Summer 2016

Comparing Soil Datasets with the APEX Model: Calibration and Validation for Hydrology and Crop Yield in Whatcom County, Washington

Andrew M. Monks
Western Washington University, dmmonks@gmail.com

COMPARING SOIL DATASETS WITH THE APEX MODEL: CALIBRATION AND VALIDATION FOR HYDROLOGY AND CROP YIELD IN WHATCOM COUNTY, WASHINGTON

By

Andrew M. Monks

Accepted in Partial Completion
Of the Requirements for the Degree
Master of Science

Kathleen L. Kitto, Dean of the Graduate School

ADVISORY COMMITTEE

Chair, Dr. David U. Hooper

Dr. Benjamin Miner

Dr. James Helfield

Dr. David Wallin

COMPARING SOIL DATASETS WITH THE APEX MODEL: CALIBRATION AND VALIDATION FOR HYDROLOGY AND CROP YIELD IN WHATCOM COUNTY, WASHINGTON

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By

Andrew M. Monks

June 2016

ABSTRACT

Controlling pollution from agricultural lands is a priority for improving watershed health. Best management practices (BMPs) recommend strategies such as riparian buffers and altered fertilizer application timing and rates for reduction of nutrient and sediment export from agricultural watersheds, but BMP effectiveness in nutrient retention can vary greatly depending on differences in crops, soils, and topography.

Conducting nitrogen (N) and phosphorus (P) measurements in all BMP projects is generally not feasible, so well-validated models can help estimate benefits on the watershed scale. This project uses the Agricultural Policy/ Environmental Extender (APEX) model to simulate crop yield, streamflow, and surface runoff in a small watershed in Whatcom County, Washington, to prepare the model for future use in estimating nutrient and sediment retention benefits by BMPs. The APEX model requires detailed inputs for soils, climate, cropping system, and agricultural management; outputs must be calibrated and validated against existing environmental data. No current consensus exists as to the ideal set of soil data for the APEX model. I tested the APEX model for three different soils datasets: the Soil Survey Geographic Database (SSURGO), the National Cooperative Soil Survey (NCSS), and the Nutrient Tracking Tool (NTT), to determine the best dataset to use in terms of ease of use and model fit.

I modeled the northern Kamm Creek watershed, a 227 hectare watershed that contains a diverse representation of Whatcom County cropping systems. As the first APEX modelling effort in western Washington, this study investigated parameters for blueberry and raspberry, two crops new to the APEX model, while testing model performance with three

different sets of soils data. I manipulated key parameters in two of the datasets to evaluate their effects on hydrology and yield.

The model performed well for streamflow and surface runoff across all soils during calibration, with satisfactory validation for surface runoff, but not streamflow. Performance for crop yields, however, varied across both crop type and soil data sets. Simulated crop yields fell within 10% of county-reported average yields for four of the five soils for blueberry, raspberry, and corn silage crops, whereas NTT soils drastically underestimated yields of both berry crops.

I recommend applying the SSURGO soils dataset to future APEX modelling in Whatcom County, as it had the best model fit for hydrology and crop yields. Further recommendations are made for obtaining data to parameterize, calibrate, and validate the model to assure accuracy for future APEX modelling efforts.

ACKNOWLEDGEMENTS

This project would not have been possible without the help of several others over the past three years. First, thank you to my advisor, Dr. David Hooper, for all of the guidance and advice you provided throughout my entire career here at WWU. Your tutelage continues to shape my skills and abilities as a teacher, communicator, and an ecologist. Many thanks go out to my thesis committee: Dr. Jim Helfield, Dr. Ben Miner, and Dr. Dave Wallin for input and support on this project. Thanks also to Mary Ann Merrill, Peter Thut, Kendra Bradford, and Mark Price for keeping the WWU Biology Department running smoothly. Thank you to my fellow graduate students who kept me company throughout this process, especially fellow APEX pioneer Bridger Cohan.

Thank you to all of the APEX developers, land managers, and agronomists that assisted me and made data available for this formidable modelling effort. Special thanks to Dr. Lisa DeVetter for giving me access to berry fields and management data to introduce two new crops to the APEX model, along with Jim Kiniry and Amber Williams who loaned me equipment and provided guidance with crop modelling. Thanks to the APEX development team for training and expertise on this complex model: Evelyn Steglich, Susan Wang, Jaehak Jeong, Luca Dora, Daniel Moriasi, Stephen Teet, Lee Norfleet, and Mari-Vaughn Johnson. Thank you to all those in Whatcom County and beyond who gave me invaluable advice and data on farming practices in the Pacific Northwest: Chuck Timblin, Chris Benedict, David Poon, Sam Baraso, David Bryla, and Perry Beale. Finally, thank you to those who helped fund this project: the Hodgson family, the Ross family, and Truc and Jerry Thon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTARY DATA

INTRODUCTION

Overview

Food security is crucial to society but the environmental impacts of agriculture can be costly. As demand for food increases with the world's population, land managers are seeking strategies to limit environmental impacts of agricultural non-point source (NPS) pollution (Chen *et al.*, 2014; Cui *et al.*, 2016). Sediment erosion and NPS pollution of nitrogen (N) and phosphorus (P) from agriculture costs an estimated $210 billion per year in the United States from losses in recreation, property values, biodiversity, and fouling of drinking water (Sobota *et al.*, 2015). Agricultural NPS pollution in runoff is implicated in soil acidification, biodiversity loss, and eutrophication and subsequent hypoxia of both freshwater and coastal marine environments (Schindler & Fee, 1974; Sobota *et al.*, 2015; Vitousek *et al.*, 1997). Globally, formation of hypoxic "dead zones" in marine environments caused by NPS pollution from heavy agriculture has led to the death of benthic organisms and disruption of ecosystem productivity (Diaz & Rosenberg, 2008; Turner & Rabalais, 2003). Nitrate pollution in drinking water is toxic to humans and can lead to methhemoglobinemia (blue baby syndrome) and increased risk of cancer (Sobota *et al.*, 2015). Agriculture accounts for over 40% of atmospheric emissions of nitrous oxide, a potent greenhouse gas (Van Grinsven *et al.*, 2013). A cost-benefit analysis of N application in European agriculture showed that costs incurred from degradation of human health and the environment might outpace benefits associated with enhanced crop yields (Van Grinsven *et al.*, 2013). Alternatively, through careful farm management and fertilizer use, crops yields can increase while decreasing environmental impact (Chen *et al.*, 2014; Cui *et al.*, 2016). Widespread adoption of farm practices that reduce NPS pollution can both combat environmental problems associated with agricultural NPS pollution and enhance economic prosperity (Maringanti, Chaubey & Popp, 2009).

Incentivizing best management practices

Best management practices (BMPs) provide guidelines for two main strategies for reduction of nutrient exports from agricultural lands: reducing fertilizer application rates and intercepting excess nutrients lost from fields (Ribaudo *et al.*, 2001). Fertilizer reduction can involve decreasing amounts of fertilizer used, applying fertilizer at strategic times, or changing to less fertilizer-intensive cropping systems (Mitsch *et al.*, 1999). On the other hand, planting riparian buffers, grassed waterways, and restoring wetlands can intercept N and P from overland runoff and lateral subsurface flow (Morris, 2014; Peterjohn & Correll, 1984; USDA, 2014). BMP effectiveness in nutrient retention, however, can vary greatly depending on differences in crops, soils, topography, climate, season, and agricultural management (Brooks *et al.*, 2015; Mulla, 2008)**.** In a study of several riparian buffers in Ontario, Vidon and Hill (2004) found that the buffer width required to remove 90% of upland nitrate inputs at different sites varied from less than 10 meters to more than 150 meters depending on soil porosity and riparian sediment depth. Nutrient trapping efficiency of riparian buffers also depends on buffer vegetation type, buffer width, slope, and time since establishment (Dosskey, Hoagland & Brandle, 2007; Fennessy & Cronk, 1997; Ranalli & Macalady, 2010). More research is needed to fully understand spatial and temporal variability in BMP effectiveness, especially at the watershed scale (Mulla, 2008).

For all BMP implementation, associated costs can hinder farmers' willingness to improve agricultural practices (Osmond *et al.*, 2015). Solutions that are economically viable for farmers stand the greatest chance of widespread adoption and success (Osmond *et al.*, 2015). Providing financial incentives for projects that enhance ecosystem services, such as clean air and water, habitat, and flood protection, is a promising avenue for improving agricultural watershed management. Ecosystem service markets can monetize these benefits in the form of credits (Chesapeake Bay Commission, 2012; Electric Power Research Institute, 2014; Willamette Partnership, 2015). Dozens of instances of nutrient credit accounting and trading currently exist worldwide (Greenhalgh & Selman, 2012). Awarding credits for nutrient reduction projects requires accurate assessment of the biological processes affected by BMPs and effective accounting of any resulting nutrient reductions. Uncertainty in BMP effectiveness, however, can undermine this effort, leading to high trading ratios, where a polluter must purchase more credits from a manager looking to reduce nutrient loads (Olander *et al.*, 2014).

The APEX Model

Conducting N and P measurements in all projects in an ecosystem service trading market is generally infeasible, so developing well-validated models is essential. The Agricultural Policy/ Environmental Extender (APEX) is a computer model developed by the United States Department of Agriculture- Agricultural Research Service (USDA-ARS) in Temple, Texas, to simulate agricultural watersheds using algorithms from the Environmental Policy Impact Climate (EPIC) model and the Soil and Water Assessment Tool (SWAT) (Williams *et al.*, 2008). The model was designed to bridge the spatial gap between existing models that focus on impacts of agricultural practices on single farm (EPIC) and whole watershed (SWAT) scales (Gassman *et al.*, 2010). The major components of the EPIC model are climate, hydrology, crop growth, nutrient cycling, soil erosion, carbon cycling, and agricultural management practices (Steglich & Williams, 2013). EPIC is capable of simulating a single field, while APEX has an additional component of subarea routing that can link several fields within a watershed and generate simulations on the sub-watershed and watershed scales. While SWAT operates over a regional scale, APEX can track movement of water, sediment, and nutrients (in the form of sediment-bound and dissolved species) from agricultural fields to streams in watersheds ranging from single fields to multi-state agricultural regions (Gassman *et al.*, 2010). The model is an important tool for estimating efficacy of BMPs at both the field scale and the watershed scale (Francesconi *et al.*, 2015; Plotkin *et al.*, 2013; Qiu *et al.*, 2002; Williams *et al.*, 2006). APEX is currently in use to model potential benefits of various BMPs as a part of the USDA's national Conservation Effects Assessment Program (CEAP; Santhi *et al.*, 2014; USDA, 2014). The spatial flexibility of APEX and its ability to simulate N and P migration in sediment, runoff, and subsurface flow make it potentially useful as a tool in nutrient trading in the Pacific Northwest. Before APEX can be reliably used in a given region, however, it must be calibrated against a set of existing measured data for all simulated outputs (crop yield, streamflow, runoff, N and P loss, etc.). Calibration is an iterative process that continues until model fit is deemed acceptable, at which point the model must be validated against a separate, untrained set of data (Wang, Kemanian & Williams, 2011).

Despite the wide applications and uses of APEX, however, it had not been explicitly calibrated and validated in this region. APEX requires detailed inputs of climatic, soil, and hydrologic data and can simulate a wide variety of agricultural management scenarios and cropping systems. Several issues needed resolution for its successful use at the small watershed scale, the critical scale for evaluating water quality benefits from BMPs (Strauss *et al.*, 2007). Understanding how soil properties affect dominant hydrologic pathways in a watershed is fundamental to predicting BMP effectiveness for nutrient and sediment retention (Brooks *et al.*, 2015). Soil properties determine how APEX simulates hydrology, crop growth, erosion, and nutrient flow (Steglich & Williams, 2013). There is no consensus in the modelling community on what soils database interfaces best with the APEX model. For this modeling effort, I explored three separate sources of soil data. The USDA-NRCS developed and maintains two of these sources, the Soil Survey Geographic Database (SSURGO) and National Cooperative Soil Survey (NCSS; Dr. Nathan Nelson, Kansas State University, pers. comm., Dr. M. Lee Norfleet, USDA-NRCS pers. comm.). The third source was from the Nutrient Tracking Tool (NTT), an agricultural model that interfaces with APEX (Dr. Ali Saleh, Tarleton State University, pers. comm.). All three soils datasets contained detailed information on soil physical characteristics and nutrient content organized by soil layer, including layer depth, bulk density, sand, silt, and clay content, pH, percent organic carbon, and saturated hydraulic conductivity. These data sets differed in readiness for input into APEX, as well as public accessibility. For example, while soil layers from NTT were recommended by some APEX users, incorporating them requires sending spatial data from the modeled watershed to the developers of the NTT model so they can extract APEX-ready soils data. These issues could have a large impact on ease of adoption of APEX in new locations. Testing the APEX model using different soils data sources allowed me to determine which soils produced the most accurate hydrologic outputs and crop yields for my area, to investigate what soil properties affected these outputs, and to make recommendations on which data sets are most practicable.

Accurate parameterization of crop growth and management in APEX drives key processes for modeling hydrologic and nutrient fluxes. Most crops modeled in APEX are annuals; however, perennial shrubs dominate cropping in some regions (e.g., blueberries and raspberries in western Washington) and are important components of riparian restoration projects aimed at reducing nutrient loading in the Pacific Northwest. Including blueberry and raspberry crops in this project required working with the developers of the ALMANAC model (Agricultural Land Management Alternative with Numerical Assessment Criteria), which is used by APEX to simulate crop growth, nutrient uptake, and yield (USDA-ARS). Furthermore, limited access to detailed farm management data, including crop rotations, fertilizer application strategies, and timing of tillage, planting, and harvest, can impede accurate parameterization of APEX, as most of these field-level data are confidential. Finally, locating calibration and validation data for hydrology and crop yield pose additional challenges. Using APEX to evaluate BMPs requires accounting for seasonal variability in climate and watershed hydrology. This means that calibration data must have a monthly timestep or less. Furthermore, obtaining local crop yield data for several crops on a yearly basis is required to capture year-to-year variability in crop yields in a simulated watershed.

Study overview

I applied the APEX model to an agricultural watershed, Kamm Creek, in northern Whatcom County, Washington, to test using APEX to estimate nutrient retention benefits of various BMPs. Whatcom County's agricultural economy has a market value of over $350 million per year, but like the rest of the nation, faces many issues with NPS pollution (USDA-NASS, 2014). Dairy products account for over 50% of the agricultural economy, but are connected with N, P, and fecal coliform pollution of surface and groundwater from manure production and spreading on grass hay and silage corn fields (Carey & Cummings, 2012; Olander *et al.*, 2014; Rosenstock *et al.*, 2014). Many of Whatcom County's agricultural waterways are listed on the Environmental Protection Agency's 303 (d) list of impaired waterways due to fecal coliform, sediment pollution, and reduced dissolved oxygen (Bandaragoda *et al.*, 2012). Many of these streams are critical to migration, spawning, and rearing for chinook salmon (*Oncorhynchus tshawytscha*), steelhead (*O. mykiss*) and bull trout (*Salvelinus confluentus*), three salmonid fish species listed as threatened under the Endangered Species Act (Shared Strategy Development Committee, 2007). Furthermore, fecal coliform from streams has forced multiple closures of Whatcom County's near-shore commercial shellfish harvest, which generates $79 million per year (Snyder, 2015; Whatcom County Public Works-Natural Resources, 2015). State and federal agencies, local governments, and Western Washington University are currently collaborating with landowners to test payments for agricultural practices that promote ecosystem function and agricultural economies (MacKay, 2013).

Calibration and validation of APEX in the Kamm Creek watershed proceeded in a stepwise manner that involved data collection and model testing against existing environmental datasets (Figure 1). I built different APEX models based on three different soil data sources: SSURGO, NCSS, and NTT. I also developed growth models for blueberry and raspberry, perennial berry crops not already covered by APEX by measuring growth of these crops in the field to parameterize APEX (Dr. James Kiniry, USDA-ARS pers. comm.). Following initial model runs with each of the soil data sets, I created additional soil data sets by varying key soil parameters to better understand what factors drove differences among the three original data sets. These initial calibration and validation steps will allow future testing of the model for simulating nutrient fluxes under a variety of conservation practices.

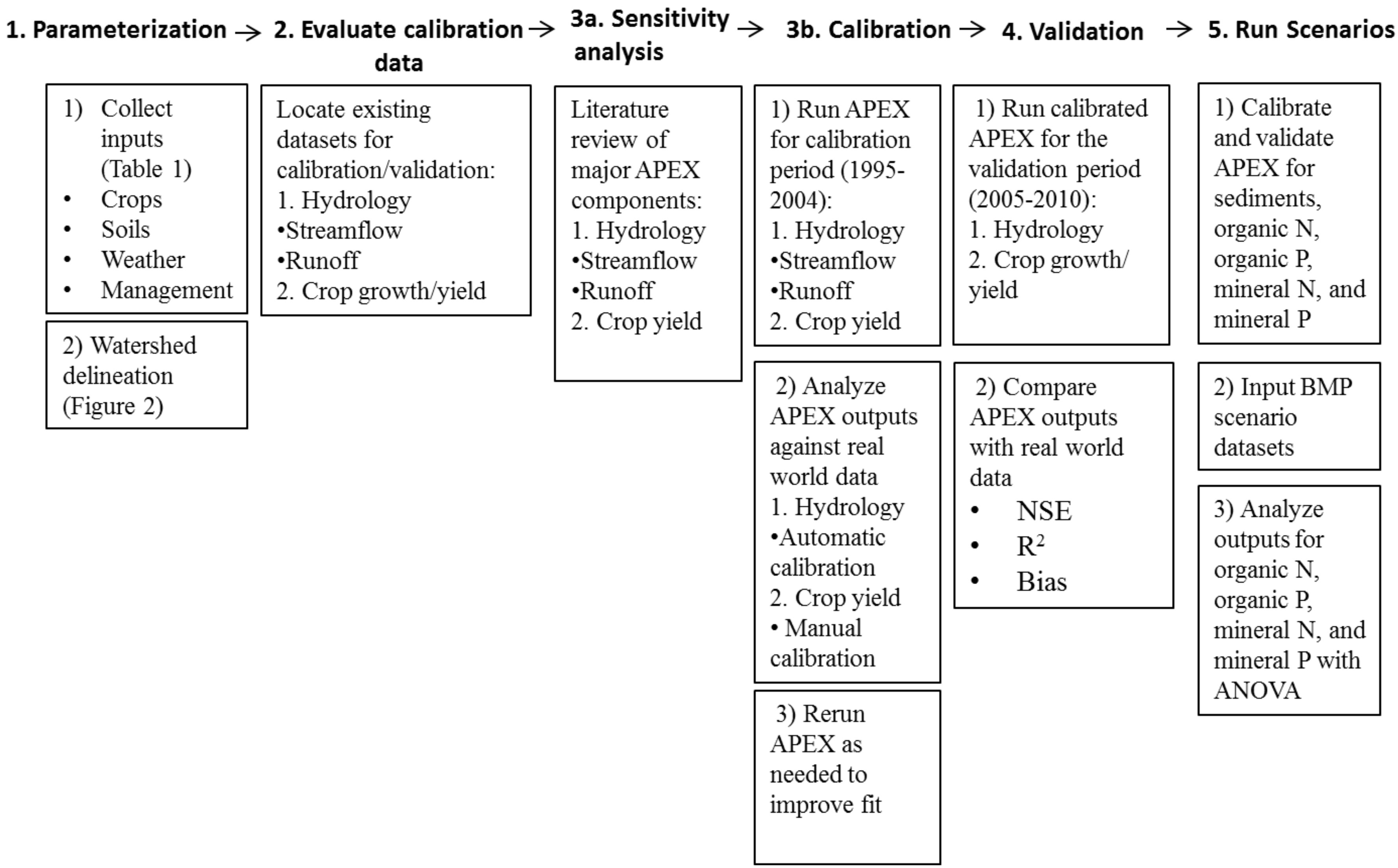

Figure 1. Flow chart of experimental approach to APEX model application in Whatcom County, WA. Model setup proceeds in a stepwise direction. This project covers steps one through four.

OBJECTIVES

Below, I enumerate the key progression of goals necessary to achieve calibration and validation of APEX in this project (Fig. 1). First, I gathered relevant soil, cropping system, management, and climate data as inputs to the APEX model. Second, I gathered existing data for the watershed of interest for calibration and validation of hydrology and crop yield. I calibrated the model by manipulating sensitive parameters (as determined by a literature search) and analyzing model fit to calibration data for watershed hydrology and crop yield. To calibrate for hydrology, I used an automatic calibration tool, APEX- auto-Calibration and UncerTainty Estimator (APEX-CUTE) that repeatedly runs the model in search of optimal combinations of sensitive parameters to maximize model accuracy. I manually calibrated crop yield by adjusting sensitive parameters. Finally, I validated the model against a separate sent of data to test the efficacy of model calibration. This study aimed to:

1) Determine which of the three available soil datasets produce the closest model fit for hydrology and crop yield.
2) Determine which soil parameters are critical drivers of hydrology and yield outputs.
3) Parameterize two perennial shrubs, blueberry and raspberry, for the first time in APEX, while evaluating the model's performance in simulating shrub growth and annual yield.
4) Make recommendations for future use of the APEX model locally and regionally by outlining key steps in model preparation that are not covered in the documentation, to enhance the ability of APEX to simulate effects of BMPs.

METHODS

Study system

Site Description

This study took place in the northern portion of the Kamm Creek watershed, a lowland agricultural watershed located in northern Whatcom County, Washington. The modeled watershed covers 227 ha and is relatively flat, with slopes between zero and four percent, and an average slope of two percent (Goldin, 1992). The lowlands of Whatcom County have a mild climate that is heavily influenced by proximity to the Pacific Ocean, with an average yearly temperature of ~10ºC. January is the coldest month, with an average temperature of ~3ºC, and July is the warmest month with an average temperature of ~17ºC (Goldin, 1992). The annual average precipitation between 1960 and 2010 was 1400 mm, with 70% falling as rainfall between October and March. Lowland soils are diverse, but generally consist of well-drained silt loam (Goldin, 1992). Whatcom County's primary agricultural products are dairy, raspberries, blueberries, and strawberries (USDA-NASS, 2014).

I chose the northern Kamm Creek watershed for this study for several reasons. The watershed contains a variety of crop types that are representative of Whatcom County agriculture including blueberry, raspberry, silage corn, and pasture (orchard grass). Silage corn and orchard grass are associated with dairy operations. 58% of the watershed is used for agriculture, while the other 42% of the watershed includes alternative land uses: forests, residential areas, and fallow fields. This watershed contains no tile drainage, though tile drainage is present downstream in the greater Kamm Creek watershed. APEX can simulate tile drainage, but this further complicates model parameterization and applicability to other watersheds. Finally, presence of calibration and validation data is essential to applying APEX in a new region. Within the Kamm Creek watershed, several environmental studies have monitored both hydrologic and nutrient conditions at multiple points in the watershed (Bandaragoda *et al.*, 2012; Matthews & Vandersypen, 1998), although using these data required some assumptions, as described below.

Table 1. Data needs and sources for APEX parameterization, calibration, and validation. Inputs refer to parameterization components and outputs to calibration components.

Component	APEX inputs & outputs	Source(s)
Lowland Whatcom County climate data	Daily minimum/maximum temperature, daily precipitation	National Climate Data Center, Natural Resource Conservation Service (NRCS) Snotel stations (WRIA1 2012)
Lowland Whatcom County soils data (by layer)	Layer depth, % silt/sand/clay, bulk density, saturated hydraulic conductivity, soil water content, organic content	1. NRCS: Soil Survey Geographic Database (SSURGO) 2. Nutrient Tracking Tool (NTT) 3. National Cooperative Soil Survey (NCSS)
Cropping system/ Management data: 1. Berry crops 2. Grass/ Hay/ Corn/ Dairy	Crop type, fertilizer application, irrigation, planting/harvest, tillage	1. WSU-Extension Mt. Vernon: Dr. Lisa Wasko-DeVetter 2. Whatcom Conservation District: Chuck Timblin, WSU-Extension Bellingham: Chris Benedict
Crop Yield Data (for calibration) 1. Silage Corn 2. Orchard Grass 3. Blueberry 4. Red Raspberry	Dry kg/ha/yr	1. NASS: Whatcom County (1995-2008) 2. NASS: Whatcom County (2002, 2007) 3. NASS: Whatcom County (1987-2002) NASS: Washington State (1985-2011) 4. NASS: Whatcom County (1985-2002) NASS: Washington State: (1985-2011)
Streamflow Data (for calibration)	mm/month	Topnet Water Management Model (WRIA1 2012)
Surface runoff Data (for calibration)	mm/month	Topnet Water Management Model (WRIA1 2012)
2013 National Agriculture Imagery Program Mosaic	Field area, crop type, water routing	NRCS
3M Digital Elevation Map	Watershed area, slope, water routing, reach length	NRCS
Washington State crops: Whatcom County	Crop type, field size	Washington State Department of Agriculture

Calibration and Validation Data

APEX calibration and validation required existing measured data against which to compare model outputs, which were not readily available as direct measurements for my test watershed under current land use conditions (Table 1). The northern Kamm Creek watershed has changed dramatically since the 1980s, with a marked decrease in lands used for dairy farms, an increase in blueberry and raspberry cultivation, and an increase in residential land (Matthews & Vandersypen, 1998; C. Timblin, Whatcom Conservation District, pers. comm.). Hydrologic processes required calibration first, as they drive the transport of sediment and nutrients (Wang, Kemanian & Williams, 2011). The Washington State Water Resource Inventory Area No. 1 (WRIA1) Joint Board provided data on the water cycle for all watersheds and subbasins within the agricultural portions of Whatcom County (Bandaragoda *et al.*, 2012). This water budget used the TOPNET hydrologic model to simulate major processes including streamflow, runoff, rainfall interception by vegetation, evapotranspiration, and snow accumulation. Like APEX, TOPNET uses input data on soil type, climate, land use, vegetation, and artificial drainage to model hydrology. The streamflow component of the TOPNET model was directly calibrated and validated against streamflow data from several U.S. Geological Survey (USGS) and Washington Department of Ecology (WA DoE) stream gauges in the Lower Nooksack Basin. Although none of these gauges were located along Kamm Creek, two were along the mainstem of the Nooksack River, into which Kamm Creek flows, and several were in smaller watersheds in Whatcom County, including Fishtrap and Bertrand Creeks, two agricultural watersheds adjacent to Kamm Creek (Bandaragoda *et al.*, 2012; Tarboton, 2007). One subbasin in particular within the Kamm Creek watershed, node 173, contained the general area of the northern Kamm

Creek watershed, so I used TOPNET hydrologic outputs at this location to calibrate and validate my model. Use of TOPNET-modeled data rather than direct measurements to calibrate APEX was not ideal. However, very little hydrologic data exist for current conditions on the subbasin scale for Whatcom County. TOPNET hydrologic outputs were generated at the daily scale from 1952 to 2011, providing continuous data for calibration. A calibration dataset that included daily data offered flexibility to look at hydrologic trends on a daily, monthly, or annual scale (Sudheer *et al.*, 2007).

To calibrate for crop yield, I would ideally use yield data from individual fields within my watershed but these data were confidential. Instead, I obtained annual crop yield data for calibration from the National Agricultural Statistical Service (Table 1; USDA-NASS, 2014). While I found county-wide average yields from 1985-2002 for raspberry and 1987-2002 for blueberry crops, only statewide averages were available after 2002. I located county-specific corn silage data from 1985-2008, but hay yields were available from only census years 2007 and 2012.

Model Parameterization

APEX Overview

I gathered data from several sources to construct a watershed that represented actual conditions in the northern Kamm Creek watershed as of 2012 (Table 1). I compiled these datasets in ArcMap 10.2 (ESRI Corporation, Redlands, California, U.S.A.) to delineate the watershed and subareas, and to characterize soils and cropping systems within subareas. APEX watersheds are constructed of discrete, interconnected plots called “subareas,” each of which has homogeneous plant cover, management regime, soil type, slope, and climate. Water, sediment, and nutrients are routed from the edge of each subarea to the next, ultimately resulting in outputs calculated at the watershed outlet. APEX is driven by a series of interconnected text files containing input data for all subareas along with background parameters that determine model output calculations. To simplify the modeling process, I built my watershed using the winAPEX interface (Steglich, 2014b), creating 36 interconnected subareas (Fig. 2). APEX offers several process models for simulating hydrology, soil erosion, plant growth, and nutrient cycling, depending on available data and project scope (Wang, Kemanian & Williams, 2011). I chose the Hargreaves method to estimate potential evapotranspiration as it is the most robust method and does not require solar radiation or wind inputs (Ford *et al.*, 2015; Francesconi *et al.*, 2014; Hargreaves & Samani, 1985). For surface runoff estimates, I chose the variable daily curve number method that varies based on daily soil moisture conditions. I adjusted the atmospheric CO_2 concentration to 400 ppm to reflect current conditions (IPCC, 2013). I used default settings for all other control file parameters not mentioned above (Evelyn Steglich, USDA-NRCS, pers. comm.).

Watershed and subarea delineation

I delineated the northern Kamm Creek watershed using the hydrology toolbox in the ArcMap 10.2 spatial analyst extension (ESRI Corporation, Redlands, California, U.S.A.; Merwade, 2012). I first constructed flow direction and flow accumulation rasters from a three meter digital elevation model of the greater Kamm Creek watershed (USDA-NRCS). I used the “watershed” tool to delineate the watershed from the flow direction raster, assigning the outlet point to node 173 from the WRIA1 study to assure that my watershed matched the calibration watershed.

I defined APEX watershed subareas primarily by spatial cropping system data obtained from the Washington Department of Agriculture (Perry Beale, pers. comm.). I combined adjacent fields of the same cropping system to simplify the watershed, and delineated residential areas and forests separately from agricultural fields. I also delineated subareas by soil type, splitting fields that contained distinct soils at the soil boundary from the SSURGO soils layer. All soil datasets contained the same spatial boundaries as the SSURGO dataset. I used aerial imagery to determine percent impervious surfaces (parking lots, roads, and homesteads). I used the GIS spatial analyst to assign minor stream paths to devise a routing scheme for my subareas by creating streams from the flow accumulation raster at a threshold contributing area of one hectare, determined by trial and error as the threshold where each subarea had at least one minor stream path present. To assign average slopes to individual subareas, I used the DEM to create one meter contour lines for my watershed and measured the average distance between contours.

Weather

I compiled daily weather data on precipitation and minimum and maximum temperature using methods similar to those employed by the TOPNET model (Bandaragoda *et al.*, 2012). Briefly, I used a gridded dataset of weather data maintained by the University of Washington Surface Hydrology Research Group (Hamlet & Lettenmaier, 2005). I located daily weather files from four stations near my site used in the WRIA1 report and generated spatially-weighted averages of daily precipitation, minimum temperature, and maximum temperature, which I converted to an APEX-ready daily weather file using the “Weather Import” program (Steglich, 2014a).

Soils

I initially tested the response of APEX to inputs from three different soil data sets (Table 1). The names and spatial extents of the soil types are identical across these data sets. Differences among soil data sets lie in how soil properties are organized by layers for APEX inputs. The NCSS data splits soil into layers by horizon (A, B, and C) and subhorizon (NCSS). NTT and SSURGO soils are split into fewer horizons, and often soil subhorizons are combined (USDA-NRCS). The NTT soils split the A horizon into two layers, with the upper layer containing the top ten cm of soil, so they have one more layer than the SSURGO data. Furthermore, all three soil datasets contain slightly different parameters for percent organic matter and soil texture (percent sand, silt, and clay). NCSS soils had an additional percent rock parameter that was not specified in the SSURGO or NTT data. I extracted APEX parameters from SSURGO and NCSS soil databases using MS Access (Dr. Nathan Nelson, Kansas State University, pers. comm., Carrie-Ann Houdeshell, USDA-NRCS pers. comm.). For NTT soils, I loaded my APEX watershed into an unreleased update of the NTT model and model developers extracted soil data from the resulting watershed in APEX format (Dr. Ali Saleh, Tarleton State University, pers. comm.). I also modified two of the soils datasets to investigate the effects of specific soil parameters on crop yield and hydrology:

1. NTT soils had very low percent carbon in the A layers, so I modified these values using percent carbon data from the SSURGO and NCSS datasets, creating the “NTT Normal C” soil data (Goldin, 1992).

2. NCSS soils contained 5-10% rock in their surface layers and over 33% rock in subtending layers, a parameter not included in the SSURGO and NTT soils. I set the percent rock parameter of the NCSS soils to 0%, to make “NCSS No Rock” soils (USDA-NRCS).

In all, I tested the APEX model using these five separate soil datasets: SSURGO, NCSS, NTT, NTT Normal C, and NCSS No Rock. For all five soils datasets, I kept all values for wilting point, field capacity, and saturated hydraulic conductivity at zero to allow the model to simulate these parameters (Dr. Lee Norfleet, USDA-ARS pers. comm.).

Crops

Bringing the APEX model to a new area required parameterizing growth of additional crops for the model. Though APEX could simulate silage corn and orchard grass fields, the existing model did not include blueberry or raspberry crops. The APEX model simulates plant growth based on biomass accumulation by photosynthesis. APEX uses heat unit accumulation as a proxy for solar energy, which is assimilated based on a specific plant Leaf Area Index (LAI). LAI is the dimensionless amount of leaf area in a field exposed to sunlight at any given time (LAI= one-sided leaf area/ ground area). APEX simulates growth over time using a sigmoidal curve that has crop-specific values of LAI increase over the course of the growing season (Steglich & Williams, 2013).

To populate these inputs, I used experimental blueberry and raspberry fields at the Washington State University Extension (WSUE)-Mount Vernon. I investigated “Duke” blueberries and “Meeker” raspberries, two common varietals grown in Whatcom County (Dr. Lisa DeVetter, WSUE pers. comm.). I first measured “Duke” blueberry LAI in September, 2014, to capture the LAI of the plant at its maximum. I made subsequent measurements in April, May, and August 2015 to capture LAI development throughout the growing season for both "Duke" blueberries and "Meeker" raspberries.

To calculate LAI, I followed a protocol specified by the Temple, TX, ARS station using an AccuPAR LP-80 ceptometer (Decagon Devices, Pullman, WA) with an additional sensor to compare photosynthetically active radiation above and below the canopy (Williams, 2014). For each LAI sampling of raspberries and blueberries, I set up a three meter transect (the row spacing between plants), perpendicular to the plant rows with the midpoint of the transect at the center of a haphazardly-chosen plant. I took seven evenly-spaced measurements with the light bar running parallel to the row of berries (Fig. S1), assuring that I sampled the LAI both below plants and between plant rows (Johnson, Kiniry & Burson, 2010). I repeated this for four plants of each berry varietal. I also destructively sampled blueberry and raspberry plants and shipped the samples to the Temple, TX, ARS station to measure biomass per plant and N and P content in leaves and shoots. The ARS also measured physical leaf area using a LICOR LI-2100 (LICOR, Inc., Lincoln, NE) leaf area meter to estimate an extinction coefficient, K, which we used in Beer's law to convert the fraction of photosynthetically active radiation absorbance by the canopy into absolute LAI (Johnson, Kiniry & Burson, 2010; Amber Williams, USDA-ARS pers. comm).

Another critical crop parameter was harvest index (HI), which is the amount of fruit biomass produced in a growing season divided by the total above ground plant biomass. Through repeated calibration with observed blueberry yield, I found an HI value of 0.30, which was consistent with reported values in the literature (published mean= 0.39, standard deviation=0.25; Bryla *et al.*, 2012; Pritts & Hancock, 1985; Strik & Buller, 2005). For raspberry, I used an HI value of 0.30 as well, close to HI values arising from a single study and which ranged from 0.27-0.32 (Bryla, unpublished).

One major challenge of modelling blueberries and raspberries was that APEX is rarely used to model growth of perennial shrubs. I used information from WSU Ag extension to adjust blueberry and raspberry-specific parameters for the simulated crop including crop height, rooting depth, and optimal and minimum temperatures for growth (Dr. Lisa DeVetter, WSUE pers. comm.). Plant growth form is an important parameter in the APEX model, but unfortunately, perennial shrub is not a growth form currently supported by the model. After calibrating the crops against county yield data, I chose to model blueberries as a “deciduous tree” and raspberries as a “perennial” to maximize model fit (Dr. Dr. James Kiniry, USDA-ARS pers. comm.). Blueberry plants are kept in agricultural fields for 15 to 20 years before being replaced, so I checked that APEX was properly modelling year-to-year development in the plants by comparing APEX biomass outputs to biomass data from both mature and two year old blueberry plants (Bryla, Gartung & Strik, 2011; Bryla *et al.*, 2012).

Agricultural Management

I was unable to obtain field-specific crop management regimes for the northern Kamm Creek watershed due to farmer confidentiality. Instead, I obtained crop-specific, county-wide estimates from experts at local agricultural agencies (Table 1). I used a nutrient management plan for a corn silage and orchard grass hay operation that was representative of northern Kamm Creek watershed practices from the Whatcom Conservation District (Chuck Timblin pers. comm.). Because both of these crops are associated with dairy cow operations, all nutrients applied to fields came as manure, and I adjusted manure N and P content according to data in the management plan. I used approximate planting and harvesting dates for silage corn and orchard grass, which is harvested multiple times per year, from the WSUE office in Whatcom County (Chris Benedict, pers. comm.). These dates vary year to year, but for simplicity, I kept them fixed for my model. Raspberry and blueberry management regimes were based on recommendations from Oregon State University Extension, WSUE Mt. Vernon, and the British Columbia Ministry of Agriculture (Hart, Strik & Rempel, 2006; Hart *et al.*, 2006; Dr. David Bryla pers. comm.; Dr. Lisa DeVetter pers. comm.; David Poon pers. comm.). Because I was unable to obtain irrigation data, I used the APEX automatic irrigation function that triggers irrigation when plants experience drought stress. I programmed blueberry and raspberry rotations to last 15 and eight years, respectively, to reflect the average field lifespan for each plant (Dr. Lisa DeVetter, WSUE Mt. Vernon pers. comm.). Because I used county-wide averages, I could not incorporate variation in field-to-field management practices. Though farmers are instructed to follow crop management guidelines as outlined by local agricultural agencies, there is no guarantee that they do. Variations in irrigation, fertilizer application, or timing of planting or harvest may affect crop growth and nutrient leaching dynamics within the watershed.

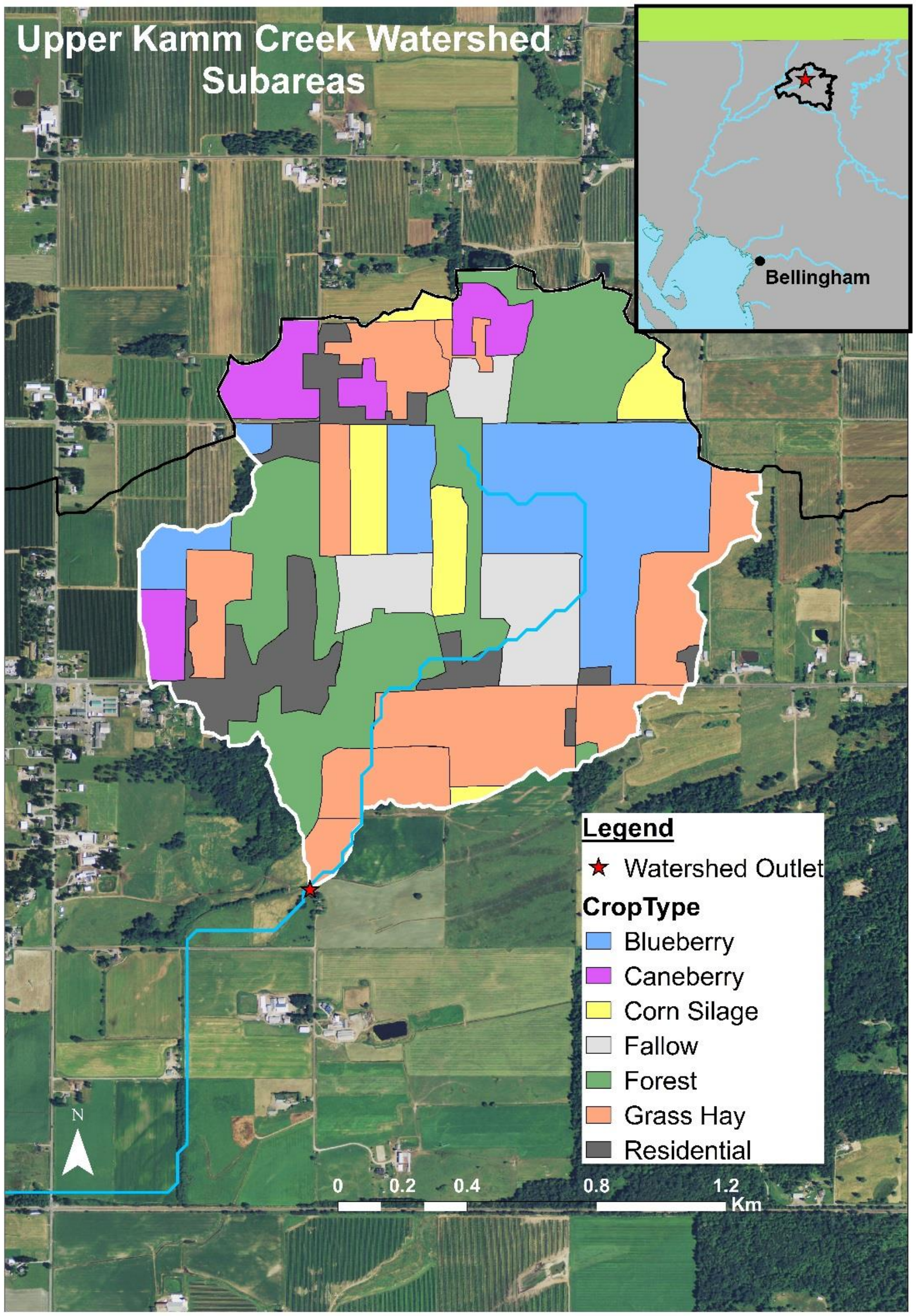

Figure 2. Subareas for the APEX model by crop type in the northern Kamm Creek watershed, Whatcom County, WA.

Calibration and Validation

To account for tree growth in forested subareas, I included a 35 year run-up period on the model, starting in 1960. I used monthly streamflow and surface runoff outputs from TOPNET from 1995-2004 to calibrate APEX and from 2005-2010 to validate APEX. I used a longer calibration period to better capture annual variability in weather before validation (Moriasi *et al.*, 2007). I calibrated hydrology in the model with the APEX-CUTE autocalibration tool developed by the USDA-ARS (Wang *et al.*, 2014). APEX-CUTE uses a dynamically-dimensioned search algorithm to fine tune background parameters to maximize model fit based on two statistical measures recommended by Moriasi *et al.* (2007):

- Nash-Sutcliffe Efficiency (NSE) evaluates model fit compared to variability of observed data, where P_i is the model-generated output, O_i is observed data, and $\bar{O}$ is the mean of observed data. NSE ranges from negative infinity to 1, with 1 being a perfect model fit.

$$\mathrm{NSE} = 1 - \left[\sum_{i=1}^{n}(P_i - O_i)^2 / \sum_{i=1}^{n}(O_i - \bar{O})^2\right]$$

- Percent bias (PBIAS) expresses deviation between mean observed ($\bar{O}$) and mean predicted ($\bar{P}$, modelled) data, where lower percentages indicate closer fit. A positive PBIAS indicates model underestimation, while a negative PBIAS indicates overestimation.

$$\mathrm{PBIAS} = \left[\frac{\bar{O} - \bar{P}}{\bar{O}}\right] * 100$$

- The two measures are combined in an objective function (OF). APEX-CUTE runs the model hundreds of times to minimize this function.

$$OF_i = [(1 - NSE)^2 + (|PBIAS_i| + 0.5)^2]^{1/2}$$

Guidelines suggest that satisfactory model fit for hydrologic processes occurs with NSE ≥ 0.5 and PBIAS ≤ ±25% (Moriasi *et al.*, 2007; Table 2). I compiled previous sensitivity analyses of streamflow and runoff in APEX to determine hydrology-related APEX parameters to adjust, default values, and the ranges of those parameters (Francesconi *et al.*, 2014; Kumar *et al.*, 2010; Wang & Jeong, 2015; Wang *et al.*, 2006b; Wang *et al.*, 2012; Table S2). For each soil type, APEX-CUTE generated 250 model runs; I chose the optimal model run for determining parameters by first looking at the OF, then comparing NSE and PBIAS values to find optimal model fit.

I calibrated annual crop yields manually on a crop by crop basis (Table 1). I was able to use NSE and PBIAS as guidelines to calibrate and validate all crops except hay, which only had data available for 2 years of NASS censuses (2002, 2007). For crop yield, a PBIAS ≤ ±25% is considered satisfactory (Wang *et al.*, 2012). Yield does not have established guidelines for NSE, but they are typically less stringent than those for hydrology (Dr. Daniel Moriasi, USDA-ARS pers. comm.). Since blueberry and raspberry yields vary by crop age, I randomized time of planting for fields, staggering them every two years for raspberry and every four years for blueberry. Since county-level yield data for berries were only available from 1985-2002, I ran a regression between state and county numbers to estimate yield data from 2002-2009 (state data were available annually to the current year). I found a significant, positive linear relationship between state and county yields for both blueberry (R^2=0.72, $p<0.001$) and raspberry (R^2=0.97, $p<0.001$). Because I estimated berry yield data from 2002-2009, I calibrated using even years and validated using odd years from 1995-2009. For silage corn, county-specific yield data were available annually up until 2008, so I calibrated from 1995-2001 and validated from 2002-2008. I calculated spatially weighted average yields by field area for each crop simulated in APEX.

Table 2. Performance ratings as recommended by Moriasi and others (2007) for streamflow on a monthly time step for APEX model evaluation.

Performance Rating	**NSE**	**PBIAS (%)**
Very Good	$0.75 < NSE \leq 1.00$	$PBIAS < \pm 10$
Good	$0.65 < NSE \leq 0.75$	$\pm 10 \leq PBIAS < \pm 15$
Satisfactory	$0.50 \leq NSE \leq 0.65$	$\pm 15 \leq PBIAS < \pm 25$
Unsatisfactory	$NSE < 0.50$	$PBIAS \geq \pm 25$

RESULTS

Hydrology

The APEX model closely predicted monthly runoff and streamflow values when compared to TOPNET data during calibration (1995-2004) for all soil types (Figs. 3-4, S2-S6). During calibration, APEX had a “very good” fit with TOPNET for monthly runoff for all soils, while streamflow had a “good” fit for SSURGO soils and a “satisfactory” fit for all other soils (Table S1; Figs. 3-4, S2-S6). For all soils, APEX under-predicted both streamflow and surface runoff during low-flow events, and over-predicted those fluxes during high flow events, illustrated by slopes significantly greater than one in all regressions (Figs. 4, S6). These trends were driven by seasonality, with high-flow events occurring during winter months and low-flow events occurring during the summer (Figs. 3, S2-S5). The APEX-simulated hydrologic data indicated that runoff was the primary source of streamflow for the watershed, generating 50% of annual flow volume for all soils (Table S1); this proportion was higher in the winter months than in the summer. The TOPNET data had a 20-30% higher portion of streamflow generated through runoff than APEX (Table S1). SSURGO soils performed slightly better than NCSS and NTT soils for hydrology calibration, but overall, the differences were small.

During the validation period (2005-2009), seasonal trends between APEX and TOPNET monthly runoff and streamflow data were generally similar to those in the calibration. For all soils in the validation period, APEX captured seasonal variability in flows and average measurements of streamflow and runoff were closer (lower PBIAS) to TOPNET averages than in calibration (Figs. 3-5, S2-S5, S7). On the other hand, model fit for monthly trends in both streamflow and runoff were lower in validation than during the calibration period (Figs 4-5, S6-S7). All soils had satisfactory to good fit for runoff (NSE>0.5), but not streamflow. This occurred because APEX had a stronger over-prediction at high flows and under-prediction at low flows in the validation than calibration periods (i.e., slopes much greater than 1, Figs. 4-5). During validation, SSURGO soils showed the best fit with monthly variation in TOPNET hydrologic data, with a "good" NSE for runoff and an NSE for streamflow just below the "satisfactory" threshold (Table S3; Fig. 5). For all other soils during validation, fit with monthly TOPNET streamflow and runoff data was lower than SSURGO results (Figs. 5, S2-S5, S7). Interestingly, the values of parameters that optimized model fit during autocalibration with APEX-CUTE were identical for all APEX soil types (Table S2).

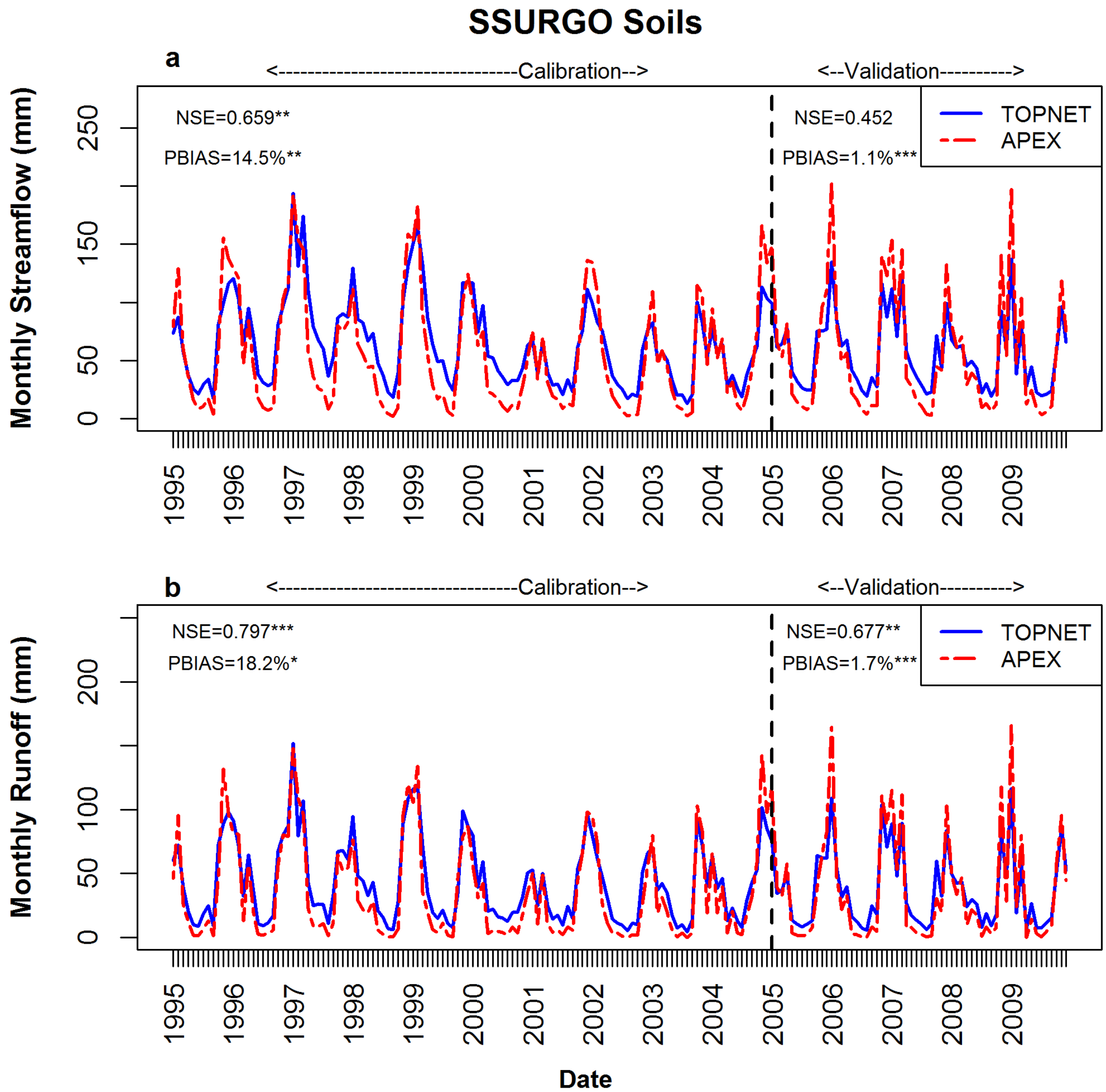

Figure 3. Predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff (a) and streamflow (b) during the calibration (1995-2004) and validation (2005-2010) periods for SSURGO soils. *Indicates “satisfactory”, ** indicates “good”, and *** indicates “very good” fits as determined by NSE and PBIAS (Table 2, Moriasi et al. 2007).

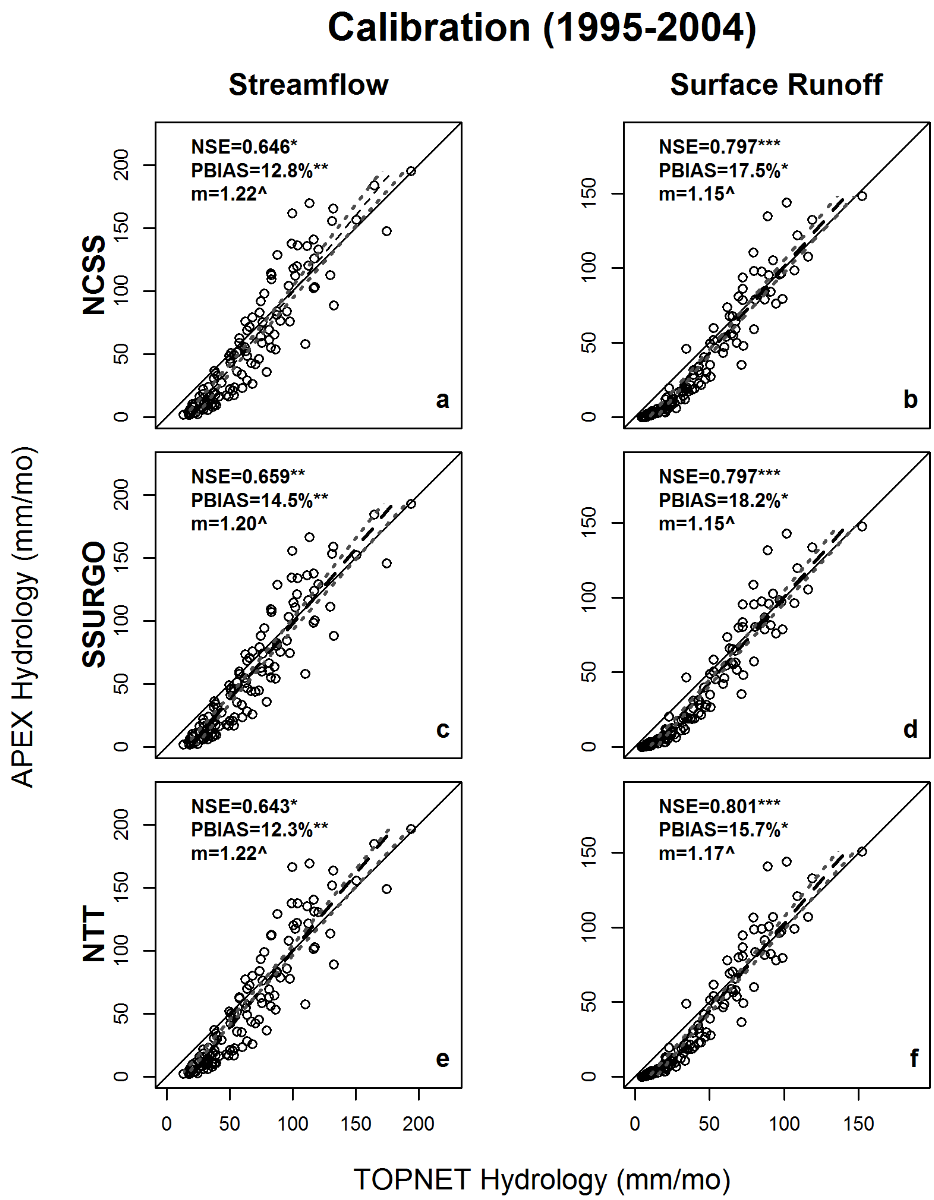

Figure 4. Linear regression of predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff and streamflow during the calibration period (1995-2004) for NCSS (a,b), SSURGO (c, d), and NTT (e, f) soils. The black line indicates a 1:1 line, the black dashed line indicates a linear regression line through the simulated points, and the gray dotted lines indicate 95% confidence intervals. *Indicates a "satisfactory", ** indicates "good", and *** indicates "very good" fit as determined by NSE and PBIAS (Moriasi et al. 2007). ^Indicates a significant linear regression ($p<0.05$). # Indicates a significant regression with a slope (m) not statistically different from l ($p>0.05$; Table S3).

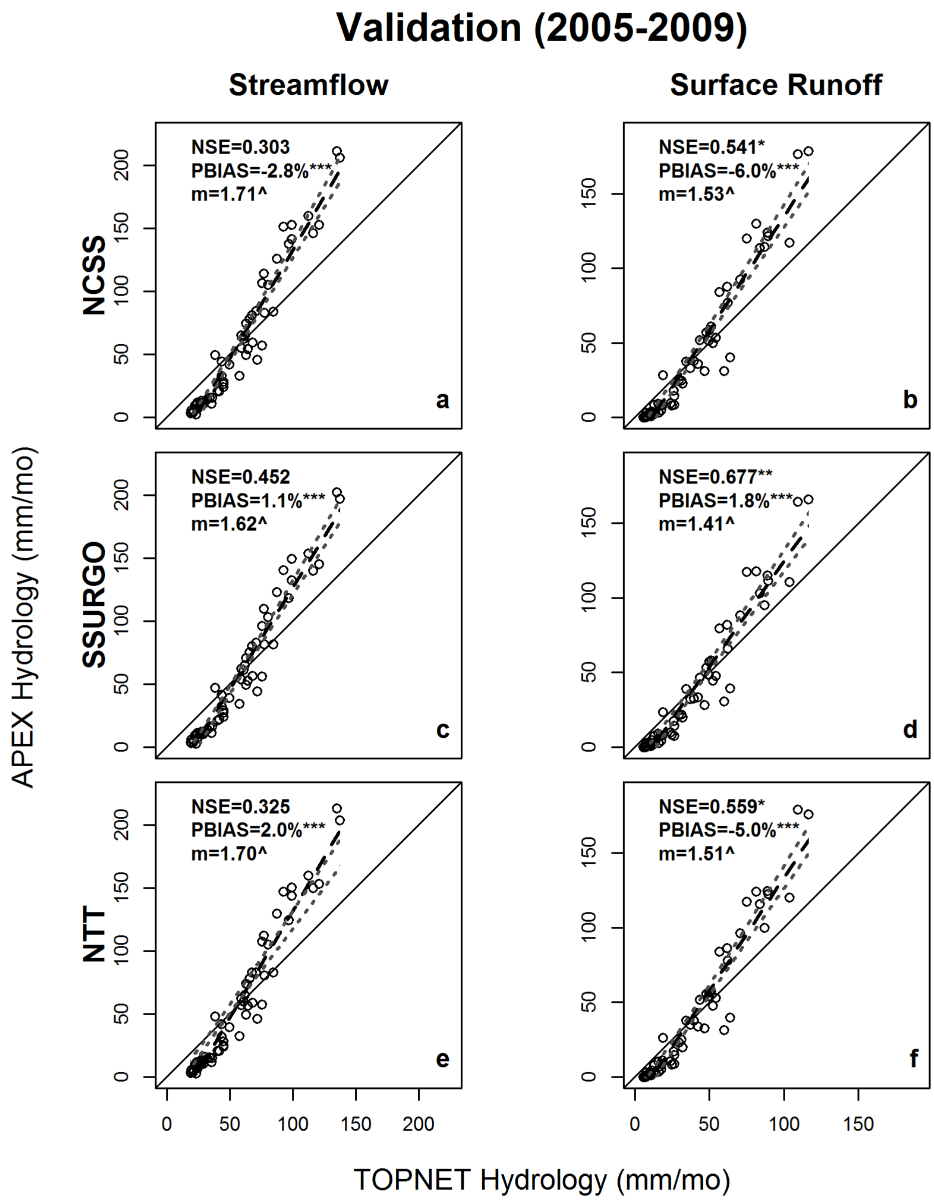

Figure 5. Linear regressions of predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff and streamflow for the North Kamm Creek Watershed during the validation period (2005-2009) for NCSS (a,b), SSURGO (b,c), and NTT (d, e) soils. Lines and symbols as in Fig. 4.

Crop Yield

APEX's ability to predict crop yields varied greatly among crop and soil types, with different results in the calibration and validation periods. NCSS and SSURGO soils showed similar trends in calibration and validation for the three primary crops in this watershed (Figs. 6-7, S8-S9). During calibration, raspberry and blueberry crops in NCSS and SSURGO soils had average yields close to reported values (PBIAS< ±10%). During calibration, APEX-predicted average raspberry yields for NCSS and SSURGO soils were within ~5% of observed values. For these soils, raspberry yields closely followed annual trends for NCSS and SSURGO soils, with regression slopes not different from one (Figs. 6,7, S8-S9). SSURGO soils had the closest fit with annual trends for observed raspberry yields. In contrast, NTT soils during calibration dramatically under-predicted raspberry yields and failed to capture annual variation (Fig. 6).

Blueberry yields during calibration were very close, on average, to observed values for NCSS, SSURGO soils (PBIAS < ±1%; Figs. 6, S8). Blueberry yields satisfactorily followed annual trends for NCSS and SSURGO soils (NSE>0.4), though not as closely as in raspberry, as slopes were significantly less than one (Figs. 6, S8). Similar to raspberry, NTT soils dramatically under-predicted yields and failed to capture annual variation (Fig. 6).

During the calibration period, APEX fit of silage corn yields differed from results of the berry crops (Figs. 6, S8). For NCSS soils, APEX dramatically under-predicted yields and failed to capture yearly variability (Fig. 6, S8). Although the slope of the regression for NCSS soils is close to one, the regression is not significant due to low interannual variation in observed yield data and a large difference between observed and simulated yields (Table S3; Fig. 6). On the other hand, yields for corn silage during calibration were well within the range of observed yields for both SSURGO and NTT soils, with PBIAS values all below

10% (Figs. 6, S8). Actual yields of silage corn varied less than 7% across all years, so NSE values and regression fits were correspondingly low for all soils.

Because initial crop calibration was poor for NTT soils in both berry crops and NCSS soils for silage corn, I investigated which soil parameters caused anomalous results. NTT soils had much smaller values for percent organic carbon, often close to 0% in the top two layers of soil for Kickerville silt loam and Whatcom silt loam, the two most common soils in this watershed. The NTT Normal C soil had much better calibration results for berry yields, which mirrored NCSS and SSURGO soil results (Fig. S8). Compared to observed values, average raspberry yields and annual trends improved dramatically from NTT to NTT Normal C in terms of PBIAS, NSE, and slope (Figs. 6, S8). Similarly, estimates of blueberry crop yields improved dramatically from NTT to NTT Normal C soils, though slope for NTT Normal C was still significantly lower than one (Figs. 6, S8). On the other hand, silage corn improved only slightly in annual fit from NTT to NTT Normal C soils, though average yields did increase (Figs. 6, S8). NCSS soils were the only ones that contained data on percent rock content (this parameter was left at zero for all other soils). These values were not trivial, with percent rock of 9 and 10% in the surface soil layers of Kickerville silt loam and Whatcom silt loam, respectively. Below the top two soil layers, both soils had percent rock of up to 33%. Rock-containing soils in the NCSS database make up 70% of the simulated watershed area. When I eliminated percent rock data from NCSS soils ("NCSS No Rock"), silage corn yield results closely matched other soils and average values clustered around the mean measured values (Figs. 6, S8). Compared to NCSS soils, NCSS No Rock blueberry crops saw no change in average yield nor interannual fit, while raspberry crops saw a slight improvement in both fit and average yield (Figs. 6, S8).

For all soils, fit for yield of all crops worsened in the validation period compared to the calibration period. In contrast to calibration results, for NCSS, SSURGO, and NTT Normal C soils, APEX failed to capture year-to-year variation in blueberry and raspberry yields during validation (NSE<0, $m \neq 1$; Figs. 7, S9). Average yields, compared to observed data, increased slightly for raspberry crops in validation compared to calibration, indicating a minor increase in model over-prediction for all soils. Additionally, model fit of both raspberry and blueberry yields decreased compared to observed data for all soil types (Figs. 6-7, S8-9). This pattern persisted for silage corn, which saw a decrease in model fit along with only slight changes in average yields compared to observed data from calibration to validation (Figs. 6-7, S8-S9).

I could not calculate NSE and PBIAS statistics for orchard grass hay yields since only two years of county-level data existed for the calibration and validation period. APEX-predicted yields were lower than county averages for all three soils, but within ~20% (Table 3). As with the actual county data, APEX estimated higher yields for 2002 than 2007 for all five soil data sets.

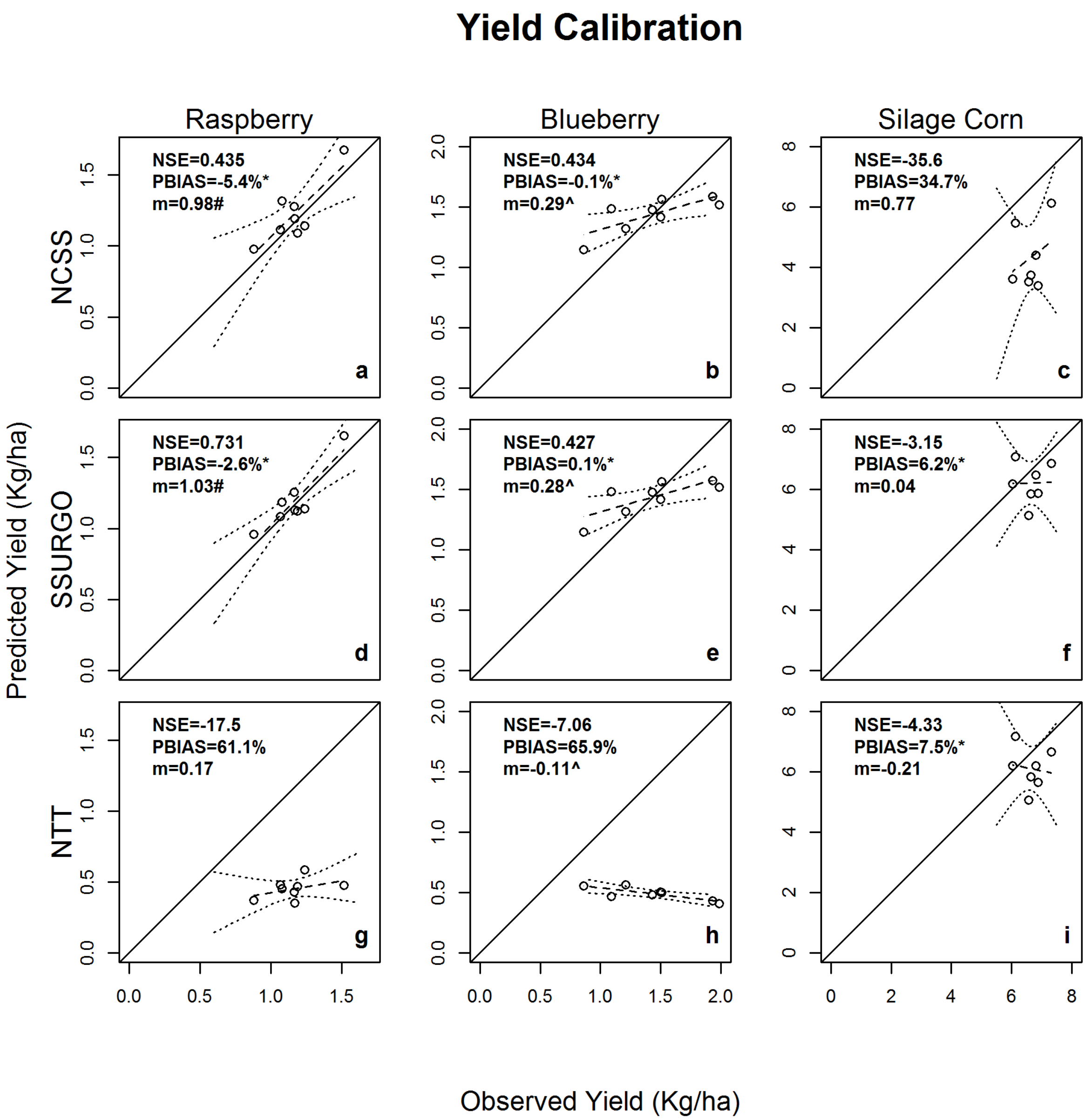

Figure 6. Linear regressions of predicted (APEX generated) versus observed yields for three major crops in the North Kamm Creek Watershed during the calibration period for NCSS (a,b,c), SSURGO (d,e,f), and NTT (g,h,i) soils. The black line indicates a 1:1 line, the black dashed line indicates a linear regression line through the simulated points, and the black dotted lines indicate 95% confidence intervals. ^Indicates a significant linear regression (p<0.05). # Indicates a significant regression with a slope (m) not statistically different from 1 (p>0.05). *Indicates a "satisfactory" fit for PBIAS as determined by Wang and others (2012).

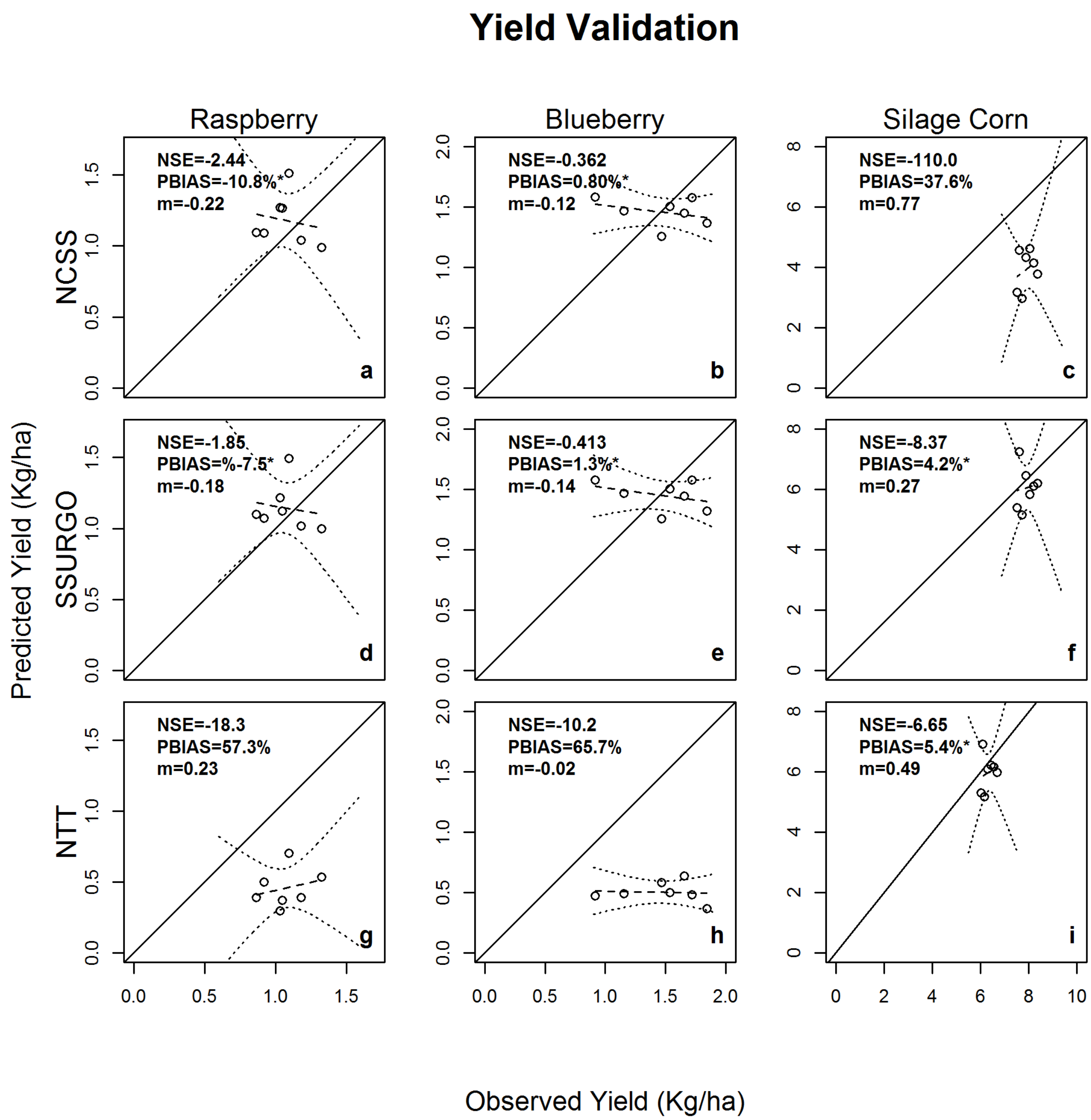

Figure 7. Linear regression plots of predicted (APEX generated) vs observed yields for three major crops in the north Kamm Creek watershed during the validation period for NCSS (a,b,c), SSURGO (d,e,f), and NTT (g,h,i) soils. Lines and symbols as in Fig. 6.

Table 3. Orchard grass hay yields (dry kg/ha) compared to county averages for each soil run. Numbers in parentheses are percent differences from Whatcom County reported yields (USDA-NASS, 2014).

Year	Whatcom County	NCSS	SSURGO	NTT	NTT Normal C	NCSS No Rock
2002	11.9	9.6 (19%)	9.8 (18%)	9.5 (20%)	9.8 (18%)	9.6 (19%)
2007	13.0	10.6 (18%)	10.9 (16%)	10.3 (21%)	10.8 (17%)	10.6 (18%)

DISCUSSION

Overview

Several studies have used APEX to model BMP performance in single fields (Plotkin *et al.*, 2013; Wang *et al.*, 2008; Williams *et al.*, 2006; Yin *et al.*, 2009; Mudgal *et al.*, 2010), and large watersheds, ranging up to the regional scale for the national CEAP report (Santhi *et al.*, 2014; Tuppad *et al.*, 2010; Wang *et al.*, 2014; Williams *et al.*, 2010), but few exist at the sub-watershed scale, which is critical to assess the conservation value of BMPs (Strauss *et al.*, 2007). For example, successful nutrient trading requires involvement of multiple stakeholders within the same watershed, allowing in-kind payments from polluters to landowners that establish BMPs to reduce overall pollutant export from that watershed (Greenhalgh & Selman, 2012). Thus, predicting benefits of various BMPs with a properly validated APEX model on the sub-watershed scale shows promise for nutrient credit calculation, which is an overarching goal of the national CEAP program (Santhi *et al.*, 2014; USDA, 2014; Williams *et al.*, 2010). The results of this study will inform future use of the APEX model in Whatcom County, the Pacific Northwest, and nationally, to simulate sediment and nutrient dynamics. Four general findings emerged from this study. First, soil played a more important role in model fit for crop yield than for hydrology. Second, validation of two perennial shrubs, blueberry and raspberry was limited by the APEX's ability to simulate shrub growth and development. Third, crop yield is sensitive to two critical soil parameters: percent organic carbon, and soil rock content. Finally, this study makes recommendations on data requirements and quality for future parameterization, calibration, and validation of the APEX model.

Hydrology

Overall, all soil data sets performed similarly in simulating hydrology in APEX when compared to TOPNET data. Statistically, all soil scenarios calibrated and validated surface runoff and calibrated streamflow satisfactorily to very well, depending on soil data set. All soils exaggerated seasonal variation in streamflow during validation due to overprediction of high flows and underprediction of low flows, but average predicted streamflow values were reasonably close to TOPNET values (Figs. 5, S7). The tendency of APEX to overestimate and underestimate streamflow and runoff in extreme wet and dry events is documented in the literature and may also pose a problem when applying the model to nutrient and sediment data (Ford *et al.*, 2015; Kumar *et al.*, 2010; Plotkin *et al.*, 2013; Wang *et al.*, 2014). Runoff was a major contributor to streamflow in my watershed, accounting for around half of streamflow annually (Table S1). Predominance of runoff was most pronounced during winter, when Whatcom County receives most of its precipitation. Runoff is a major contributor to soil erosion and nutrient loss in agricultural watersheds, so accurately simulating it is an important first step to future modelling efforts focused on sediment and nutrient simulations (Brooks *et al.*, 2015). For all soils in both calibration and validation, however, APEX consistently attributed less streamflow to runoff (as a percentage) than did the TOPNET model (Table S1). Overestimates in runoff would cause overestimation of erosion and nutrient loss in winter when fields are bare and susceptible to erosion, while underestimates may fail to generate erosion and nutrient loss in summer, particularly in the shoulder months (April-June, September; Wang, Kemanian & Williams, 2011).

After autocalibration, all soil scenarios had the same set of optimal parameters, but SSURGO soils showed the best model fit to TOPNET data, especially during validation, where SSURGO had a good fit with monthly trends in runoff (Table S2). Any slight differences in model fit among soil datasets for hydrology were likely from differences in soil bulk density, texture, and rock content, as these differed throughout data sources. For example, percent silt content and bulk density of the top layer of Kickerville silt loam ranged from 55 to 68% and 1.0 to 1.15 Mg/m^3 for SSURGO and NCSS soils, respectively. Differences among soils in layer number and depth of layers were probably less important, because the APEX model automatically splits soil layers up to more accurately model water percolation and nutrient availability to plant roots (Williams, Izaurralde & Steglich, 2012).

I found differences between model fit in the calibration and validation periods for all five soil types. APEX predictions for watershed hydrology showed better fit during calibration than validation (Figs. 4-5, S6- S7). This result was not unexpected, and was found in other APEX modelling studies (Gitau, Veith & Gburek, 2004; Wang, Kemanian & Williams, 2011; Wang *et al.*, 2012; Yin *et al.*, 2009). In a study using the APEX-CUTE autocalibration tool to calibrate and validate streamflow in a large watershed in Inner Mongolia, China, researchers attributed decreased model fit during validation to climactic differences between the calibration and validation period (Wang *et al.*, 2014). Specifically, the authors found that precipitation during the validation period was much lower than in calibration, limiting the model's ability to simulate low-flow events (Wang *et al.*, 2014). In my study, average annual precipitation was 10% lower during the validation period than in calibration, which may have contributed to decreased model fit (Moriasi *et al.*, 2007). In particular, average July and August precipitation were 68% and 21% lower, respectively, during validation than during calibration. (Figs. 6, S1-S4).

APEX model fit was poorer for streamflow than runoff for all soils in both calibration and validation (Figs. 4-5, S6-S7). Finding weaker fits for streamflow than runoff was not surprising, as streamflow is a complex process that involves soil storage of water, infiltration to groundwater, lateral subsurface flow, and return flow to streams from groundwater (Gassman *et al.*, 2010; Williams, Izaurralde & Steglich, 2012). In a study calibrating APEX simultaneously for surface runoff and subsurface lateral flow investigating CEAP BMP application, Plotkin and others (2013) found very good fit for surface runoff (NSE=0.80), but poor fit for subsurface lateral flow (NSE=0.16). The similarly weaker fit in streamflow may also indicate that APEX parameters governing soil moisture and groundwater storage did not mirror conditions in the TOPNET model.

Lack of availability of directly-measured hydrologic data in the northern Kamm Creek watershed was a source of uncertainty for this project. While the TOPNET model allowed for continuous monthly calibration and validation of both runoff and streamflow in the North Kamm Creek watershed, these data are also modeled estimates. TOPNET streamflow was calibrated and validated against real-world streamflow data at several USGS and WA DoE gauging stations in Whatcom County, including some on the mainstem of the Nooksack River, to which Kamm Creek is a tributary. None of those gauges, however, lay on Kamm Creek itself (Tarboton, 2007). Furthermore, surface runoff data were not calibrated in TOPNET (Tarboton, 2007). Both gauges on the Nooksack River downstream of Kamm Creek were satisfactorily calibrated for streamflow (with less than a 5% deviation from observed data), however, these gauges served a much larger catchment area than that of the northern Kamm Creek watershed. Tarbaton (2007) also calibrated TOPNET at several gauging sites along creeks serving much smaller watersheds, including Fishtrap Creek, a watershed with similar land use to Kamm. TOPNET modeled flow was consistently 1.5 times higher than measured flow at a gauging station on Fishtrap Creek near Lynden, WA. Such an overestimate of TOPNET data in Kamm Creek would indicate that overestimates by the APEX model during high flow events were even more severe when compared to measured flows. Unlike APEX, TOPNET data did not appear to underestimate flow during low-precipitation events on Fishtrap Creek. Future research will evaluate the accuracy of APEX-simulated hydrology in the northern Kamm Creek watershed by comparing APEX outputs against measured flows.

Using appropriate soils data is essential to modelling with APEX on any scale. APEX modelling efforts on the single-field scale often include collecting soil samples and calculating physical soil properties in a lab (Kumar *et al.*, 2010; Plotkin *et al.*, 2013). However, in large watersheds with several different soil types, this is impractical. Most APEX modelling studies recommend using data from publicly available sources (Ford *et al.*, 2015; Wang, Kemanian & Williams, 2011; Francesconi *et al.*, 2014). In using APEX to model several contiguous ~10 ha watersheds to evaluate effectiveness of riparian buffers, Qiu and others (2002) obtained soils data from the SSURGO database. Another more recent study used SSURGO soils data in an APEX modelling effort to evaluate nutrient and sediment retention by various BMPs in several watersheds ranging from 3,426 to 14,082 ha (Tuppad *et al.*, 2010). SSURGO data had the best fits for hydrology and crop growth in this study, were publically available, and are supported nationally, making this data set a good choice for APEX projects moving forward. Locally, land managers in Whatcom County mentioned SSURGO as a preferred source of soil data (Heather MacKay, FHB Consulting, pers. comm.).

Crops

In contrast to hydrology, soil type was a major factor in simulating crop growth. Causes of this variation depended on the soil data source and the crop of interest. For both berry crops, yields were close to expected average yields during the calibration and validation periods for all soils except NTT, however, yearly variation in berry yields was not captured during the validation period (Figs. 6,7). For silage corn, yield fit also differed among soil types during calibration and validation. For berries, NTT soils lacked sufficient organic carbon content for proper berry growth. Organic carbon percentage determines N and P availability to crops in APEX (Williams, Izaurralde & Steglich, 2012), and values as low as 0% (in the top two soil layers of Whatcom and Kickerville silt loam) caused both berry crops to exhibit stunted growth due to N and P stress. Organic carbon percentage values of SSURGO and NCSS are more consistent with published data for Whatcom County (Goldin, 1992). For silage corn, yield fit also differed among soil types during calibration and validation. While average corn silage yields were within 10% of county-wide yield data, simulated yields did not reflect year-to-year variation in observed yield (NSE<0; Figs. 6,7). Both measured and modelled silage corn yields showed little year-to-year variation for all soils. For modeling, this pattern may result from having fixed annual planting, harvesting, and fertilizer timing. Realistically, this management would vary considerably from year to year and from field to field (Chuck Timblin, WCD pers. comm.). NCSS soils with percent rock data had particularly low yields (Figs. 6,7). Interestingly, this percent rock data did not affect blueberry, raspberry, or orchard grass growth, likely because these perennial crops have higher root strength than does silage corn, an annual with growth limited by soil density (Cresswell & Kirkegaard, 1995; Laboski *et al.*, 1998). In a sensitivity analysis of the APEX model for the national CEAP report, modelled corn, soybean, and winter wheat yield were especially sensitive to the "root growth soil strength" parameter that can cause crop growth stress in compact soils (Wang *et al.*, 2006b), similar to the pattern with saw with high rock content. Alternatively, soils with high rock content in the northern Kamm Creek watershed may benefit crop yields, as soils such as Kickerville silt loam and Whatcom silt loam have a gravel layer subtending their loamy surface, allowing the soils to drain well while retaining moisture effectively in surface layers (Chuck Timblin, WCD pers. comm.).

This study was the first to calibrate and validate yields for blueberry and raspberry crops, so modelling these unique crops presented some challenges. Unfortunately, no literature calibrating yields from fruit bearing shrubs with APEX exists. My inability to predict the full range of variation in yields for blueberry and raspberry crops likely stemmed from two main issues (Fig. 7). First, to simulate crop growth, APEX does not incorporate plant storage: the LAI development curve assumes all biomass accumulation (including fruit production) from a certain year comes entirely from photosynthesis in leaves from that same year (Williams, Izaurralde & Steglich, 2012). In both blueberry and raspberries, however, buds develop during the growing season before fruiting, indicating that photosynthesis in the previous year is storing energy to allow next year's fruit development (Hart, Strik & Rempel, 2006; Pritts & Hancock, 1985). The inability of APEX to model this type of fruit development likely affected my ability to capture yearly variation in yields for both blueberry and raspberry. Second, like with silage corn, my berry management regime assumed identical timing of planting, pruning, harvesting, and fertilizing each year for both blueberry and raspberry. Field and year specific yield calibration and management data would likely improve year to year model fit for berry yield.

Literature documenting APEX model fit for crop yield is sparse and generally focuses only on major crops such as corn (non-silage), soybeans, and wheat (Gassman *et al.*, 2010). Statistical analyses in these studies are limited to basic comparisons of percent error and PBIAS, often finding yields within 10% of observed values (Harman, Wang & Williams, 2004; Wang *et al.*, 2006a; Wang *et al.*, 2008). A recent study applying APEX to two small (~2 ha) watersheds in the Midwest found good fit for corn yield, but overestimated soybean yield by 60-70%; less accurate than my model fit with SSURGO soils for all crops (Francesconi *et al.*, 2014). One paper using the SWAT model, a similar model to APEX that is applied to whole watersheds, did calculate NSE as a goodness-of-fit statistic for corn, soybean, and winter wheat, and found model fit for all three crops (NSE: 0.5-0.7) comparable to my yield calibration results for blueberry and raspberry with all soils except NTT (Nair *et al.*, 2011). Overall, calibration of silage corn along with two new crops in this APEX watershed was well in line with reported ranges in the literature, but fit for validation was worse.

For all crops, additional issues affecting yield calibration and validation arose. First, yield statistics were based on annual data, thus there were fewer points contributing to NSE and PBIAS statistics than with monthly hydrology. Because of this, a single year where predicted yield differs highly from observed yield has the ability to skew goodness of fit statistics in either calibration or validation. Furthermore, NSE delivers higher scores in scenarios where observed values vary widely, so crops that have little variability in observed yields over the calibration and validation periods would deliver low NSE scores unless modelled values were very close to observed values (Krause, Boyle & Bäse, 2005). Since NSE measures the squared differences between observed and predicted values, the statistic will exaggerate differences between crops with higher yields (silage corn) compared to crops with lower yields if variance is proportional to the mean (blueberry and raspberry; Krause, Boyle & Bäse, 2005). Additionally, for both blueberry and raspberry crops, half of the observed calibration and validation points were estimated based on state to county regressions. Any instance where state yields are decoupled with Whatcom County yields could skew my calibration data and affect results. This was especially possible for blueberries, which had a weaker state-to-county correlation than did raspberry. Much of the state raspberry crop, on the other hand, comes from Whatcom County. More detailed county or farm-scale yield data would improve this fit analysis. Finally, all yield calibration and validation data were taken from county-wide averages, which may not be representative of crop yields in the northern Kamm Creek watershed. This watershed contains some soils that are poorly drained and less suitable for corn silage growth, which could have driven modeled yields below county averages (Chuck Timblin, WCD pers. comm.). Future success in calibrating crop yields with APEX would benefit from a long time period of accurate, site-specific management and crop yield data.

Recommendations

The results of this study informed specific recommendations on future use of APEX in Whatcom County and elsewhere. First, I recommend using SSURGO soil data, as data were easily accessible and produced the closest model fit for both hydrology and crop yield. When examining soil data, modelers should pay special attention to percent carbon and percent rock content, as these soil parameters were critical to crop growth and yield. Second, to improve APEX simulation of perennial shrubs like blueberry and raspberry, the model must account for energy storage from the previous year to determine the next year's yield.

Improving how APEX simulates perennial shrubs also has implications for simulating forested riparian buffers, as these consist mainly of trees and shrubs.

More generally, all future APEX modelling projects should prioritize data quality for parameterization, calibration, and validation. This project by necessity relied on TOPNET-simulated calibration and validation data, so assumptions on model closeness of fit were not based on direct measurements. I recommend that future APEX modelling efforts use actual measured streamflow (and runoff, if possible) on the daily or monthly timestep. Furthermore, because I calibrated watershed hydrology on a monthly time scale, the model should not be used to infer daily or weekly trends in runoff or streamflow (Sudheer *et al.*, 2007). Also, to improve water balance simulation, I recommend measuring water table depth and/or potential evapotranspiration, which both can be calibrated in the APEX model (Cavero *et al.*, 2012; Wang *et al.*, 2014; Williams *et al.*, 2013). Calibrating and validating the APEX model for crop yield is essential, as crop evapotranspiration and plant interception of moisture are critical elements of watershed hydrology (Nair *et al.*, 2011). To parameterize APEX with field-specific management data and calibrate against field specific yield data, I recommend working with farmers through conservation district liaisons. I inferred agricultural management data from state and county recommendations, which may not accurately reflect management practices in my watershed. While I expect farm management and crop yield to vary from field to field and year to year, my simulated management regimes were constant through time for each crop type. Time of planting, fertilizer types, amounts, irrigation, and soil tillage are all aspects of farm management that will vary. These aspects have an important effect on watershed hydrology and crop yields. Moving forward, they are even more critical when examining soil erosion and nutrient loss (Moriasi *et al.*, 2014; Wang, Kemanian & Williams, 2011). Determining the farm management aspects to which APEX outputs for hydrology, crop yield, and nutrients are most sensitive is critical for future model applications.

To use APEX results to support farm management and conservation decisions, all stakeholders involved must have confidence in the model. APEX simulation accuracy depends on a partnership between scientists and farmers to accurately simulate complex farm management processes (Moriasi *et al.*, 2014). After watershed hydrology, sediment dynamics, and nutrient cycling are satisfactorily validated, managers can use the model to simulate potential effectiveness of various BMPs. Use of a properly validated APEX watershed can lead to accurate accounting of nutrient savings from BMPs, allowing land managers to cost-effectively prioritize BMP type and placement to reduce nonpoint source pollution.

REFERENCES

Bandaragoda C., Greenberg J., Dumas M. & Gill P. (2012) Lower Nooksack Water Budget Whatcom County, WA: WRIA1 Joint Board, Bellingham, WA.

Brooks E.S., Saia S.M., Boll J., Wetzel L., Easton Z.M. & Steenhuis T.S. (2015) Assessing BMP Effectiveness and Guiding BMP Planning Using Process-Based Modeling. *Journal of the American Water Resources Association,* **51,** 343-358.

Bryla D.R., Gartung J.L. & Strik B.C. (2011) Evaluation of irrigation methods for highbush blueberry—I. Growth and water requirements of young plants. *HortScience*, **46,** 95-101.

Bryla D.R., Strik B.C., Bañados M.P. & Righetti T.L. (2012) Response of highbush blueberry to nitrogen fertilizer during field establishment—II. plant nutrient requirements in relation to nitrogen fertilizer supply. *HortScience,* **47,** 917-926.

Carey B. & Cummings R. (2012) Sumas-Blaine Aquifer Nitrate Contamination Summary (Ed WA DoE), Olympia, WA.

Cavero J., Barros R., Sellam F., Topcu S., Isidoro D., Hartani T., Lounis A., Ibrikci H., Cetin M., Williams J.R. & Aragüés R. (2012) APEX simulation of best irrigation and N management strategies for off-site N pollution control in three Mediterranean irrigated watersheds. *Agricultural Water Management,* **103,** 88-99.

Chen X., Cui Z., Fan M., Vitousek P., Zhao M., Ma W., Wang Z., Zhang W., Yan X. & Yang J. (2014) Producing more grain with lower environmental costs. *Nature,* **514,** 486-489.

Chesapeake Bay Commission. (2012) *Nutrient credit trading for the Chesapeake Bay: An economic study,* Chesapeake Bay Commission, Research Triangle Park, NC.

Cresswell H. & Kirkegaard J. (1995) Subsoil amelioration by plant-roots-the process and the evidence. *Soil Research,* **33,** 221-239.

Cui Z., Vitousek P.M., Zhang F. & Chen X. (2016) Strengthening Agronomy Research for Food Security and Environmental Quality. *Environmental Science & Technology,* **50,** 1639-1641.

Diaz R.J. & Rosenberg R. (2008) Spreading dead zones and consequences for marine ecosystems. *Science,* **321,** 926-929.

Dosskey M., Hoagland K. & Brandle J. (2007) Change in filter strip performance over ten years. *Journal of Soil and Water Conservation,* **62,** 21-32.

Electric Power Research Institute. (2014) What is Water Quality Trading? In: *Ohio River Basin Water Quality Project*, Palo Alto, CA.

Fennessy M. & Cronk J. (1997) The effectiveness and restoration potential of riparian ecotones for the management of nonpoint source pollution, particularly nitrate. *Critical Reviews in Environmental Science and Technology,* **27,** 285-317.

Ford W., King K., Williams M., Williams J. & Fausey N. (2015) Sensitivity analysis of the Agricultural Policy/Environmental eXtender (APEX) for phosphorus loads in tile-drained landscapes. *J Environ Qual,* **44,** 1099-1110.

Francesconi W., Smith D.R., Flanagan D.C., Huang C.-H. & Wang X. (2015) Modeling conservation practices in APEX: from the field to the watershed. *Journal of Great Lakes Research,* **41,** 760-769.

Francesconi W., Smith D.R., Heathman G.C., Wang X. & Williams C.O. (2014) Monitoring and APEX Modeling of No-Till and Reduced-Till in Tile-Drained Agricultural Landscapes for Water Quality. *Transactions of the ASABE,* **57,** 777-789.

Gassman P., Williams J., Wang X., Saleh A., Osei E., Hauck L., Izaurralde R. & Flowers J. (2010) Invited Review Article: The agricultural policy/environmental eXtender (APEX) model: an emerging tool for landscape and watershed environmental analyses. *Transactions of the ASABE,* **53,** 711-740.

Gitau M., Veith T.L. & Gburek W. (2004) Farm–Level Optimization of BMP Placement for Cost–Effective Pollution Reduction. *Transactions of the ASAE,* **47,** 1923.

Goldin A. (1992) Soil Survey of Whatcom County Area, Washington. (Ed Soil Conservation Service-United States Department of Agriculture).

Greenhalgh S. & Selman M. (2012) Comparing Water Quality Trading Programs: What Lessons Are There To Learn? *Journal of Regional Analysis & Policy,* **42,** 104.

Hamlet A.F. & Lettenmaier D.P. (2005) Production of Temporally Consistent Gridded Precipitation and Temperature Fields for the Continental United States*. *Journal of Hydrometeorology,* **6,** 330-336.

Hargreaves G.H. & Samani Z.A. (1985) Reference crop evapotranspiration from temperature. *Applied engineering in agriculture,* **1,** 96-99.

Harman W.L., Wang E. & Williams J. (2004) Reducing atrazine losses. *J Environ Qual,* **33,** 7-12.

Hart J., Strik B. & Rempel H. (2006) Caneberries. Nutrient management guide. Extension Service, Oregon State University, Corvallis, OR.

Hart J.M., Strik B., White L. & Yang W. (2006) Nutrient management for blueberries in Oregon. Extension Service, Oregon State University, Corvallis, OR.

IPCC. (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (Eds T.F. Stocker & D. Qin & G.-K. Plattne & M. Tignor & S.K. Allen & J. Boschung & A. Nauels & Y. Xia & V. Bex & P.M. Midgley), Cambridge, United Kingdom and New York, NY, USA.

Johnson M.-V.V., Kiniry J.R. & Burson B.L. (2010) Ceptometer deployment method affects measurement of fraction of intercepted photosynthetically active radiation. *Agronomy journal,* **102,** 1132-1137.

Krause P., Boyle D. & Bäse F. (2005) Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences,* **5,** 89-97.

Kumar S., Udawatta R.P., Anderson S.H. & Mudgal A. (2010) APEX model simulation of runoff and sediment losses for grazed pasture watersheds with agroforestry buffers. *Agroforestry Systems,* **83,** 51-62.

Laboski C., Dowdy R., Allmaras R. & Lamb J. (1998) Soil strength and water content influences on corn root distribution in a sandy soil. *Plant and Soil,* **203,** 239-247.

Mackay H. (2013) Whatcom County Ag-Watershed Project: Program Strategy Report (Rev 1) Technical Memorandum #2. (Ed H. Mackay), Lynden, WA.

Maringanti C., Chaubey I. & Popp J. (2009) Development of a multiobjective optimization tool for the selection and placement of best management practices for nonpoint source pollution control. *Water Resources Research,* **45**.

Matthews R.A. & Vandersypen J.P. (1998) Kamm Creek Watershed Monitoring Project: Final report 1993-1998. Institute for Watershed Studies, Huxley College of Environmental Studies, Western Washington University.

Merwade V. (2012) Stream Network Delineation using Spatial Analyst Hydrology Tools Vol. Oct/14/2014. https://web.ics.purdue.edu/~vmerwade/education/hydrology.pdf.

Mitsch W.J., Day Jr J.W., Gilliam J.W., Groffman P.M., Hey D.L., Randall G.W. & Wang N. (1999) *Reducing nutrient loads, especially nitrate-nitrogen, to surface water, groundwater, and the Gulf of Mexico.* NOAA Coastal Ocean Office.

Moriasi D.N., Arnold J.G., Van Liew M.W., 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.

Moriasi D.N., Bosch D.D., Bjorneberg D.L. & King K.W. (2014) A literature review of current parameterization, calibration, and validation approaches of APEX to support NTT. *Report Submitted to USDA Office of Environmental Markets*.

Morris C. (2014) Management Options for Animal Operations to Reduce Nutrient Loads In: *NI R 14-03 Supplemental Paper 3,* Duke University, Durham, NC.

Mudgal A., Baffaut C., Anderson S.H., Sadler E.J. & Thompson A. (2010) APEX model assessment of variable landscapes on runoff and dissolved herbicides. *Transactions of the ASABE,* **53,** 1047-1058.

Mulla D.J. (2008) 14. Limitations of Evaluating the Effectiveness of Agricultural Management Practices at Reducing Nutrient Losses to Surface Waters, 189-212.

Nair S.S., King K.W., Witter J.D., Sohngen B.L. & Fausey N.R. (2011) Importance of Crop Yield in Calibrating Watershed Water Quality Simulation Tools. *Journal of the American Water Resources Association,* **47,** 1285-1297.

NCSS. National Cooperative Soil Survey Soil Characterization Database. http://ncsslabdatamart.sc.egov.usda.gov/.

Olander L., Walter T., Vadas P., Heffernam J., Kebriab E., Ribaudo M. & Harter T. (2014) Refining models for quantifying the water quality benefits of improved animal management for use in water quality trading. NI R 14-03, Duke University, Durham, NC.

Osmond D.L., Hoag D.L., Luloff A.E., Meals D.W. & Neas K. (2015) Farmers' use of nutrient management: lessons from watershed case studies. *J Environ Qual,* **44,** 382-390.

Peterjohn W.T. & Correll D.L. (1984) Nutrient dynamics in an agricultural watershed: observations on the role of a riparian forest. *Ecology,* **65,** 1466-1475.

Plotkin S., Wang X., Potter T., Bosch D., Williams J., Hesketh E. & Bagdon J. (2013) APEX calibration and validation of water and herbicide transport under US southern Atlantic coastal plain conditions. *Transactions of the ASABE,* **56,** 43-60.

Pritts M.P. & Hancock J.F. (1985) Lifetime biomass partitioning and yield component relationships in the highbush blueberry, Vaccinium corymbosum L.(Ericaceae). *American journal of botany***,** 446-452.

Qiu Z., Prato T., Godsey L. & Benson V. (2002) Integrated Assessment of Uses of Woody Draws in Agricultural Landscapes. *Journal of the American Water Resources Association*, **38**, 1255-1269.

Ranalli A.J. & Macalady D.L. (2010) The importance of the riparian zone and in-stream processes in nitrate attenuation in undisturbed and agricultural watersheds – A review of the scientific literature. *Journal of Hydrology,* **389,** 406-415.

Ribaudo M.O., Heimlich R., Claassen R. & Peters M. (2001) Least-cost management of nonpoint source pollution: source reduction versus interception strategies for controlling nitrogen loss in the Mississippi Basin. *Ecological Economics,* **37,** 183-197.

Rosenstock T.S., Liptzin D., Dzurella K., Fryjoff-Hung A., Hollander A., Jensen V., King A., Kourakos G., Mcnally A., Pettygrove G.S., Quinn J., Viers J.H., Tomich T.P. & Harter T. (2014) Agriculture's Contribution to Nitrate Contamination of Californian Groundwater (1945-2005). *J Environ Qual,* **43,** 895-907.

Santhi C., Kannan N., White M., Di Luzio M., Arnold J., Wang X. & Williams J. (2014) An integrated modeling approach for estimating the water quality benefits of conservation practices at the river basin scale. *Journal of Environmental Quality,* **43,** 177-198.

Schindler D. & Fee E. (1974) Experimental lakes area: whole-lake experiments in eutrophication. *Journal of the Fisheries Board of Canada,* **31,** 937-953.

Shared Strategy Development Committee. (2007) Puget Sound Salmon Recovery Plan. NOAA Fisheries, Olympia, WA.

Snyder J. (2015) Addendum to the 2009 Sanitary Survey Report of Portage Bay. Washington State Department of Health: Office of Shellfish and Water Protection Olympia, WA.

Sobota D.J., Compton J.E., Mccrackin M.L. & Singh S. (2015) Cost of reactive nitrogen release from human activities to the environment in the United States. *Environmental Research Letters,* **10,** 025006.

Steglich E.M. (2014a) Weather Import for winAPEX. Blackland Research and Extension Center, Temple, TX.

Steglich E.M. (2014b) WinAPEX: An APEX Windowns Interface Users Guide. Blackland Research and Extension Center, Temple, TX.

Steglich E.M. & Williams J.R. (2013) Agriclutural Policy/ Environmental EXtender Model: User's manual version 0806. Blackland Research and Extension Center, Temple, TX.

Strauss P., Leone A., Ripa M., Turpin N., Lescot J.M. & Laplana R. (2007) Using critical source areas for targeting cost-effective best management practices to mitigate phosphorus and sediment transfer at the watershed scale. *Soil Use and Management,* **23,** 144-153.

Strik B. & Buller G. (2005) The impact of early cropping on subsequent growth and yield of highbush blueberry in the establishment years at two planting densities is cultivar dependant. *HortScience,* **40,** 1998-2001.

Sudheer K.P., Chaubey I., Garg V. & Migliaccio K.W. (2007) Impact of time-scale of the calibration objective function on the performance of watershed models. *Hydrological Processes,* **21,** 3409-3419.

Tarboton D.G. (2007) Surface Water Quality Model Development and Calibration, WRIA 1 Watershed Management Project Phase III, Task 4.1 Report Utah Water Research Laboratory, Utah State University.

Tuppad P., Santhi C., Wang X., Williams J., Srinivasan R. & Gowda P. (2010) Simulation of conservation practices using the APEX model. *Applied engineering in agriculture,* **26,** 779-794.

Turner R.E. & Rabalais N.N. (2003) Linking landscape and water quality in the Mississippi River basin for 200 years. *BioScience,* **53,** 563-572.

USDA-ARS. Agricultural Land Management Alternative with Numerical Assessment Criteria (ALMANAC). United States Department of Agriculture- Agricultural Research Service. Temple, TX.

USDA-NASS. (2014) 2012 Census of Agriculture. United States Department of Agriculture. Washington, D. C.

USDA-NRCS. Geospatial Data Gateway: 3 meter digital elevation map. https://gdg.sc.egov.usda.gov/.

USDA-NRCS. Soil Survey Geographic (SSURGO) Database. http://sdmdataaccess.nrcs.usda.gov/.

USDA. (2014) Assessment of the Effects of Conservation Practices on Cultivated Cropland in the Pacific Northwest Basin. United States Department of Agriculture Natural Resources Conservation Service Conservation Effects Assessment Program. Washington, D.C.

Van Grinsven H.J., Holland M., Jacobsen B.H., Klimont Z., Sutton M.A. & Jaap Willems W. (2013) Costs and benefits of nitrogen for Europe and implications for mitigation. *Environ Sci Technol,* **47,** 3571-3579.

Vidon P.G. & Hill A.R. (2004) Landscape controls on nitrate removal in stream riparian zones. *Water Resources Research,* **40,** W03201.

Vitousek P.M., Mooney H.A., Lubchenco J. & Melillo J.M. (1997) Human domination of Earth's ecosystems. *Science,* **277,** 494-499.

Wang E., Xin C., Williams J.R. & Xu C. (2006a) Predicting soil erosion for alternative land uses. *J Environ Qual,* **35,** 459-467.

Wang X., Gassman P., Williams J., Potter S. & Kemanian A. (2008) Modeling the impacts of soil management practices on runoff, sediment yield, maize productivity, and soil organic carbon using APEX. *Soil and Tillage Research,* **101,** 78-88.

Wang X. & Jeong J. (2015) APEX-CUTE 2.0 User's Manual. Temple, TX.

Wang X., Kemanian A.R. & Williams J.R. (2011) Special features of the EPIC and APEX modeling package and procedures for parameterization, calibration, validation, and applications. *Methods of Introducing System Models into Agricultural Research***,** 177-207.

Wang X., Potter S., Williams J., Atwood J.D. & Pitts T. (2006b) Sensitivity analysis of APEX for national assessment. *Transactions of the ASABE,* **49,** 679-688.

Wang X., Williams J., Gassman P., Baffaut C., Izaurralde R., Jeong J. & Kiniry J. (2012) EPIC and APEX: Model use, calibration, and validation. *Transactions of the ASABE,* **55,** 1447-1462.

Wang X., Yen H., Liu Q. & Liu J. (2014) An auto-calibration tool for the Agricultural Policy Environmental eXtender (APEX) model. *Transactions of the ASABE,* **57,** 1087-1098.

Whatcom County Public Works-Natural Resources. (2015) Shellfish of Whatcom County Vol. 11/30/2015. Whatcom County Public Works- Natural Resources.

Willamette Partnership W.R.I., And the National Network on Water Quality Trading. (2015) Building a Water Quality Trading Program: Options and Considerations.

Williams A. (2014) Initial Sampling Protocol using the AccuPAR LP-80. Vol. Sep/26/2014.

Williams C.O., Lowrance R., Bosch D.D., Williams J.R., Benham E., Dieppa A., Hubbard R., Mas E., Potter T., Sotomayor D., Steglich E.M., Strickland T. & Williams R.G. (2013) Hydrology and water quality of a field and riparian buffer adjacent to a mangrove wetland in Jobos Bay watershed, Puerto Rico. *Ecological Engineering,* **56,** 60-68.

Williams J., Arnold J., Kiniry J., Gassman P. & Green C. (2008) History of model development at Temple, Texas. *Hydrological sciences journal,* **53,** 948-960.

Williams J., Harman W., Magre M., Kizil U., Lindley J., Padmanabhan G. & Wang E. (2006) APEX feedlot water quality simulation. *Transactions of the ASABE,* **49,** 61-73.

Williams J., Potter S., Wang X., Atwood J., Norfleet L., Gerik T., Lemunyon J., King A., Steglich E. & Wang C. (2010) APEX model validation for CEAP. United States Department of Agriculutre-Natural Resources Conservation Service, Temple, TX**,** 1-13.

Williams J.R., Izaurralde R.C. & Steglich E.M. (2012) Agricultural Policy/ Environmental EXtender Model: Theoretical Documentation version 0806. Blackland Research and Extension Center Temple, TX.

Yin L., Wang X., Pan J. & Gassman P. (2009) Evaluation of APEX for daily runoff and sediment yield from three plots in the Middle Huaihe River Watershed, China. *Transactions of the ASABE,* **52,** 1833-1845.

SUPPLEMENTARY DATA

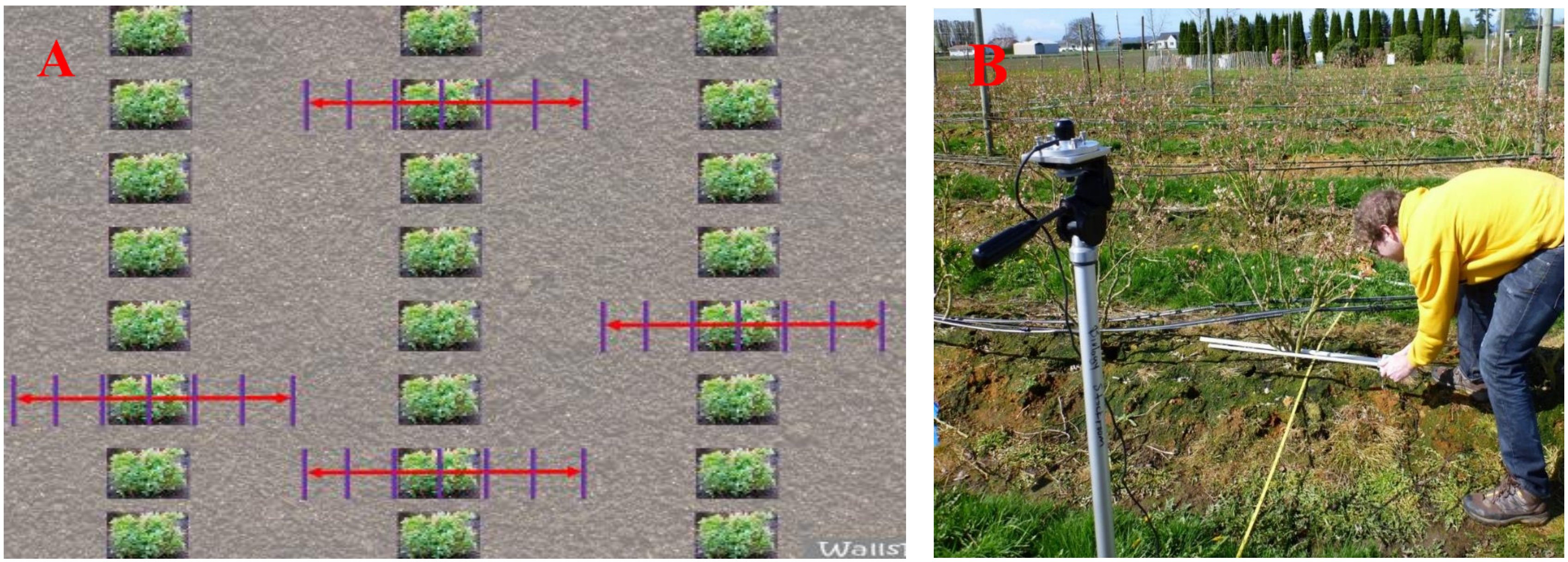

Figure S1. (A) Schematic of LAI sampling for blueberry and raspberry fields. Each transect (red) was sampled seven times using the ceptometer (purple) to capture LAI of the plants and spaces between rows. **(B)** Measuring light interception beneath a blueberry (*Vaccinium corymbosum*, Duke variety) canopy in April, 2015 to develop growth curves of leaf area index, the fundamental parameter APEX uses to simulate plant growth rate.

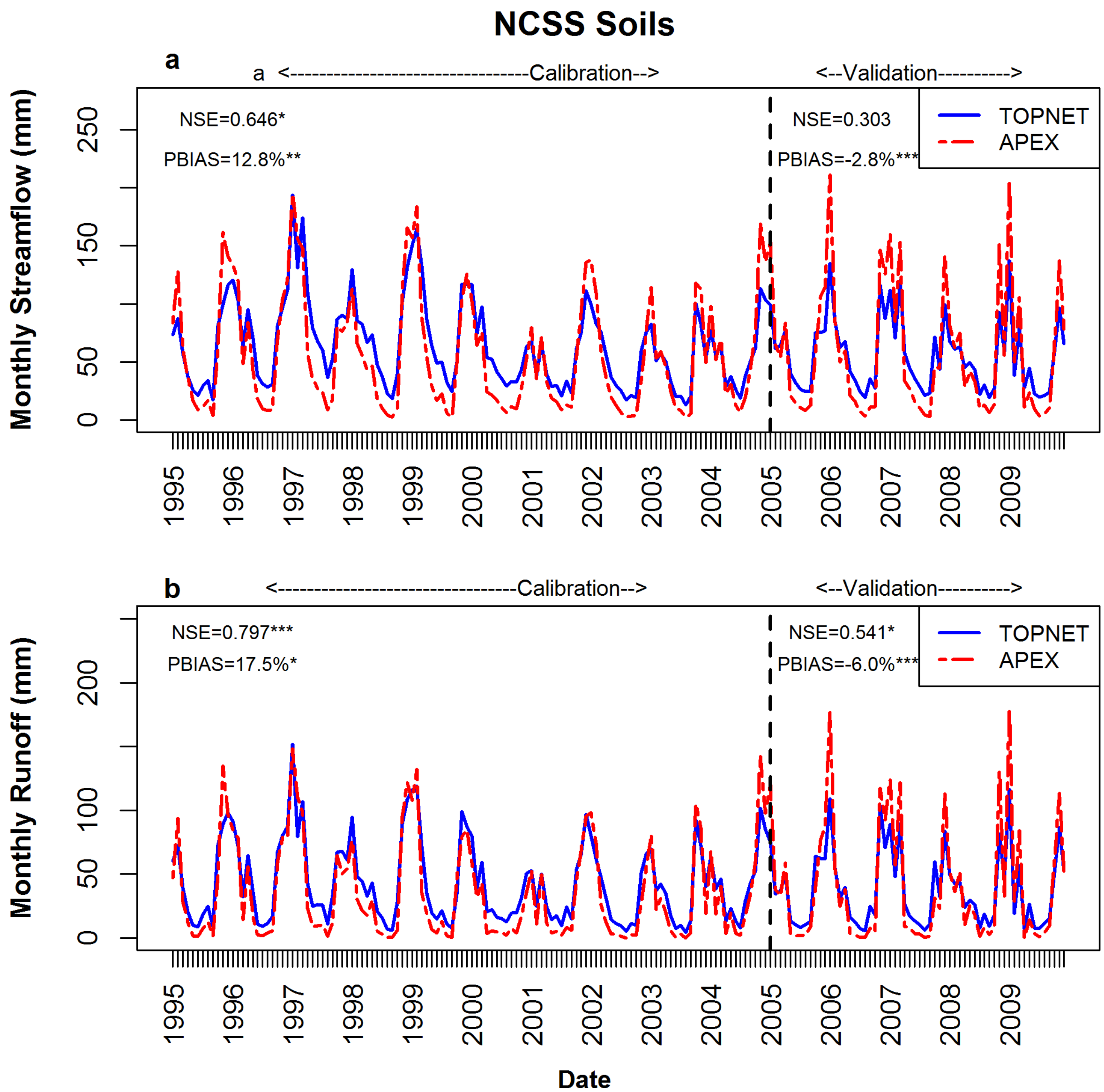

Figure S2. Predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff (a) and streamflow (b) during the calibration (1995-2004) and validation (2005-2010) periods for NCSS soils. Lines and symbols as in Fig. 3.

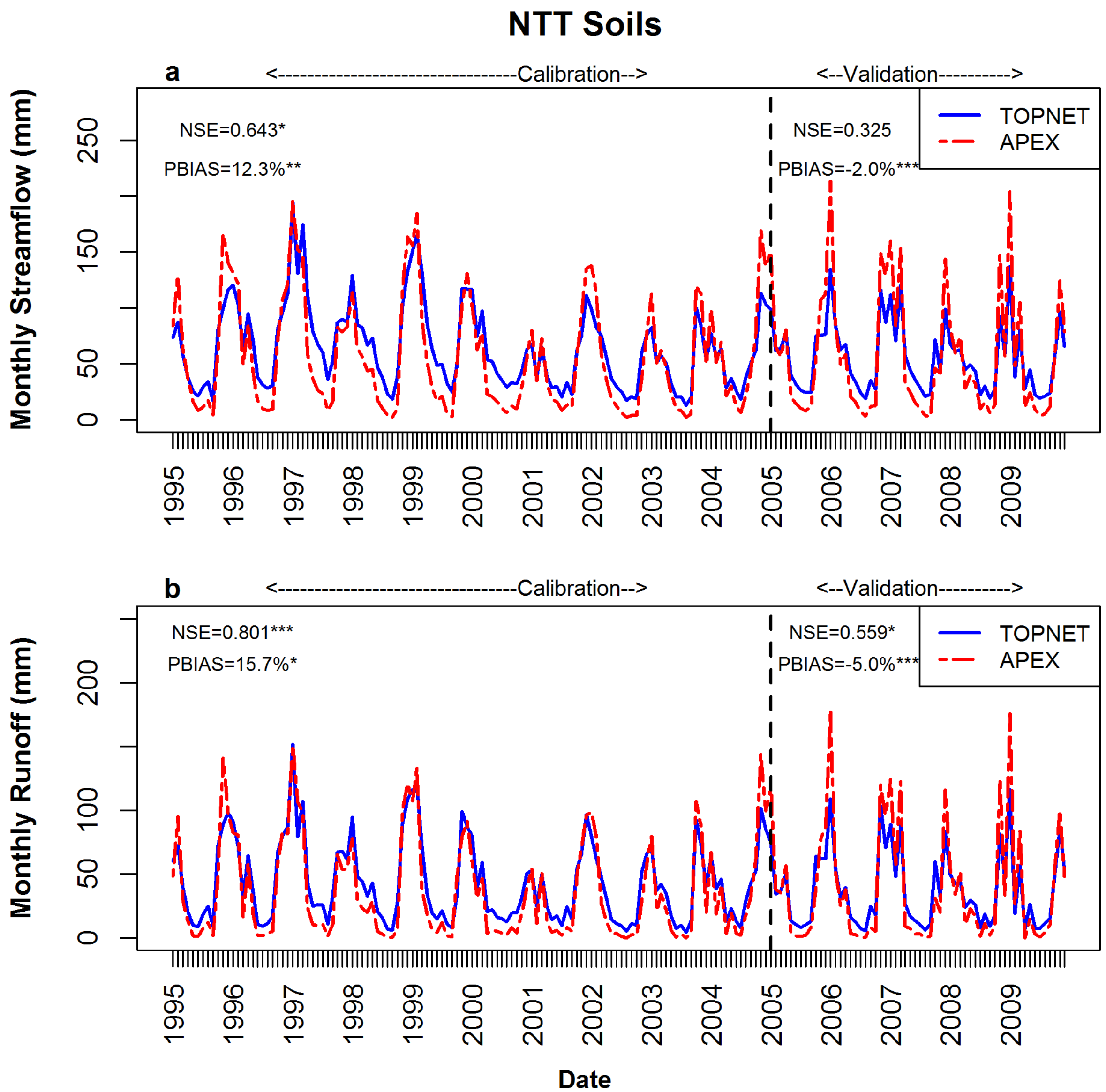

Figure S3. Predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff (a) and streamflow (b) during the calibration (1995-2004) and validation (2005-2010) periods for NTT soils. Lines and symbols as in Fig. 3.

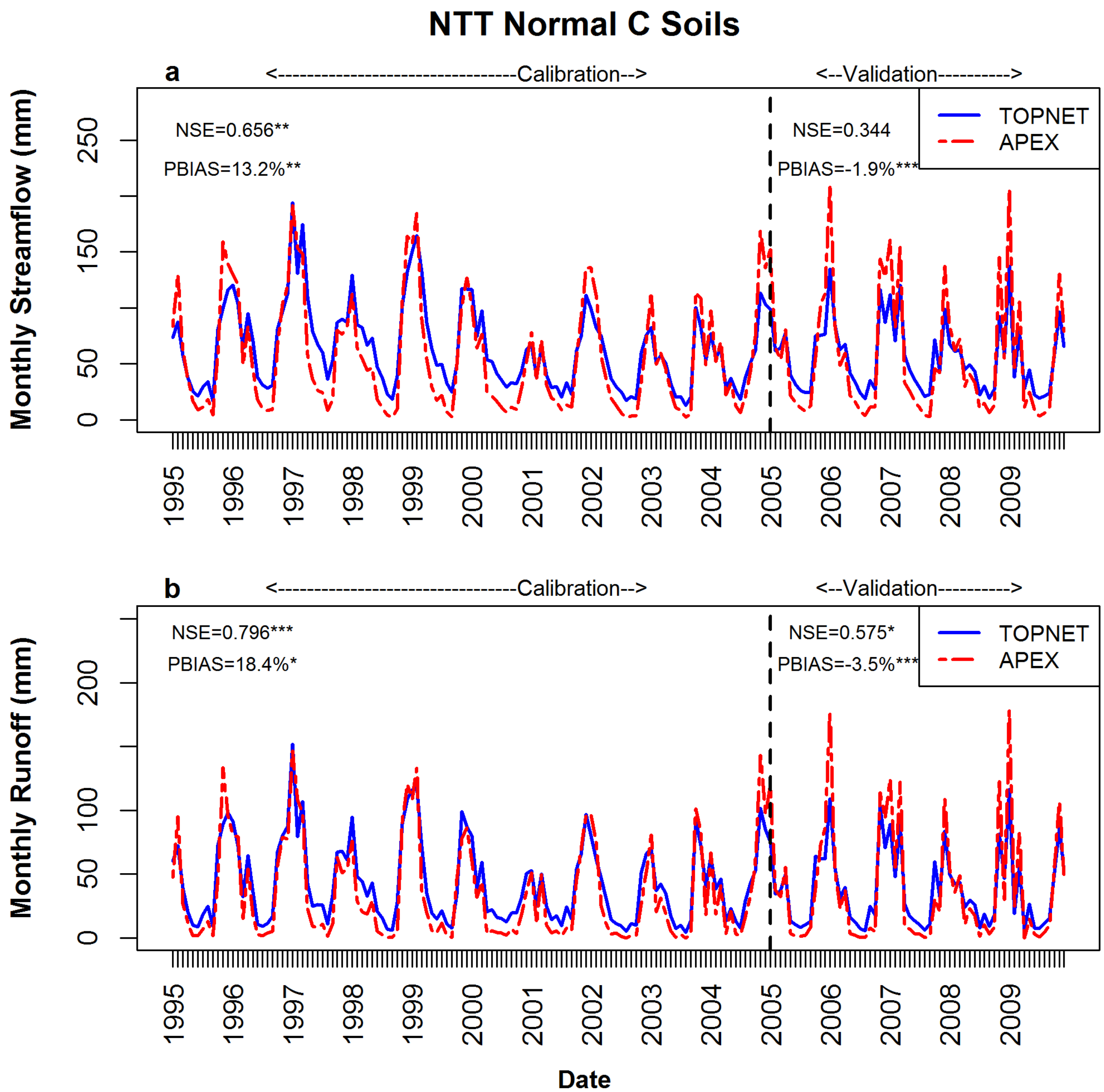

Figure S4. Predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff (a) and streamflow (b) during the calibration (1995-2004) and validation (2005-2010) periods for NTT Normal C soils. Lines and symbols as in Fig. 3.

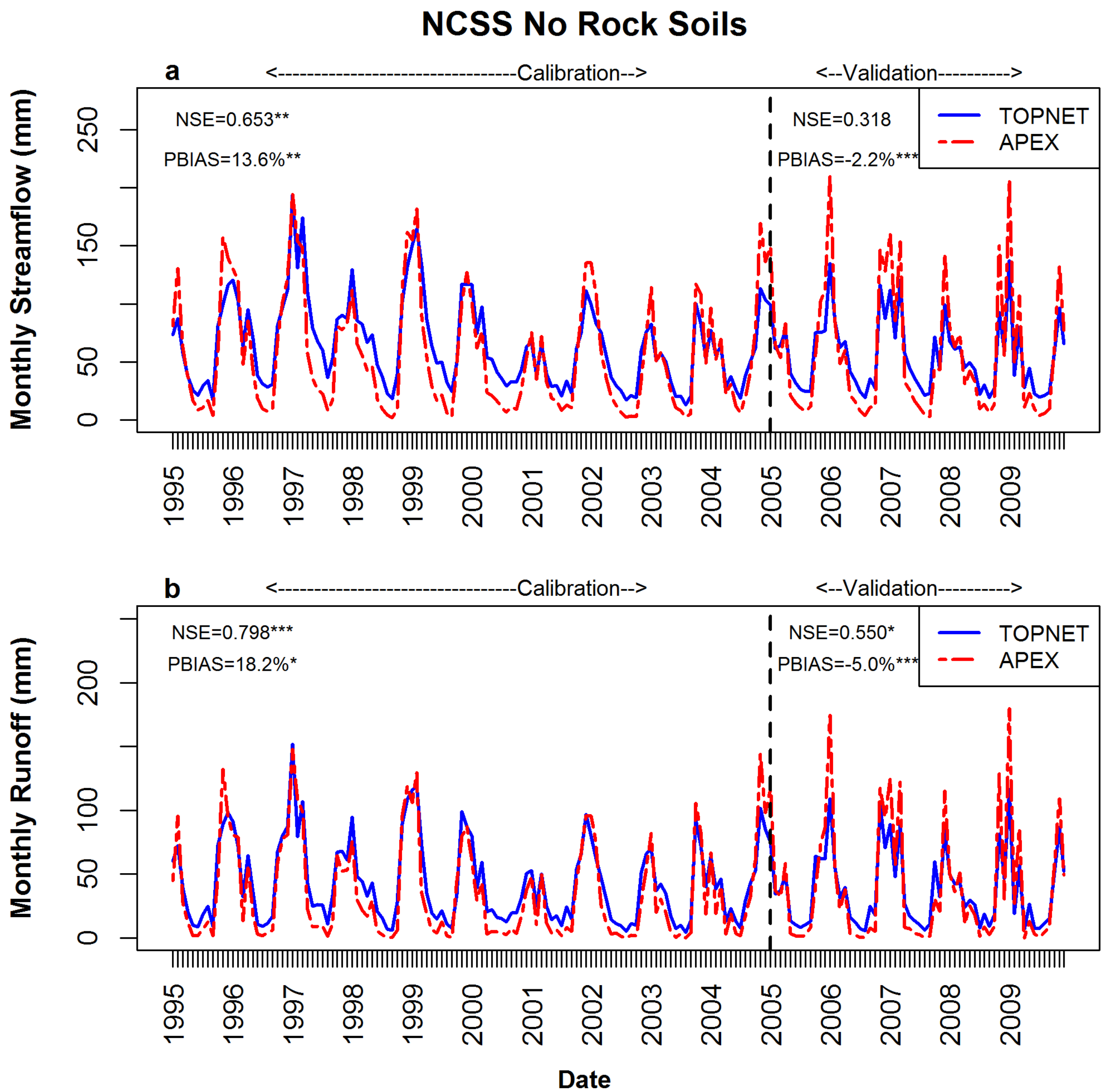

Figure S5. Predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff (a) and streamflow (b) during the calibration (1995-2004) and validation (2005-2010) periods for NCSS No Rock soils. Lines and symbols as in Fig. 3.

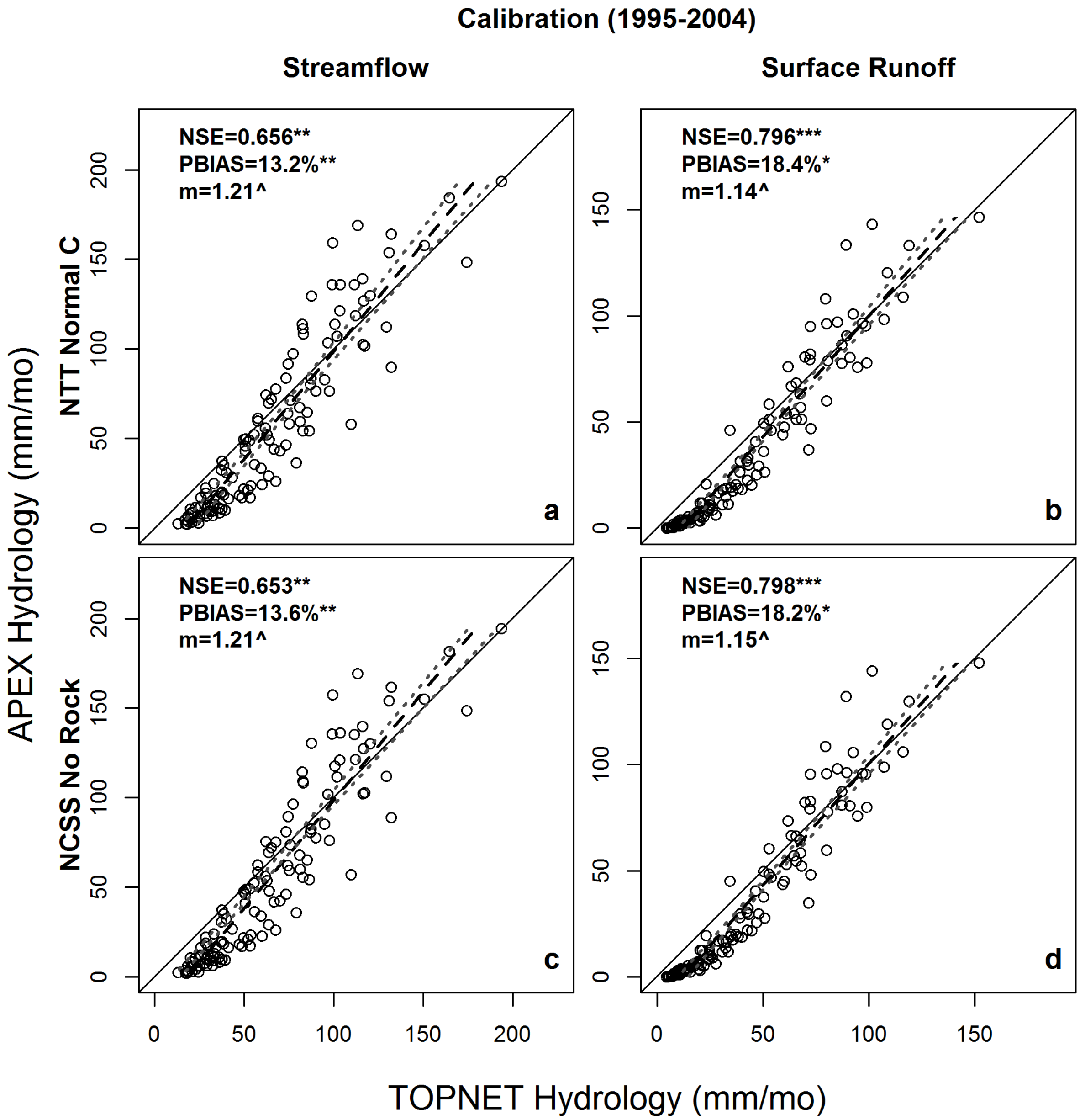

Figure S6. Linear regressions of predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff and streamflow for the North Kamm Creek Watershed during the calibration period (1995-2004) for NTT Normal C (a,b) and NCSS No Rock (c, d) soils. Lines and symbols as in Fig. 4.

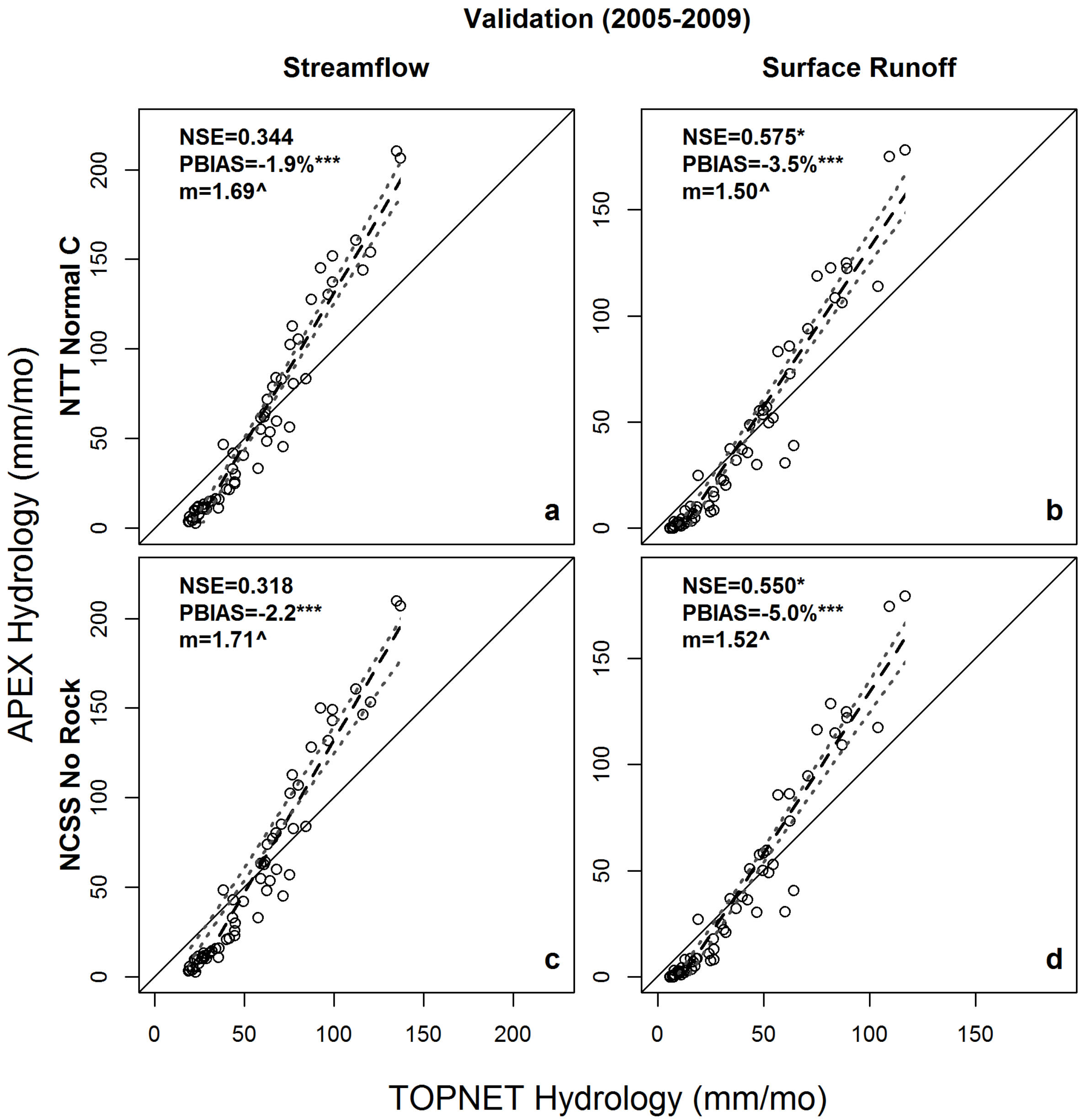

Figure S7. Linear regressions of predicted (APEX generated) vs observed (TOPNET generated) monthly surface runoff and streamflow for the North Kamm Creek Watershed during the validation period (2005-2009) for NTT Normal C (a,b) and NCSS No Rock (b,c) soils. Lines and symbols as in Fig. 4.

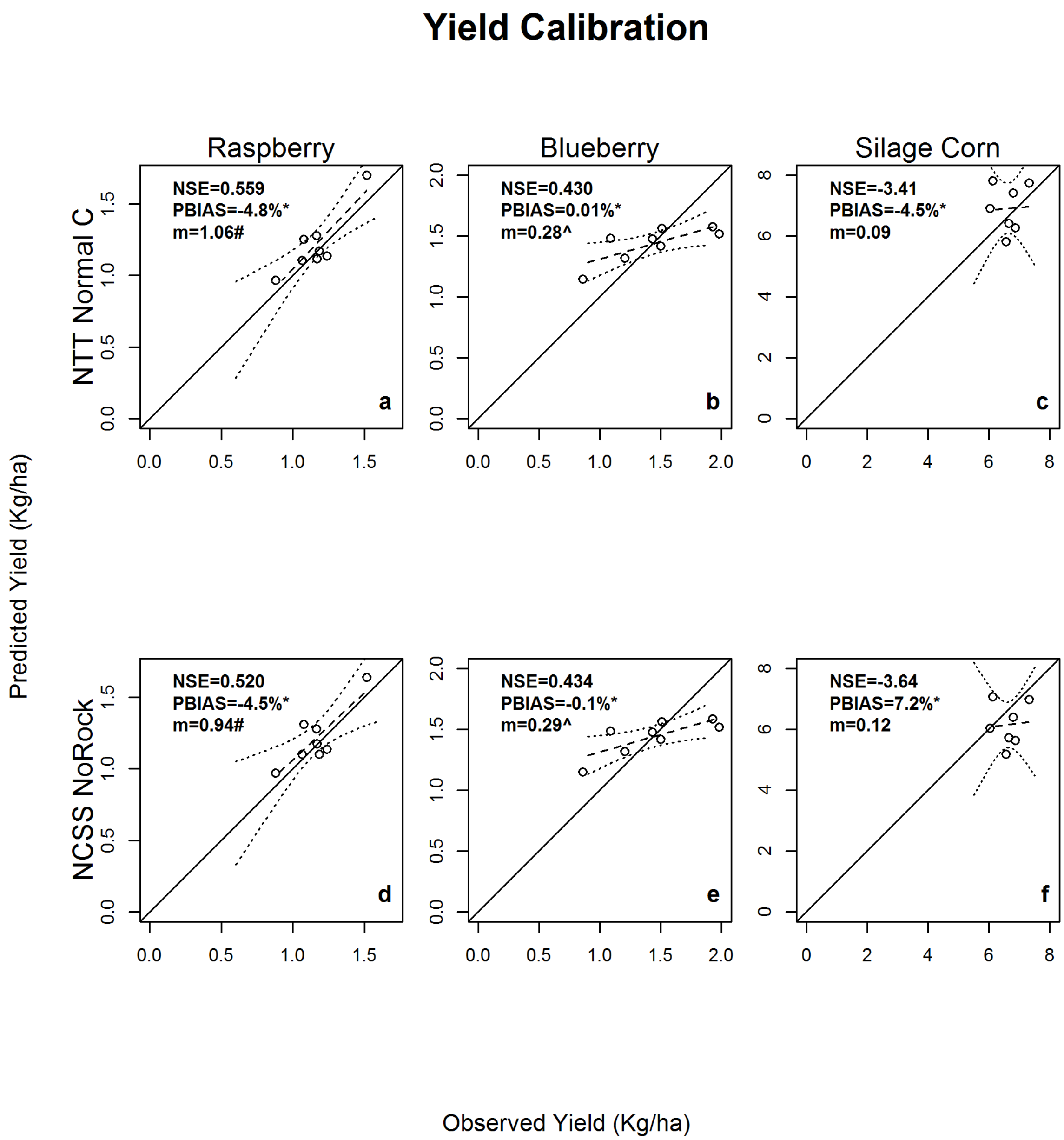

Figure S8. Linear regressions of predicted (APEX generated) vs observed yields for three major crops in the North Kamm Creek Watershed during the calibration period for NTT Normal C (a,b,c) and NCSS NO Rock (d,e,f) soils. Lines and symbols as in Fig. 6.

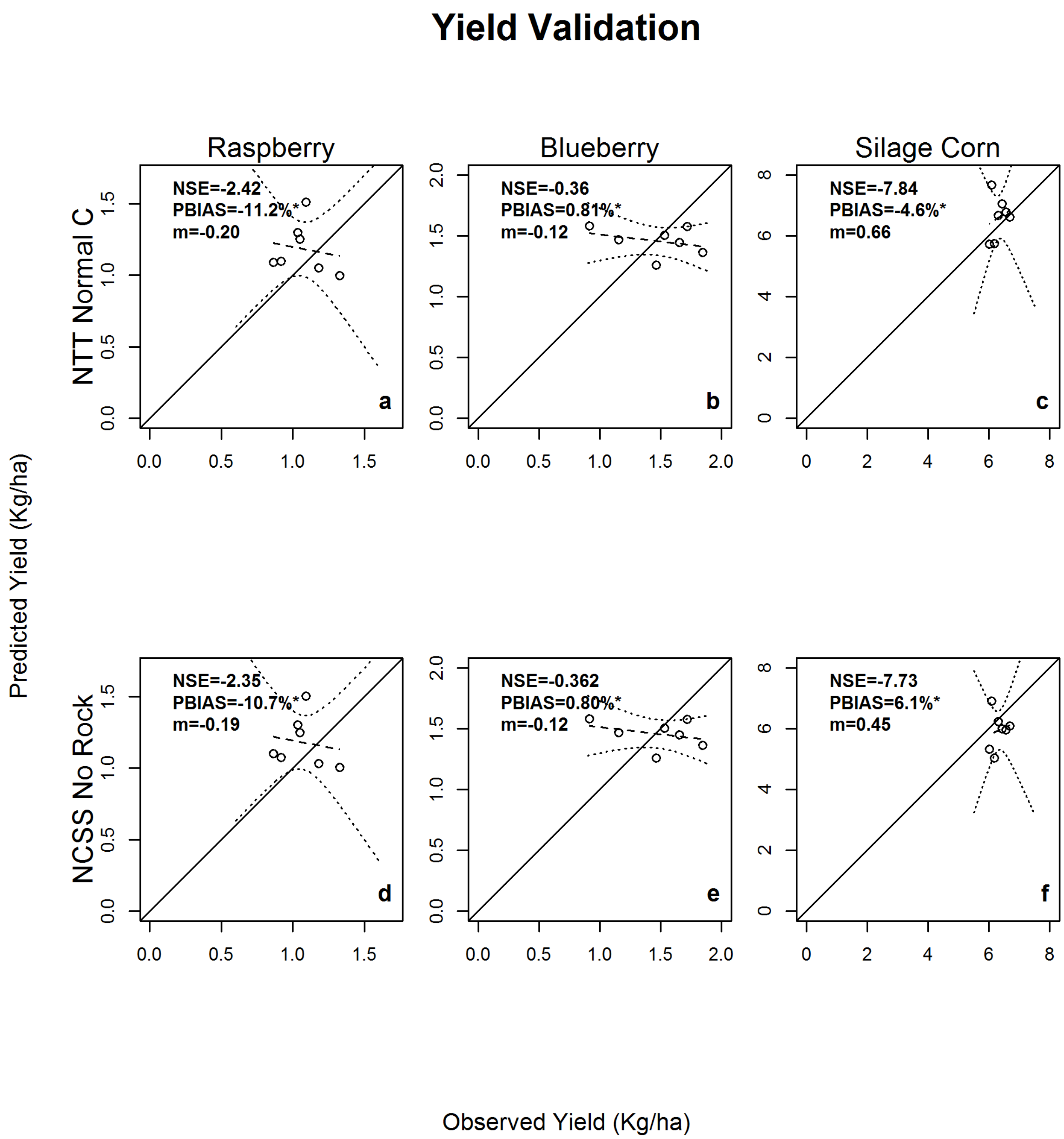

Figure S9. Linear regressions of predicted (APEX generated) vs observed yields for three major crops in the North Kamm Creek Watershed during the validation period for NTT Normal C (a,b,c) and NCSS No Rock (d, e, f) soils. Lines and symbols as in Fig. 6.

Table S1. Percent streamflow derived from surface runoff for model calibration (1995-2004) and validation (2005-2009) periods. “Summer” covers April through September while “Winter” covers October through March. ± Values are standard deviations. “Observed” data were generated by the TOPNET model.

Soil Type:	Observed	SSURGO	NCSS	NTT	NTT Normal C	NCSS No Rock
Overall Calibration	61 ±19	50 ±24	50 ±24	51 ±23	49 ±24	50 ±24
Overall Validation	60 ±19	51 ±27	53 ±27	52 ±27	52 ±27	52 ±27
Summer Calibration	47 ±14	35 ±22	36 ±22	36 ±21	34 ±21	36 ±22
Winter Calibration	75 ±11	66 ±16	65 ±15	66 ±15	65 ±15	65 ±16
Summer Validation	45 ±13	34 ±27	35 ±26	35 ±26	34 ±26	35 ±27
Winter Validation	75 ±11	68 ±13	70 ±12	69 ±13	69 ±13	69 ±13

Table S2. APEX Parameter choices, defaults, ranges, and descriptions for runoff and streamflow autocalibration with APEX-CUTE. I determined optimal parameter values by evaluating the best combinations of PBIAS and NSE after 250 runs. All soil types had the same set of optimal parameters. “n/a” indicates a unitless parameter. Adopted from Wang and Jeong (2015).

APEX Parameter	Default	Minimum	Maximum	Optimal	Units	Description
RFTO	10	0	50	48	days	Groundwater residence time
RFPO	0.95	0.05	0.98	0.44	n/a	Return flow ratio: (return flow)/ (return flow+ deep percolation)
PARM17	0.1	0	0.5	0.207	n/a	Evaporation plant cover factor
PARM20	0.2	0.05	0.4	0.202	n/a	Runoff curve number initial abstraction
PARM23	0.0032	0.0023	0.0032	0.003	n/a	Hargreaves equation coefficient
PARM34	0.6	0.5	0.6	0.532	n/a	Hargreaves equation exponent
PARM40	0.001	0.001	1	0.092	n/a	Groundwater storage threshold: fraction of groundwater storage that initiates return flow
PARM42	0.5	0.3	2.5	2.239	n/a	Curve Number index coefficient
PARM46	0.5	0.5	1.5	1.124	n/a	RUSLE c factor coefficient in exponential crop height function in biomass factor
PARM49	0	0	15	13.774	mm	Maximum rainfall interception by plant canopy (mm)
PARM92	1	0.1	2	0.283	n/a	Curve number retention parameter coefficient

Table S3. Linear regression results of predicted (APEX generated) vs observed (TOPNET generated) hydrology and yield for all soils. Slope values are ± standard error. ^Indicates a significant regression ($p<0.05$). # Indicates a significant regression with a slope (m) not statistically different from l ($p>0.05$).

	Calibration		***Validation***	
	R^2	**Slope**	**R^2**	**Slope**
Hydrology				
SSURGO streamflow	0.86^	1.2 ± 0.05	0.94^	1.62 ± 0.05
SSURGO runoff	0.92^	1.14 ± 0.03	0.93^	1.41 ± 0.05
NCSS streamflow	0.85^	1.22 ± 0.05	0.94^	1.71 ± 0.06
NCSS Runoff	0.91^	1.15 ± 0.03	0.93^	1.53 ± 0.06
NTT streamflow	0.85^	1.22 ± 0.05	0.94^	1.7 ± 0.06
NTT runoff	0.91^	1.17 ± 0.03	0.93^	1.51 ± 0.05
NTT Normal C streamflow	0.85^	1.21 ± 0.05	0.94^	1.69 ± 0.06
NTT Normal C runoff	0.91^	1.14 ± 0.03	0.93^	1.5 ± 0.05
NCSS No Rock streamflow	0.85^	1.21 ± 0.05	0.94^	1.71 ± 0.06
NCSS No Rock runoff	0.92^	1.14 ± 0.03	0.93^	1.52 ± 0.05
Crop Yield				
SSURGO raspberry	0.82^	1.03 ± 0.2#	0.03	-0.18 ± 0.48
SSURGO blueberry	0.58^	0.28 ± 0.1	0.14	-0.14 ± 0.15
SSURGO silage corn	0.00	0.04 ± 0.67	0.01	0.27 ± 1.23
NCSS raspberry	0.69^	0.98 ± 0.27#	0.03	-0.22 ± 0.51
NCSS Blueberry	0.59^	0.29 ± 0.1	0.11	-0.12 ± 0.15
NCSS Silage corn	0.10	0.77 ± 1.03	0.09	0.77 ± 1.11
NTT raspberry	0.18	0.17 ± 0.15	0.07	0.23 ± 0.38
NTT blueberry	0.64^	-0.11 ± 0.03	0.00	-0.02 ± 0.12
NTT silage corn	0.02	-0.21 ± 0.69	0.04	0.49 ± 1.03
NTT Normal C Raspberry	0.77^	1.06 ± 0.23#	0.03	-0.19 ± 0.51
NTT Normal C blueberry	0.59^	0.28 ± 0.1	0.11	-0.12 ± 0.15
NTT Normal C silage corn	0.00	0.09 ± 0.78	0.06	0.66 ± 1.2
NCSS No Rock raspberry	0.70^	0.94 ± 0.25#	0.03	-0.19 ± 0.51
NCSS No Rock blueberry	0.59^	0.29 ± 0.1	0.11	-0.12 ± 0.15
NCSS No Rock silage corn	0.01	0.12 ± 0.71	0.03	0.45 ± 1.07