, northeastern watershed with detailed, field-level data. *Transactions of the ASABE*. 51, 471-483; doi: 10.13031/2013.24389.

Vitosh, M.L., Johnson, J.W., Mengel, D.B., 1995. Tri-state fertilizer recommendations for corn, soybeans, wheat and alfalfa. Extension Bulletin E-2567.

Williams, M.R., King, K.W., Dayton, E., LaBarge, G.A., 2015. Sensitivity analysis of the Ohio phosphorus risk index. *Transactions of the ASABE*. 58, 93–102; doi:doi: 10.13031/trans.58.10778.

Williams, M.R., King, K.W., Fausey, N.R., 2016. Edge-of-field research to quantify the impacts of agricultural practices on water quality in Ohio. Soil and Water Conservation. 71, 9A–12A. https://doi.org/10.2489/jswc.71.1.9A.

Wilson, R.S., Burnett, L., Ritter, T., Roe, B., Howard, G., 2013. Farmers, Phosphorus and Water Quality: A Descriptive Report of Beliefs, Attitudes and Practices in the Maumee Watershed of Northwest Ohio. The Ohio State University, School of Environment & Natural Resources, Columbus, OH.

Yen, H., Bailey, R.T., Arabi, M., Ahmadi, M., White, M.J., Arnold, J.G., 2014. The role of interior watershed processes in improving parameter estimation and performance of watershed models. J. Environ. Qual. 43, 1601–1613. https://doi.org/10.2134/jeq2013.03.0110.