Coupling Physical, Statistical, and Machine Learning Models with High-Resolution Information Transfer and Rapid-Update Frameworks for Environmental Applications

Andrew Richard Sommerlot

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biological Systems Engineering

Zachary M. Easton, Chair

William C. Hession

James B. Campbell

Stephen H. Schoenholtz

September 20th, 2017

Blacksburg, Virginia

Keywords: Critical source area, deep learning, decision support system, forecast, soil moisture

Coupling Physical, Statistical, and Machine Learning Models with High-Resolution Information Transfer and Rapid-Update Frameworks for Environmental Applications

Andrew Richard Sommerlot

ABSTRACT (Academic)

Few current modeling tools are designed to predict short-term, high-risk runoff from critical source areas (CSAs) in watersheds which are significant sources of non point source (NPS) pollution. This study couples the Soil and Water Assessment Tool-Variable Source Area (SWAT-VSA) model with the Climate Forecast System Reanalysis (CFSR) model and the Global Forecast System (GFS) model short-term weather forecast, to develop a CSA prediction tool designed to assist producers, landowners, and planners in identifying high-risk areas generating storm runoff and pollution. Short-term predictions for streamflow, runoff probability, and soil moisture levels were estimated in the South Fork of the Shenandoah river watershed in Virginia. In order to allow land managers access to the CSA predictions a free and open source software based web was developed. The forecast system consists of three primary components; (1) the model, which preprocesses the necessary hydrologic forcings, runs the watershed model, and outputs spatially distributed VSA forecasts; (2) a data management structure, which converts high resolution rasters into overlay web map tiles; and (3) the user interface component, a web page that allows the user, to interact with the processed output. The resulting framework satisfied most design requirements with free and open source software and scored better than similar tools in usability metrics. One of the potential problems is that the

CSA model, utilizing physically based modeling techniques requires significant computational time to execute and process. Thus, as an alternative, a deep learning (DL) model was developed and trained on the process based model output. The DL model resulted in a 9% increase in predictive power compared to the physically based model and a ten-fold decrease in run time. Additionally, DL interpretation methods applicable beyond this study are described including hidden layer visualization and equation extractions describing a quantifiable amount of variance in hidden layer values. Finally, a large-scale analysis of soil phosphorus (P) levels was conducted in the Chesapeake Bay watershed, a current location of several short-term forecast tools. Based on Bayesian inference methodologies, 31 years of soil P history at the county scale were estimated, with the associated uncertainty for each estimate. These data will assist in the planning and implantation of short term forecast tools with P management goals. The short term modeling and communication tools developed in this work contribute to filling a gap in scientific tools aimed at improving water quality through informing land manager's decisions.

Coupling Physical, Statistical, and Machine Learning Models with High-Resolution Information Transfer and Rapid-Update Frameworks for Environmental Applications

Andrew Richard Sommerlot

ABSTRACT (General Audience)

Water pollution in the United States costs billions of dollars every year. Surface water pollution is caused by excess nutrients and effects the value of fisheries, recreational activities, and commercial operations, and can even lead to health hazards such as dangerous algal blooms. A major source of water pollution is from agricultural activities such as fertilizing crops. This type of pollution is called non-point source, as there is no obvious point where excess nutrients from fertilizers or manure enters the water, such as a discharge pipe, instead the pollution flows over the land first and then into the waterways following the rainfall-runoff patterns. One way to prevent non-point source pollution from agricultural activities is to give farmers tools to optimize operations in a way that can help lower the chance that pollution will occur. Scientific models, like a weather forecast, can help, but there is a lack of tools made specifically for reducing water pollution that are available to farmers. This work focuses on creating a free to use, high resolution and rapid update forecast delivered over the internet, capable of informing agricultural management practices to reduce water pollution. Over the course of this study, published advances in watershed modeling were made filling gaps in existing forecast technology. The final product combines cutting edge watershed science, machine learning and statistical models, web mapping tools, and terabytes of data to meet design goals.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. Zach Easton for being a challenging and valuable mentor through this process and especially for his understanding over the last year. I would also like to thank my committee, Dr. Cully Hession, Dr. Stephen Schoenholtz, and Dr. Jim Campbell for taking time out of their schedules to offer constructive critiques and guidance.

I would also like to acknowledge high-performance computing support from Yellowstone (http://n2t.net/ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, support from the National Science Foundation under award numbers 1360415 and 1343802, and funding support from the USDA under project number 2012-67019-19434.

Table of Contents

List of Tables

List of Figures

AUTHOR'S PREFACE

The following dissertation is comprised of six chapters, four being standalone research studies submitted for publication in peer reviewed scientific journals in the fields of water resource management, environmental modeling, and hydrology or in review processes. The first chapter serves as an introduction to the research described in detail in the remaining chapters and a review of existing literature, the second chapter reviews current literature describing the current state of short term forecasting and the direction for the following research. Chapters 3, 4, 5, and 6 are research studies describing work towards filling the gaps highlighted in the review of current knowledge. The final chapter of this manuscript draws from all preceding chapters to form a statement of conclusions and recommendations looking forward in the field of short-term hydrologic forecasting and operational decision support in surface water protection application.

CHAPTER 1. Introduction

A major problem arising from prolific agriculture is pollution entering the waterways from production practices on land (Carpenter et al., 1998). Protecting surface water quality from this pollution is advantageous for ecosystem health and biodiversity, an economic necessity, and even a matter of national security (Zhang et al., 2007; Shortle et al., 2012; Gleick, 2016). A major threat to the quality of these waters is categorized as non-point source (NPS) pollution from agricultural activities with associated costs estimated in excess of 16 Billion dollars per year and rising (Shortle et al., 2012). Sediment and nutrients, primarily excess nitrogen (N) and phosphorus (P) are the major contributors to this pollution (Shortle et al., 2012). Over the last few decades, various government programs and efforts designed to address the problem of NPS pollution have been implemented; although billions have been spent through these programs, mitigation goals remain difficult to achieve after decades of effort (Sharpley et al., 2009). One reason for this challenge is convincing those responsible for polluting activities to adopt practices which reduce NPS pollution, or best management practices (BMPs) is difficult (Shortle et al., 2012).

Surface runoff is a critical pathway of NPS pollutant transport to streams and often it is small and variable sub-field-level areas of land that are the primary source of surface runoff and water quality degradation (Walter and Walter, 1999). These so called variable source areas (VSAs) as defined by Hewlett and Hibbert (1967) often occur in areas of watersheds that have a shallow top soil layer with high-infiltration rates overlaying a dense subsoil layer which restricts percolation out of the vadose zone (Dunne and Leopold, 1978; Easton et al., 2008; Walter and Walter, 1999). When these features are located in convergent areas of the landscape (e.g., large

upslope contributing areas) the soil must be able to conduct large volumes of water through the profile rapidly, or else it saturates and becomes a runoff source area. These factors result in some areas of the landscape having low soil water storage capacity and are thus easily saturated during precipitation events. The location, size, and duration of saturation of these VSAs depend on variable spatial and temporal factors including soil depth and conductivity, landscape position, antecedent soil moisture and climate (Needelman et al., 2004). Consequently, these VSAs are difficult to model using traditional watershed models (e.g., SWAT, GWLF, HSPF), making the prioritization of landscape management practices challenging (Brooks et al., 2015; Dahlke et al., 2012; Dunne and Leopold, 1978; Marjerison et al., 2011).

Walter et al. (2000) introduced the concept of hydrologically sensitive areas (HSAs) to refer to parts of the landscape where VSAs form and a pollutant source exists (e.g., manure, fertilizer). Knowledge of when and where these HSAs form could be used to prioritize landscape management strategies. For example, a nutrient management plan, often written to direct manure spreading on a seasonal basis, may suggest an operations schedule that does not take into account the location of HSAs within the target fields. Information about where these HSAs exist could be used to modify the nutrient management plan and avoid pollution causing activities during sensitive times or in sensitive areas (Hanrahan et al., 2004; Smith et al., 2007; Vadas et al., 2011).

A major challenge in modeling VSAs is their temporal variability; VSAs can evolve over a single precipitation event, as well as shift due to seasonal fluctuations in precipitation and evapotranspiration. Attempts to model this phenomenon include the Soil and Water Assessment Tool-VSA adaption (SWAT-VSA), a modification of the popular SWAT model that uses topographically derived parameters designed to better quantify the distribution of surface runoff

and soil moisture from VSAs (Easton et al., 2008, 2011, 2010; White et al., 2011; Woodbury et al., 2014). The base SWAT model is a widely used watershed scale model with a variety of successful applications including sub field scale hydrology outputs and sub-daily temporal resolution (Collick et al., 2014; Douglas-Mankin et al., 2010; Easton et al., 2008; Gassman et al., 2014; Sommerlot et al., 2013). SWAT employs the curve number (CN) equation to predict runoff and the way the CN is applied in SWAT implicitly assumes an infiltration-excess (or Hortonian) (Horton, 1933) response to rainfall. However, in humid, well-vegetated regions, especially those with permeable soils underlain by a shallow restricting layer, this formulation fails to capture the observed runoff generating processes (Beven, 2001; Dunne and Black, 1970; Dunne and Leopold, 1978; Needelman et al., 2004). For instance, many soils in the Mid-Atlantic region have a substantially higher infiltration rate than the total precipitation depth of all but the most intense, convective type storms, thus runoff is generated from areas of the landscape that saturate and can no longer store more moisture (Easton et al., 2008). For this reason, the standard CN method may be insufficient for providing detailed information about field scale hydrologic responses. The SWAT-VSA model is more appropriate as it is designed specifically to address areas of short-term high-risk runoff generation and has shown promising results in the estimation of distributed hydrologic processes (Easton et al., 2008). In addition, SWAT-VSA has been shown to provide more accurate predictions of soil moisture and runoff generation than the unmodified SWAT model in watersheds with similar physical characteristics and climate to the study area, located in the Shenandoah Valley in Virginia, United States (Easton et al., 2008; Woodbury et al., 2014).f

In order to use SWAT-VSA in forecast capacity it needs to be coupled with some type of short-term weather model. Short- and medium-range (1-16 day) weather forecasts assimilate

remote and ground observations into an atmospheric model analysis time step to initialize physically based solutions of future surface weather conditions (GFS, 2015; NCEP, 2003). These predictions of future surface weather conditions can then be directly used to force the SWAT-VSA model to provide forecasts of hydrological processes. The National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS, 2015; NCEP, 2003) formerly known as the Medium Range Forecast model (MRF), initially developed by Sela (1982, 1979), provides forecasted surface predictions out 16 days for every location on the planet, four times a day, every single day of the year. This reliability and global coverage makes it an ideal candidate for using as the basis of management planning tool requiring reliable data sources. The GFS is also complemented by the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) that provides long histories with similar physical solutions, helping to mitigate solution based biases that exist between forecasting systems (Hamill et al., 2008). These datasets, when coupled, provide contiguous estimates of weather variables, including precipitation and temperature, for any terrestrial location worldwide from 1979 to 16 days into the future. Both the GFS and the CFSR have been separately evaluated as drivers for the SWAT model with Fuka et al. (2013) demonstrating that CFSR precipitation and temperature inputs resulted in simulated daily discharge that was at least as accurate as those based on records from weather stations located more than 10km away from the geographic center of watersheds in California, New Mexico and New York, USA.

This work demonstrates coupling SWAT-VSA with global weather forecast (GFS) products to provide short-term hydrologic forecasts capable of guiding landscape management process in real-time. This tool can be used for planning or directing operational, short-term field management. Short term hydrologic forecast frameworks have been described in recent studies

using rainfall-runoff and time-series models, but do not specifically address predictions intended to inform field level decision making (Abaza et al., 2014; Dutta et al., 2012; Tsai et al., 2014). The SWAT-VSA model provides two primary forecasts: hydrologic in the form of daily volumetric stream flow estimates and distributed surface hydrology in the form of sub-field level predictions of surface runoff risk probability and soil moisture.

This framework improves upon currently implemented tools such as the P-Index which are annual and developed at a field level resolution. The spatial and temporal scales used to identify areas at a high risk of contributing to NPS pollution often fail to provide information at a high enough spatial and temporal resolution to consistently allow targeted landscape management (Easton et al., 2008; Marjerison et al. 2011; Dahlke et al., 2012), and effectively communicating the results remains a challenge (Easton et al., 2017).

As an example, the P Index is widely used to direct manure and fertilizer management but may fail to account for short duration high-risk periods, such as manure application to saturated areas, so called variable source areas (VSAs). Effectively communicating when and where VSAs form could be used to further optimize agricultural nutrient management operations (Hanrahan et al., 2004; Smith et al., 2007; Vadas et al., 2011). While the capability exists to model VSAs, the results are not immediately available to users or in a format conducive to decision making. Agricultural producers (end users) need a model to run and update in real-time with necessary spatial and temporal resolution imagery to provide insight into the nature of near-future conditions on the ground. In the surface hydrology modeling and land management communities there is a gap between the predictive ability of the modeling systems and how that information is delivered to end users (Jones et al., 2016). Developing policy or management with annual or seasonal averages or at a larger (e.g., watershed) scale is like planning an out-door

event for this upcoming weekend based on average monthly rainfall estimates; although there are data involved and some analysis can be applied, most would agree such a planning process carries considerable uncertainty.

Real time decision support systems (DSS) are an effective approach used to disseminate information by leveraging internet connected devices to directly deliver relevant information to end users (Bharati and Chaudhury, 2004; Durmuşoğlu and Barczak, 2011; Power and Phillips-Wren, 2011). In order to provide VSA forecasts to a large audience of end users and tools to interested scientist developing solutions to similar problems, this study introduces a DSS that provides real-time and spatially detailed, short-term forecasts in a web based interface built on free and open source software (FOSS). The DSS includes an interface and data management structure component designed to disseminate high resolution spatial forecasts in an effort to bridge the gap between modeling capabilities and application to land management decisions made by end users. The method described here can display any georeferenced raster model output developed anywhere in the world at very fine resolutions as an interactive map. Although DSS platforms with similar goals have been introduced (e.g., predicting streamflow or reservoir status (Choi et al; 2005; Dutta et al., 2012; Abaza et al., 2014; Tsai et al., 2014) and software tools have been designed to communicate geospatial information including real time data and forecasts from earth systems models (Kulawiak et al., 2010; Swain et al., 2015; Snow et al, 2016; Tayyebi et al., 2016) we describe a system that is tailor-made for the particular problem of rapid refresh, high resolution, scalable extent native forecast products. We postulate that a web based interface built entirely with open source software can meet the end user requirements of the target audience while being flexible and scalable enough to provide scientists working on similar

problems a method to communicate real time, spatially detailed model results to land managers and decision makers.

The web map framework described above proved to be fast and reliable, and highlighted short-comings of using a cumbersome process based model with a complex input-output structure and no tools for live deployment. A plethora of different models have been applied to hydrologic problems and include a wide range of frameworks. Hydrologic models are often classified into two groups: stochastic or data driven models, which are based on mathematic or statistical data representations, such as regression or neural networks, and process-based models, which represent the physical process occurring, such as rainfall-runoff responses or channel flow (Daniel et al., 2011). Processed-based models are routinely used in agricultural science and water resources fields (Daniel et al., 2011). An emerging application of these models in water resources is rapid update, spatial forecasts of distributed hydrologic parameters, such as soil moisture levels or runoff response (Easton et al., 2017). For these applications, various spatial discretization schemes exist, broadly falling into fully distributed, semi-distributed, and lumped categories (Daniel et al., 2011). Although benefits and costs exist with each method, the computational efficiency and scientific appeal due to the widely applicable characterization schemes of semi-distributed models have led to research and applications of these methods to predict distributed hydrology (Easton et al., 2008; Dalhke et al., 2011; Easton et al., 2017).

The basic structure of Neural Networks (NN), appearing in literature as Artificial Neural Networks (ANN), Multi-Layer Perceptions (MLP), and Deep Learning (DL), have been previously used in hydrology (Hsu, et al., 1995; Dawson and Wilby, 1998; Sajikumar and Thandaveswara, 1999; Abrahart and See, 2000; Luk et al., 2000; Dawson and Wilby 2001; Nagy et al., 2002; Rajurkar et al., 2004; Chen et al., 2012; Valipour et al., 2013; Yuan and Jia, 2016).

"Deep learning" does not yet have a consensus-supported definition, however Schmidhuber (2015) suggests DL can be defined as a NN structure with three or more hidden layers. Applications of NN include reservoir inflow forecasting (Bai et al., 2016), rainfall-runoff response modeling (Wilby et al., 2003), groundwater dynamics (Nayak et al., 2006), soil moisture estimation (Jiang et al., 2004) and forecasting extreme events (Coulibaly et al., 2001). Although these methods show promise, DL models are used less often than process-based models to drive decisions in water resource management (Fatichi et al., 2016). An often-cited reason is that hydrologic models for water resource management must include attempts to describe physical processes not dependent on input data in order to create a "virtual laboratory" where scenarios can be assessed and processes understood (Fatichi et al., 2016). Though long-term scenario modeling is important for water resource management, there has been a recent interest in real-time and short-term forecasting and modeling as components of decision support systems (DSS) designed for water resource management (Easton et al., 2017). In these types of applications, it more important to provide accurate forecasts than to have the ability to model scenarios or fully understand the underlying processes. These short-term forecasting applications present an opportunity to benefit from the predictive power of DL. Ideally, in addition to providing high performance estimations, DL could be used to develop new rules and equations that describe physical processes (Jain et al., 2004). Deep Learning has a proven track record of high performance in a variety of fields and with recent improvements in computation and software structure, these methods are being widely adopted to solve difficult and complex problems (LeCun et al., 2015). Additional improvements in DL platforms have made implementation, training, and in-depth exploration of the final model relatively straightforward for researchers and practitioners (Candel et al., 2016; LeCun et al., 2015; Daniel et al., 2011).

This study demonstrates an implementation of DL designed for enterprise-like applications in providing rapid update, high resolution soil moisture forecasts. A data driven DL model is trained with an input dataset containing parameters that relate to hydrologic processes. The input dataset required to train the DL model is built by applying scientific principles of hydrology in a preprocessing step, guided by what is known and assumed about hydrology in the region of application. In this way, decades of scientific knowledge and study are used to create an input dataset which provides a physical context built with known processes for training the DL model. Based on this input data and the dependent variable (soil moisture), the DL model is then free to learn while optimizing a loss function during the training process. The DL model learns both major driving relationships that may be similar to process-based models, and additional intricacies represented by data driven relationships. Additionally, the DL model can be trained on higher discretization resolution where initialization of the process based model fails or becomes too cumbersome for calibration or rapid-update implementation.

The objective of this work is to fill the gaps in the development and application of rapid update, high resolution soil moisture forecasts intended to support agricultural management decisions. The following chapters describe the development and testing of technologies designed address 1) the modeling limitations of short-term forecasting of overland hydrology 2) lacking frameworks to communicate high resolution, rapid update forecasts, 3) limited ability of hydrology models to scale to very large areas when high resolution inputs are used, and 4) fill data gaps in the soil nutrient record in an effort to provide supporting information to future work connecting soil moisture with NPS risk from agricultural land.

CHAPTER 2. Literature Review

Agriculture in the United States

The gross production value of agricultural products in the United States (US) was an estimated $252 billion in 2014, accounting for over one percent of the gross domestic product (GDP) (CIA factbook, 2017; FAO stat, 2017). The US is the world's largest agricultural exporter and provides over half of the total exported corn in the world (CIA factbook, 2017). Over 85% percent of households in the US were considered food secure in 2014, and the US received a top score in the Global Food Security Index in 2016 (Coleman-Jensen et al., 2014; The Economist Intelligence Unit, 2016). The success of agriculture in the US is very important to food and energy security, and economic development in the US (Joint Economic Committee, 2013), however, industrialized agriculture can cause many negative environmental consequences, including habitat destruction, greenhouse gas emissions, and perhaps most critically, water quality degradation (EPA, 2005).

2.1 State of Water Quality in the United States

Non-point source (NPS) pollution from agricultural lands is a significant problem that threatens water quality throughout the United States (EPA, 2005). Excess sediment load is a major portion of this NPS pollution and directly links agricultural landuse to water quality (Bossio et al., 2010). Publically sponsored programs exist that have a goal of reducing NPS pollution from agricultural lands (Shortle, 2012). These programs usually experience mixed success, and have difficulty meeting their goals of NPS pollution reduction, in part due to a lack of quality data about the sources and amounts of excess sediment loads (Thomas and Froemke, 2012). By collecting more data describing a watershed, water quality programs can be improved and experience a higher level of effectiveness. Monitoring sediment loads in a watershed is not always feasible due to economic and time constraints. Watershed models can fill the gap in available data for decision makers without being a heavy economic or time burden.

2.2 Water Quality Management Programs

Modern water quality management practices or best management practices (BMPs) stemmed from the Clean Water Act, or CWA (Shortle et al., 2012). The CWA defined the government responsibility structure for surface water pollution and categorized point-source (PS) pollution as a federal responsibility and NPS pollution as a state-level responsibility (Shortle et al., 2012). A complex system of initiatives and efforts has risen from the CWA and NPS management programs vary between states, and often include or are limited to voluntary programs (Shortle et al., 2012, NRCS, 2012a).

2.2.1 Voluntary Programs

Multiple government agencies sponsor many water quality programs; some of the major efforts are sponsored by the United States Department of Agriculture (USDA), including the Conservation Reserve Program (CRP), the Environmental Quality Incentive Program, (EQIP), and the Conservation Stewardship program (CSP) (NRCS, 2012b). These programs provide monetary and/or technical assistance in an attempt to encourage land owners to implement best management practices (NRCS, 2012b). Other programs exist which may have a more specific focus, many of which were derived from the 2008 Farm Bill (USDA, 2008a) including the Agricultural Management Assistance Program (AMA) Cooperative Conservation Partnership Initiative (CCPI) Conservation of Private Grazing Land Program (CPGL) Agricultural Water Enhancement Program (AWEP) Conservation Innovation Grants (CIG) Farm and Ranch Lands Protection Program (FRPP) Wetlands Reserve Program (WRP) **(**NRCS, 2008; NRCS, 2012b; USDA, 2008a; USDA, 2008b; USDA, 2008c; USDA, 2008d, USDA, 2008e, USDA, 2008f, USDA, 2008g; USDA, 2008h; USDA, 2008i)

2.2.2 Total Maximum Daily Load

The total maximum daily load (TMDL) is a regulatory instrument defined by the CWA section 303(d) which sets the limit of pollution that may enter a body of water which is deemed impaired and still meet water quality standards (USPEA, 2008; 2017). The process defined in section 303(d) is comprised of multiple steps beginning with water quality standards (WQS) development the process then moves through reporting and defining which water bodies do not meet water quality standard, labeling them impaired waters. A TMLD is developed for impaired waters, guiding issuance of discharge permits and implementation of NPS management programs

(USEPA, 2017). Many TMDLs are currently in effect in the US, the largest of which is the Chesapeake Bay TMDL, a key component of the Chesapeake Bay Program (CBP) (USDA, 2008b).

2.3 Role of Hydrologic Modeling in Watershed Management

Hydrologic models often provide analysis for government sponsored water quality programs; models in the literature include the Soil and Water Assessment Tool (SWAT), the Water Erosion Prediction Project (WEPP), the Annualized Agricultural Non-Point Source (AnnAGNPS) model, the Long-Term Hydrologic Impact Assessment model (L-THIA), the PLOAD model, and the HSPF model (Shen et al., 2009; Parajuli et al., 2009; Im et al., 2009; Nejadhashemi et al., 2011; Giri et al., 2012). Tools like these and their predecessors have long been used for scenario analysis focused on potential changes in water quality in the long term future based on current or simulated management actions, and to assess progress in regional programs such as initiatives or TMDLs (Thomann, 1972; Mitsch and Wang, 2000; National Research Council, 2001; Chapra, 2008; Shenk and Linker, 2013). Much modeling work focuses on methods that estimate physical processes and are capable of creating a "virtual laboratory" where scenarios can be assessed, processes understood, and key findings easily interpreted (Fatichi et al., 2016).

Hydrologic models and a subset of this category, water quality models, have various use cases in watershed management; as categorized by Harmel et al (2014) they are exploratory, planning, and regulatory for legal purposes, with models for planning purposes mainly including processes that are capable of capturing scenario differences. Although this may be a reasonably complete categorization of the current state of use cases for hydrologic models

(Daniels et al, 2011), Grey et al., (2013) claim hydrologic and water quality science must play a more direct role in advancing stakeholder understanding of well known science concerning water quality and take more direct action in bridging the gap between scientific knowledge and management actions. One promising vector for working to close this gap is "…through informing decisions using simple, quantitative spatial tools within a clear decision support process…" which could be a valuable asset in informing management with best science principles directly from the source of research or as part of larger programs or officially defined partnerships (Ruckelshaus et al., 2015).

2.4 Application of Short Term Forecasting Tools in Watershed Management

Tools designed to inform short term decisions related to agricultural management are prevalent in the literature, and examples exist discussing implementations across the globe (Quinn, 2000; De la Rosa et al., 2004; Naab, 2004; Nielsen et al., 2005; Stefanski and Sivakumar, 2009; Liu et al., 2011; Saseendran et al., 2013; Lee et al., 2016). Over 50 applications of published frameworks were reviewed, ranging from simple to complex, but sharing a goal of informing short term decision making that discussed applications in an agricultural management situation utilizing weather. These were separated into seven broad categories: 1) agricultural risk (Gutenson, 2016; Baig et al., 2015; Whateley et al., 2015; Wilkinson et al., 2012; Zhan et al., 2012; Wilhite, 2011; Hong et al., 2010; Funk and Verdin, 2010; Stefanski, and Sivakumar, 2009; Cohen et al., 2008), 2) crop yield (Naab et al., 2004; Becker-Reshef et al., 2010; Liu et al., 2011; Saseendran et al., 2013) 3) general purpose, including water quantity management and those who discussed multiple or broadly defined use cases (Quinn, 2000; Matthews, 2000; Quinn et al., 2001; Giacomelli et al., 2001; Levy 2001;

Mastin and Vaccaro, 2002; Murphy, 2003; Quinn et al., 2004; Martius et al., 2004; Ouda, 2004; Riquelme et al., 2005; Nielsen et al., 2005; Neumann et al., 2006; Regonda et al., 2006; McColl and Aggett, 2007; De Kort and Booi, 2007, Garbrecht and Schneider, 2007; Thorp et al., 2008; Kumar et al., 2008; Sankarasubramanian et al., 2009; Quinn et al., 2010; Ziervogel et al., 2010; Chen et al., 2014; Lee et al., 2016), 4) irrigation (Vick, 2016; Tapsuwan et al., 2015; Duan, 2005; Bazzani, 2005; Leenhardt et al., 2004; Bazzani et al., 2004), 5) livestock management (Nienaber and Hahn, 2007; Clements et al., 2006), 6) nutrient management (Trepel and Kluge, 2004; De and Bezuglov, 2006; Bruges and Smith, 2008; Hickey and Gibbs, 2009), and 7) soil conservation (De la Rosa et al, 2004; Tang et al., 2004; Chen et al., 2004; Lim et al., 2005; Hunink et al., 2012; Zanuttigh et al., 2014). Numbers of these studies are summarized by topic in Figure 2.1. The general-purpose category was most common, where authors presented their decision support system as having wide ranging water quality or/or quantity applications in agriculture, the agricultural risk category was next, including studies presenting work aimed at users attempting to avoid or prepare for events that pose threats to agricultural productivity, such as extreme whether events or pests. The remaining categories included very specific use case demonstrations where the tool was designed for a niche application. Of the reviewed studies, claims of widely applicable frameworks are most common, with those focusing on specific use cases less common.

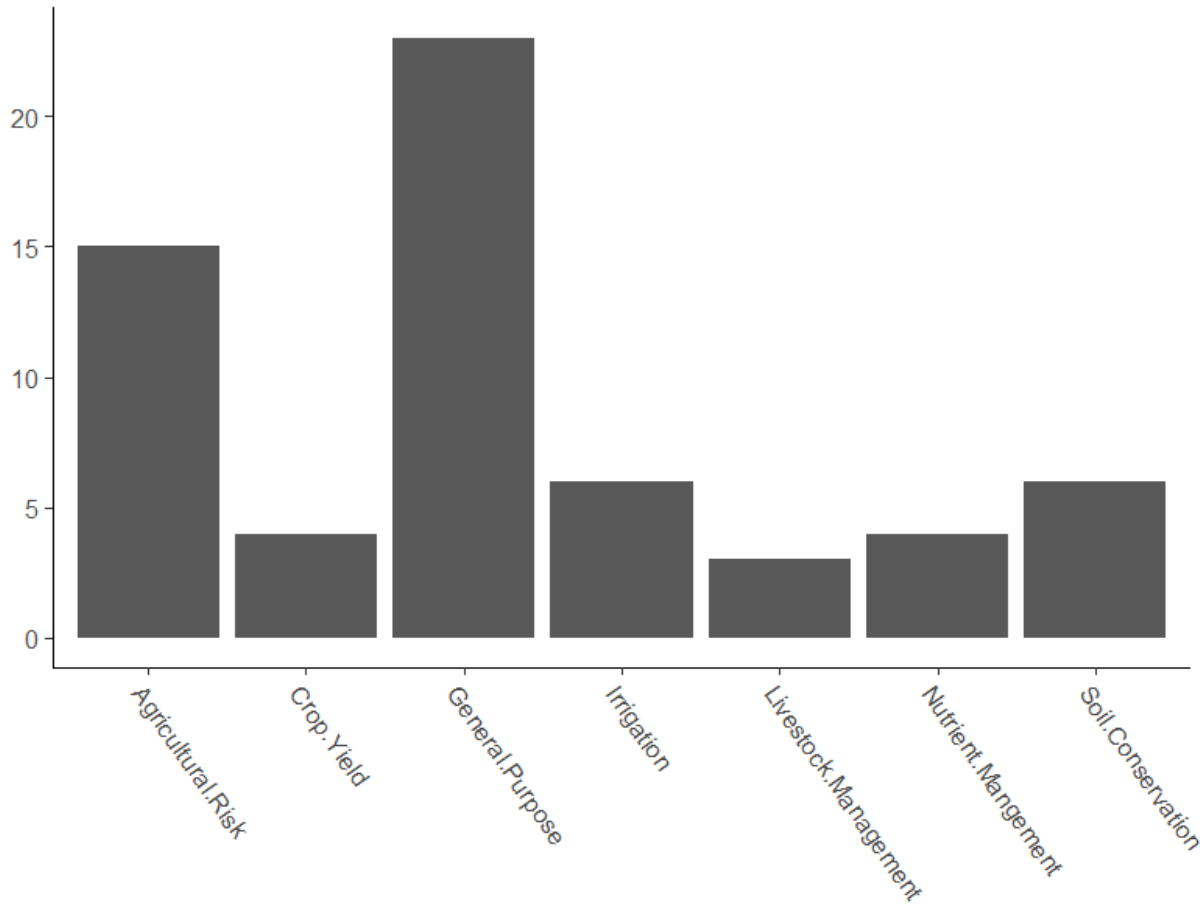


Figure 2.1: Number or Quantity of use cases given for published tools implementing short term modeling and/or monitoring methods to inform strategic agricultural decision support.

The global interest in short term decision support is prevalent in the literature, with application areas ranging worldwide (Lee et al., 2016; Verdin et al., 2016; Vick, 2016; Tapsuwan et al., 2015; Chen et al., 2014; Hunink et al., 2012). Of the studies reviewed, the US had the largest number of studies, with areas around the Mediterranean and the United Kingdom being the second most common location (Figure 2.2). Irrigation was a common use case among

Mediterranean countries, the Middle East, and Australia, agricultural risk in Central America and the Middle East, and soil conservation in the North America.

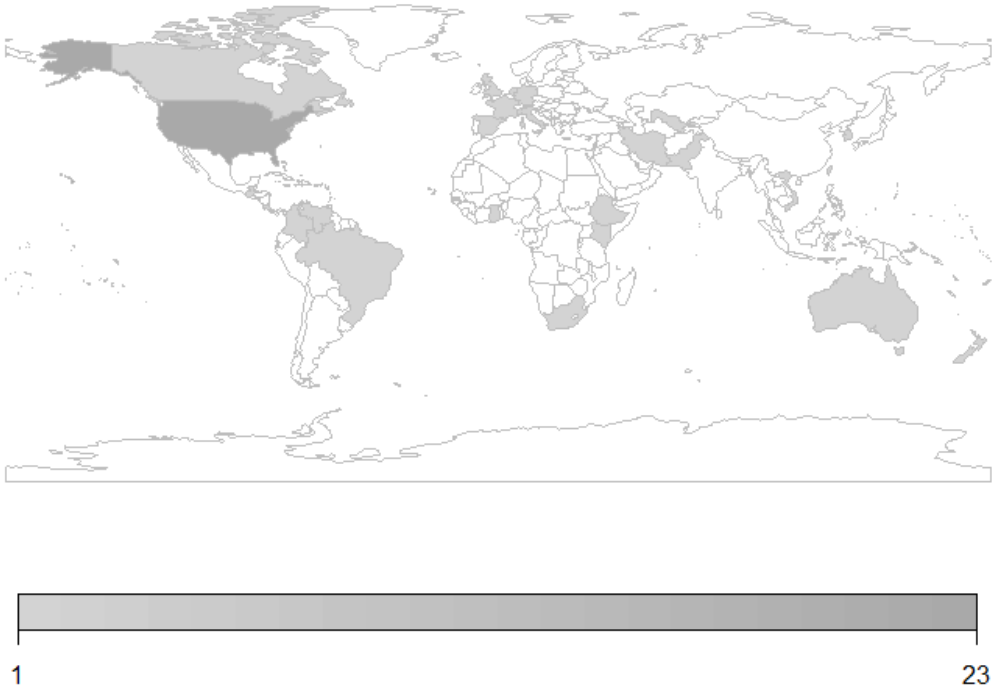


Figure 2.2. Choropleth map of the number of reviewed decision support systems discretized by country of application.

In the US, a trend in watershed management involving "bottom up" management plans has emerged, which focuses on cultivating regular, active stakeholder decisions to increase engagement and communication among decision makers, stakeholders, and the academic community (Fraser et al., 2006; Voinov and Gaddis, 2008; Koontz and Newig, 2014). Short-term forecast tools have the capability to inform regular decision making from stakeholders without

direct input from policy makers concerning each individual decision (Sommerlot et al., 2016). An effectively implemented and employed short-term forecast tool would sit at the nexus of applied research, best-science policy approaches, and active stakeholder engagement, demonstrating a collaborative initiative in practice among academics, stakeholders, and policy makers. This would not be a large conceptual step from some tools and services that already exist; for example, government agencies with a focus on research and data analysis have provided forecast tools to a wide audience of agricultural stakeholders for years including crop yield, economic-, and weather forecasts concerning agricultural products (Frisvold and Murugesan, 2013; Irwin and Good, 2015; USDA, 2017). Products like these are used regularly by agricultural stakeholders and have a modicum of trust amongst the agriculture community (Frisvold and Murugesan, 2013). Implementing niche forecast products designed with the specific needs of water quality in geographic areas improving on currently available spatial and temporal resolution through a useful platform to producers and informed by policy makers is a good fit for academic research and applied science (Grey et al., 2013; Ruckelshaus et al., 2015). Such an endeavor may not be lucrative enough for businesses to venture on, and government agencies may not move fast enough to pilot new, specific decision tools released under their preview (Grey et al., 2013). With the advent and continual improvement of open source statistical and scientific models, communication tools, and data structures, academics can research, design, launch, maintain, and iterate tools for specific decision contexts like short term forecasts. The resulting tools could become trusted and useful tools to producers as well as contribute to water quality improvement goals of policy makers, improving communication between academics, policy makers, and producers (Ruckelshaus et al., 2015). These tools could play a valuable role as part of larger, hybrid programs employing a mix of incentives, regulatory

instruments, and available technology and data working towards the improvement of water quality in the US through filling the critical gap in converting scientific knowledge into action (Grey et al., 2013; Guerry et al., 2015; Ruckelshaus et al., 2015).

2.5 Current Short-Term Forecast Tools Intended for Agriculture Decision Support

Short term forecasts designed as decision support systems (DSS) exist in the literature with goals including predicting streamflow and reservoir status (Choi et al; 2005; Dutta et al., 2012; Abaza et al., 2014; Tsai et al., 2014). Additionally, communication tools have been described that are capable of real time updates and display forecasts from model platforms like those cited above (Kulawiak et al., 2010; Swain et al., 2015; Snow et al, 2016; Tayyebi et al., 2016). A number of short term forecasts have been published specifically for agriculture (Easton et al., 2017). However, these forecast tools are largely not designed for a modern audience using a variety of internet connected devices, they often lack sub-field level resolution, and employ proprietary mapping or server software to deliver forecasts (Sommerlot and Easton, 2017). There is a lack of tools in this niche with necessary technical capabilities related to forecast quality, usability and accessibility to users, and flexibility and cost to researchers developing these decision aids (Sommerlot and Easton, 2017). Many producers consider soil moisture the most important meteorological forecast for planning management actives, and actively use this information for planting, harvesting, and fertilizing among other operations (Frisvold and Murugesan, 2013).

2.6 References

Baig, M.A., Adamowski, J. and Prasher, S., 2015. Development and Application of a Decision Support System for Optimizing Cropping Patterns under Saline Agriculture Conditions in Rechna Doab, Pakistan. In 2015 ASABE Annual International Meeting (p. 1). American Society of Agricultural and Biological Engineers.

Baumgart-Getz, A., Prokopy, L.S. and Floress, K., 2012. Why farmers adopt best management practice in the United States: A meta-analysis of the adoption literature. Journal of environmental management, 96(1), pp.17-25.

Bazzani, G.M., 2005. A decision support for an integrated multi-scale analysis of irrigation: DSIRR. Journal of environmental Management, 77(4), pp.301-314.

Bazzani, G.M., Di Pasquale, S., Gallerani, V. and Viaggi, D., 2004. Irrigated agriculture in Italy and water regulation under the European Union Water Framework Directive. Water resources research, 40(7).

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., Small, J., Pak, E., Masuoka, E., Schmaltz, J. and Hansen, M., 2010. Monitoring global croplands with coarse resolution earth observations: The Global Agriculture Monitoring (GLAM) project. Remote Sensing, 2(6), pp.1589-1609.

Breuer, N.E., Cabrera, V.E., Ingram, K.T., Broad, K. and Hildebrand, P.E., 2008. AgClimate: a case study in participatory decision support system development. Climatic Change, 87(3-4), pp.385-403.

Bruges, M. and Smith, W., 2008. Participatory approaches for sustainable agriculture: A contradiction in terms?. Agriculture and Human Values, 25(1), pp.13-23.

Chapra, S.C., 2008. Surface water-quality modeling. Waveland press.

Chen, C.W., Wei, C.C., Liu, H.J. and Hsu, N.S., 2014. Application of neural networks and optimization model in conjunctive use of surface water and groundwater. Water resources management, 28(10), pp.2813-2832.

Chen, C.W., Weintraub, L., Olmsted, L. and Goldstein, R.A., 2004. Decision framework for sediment control in muddy Creek Watershed. JAWRA Journal of the American Water Resources Association, 40(6), pp.1553-1562.

Clements, A.C., Pfeiffer, D.U. and Martin, V., 2006. Application of knowledge-driven spatial modelling approaches and uncertainty management to a study of Rift Valley fever in Africa. International Journal of Health Geographics, 5(1), p.57.

Cohen, Y., Cohen, A., Hetzroni, A., Alchanatis, V., Broday, D., Gazit, Y. and Timar, D., 2008. Spatial decision support system for Medfly control in citrus. computers and electronics in agriculture, 62(2), pp.107-117.

Crane, T.A., Roncoli, C., Paz, J., Breuer, N., Broad, K., Ingram, K.T. and Hoogenboom, G., 2010. Forecast skill and farmers' skills: Seasonal climate forecasts and agricultural risk management in the southeastern United States. Weather, Climate, and Society, 2(1), pp.44-59.

De Kort, I.A. and Booij, M.J., 2007. Decision making under uncertainty in a decision support system for the Red River. Environmental Modelling & Software, 22(2), pp.128-136.

De la Rosa, D., Mayol, F., Díaz-Pereira, E., Fernandez, M. and de la Rosa Jr, D., 2004. A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection: With

special reference to the Mediterranean region. Environmental Modelling & Software, 19(10), pp.929-942.

De, S. and Bezuglov, A., 2006. Data model for a decision support in comprehensive nutrient management in the United States. Environmental Modelling & Software, 21(6), pp.852-867.

Duan, Y., 2005. A spatial decision support system for economic analysis of sediment control on rangeland watersheds.

USEPA, United States Environmental Protection Agency. 2008. Clean Water Act.

USEPA, Environmental Protection Agency. TMDLs: Water Regulation in the U.S. 2017. Impaired Waters and TMDLs, Environmental Protection Agency. url: www.epa.gov/tmdl/impaired-waters-and-tmdls-water-regulation-us. Accessed 1 Aug. 2017.

Fraser, E.D., Dougill, A.J., Mabee, W.E., Reed, M. and McAlpine, P., 2006. Bottom up and top down: Analysis of participatory processes for sustainability indicator identification as a pathway to community empowerment and sustainable environmental management. Journal of environmental management, 78(2), pp.114-127.

Frisvold, G.B. and Murugesan, A., 2013. Use of weather information for agricultural decision making. Weather, Climate, and Society, 5(1), pp.55-69.

Funk, C. and Verdin, J.P., 2010. Real-time decision support systems: the famine early warning system network. In Satellite rainfall applications for surface hydrology (pp. 295-320). Springer Netherlands.

Garbrecht, J.D. and Schneider, J.M., 2007. Climate forecast and prediction product dissemination for agriculture in the United States. Australian Journal of Agricultural Research, 58(10), pp.966-974.

Giacomelli, A., Giupponi, C. and Paniconi, C., 2001. Agricultural impacts on groundwater: processes, modelling and decision support. In Agricultural Use of Groundwater (pp. 35-75). Springer Netherlands.

Grey, D., Garrick, D., Blackmore, D., Kelman, J., Muller, M. and Sadoff, C., 2013. Water security in one blue planet: twenty-first century policy challenges for science. Phil. Trans. R. Soc. A, 371(2002), p.20120406.

Guerry, A.D., Polasky, S., Lubchenco, J., Chaplin-Kramer, R., Daily, G.C., Griffin, R., Ruckelshaus, M., Bateman, I.J., Duraiappah, A., Elmqvist, T. and Feldman, M.W., 2015. Natural capital and ecosystem services informing decisions: From promise to practice. Proceedings of the National Academy of Sciences, 112(24), pp.7348-7355.

Gutenson, J.L., 2016. Rapid flood damage prediction in short lead-time scenarios. The University of Alabama.

Harmel, R.D., Smith, P.K., Migliaccio, K.W., Chaubey, I., Douglas-Mankin, K.R., Benham, B., Shukla, S., Muñoz-Carpena, R. and Robson, B.J., 2014. Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. Environmental modelling & software, 57, pp.40-51.

Hickey, C.W. and Gibbs, M.M., 2009. Lake sediment phosphorus release management—decision support and risk assessment framework. New Zealand Journal of Marine and Freshwater Research, 43(3), pp.819-856.

homann, R.V., 1972. Systems analysis and water quality management. New York.

Hong, Y., Adler, R.F., Huffman, G.J. and Pierce, H., 2010. Applications of TRMM-based multi-satellite precipitation estimation for global runoff prediction: Prototyping a global flood modeling system. In Satellite Rainfall Applications for Surface Hydrology (pp. 245-265). Springer Netherlands.

Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. Agricultural and forest meteorology, 103(1), pp.137-157.

Hunink, J.E., Droogers, P., Kauffman, S., Mwaniki, B.M. and Bouma, J., 2012. Quantitative simulation tools to analyze up-and downstream interactions of soil and water conservation measures: Supporting policy making in the Green Water Credits program of Kenya. Journal of environmental management, 111, pp.187-194.

Irwin, S. and Good, D., 2015. Long-Term Corn, Soybeans, and Wheat Price Forecasts and the Farm Bill Program Choice. Farmdoc daily, 20(5).

Koontz, T.M. and Newig, J., 2014. From Planning to Implementation: Top-Down and Bottom-Up Approaches for Collaborative Watershed Management. Policy Studies Journal, 42(3), pp.416-442.

Kumar, S., Peters-Lidard, C., Tian, Y., Reichle, R., Geiger, J., Alonge, C., Eylander, J. and Houser, P., 2008. An integrated hydrologic modeling and data assimilation framework. Computer, 41(12).

Lee, C.M., Serrat-Capdevila, A., Iqbal, N., Ashraf, M., Zaitchik, B., Bolten, J., Melton, F. and Doorn, B., 2016. Applying Earth Observations to Water Resources Challenges. In Earth Science Satellite Applications (pp. 147-171). Springer International Publishing.

Leenhardt, D., Trouvat, J.L., Gonzalès, G., Pérarnaud, V., Prats, S. and Bergez, J.E., 2004. Estimating irrigation demand for water management on a regional scale: I. ADEAUMIS, a simulation platform based on bio-decisional modelling and spatial information. Agricultural Water Management, 68(3), pp.207-232.

Levy, J.K., 2001. Computer support for environmental multiple criteria decision analysis under uncertainty.

Lim, K.J., Sagong, M., Engel, B.A., Tang, Z., Choi, J. and Kim, K.S., 2005. GIS-based sediment assessment tool. Catena, 64(1), pp.61-80.

Liu, H.L., Yang, J.Y., Tan, C.S., Drury, C.F., Reynolds, W.D., Zhang, T.Q., Bai, Y.L., Jin, J., He, P. and Hoogenboom, G., 2011. Simulating water content, crop yield and nitrate-N loss under free and controlled tile drainage with subsurface irrigation using the DSSAT model. Agricultural Water Management, 98(6), pp.1105-1111.

Lynch, T., Gregor, S. and Midmore, D., 2000. Intelligent support systems in agriculture: how can we do better?. Australian Journal of Experimental Agriculture, 40(4), pp.609-620.

Malekmohammadi, B., Kerachian, R. and Zahraie, B., 2009. Developing monthly operating rules for a cascade system of reservoirs: application of Bayesian networks. Environmental Modelling & Software, 24(12), pp.1420-1432.

Martius, C., Lamers, J., Wehrheim, P., Schoeller-Schletter, A., Eshchanov, R., Tupitsa, A., Khamzina, A., Akramkhanov, A. and Vlek, P.L., 2004. Developing sustainable land and water management for the Aral Sea Basin through an interdisciplinary approach. In ACIAR PROCEEDINGS (pp. 45-60). ACIAR; 1998.

Mastin, M.C. and Vaccaro, J.J., 2002. Watershed models for decision support in the Yakima River Basin, Washington (No. USGS-OFR-02-404). DEPARTMENT OF THE INTERIOR WASHINGTON DC.

Matthews, D., Hartzell, C., Meyer, S. and Fulp, T., 2000. Environmental Applications of Decision Support Systems for River System Management: Examples from the Colorado and Rio Grande Basins. In Watershed Management and Operations Management 2000 (pp. 1-8).

McColl, C. and Aggett, G., 2007. Land-use forecasting and hydrologic model integration for improved land-use decision support. Journal of Environmental Management, 84(4), pp.494-512.

Mitsch, W.J. and Wang, N., 2000. Large-scale coastal wetland restoration on the Laurentian Great Lakes: determining the potential for water quality improvement. Ecological Engineering, 15(3), pp.267-282.

Murphy, S., 2003. Development and assessment of a spatial decision support system for conservation planning. University of Maine Library. URL: http://digitalcommons.library.umaine.edu/etd/421/. Accessed August 1 2017

Naab, J.B., Singh, P., Boote, K.J., Jones, J.W. and Marfo, K.O., 2004. Using the CROPGRO-peanut model to quantify yield gaps of peanut in the Guinean Savanna Zone of Ghana. Agronomy Journal, 96(5), pp.1231-1242.

National Research Council, 2001. Assessing the TMDL approach to water quality management. National Academies Press.

Neumann, D.W., Zagona, E.A. and Rajagopalan, B., 2006. A decision support system to manage summer stream temperatures. JAWRA Journal of the American Water Resources Association, 42(5), pp.1275-1284.

Newman, S., Lynch, T. and Plummer, A.A., 2000. Success and failure of decision support systems: Learning as we go. Journal of Animal Science, 77(E-Suppl), pp.1-12.

Nielsen, D.C., Unger, P.W. and Miller, P.R., 2005. Efficient water use in dryland cropping systems in the Great Plains. Agronomy Journal, 97(2), pp.364-372.

Nienaber, J.A. and Hahn, G.L., 2007. Livestock production system management responses to thermal challenges. International Journal of Biometeorology, 52(2), pp.149-157.

Optimisation of Agricultural Water Use: A Decision Support System for the Gaza Strip-/von Omar Ouda. Institut für Wasserbau, Universität Stuttgart.- Stuttgart: Ins. Für Wasserbau, 2003 (Mitteilungen/ Institut für Wasserbau, Universität Stuttgart: H. 125) Zugl.: Stuttgart, Uni., Diss., 2003 ISBN 3-933761-28-X

Quinn, N.W., 2000. A Decision Support System for Real-Time Management of Water Quality in the San Joaquin River, California. In Environmental Software Systems (pp. 232-245). Springer, Boston, MA.

Quinn, N.W., Brekke, L.D., Miller, N.L., Heinzer, T., Hidalgo, H. and Dracup, J.A., 2004. Model integration for assessing future hydroclimate impacts on water resources, agricultural production and environmental quality in the San Joaquin Basin, California. Environmental Modelling & Software, 19(3), pp.305-316.

Quinn, N.W., Miller, N.L., Dracup, J.A., Brekke, L. and Grober, L.F., 2001. An integrated modeling system for environmental impact analysis of climate variability and extreme weather events in the San Joaquin Basin, California. Advances in Environmental Research, 5(4), pp.309-317.

Quinn, N.W., Ortega, R., Rahilly, P.J. and Royer, C.W., 2010. Use of environmental sensors and sensor networks to develop water and salinity budgets for seasonal wetland real-time water quality management. Environmental Modelling & Software, 25(9), pp.1045-1058.

Regonda, S.K., Rajagopalan, B., Clark, M. and Zagona, E., 2006. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resources Research, 42(9).

Riquelme, F.J.M. and Ramos, A.B., 2005. Land and water use management in vine growing by using geographic information systems in Castilla-La Mancha, Spain. Agricultural water management, 77(1), pp.82-95.

Ruckelshaus, M., McKenzie, E., Tallis, H., Guerry, A., Daily, G., Kareiva, P., Polasky, S., Ricketts, T., Bhagabati, N., Wood, S.A. and Bernhardt, J., 2015. Notes from the field: lessons learned from using ecosystem service approaches to inform real-world decisions. Ecological Economics, 115, pp.11-21.

Sankarasubramanian, A., Lall, U., Souza Filho, F.A. and Sharma, A., 2009. Improved water allocation utilizing probabilistic climate forecasts: Short-term water contracts in a risk management framework. Water Resources Research, 45(11).

Saseendran, S.A., Nielsen, D.C., Ahuja, L.R., Ma, L. and Lyon, D.J., 2013. Simulated yield and profitability of five potential crops for intensifying the dryland wheat-fallow production system. Agricultural Water Management, 116, pp.175-192.

Shenk, G.W. and Linker, L.C., 2013. Development and application of the 2010 Chesapeake Bay watershed total maximum daily load model. JAWRA Journal of the American Water Resources Association, 49(5), pp.1042-1056.

Sommerlot, A.R, and Easton, Z.M. Development of a free and open source web based interface for distributed short-term hydrologic forecasts. Water. 2017, 9, 604; doi:10.3390/w9080604.

Stefanski, R. and Sivakumar, M.V.K., 2009. Impacts of sand and dust storms on agriculture and potential agricultural applications of a SDSWS. In IOP Conference Series: Earth and Environmental Science (Vol. 7, No. 1, p. 012016). IOP Publishing.

Stone, R.C. and Meinke, H., 2006. Weather, climate, and farmers: an overview. Meteorological Applications, 13(S1), pp.7-20.

Tang, Z., Engel, B.A., Choi, J., Sullivan, K., Sharif, M. and Lim, K.J., 2004. A Web-based DSS for erosion control structure planning. Applied Engineering in Agriculture, 20(5), p.707.

Tapsuwan, S., Hunink, J., Alcon, F., Mertens-Palomares, A.N. and Baille, A., 2015. Assessing the Design of A Model-Based Irrigation Advisory Bulletin: The Importance of End-User Participation. Irrigation and Drainage, 64(2), pp.228-240.

Thorp, K.R., Jaynes, D.B. and Malone, R.W., 2008. Simulating the long-term performance of drainage water management across the Midwestern United States. Transactions of the ASABE, 51(3), pp.961-976.

Trepel, M. and Kluge, W., 2004. WETTRANS: a flow-path-oriented decision-support system for the assessment of water and nitrogen exchange in riparian peatlands. Hydrological Processes, 18(2), pp.357-371.

USDA 2008a. Farm Bill 2008 At a Glance: Agricultural Management Assistance. *In:* SERVICE, D. O. A. N. R. C. (ed.).

USDA 2008b. Farm Bill 2008 At a Glance: Agricultural Water Enhancement Program. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008c. Farm Bill 2008 At a Glance: Chesapeake Bay Watershed. *In:* AGRICULTURE, U. S. D. O. (ed.).

USDA 2008d. Farm Bill 2008 At a Glance: Conservation Innovation Grants. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008e. Farm Bill 2008 At a Glance: Conservation of Private Grazing Land Program. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008f. Farm Bill 2008 At a Glance: Cooperative Conservation Partnership Initiative. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008g. Farm Bill 2008 At a Glance: Farm and Ranch Lands Protection Program. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008h. Farm Bill 2008 At a Glance: Grassland Reserve Program. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA 2008i. Farm Bill 2008 At a Glance: Wetlands Reserve Program. *In:* SERVICE, U. S. D. O. A. N. R. C. (ed.).

USDA. 2017.World Agricultural Supply and Demand Estimates. ISSN: 1554-9089. https://www.usda.gov/oce/commodity/wasde/latest.pdf

Uusitalo, L., Lehikoinen, A., Helle, I. and Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software, 63, pp.24-31.

Valentin, L., Bernardo, D.J. and Kastens, T.L., 2004. Testing the empirical relationship between best management practice adoption and farm profitability. Review of Agricultural Economics, 26(4), pp.489-504.

Verdin, A., Funk, C., Rajagopalan, B. and Kleiber, W., 2016. Kriging and local polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems. IEEE Transactions on Geoscience and Remote Sensing, 54(5), pp.2552-2562.

Vick, R.L., 2016. Smart irrigation systems for crop production in the humid climate of the Southeastern United States. PhD, North Carolina State University, Raleigh, North Carolina.

Voinov, A. and Gaddis, E.J.B., 2008. Lessons for successful participatory watershed modeling: a perspective from modeling practitioners. Ecological Modelling, 216(2), pp.197-207.

Whateley, S., Walker, J.D. and Brown, C., 2015. A web-based screening model for climate risk to water supply systems in the northeastern United States. Environmental Modelling & Software, 73, pp.64-75.

Wiederhold, G., 2000. Information systems that really support decision-making. Journal of Intelligent Information Systems, 14(2-3), pp.85-94.

Wilhite, D.A., 2011. Breaking the hydro-illogical cycle: progress or status quo for drought management in the United States. European Water, 34, pp.5-18.

Wilkinson, M.E., Quinn, P.F. and Hewett, C.J., 2013. The Floods and Agriculture Risk Matrix: a decision support tool for effectively communicating flood risk from farmed landscapes. International Journal of River Basin Management, 11(3), pp.237-252.

Zanuttigh, B., Simcic, D., Bagli, S., Bozzeda, F., Pietrantoni, L., Zagonari, F., Hoggart, S. and Nicholls, R.J., 2014. THESEUS decision support system for coastal risk management. Coastal Engineering, 87, pp.218-239.

Zhan, Y. and Zhang, M., 2012. PURE: A web-based decision support system to evaluate pesticide environmental risk for sustainable pest management practices in California. Ecotoxicology and Environmental Safety, 82, pp.104-113.

Ziervogel, G., Johnston, P., Matthew, M. and Mukheibir, P., 2010. Using climate information for supporting climate change adaptation in water resource management in South Africa. Climatic Change, 103(3), pp.537-554.

CHAPTER 3. Coupling Short-Term Global Forecast System Weather Data with a Variable Source Area Hydrologic Model*

***Citation:** Sommerlot. A.R., M.B. Wagena, D.R. Fuka, and Z.M. Easton. 2016. Coupling the short-term Global Forecast System weather data with a variable source area hydrologic model. Environmental Modelling & Software. http://dx.doi.org/10.1016/j.envsoft.2016.09.0081364-8152.

Keywords: SWAT, Short-term Forecast, Distributed Hydrology, Non-point Source Pollution, Runoff Probability, Soil Moisture, High Performance Computing

3.0 Abstract

Few current modeling tools are designed to predict short-term, high-risk runoff from critical source areas (CSAs) in watersheds. This study couples the Soil and Water Assessment Tool-Variable Source Area (SWAT-VSA) model with the Climate Forecast System Reanalysis (CFSR) model and the Global Forecast System (GFS) model short-term weather forecast, to develop a CSA prediction tool designed to assist agaricultural producers, landowners, and planners in identifying high-risk areas generating storm runoff and pollution. Short-term predictions for stream flow, runoff probability, and soil moisture levels were estimated in the South Fork of the Shenandoah River Watershed, Virginia, USA. Daily volumetric flow forecasts were satisfactory several days into the future, and distributed model predictions accurately captured sub-field scale CSAs. The model has the potential to provide valuable forecasts that can be used to improve effectiveness of agricultural management practices and reduce the risk of non-point source pollution.

3.1 Introduction

Non-point source (NPS) pollution is a leading cause of water quality impairment in the United States and a major component of this pollution is caused by runoff from agricultural lands (Shortle et al., 2012). During the last several decades numerous environmental standards and management practices (e.g., NRCS 590 standard, P-Index) have been developed in an attempt to reduce NPS pollution but these practices are often variable in their effectiveness (Puckett, 1995). This is, to some extent, because their effectiveness is based on estimates made by models that do not consider the spatial variability of agricultural landscapes or the temporal dynamics of nutrient and pollutant transport processes (Dahlke et al., 2012; Easton et al., 2008). New tools need to be able to capture spatiotemporal variability of hydrological and biogeochemical processes in order to assist watershed managers, producers, and landowners to better plan agricultural management practices and reduce NPS pollution.

Surface runoff is a critical pathway of NPS pollutant transport to streams and often it is small and variable sub-field-level areas of land that are the primary source of surface runoff and water quality degradation (Walter and Walter, 1999). These so called variable source areas (VSAs) as defined by Hewlett and Hibbert (1967) often occur in areas of watersheds that have a shallow top soil layer with high-infiltration rates overlaying a dense subsoil layer which restricts percolation out of the vadose zone (Dunne and Leopold, 1978; Easton et al., 2008; Walter and Walter, 1999). When these features are located in convergent areas of the landscape (e.g., large upslope contributing areas) the soil must be able to conduct large volumes of water through the profile rapidly, or else it saturates and becomes a runoff source area. These factors result in some areas of the landscape having low soil water storage capacity and are thus easily saturated during precipitation events. The location, size, and duration of saturation of these VSAs depend on

variable spatial and temporal factors including soil depth and conductivity, landscape position, antecedent soil moisture and climate (Needelman et al., 2004). Consequently, these VSAs are difficult to model using traditional watershed models (e.g., SWAT, GWLF, HSPF), making the prioritization of landscape management practices challenging (Brooks et al., 2015; Dahlke et al., 2012; Dunne and Leopold, 1978; Marjerison et al., 2011).

Walter et al. (2000) introduced the concept of hydrologically sensitive areas (HSAs) to refer to parts of the landscape where VSAs form and a pollutant source exists (e.g., manure, fertilizer). Knowledge of when and where these HSAs form could be used to prioritize landscape management strategies. For example, a nutrient management plan, often written to direct manure spreading on a seasonal basis, may suggest an operations schedule that does not take into account the location of HSAs within the target fields. Information about where these HSAs exist could be used to modify the nutrient management plan and avoid pollution causing activities during sensitive times or in sensitive areas (Hanrahan et al., 2004; Smith et al., 2007; Vadas et al., 2011).

A major challenge in modeling VSAs is their temporal variability; VSAs can evolve over a single precipitation event, as well as shift due to seasonal fluctuations in precipitation and evapotranspiration. Attempts to model this phenomenon include the Soil and Water Assessment Tool-VSA adaption (SWAT-VSA), a modification of the popular SWAT model that uses topographically derived parameters designed to better quantify the distribution of surface runoff and soil moisture from VSAs (Easton et al., 2008, 2011, 2010; White et al., 2011; Woodbury et al., 2014). The base SWAT model is a widely used watershed scale model with a variety of successful applications including sub field scale hydrology outputs and sub-daily temporal resolution (Collick et al., 2014; Douglas-Mankin et al., 2010; Easton et al., 2008; Gassman et al.,

2014; Sommerlot et al., 2013). SWAT employs the curve number (CN) equation to predict runoff and the way the CN is applied in SWAT implicitly assumes an infiltration-excess (or Hortonian (Horton, 1933) response to rainfall. However, in humid, well-vegetated regions, especially those with permeable soils underlain by a shallow restricting layer, this formulation fails to capture the observed runoff generating processes (Beven, 2001; Dunne and Black, 1970; Dunne and Leopold, 1978; Needelman et al., 2004). For instance, many soils in the Mid-Atlantic region of the USA have a substantially higher infiltration rate than the total precipitation depth of all but the most intense, convective type storms; thus runoff is generated from areas of the landscape that saturate and can no longer store more moisture (Easton et al., 2008). For this reason, the standard CN method may be insufficient for providing detailed information about field scale hydrologic responses and the SWAT-VSA model is more appropriate as it is designed specifically to address areas of short-term high-risk runoff generation and has shown promising results in the estimation of distributed hydrologic processes (Easton et al., 2008). In addition, SWAT-VSA has been shown to provide more accurate predictions of soil moisture and runoff generation than the unmodified SWAT model in watersheds with similar physical characteristics and climate to the study area (Easton et al., 2008; Woodbury et al., 2014).

In order to use SWAT-VSA in forecast capacity it needs to be coupled with some type of short-term weather model. Short range (1-16 day) weather forecasts assimilate remote and ground observations into an atmospheric model analysis time step to initialize physically based solutions of future surface weather conditions (GFS, 2015; NCEP, 2003). These predictions of future surface weather conditions can then be directly used to force the SWAT-VSA model to provide forecasts of hydrological processes. The National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS, 2015; NCEP, 2003) formerly known as the Medium

Range Forecast model (MRF), initially developed by Sela (1982, 1979), provides forecasted surface predictions out 16 days for every location on the planet, four times a day, every single day of the year. This reliability and global coverage makes it an ideal candidate for using as the basis of management planning tool requiring reliable data sources. The GFS is also complemented by the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) that provides long histories with similar physical solutions, helping to mitigate solution based biases that exist between forecasting systems (Hamill et al., 2008). These datasets, when coupled, provide contiguous estimates of weather variables, including precipitation and temperature, for any terrestrial location worldwide from 1979 to 16 days into the future. Both the GFS and the CFSR have been separately evaluated as drivers for the SWAT model with Fuka et al. (2013) demonstrating that CFSR precipitation and temperature inputs resulted in simulated daily discharge that was at least as accurate as those based on records from weather stations located more than 10km away from the geographic center of watersheds in California, New Mexico and New York, USA.

This study demonstrates coupling SWAT-VSA with global weather forecast (GFS) products to provide short-term hydrologic forecasts capable of guiding landscape management process in real-time. This tool can be used for planning or directing operational, short-term field management. Short-term hydrologic forecast frameworks have been described in recent studies using rainfall-runoff and time-series models, but do not specifically address predictions intended to inform field-level decision making (Abaza et al., 2014; Dutta et al., 2012; Tsai et al., 2014). The SWAT-VSA model provides two primary forecasts: hydrologic in the form of daily volumetric stream flow estimates and distributed surface hydrology in the form of sub-field level predictions of surface runoff risk probability and soil moisture.

3.2 Materials and Methods

3.2.1 Overview: Model development followed a four-step process. First, the base SWAT-VSA model was developed for the South Fork of the Shenandoah River Watershed in Virginia (Fig. 3.1) using historic meteorological data. Second, the model was calibrated and corroborated against United States Geologic Survey (USGS) streamflow gage data using the base model output forced by the historic meteorological data available from station records included in the quality-controlled Global Historical Climatology Network (Menne et al., 2012). Third, the corroborated model was coupled with the GFS data by collecting historical archived GFS forecasts and applying these as weather forcings to the SWAT-VSA model in a hindcast procedure exactly recreating past outputs of model coupling. Finally, the GFS forced model outputs were used to predict watershed level discharge and distributed hydrologic response in the watershed (note that the model was not re-calibrated using the GFS data). The relationship between the base input data and the resulting derived data set are shown in Fig. 3.2.

3.2.2 Study Area: The SWAT-VSA model was developed and tested in the South Fork of the Shenandoah river watershed (HUC 02070005), with the outlet located at the USGS gaging station 01628500. The watershed has an area of roughly 2600 km2 located in the Shenandoah Valley in North-Central Virginia (Fig. 3.1). Land cover in the Shenandoah watershed is 50% forest, and almost entirely located in the steep mountain headwaters. Agricultural lands occupy 38% of the watershed and are located in the valley, urban areas occupy 11%, and the remaining 1% of the basin is coved by water (Homer et al., 2015). Soils are silt loams and silty clay loams, with slopes up to 60%, elevations of 310 to 1336 m and located in the humid continental climate with a yearly average rainfall 1057 mm (Mohamoud, 2004).

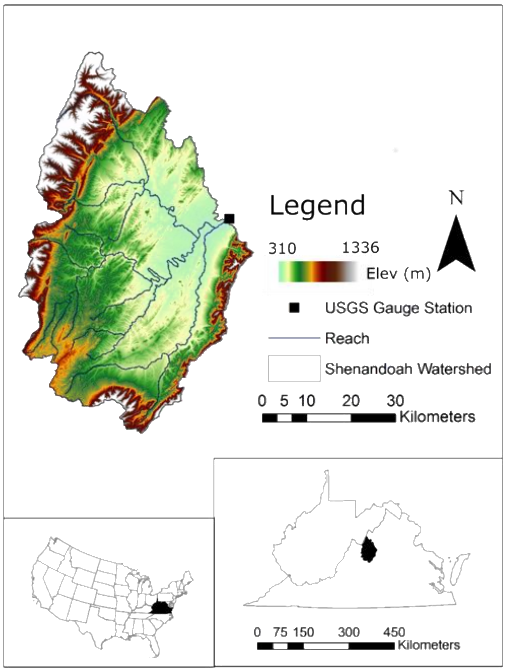


Figure 3.1. Study Area: South Fork of the Shenandoah River Watershed in north central VA with elevation features.

3.2.3 SWAT Model Description: The SWAT model, developed by USDA-ARS, is a process based, semi-distributed watershed model meant to provide predictions of the impact of agricultural landuse on water quantity and quality (Arnold et al., 2012, 1998; Gassman et al., 2007). SWAT uses weather, elevation, soil, land cover, and land management data to simulate surface and subsurface hydrology and various chemical and sediment fluxes. Required spatial data include information for soils, landuse/management, and elevation with the associated tabular look up tables that contain model required parameters. Simulations require meteorological input data including precipitation, temperature, relative humidity, wind speed, and solar radiation.

SWAT lumps unique combinations of soil type and landuse (and slope if desired) into hydrologic response units (HRUs) during the initialization process. While this lumping process reduces computational complexity, it effectively ignores the underlying spatial distribution of the input data. Thus, all HRUs containing the same soil type, landuse and slope class have identical properties irrespective of where they are located. In VSA watersheds, runoff-generating HRUs are a controlled by their location within a subbasin: VSAs are likely to occur in portions of the landscape with shallow soils underlain by some sort of restrictive layer, large upslope contributing areas, flattening slopes, or any combination of the three (Easton et al., 2008). To account for the formation and contribution of VSAs, SWAT-VSA incorporates a hybrid topographic wetness index-soil layer, which is then used in place of the standard soil layer in HRU generation (Easton et al., 2008). The topographic wetness index (TI) is a unitless metric categorizing the saturation potential of areas in a watershed and ranges from 1, the least likely to saturate, to n, the most prone to saturation (Easton et al., 2008). Maximum TI class is defined by the arbitrary decision of number of classes used to categorize TI from a continuous variable to discrete classes; unless known reasons exist for a specific number of classes, 10 is recommended (Easton et al,, 2008). Runoff and soil moisture distributions are then explicitly linked to the topographic wetness index. While the wetness classes can be used in HRU delineation instead of a soil map, SWAT still requires specific soil properties associated with the soils map (e.g., SSURGO). Thus, in SWAT-VSA soil properties are weighted and averaged for each TI class. This practice does not drastically affect model results because soil genesis is, at least partially, driven by topography (Easton et al., 2008; Page et al., 2005; Sharma et al., 2006; Thompson et al., 2006).

SWAT-VSA calculates runoff from each HRU defined by landuse and topographic index with a saturation-excess modified CN resulting in an approximation of the runoff equation described by (Schneiderman et al., 2007):

$$q_i = \frac{P_e^2}{P_e - \sigma_e} \quad (3.1)$$

where q_i is the surface runoff for any given HRU, P_e is the amount of rainfall after runoff begins (mm), and σ_e (mm) is the maximum effective soil moisture storage defined uniquely for each TI class (Easton et al., 2008).

3.2.4 Model Initialization: The SWAT-VSA model was initialized for the South Fork of the Shenandoah River watershed using ArcSWAT2012 and an ArcSWAT plugin, TopoSWAT (available from http://ww2.bse.vt.edu/eastonlab/), developed to automatically generate the TI-Soil hybrid layer used for SWAT-VSA initialization. TopoSWAT provides the necessary soil data to calculate TI via linkage to the United Nations Food and Agriculture Organization (FAO) soils database (IUSS, 2007). Other spatial inputs include landuse from the National Land Cover database (Homer et al., 2015), and a combination of 3 m and 10 m resolution DEMs from the United States Geologic Survey (GHCN) National Elevation Dataset (NED) resampled to 3 m covering the study area. No thresholds were applied during HRU definition in order to generate a full-HRU distribution. The final initialization resulted in five subbasins and 859 HRUs over the study domain.

Two distinct sets of meteorological data were used in this study. The first was historical data used for model calibration and corroboration. The second, referred to as archived forecasts, was used to validate the framework's capability to forecast without waiting months to collect data. The meteorological data set used in calibration and corroboration of the base model

consisted of redimensioned CFSR data bias corrected with data from GHCN stations. The CFSR data set provides the required meteorological parameters to force the base model: precipitation (mm), minimum and maximum temperature (deg C), percent relative humidity, wind speed (m/s), and solar radiation (MJ/m^2) at a daily time step (Fuka et al., 2013). CFSR was chosen as the base data set to make this method as widely applicable as possible, as it provides all necessary parameters on a global grid and contains no missing data. Precipitation data from five GHCN rain gauges located within the study area were used to bias correct the CFSR data as the initial analysis revealed a bias in the CFSR data. Gauge stations used were UCS00448941, USC00442208, USC00448062, USC0044322 and USC00445096, accessed through the National Centers for Environmental Information Climate Data Online search tool.

Meteorological data from both the CFSR and GHCN rain gauge data sets were interpolated to the SWAT-VSA model subbasin centroids using inverse distance weighting (IDW) squared. Bias correction of CFSR precipitation was performed by modifying functions from the QMAP R Package using methods suggested by Looper and Vieux (2012) and Gudmundsson et al. (2012). The bias correction was applied by separating each precipitation data set into unique months, and creating a Cumulative Distribution Function (CDF) for each month. A linear regression method was employed to map all future CDFs onto past CDFs by minimizing the sum of squared errors between the data sets. This method is modified from a similar process of CDF mapping described in Girvetz et al. (2013), following their suggestion to produce monthly CDF functions for daily weather data.

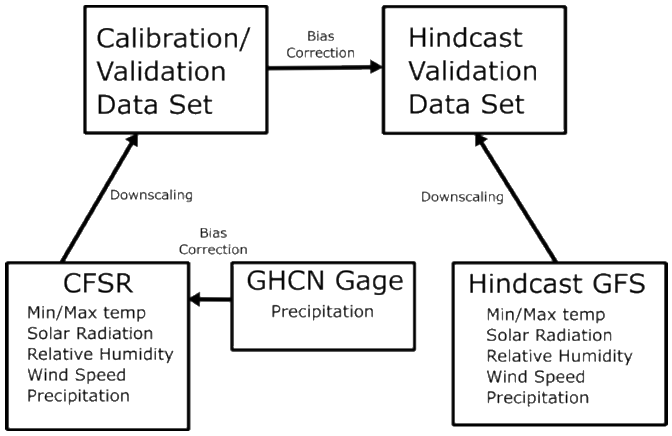


Figure 3.2. Relationship between the input meteorological data and the resulting data sets

3.2.5 Archived GFS Forecasts: The streamflow forecast corroboration was performed by collecting archived GFS forecasts through a hindcast procedure rather than waiting and collecting the results of live forecasts. Uncorrected "raw" GFS forecasts up to eight days in advance were available through NOAA in GRIB2 format through NOMADS (Rutledge et al., 2006). Each file in the data set contained forecast data at a 0.5 degree resolution. (GFS, 2015). An R script was written to catalog and organize the location of each GRIB2 file in the available data set. The result was a catalog of over 23,000 remote file locations of data storing world-wide uncorrected GFS forecasts for 1 through 8 days in the future from the GFS model run. Dates ranged from May 2013 through March 2015 at the time the catalog script was run and included forecasts updates at 6-hr intervals. A second R script optimized for parallel implementation on NCARs Yellowstone Geyser super computer cluster was used to download, convert, and summarize the uncorrected GFS data set into a daily time series of weather data required for SWAT-VSA model inputs. This script applied a bounding box around Virginia and parts of West

Virginia to reduce the computational time necessary to produce the daily time series. In addition, parallelization of this portion of the data collection and formatting reduced computational time from over 40 days to about 6 hrs. Table 3.1 summarizes the methods used to distill the full six terabyte archived GFS forecasts into daily time series files with all necessary SWAT weather input variables. In total, eight daily time series files were created from May 2013 through March 2015, one for each forecast day.

Table 3.1. Global Forecast System (GFS) model parameters used in the SWAT-VSA model.

Weather Parameter	**Provided Time Step**	**Required Time Step**	**Aggregation Method**	**Provided Units**	**Required Units**
Precipitation	4 hr Sub-daily	Daily	Summation	mm H_20	mm H_20
Minimum Temperature	4 hr Sub-daily	Daily	Minimum	Kelvin	Degrees Celsius
Maximum Temperature	4 hr Sub-daily	Daily	Maximum	Kelvin	Degrees Celsius
Relative Humidity	4 hr Sub-daily	Daily	Average	Percentage	Percentage

Wind Speed	4 hr Sub-daily	Daily	Vector Addition	u and v direction m/s	Directionless m/s
Solar Radiation	4 hr Sub-daily	Daily	Average	W/m^2	MJ/m^2

The resulting daily time-series of weather data were not contiguous due to missing files in the GFS data set and server errors during download; 41 days of data were missing in each forecast day of the data set. An imputation process was performed on the data employing time series methods from the forecast and the fGarch R package (Hyndman and Khandakar, 2008; Wuertz et al., 2009). Missing data from the original set were replaced with residual-corrected corresponding values from Autoregressive Integrated Moving Average (ARIMA) time series models. In order to use the weather data directly in the SWAT model, the data were interpolated to SWAT model subbasin centroids using the same IDW squared procedure employed for CFSR in order to keep consistency with the calibration and corroboration data sets. These methods were repeated for each of the eight weather forecast data sets.

The full forecast data sets were further preprocessed using a dynamic downscaling procedure including IDW squared and bias correction. All parameters of the archived forecast data were bias corrected using the same procedure as described for the CFSR precipitation data used for calibration. The redimensioned CFSR data set was available through August 2014, limiting the final GFS archived forecasts to the same end date for the bias correction procedure. The GFS forecasts were bias corrected using the interpolated rain gauge data for precipitation and CFSR data for the remaining parameters. Once gauge and CFSR data are collected and interpolated, the bias correction of GFS data is repeatable for actual forecasts in real-time

following this method. The GFS archived forecasts were split into two equal parts, the first representing past values and the second representing future values. The separation allowed the GFS data to be bias corrected as if the portion representing future values was actually composed of future forecasted values, while measured flow data were available. The relationship between this data set, the calibration/corroboration data set, and the input meteorological data is illustrated in Fig. 3.2.

3.2.6 Sensitivity Analysis: Prior to SWAT-VSA model calibration, a sensitivity analysis was performed using the SWATmodel (Fuka et al., 2014) and hydroPSO (Zambrano-Bigiarini and Rojas, 2013) packages in R. SWATmodel provides optimized read and write access to model files while using the lhoat function to preform sensitivity analysis from hydroPSO. The sensitivity analysis was applied on model runs from January 1 2004 through December 31 2014. hydroPSO runs a Latin-Hypercube one-at-a-time process on the selected parameters and outputs parameter sensitivity rankings (Zambrano-Bigiarini and Rojas, 2013). Flow parameters were selected based on analysis of preliminary runs of the initialized model. In addition, the preliminary analysis suggested some classically sensitive parameters be applied independently in each subbasin. The curve number (CN), channel conductivity, and Manning's n values were separated to allow unique values at the subbasin scale. Base flow and soil parameters were the most sensitive parameters (Table 3.2). Snow melt/accumulation parameters were also sensitive including snow pack lag factor, snowmelt temperature, and snowfall temperature. Surface runoff lag coefficient and ground water delay factors also ranked in the top half of the parameter list (Table 3.2). The CN, channel conductivity, and Manning's N did not all rank as high as expected since each of these parameters are unique to a subbaisn, their sensitivity was lower than if applied to the entire basin; however, the increase in spatial resolution of these parameters was

considered critical to adequate model predictions at the field scale. Thus, all parameters applied at the subbasin scale were included in the calibration. Twenty-five parameters were selected from the sensitivity analysis for model calibration (Table 3.2). The CN, channel conductivity, and Manning's N values for all five subbasins and the remaining 10 highest ranking parameters were selected.

Table 3.2. Model parameters selected for sensitivity analysis ranking, method of calibration, range, ranking and final calibrated values.

				Range		**Sensitivity**	**Calibrated**
Parameter	**Subbasin**	**File**	**Method**	**Min**	**Max**	**Ranking**	**Value**
Alpha_BF	all	*.gw	replace original with	0.5	1.5	1	0.59
Sol_AWC	all	*.sol	multiply original by	0.5	1.5	2	0.89
ESCO	all	*.hru	replace original with	0.1	1	3	0.91
Sol Depth	all	*.sol	multiply original by	0.5	1.5	4	0.56
SFTMP	all	.basin	replace original with	-5	5	5	-1.62
CN2	4	*.mgt	multiply original by	0.7	1.3	6	0.63
SMTMP	all	.basin	replace original	-5	5	7	-1.51

			with				
TIMP	all	.basin	replace original with	0	4	8	0.07
CN2	2	*.mgt	multiply original by	0.7	1.3	9	0.63
CN2	5	*.mgt	multiply original by	0.7	1.3	10	0.60
SURLAG	all	.basin	replace original with	0	15	11	14
CH_K2	3	*.rte	replace original with	0	200	12	0.936
CH_N2	3	*.rte	replace original with	0.016	0.15	13	0.087
CH_N2	2	*.rte	replace original with	0.016	0.15	14	0.090
SMFMN	all	.basin	replace original with	-5	5	15	1.09
CH_K2	5	*.rte	replace original with	0	200	16	0.051
CH_K2	2	*.rte	replace original with	0	200	17	0.082
GW_delay	all	*.gw	replace original with	0.5	2000	18	1592
CH_K2	1	*.rte	replace original	0	200	19	127

			with				
CH_N2	1	*.rte	replace original with	0.016	0.15	20	0.069
CH_K2	4	*.rte	replace original with	0	200	21	154.4
Ksat	all	*.sol	multiply original by	0.5	1.5	22	Not Calibrated
CN2	3	*.mgt	multiply original by	0.7	1.3	23	0.74
CN2	1	*.mgt	multiply original by	0.7	1.3	24	0.81
GW_REVAP	all	*.gw	replace original with	0	0.2	25	Not Calibrated
RCHRG_DP	all	*.gw	replace original with	0	1	26	Not Calibrated
CH_N2	5	*.rte	replace original with	0.016	0.15	27	0.087
GWQMN	all	*.gw	replace original with	0	500	28	471
SMFMX	all	.basin	replace original with	-5	5	29	Not Calibrated
CH_N2	4	*.rte	replace original with	0.016	0.15	30	0.090
REVAPM	all	*.gw	replace original	0	500	31	Not Calibrated

N			with				
ESCO	all	.basin	replace original with	0.1	1	32	Not Calibrated
EPCO	all	.basin	replace original with	0.1	1	33	Not Calibrated

3.2.7 Model Calibration: The initialized model was calibrated to the USGS gage flow data on a daily time step from January 1 2004 through December 31 2012, on the Yellowstone Geyser super computer cluster, a 16 node, 16,000 GB ram platform located at the National Center for Atmospheric Research Wyoming Supercomputing Center. A similar combination of the SWATmodel (Fuka et al., 2014) and hydroPSO (Zambrano-Bigiarini and Rojas, 2013) R packages as described in the sensitivity analysis was used to run the model and perform necessary statistical analysis on model results. Model calibration was run on one node of Geyser, using 40 cores and 1000 GB of memory in parallel. The SWATmodel package was used to run SWAT2012 on a Linux operating system and to provide read and write functionalities for SWAT input and output files. The read and write functions were included in an input/output wrapper that allowed the framework to utilize the output parameters of the calibration and sensitivity analysis procedures. The hydroPSO package was used as the calibration engine to maximize the fitness of the model and to apply parallel computing techniques to reduce the computation time. All R functionalities were written into a script portable to the parallel Geyser platform. The resulting parameter values following the calibration are shown in the Table 3.2.

3.2.8 SWAT-VSA Hindcast: A hydrologic hindcast was performed to investigate the model predictive performance in time-series volumetric flow and spatial soil moisture predictions without waiting for the GFS results in real-time. The meteorological data sets prepared for hindcast validation and calibration/corroboration were used to initialize a unique SWAT-VSA model run for each day in the hindcast period, recreating the live forecast procedure the model is intended to perform. Each unique model run was forced with 12 years of meteorological data, the last eight days of the run defined by the corresponding day and forecast level (i.e., 24 h, 48 h, 72 h, etc.) from the meteorological data prepared for hindcasting. The remaining input was defined by the meteorological data prepared for calibration/corroboration, just as it would be in a live forecast. The last eight days of each SWAT-VSA volumetric flow time series output, representing the hydrologic forecasts, was saved and separated to eight unique time series by forecast level. Thus, time series for all eight hindcast days could be directly compared to measured volumetric flow at the watershed outlet. In total, 462 consecutive days of hindcast runs were performed for each forecast level, from May 27, 2013, to August 31, 2014.

3.2.9 Distributed Forecast Output: The spatially distributed forecast was a binary classification procedure defining areas as saturated or unsaturated. During model initialization, a raster containing spatial references to each TI class in the SWAT-VSA model was created. The raster was then populated with binary hindcast model outputs from a single day in the spatial hindcast corroboration. To calculate the binary hindcast the HRU soil moisture level predicted by the model (in mm over the entire soil profile) was divided by the soil depth of the HRU to return the percent saturation metric. A threshold of 80% of the soil profile as saturated was used to define each HRU output as saturated (greater than or equal to 80%) or not saturated (less than

80%) (de Alwis et al., 2007). Final classifications mapped to TI class distribution were calculated through a majority-vote aggregation procedure, an ensemble of the spatial output of SWAT-VSA aggregated to the TI class distribution. Multiple HRUs correspond to each TI class in the distribution raster, thus to calculate the final saturation classification all HRUs corresponding to a unique TI class were polled and the majority vote (either saturated or not saturated) was adopted as the final label for that TI class. The resulting output was a 3 m resolution raster displaying a binary saturation classification for a single day spatial forecast. This process was repeated for two unique days, December 19 2015, and December 26 2015. Data defining saturated and unsaturated areas were collected on both of these days in various locations throughout the Shenandoah watershed for corroboration. Corresponding spatial model outputs were defined by calculating the ratio of raster cells labeled as saturated to total number of cells within each of these areas. This ratio was considered the predicted probability of the entire area to be saturated. Spatial data for a total of 49 sub field areas were collected over the two days, 24 for December 19, and 25 for December 26, with 37 total saturated areas and 12 total unsaturated areas. A simplified flow of inputs and outputs in the SWAT-VSA framework is shown in Fig. 3.3.

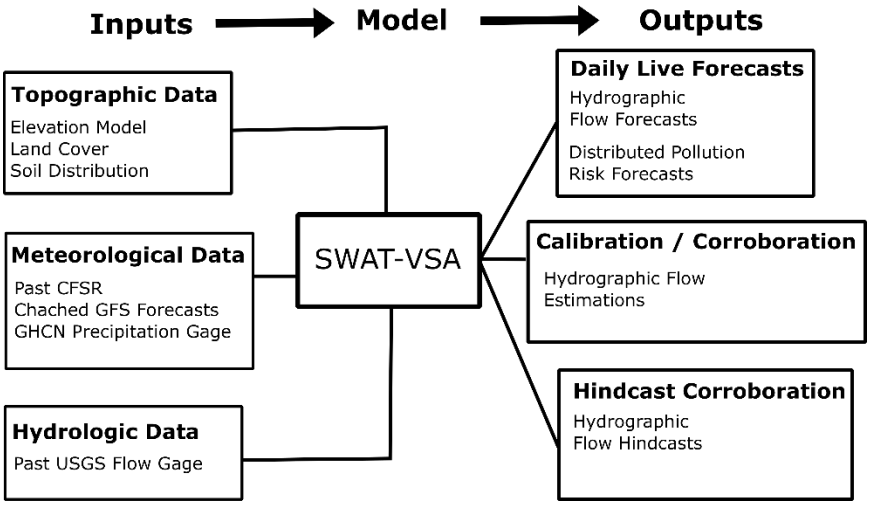


Figure 3.3. Simplified SWAT-VSA framework showing relationship between inputs and outputs

3.3 Results

3.3.1 Archived GFS Forecasts: The archived GFS forecasts were bias-corrected using station data in the watershed. The measured station data contain greater maximum values shown by the longer tails in Fig 3.4. The uncorrected GFS data increases almost linearly until the 0.9 CDF probability threshold, generally over-predicting precipitation and then, above 0.9 CDF probability, under-predicting precipitation (Fig. 3.4). The corrected GFS precipitation data more closely follow the station data shape, and captures more of the larger precipitation events, but fail to capture the largest events (>60 mm d-1-). In general, the uncorrected archived GFS forecasts have a lower probability of containing a precipitation value equal to or less than a single day value of about 8 mm, and a greater probability of containing a value equal to or less than a single day value of about 10 mm or more. In each forecast day, the maximum values are increased by the bias correction method, which extends the cumulative probability distribution curve towards the station data maximums. This suggests that the bias correction method was able to improve

the archived GFS forecasts in relation to the measured precipitation while still preserving the structure of the GFS data.

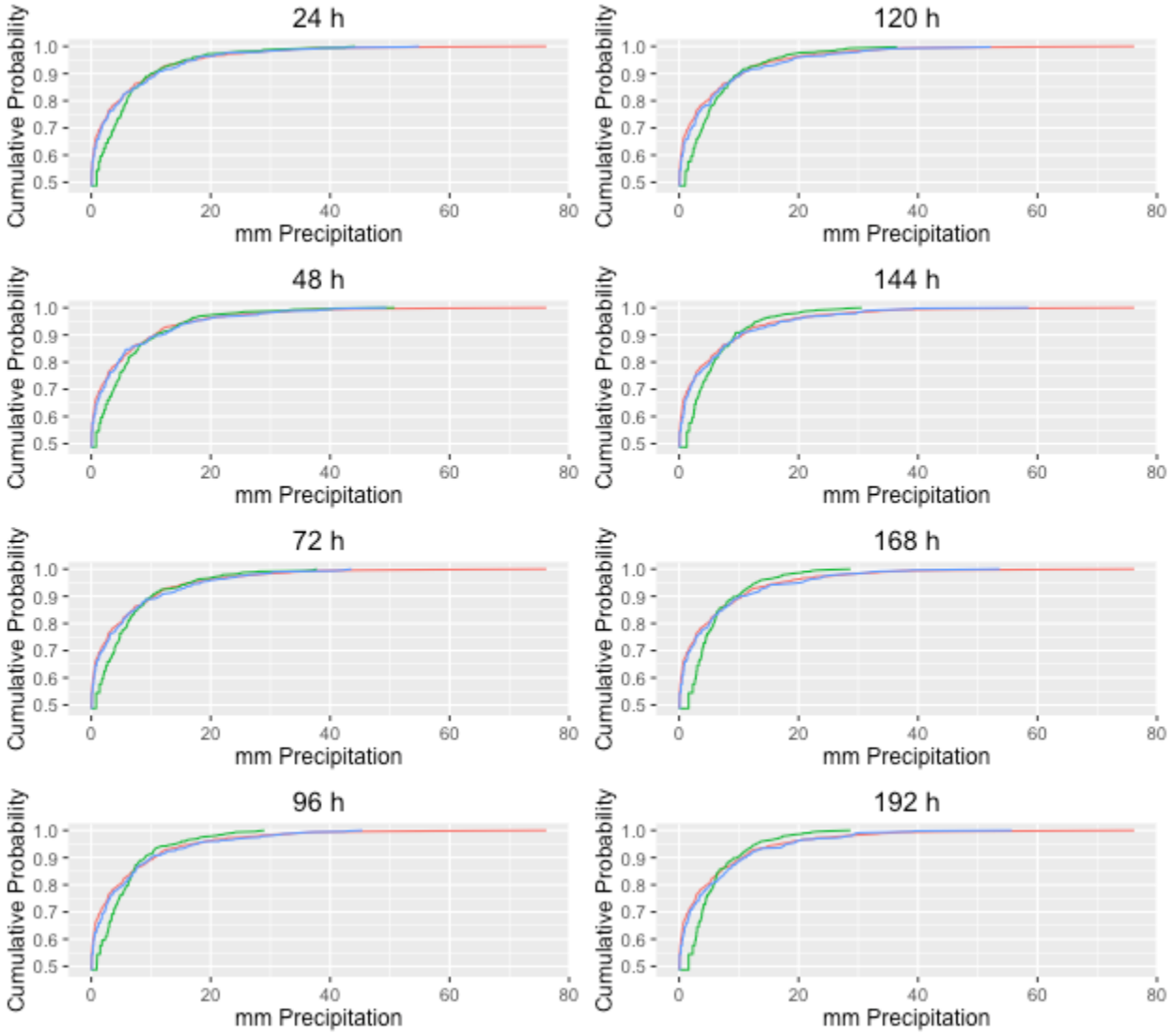


Figure 3.4. Uncorrected and bias corrected precipitation data. Red lines are station data, green lines are uncorrected, raw archived GFS forecasts, and blue lines are bias corrected archived GFS forecasts

3.3.2 Model Calibration and Corroboration Results: The SWAT-VSA model was able to predict the watershed level discharge relatively well on a daily basis over the calibration period January 1, 2005 through December 31, 2010 (Fig. 3.5), with an NSE of 0.60 in accordance with the commonly cited 'good performance' categorization (Moriasi et al., 2007). The model tended to under-predict peak flows of this period, likely due to the CFSR meteorological forcings, which tend to under predict precipitation particularly from convective type storms. The model was corroborated from January 1 2012 - August 31 2014, and again showed adequate model predictive ability on a daily basis (NSE 0.50, Fig. 3.6). This corroboration period was chosen to include that of the GFS hindcast corroboration periods, allowing for a direct comparison between standard model performance and the model performance in forecast mode.

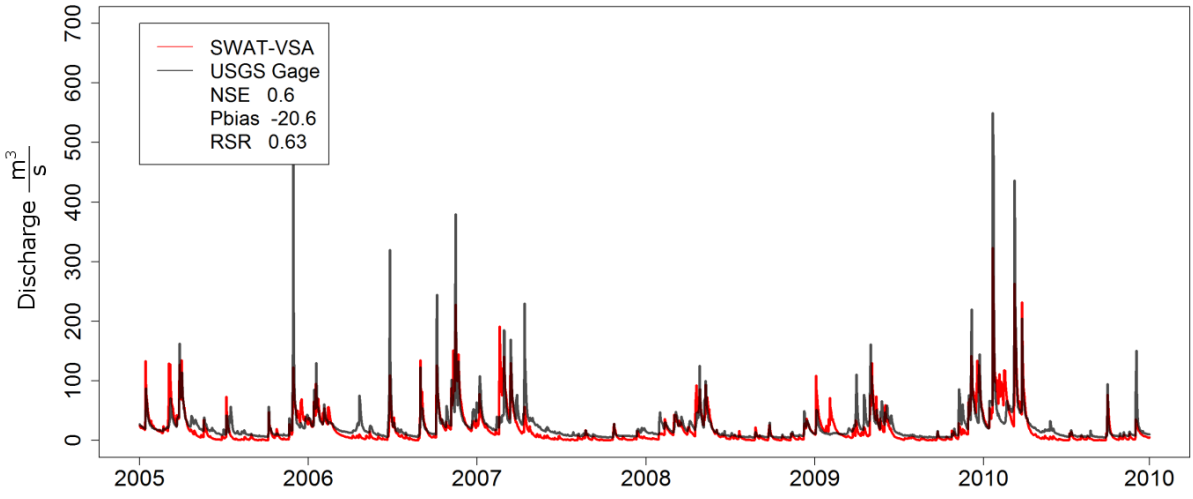


Figure 3.5. Observed and predicted stream flow at the watershed outlet during the calibration period.

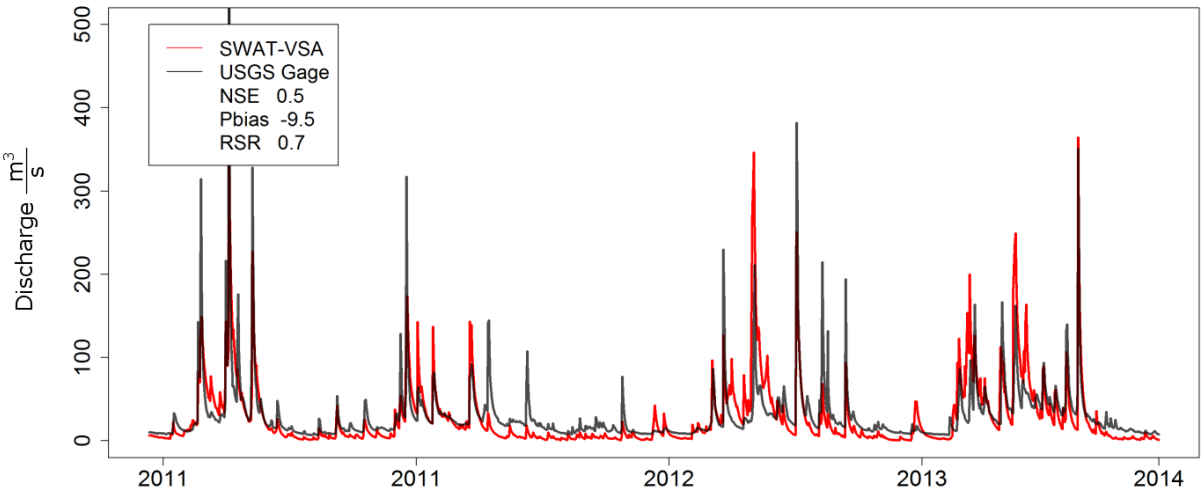


Figure 3.6. Observed and predicted stream flow at the watershed outlet during the corroboration period.

3.3.3 Hydrologic Flow Hindcast Corroboration: The hindcast corroboration was performed by comparing the daily flow predicted by the hindcast procedure to the USGS flow measurements from the gage station. The test period corresponded to the maximum amount of bias corrected archived GFS forecast data available after processing. The final corroboration range was February 1 2014 to August 30 2014. Results of the forecast corroboration showed good flow predictions for forecast days 1-2, satisfactory for days 3-4 and unsatisfactory predictions for days 5-8. The corroboration results for all flow forecasts are summarized in the Table 3.3.

Table 3.3. Predicted flow forecasts corroboration statistics.

Flow Forecast	NSE	Pbias	RSR
1 Day	0.49	-10.2	0.71

2 Day	0.52	-6.9	0.69
3 Day	0.42	-3.0	0.76
4 Day	0.40	0.2	0.77
5 Day	0.27	7.6	0.85
6 Day	0.20	7.6	0.89
7 Day	0.12	4.2	0.94
8 Day	0.07	6.4	0.93

*good forecasts are 0.50 or greater, 0.4-0.5 are satisfactory

The hydrologic hindcast procedure predicted overall flow periods with NSEs nearly or greater than 0.5 for the 1 and 2 day forecasts (Fig. 3.7). The hydrographic hincast was able to capture the timing of most storm events for all forecast days, however, there was a tendency to under predict low flow values which is quantified in the Pbias metric. This tendency to under predict low flows is likely the result of the bias correction, as CDF mapping can cause changes in variability as well as mean in precipitation, and changes in flow estimates from the forecast framework would likely be more pronounced at low flows due to greater variation in weather parameters.

The GFS hindcast flow predictions proved unsatisfactory for forecast days 5-8 (Fig. 3.8) with multi-day high flow events predicted poorly. This was due to the precipitation hindcast data under predicting multiple rainfall events within one or two weeks of each other, especially in the late spring when convective thunder storms contribute to high runoff variability. The abrupt reduction in accuracy in the weather hindcast data shows the GFS is not able to effectively capture high precipitation variability beyond a 96-hr period. It is also worth mentioning that non 24-hr time step aggregates of the hindcast data could possibly provide satisfactory flow forecasts.

This ‘drop-off’ in NSE after the 48-hr forecast could be misleading as there are multiple sub-daily forecasts in-between the 24 and 120-hr data sets that could yield satisfactory results. In addition, the NSE metric weights good predictions of extreme values higher than good predictions of low values and since the hydrologic predictions beyond 96 hrs could not predict high flows very well, their accuracy was much lower as quantified by the NSE metric. The hindcast overall shows an increase in extremes: many low flows are under predicted, and many high flows are over predicted. This is likely due the difference in variance in the archived GFS forecast and CFSR precipitation sets. GFS tends to contain a more volatile prediction of precipitation, and the hydrologic model response is influenced by this difference. Each hydrologic prediction is dependent on all previous forecasts from each model run, i.e., each 72 hr forecast is dependent on the previous 24 and 48 hr forecasts. In this way, the hindcast procedure shows error propagation through the predictions which causes increased volatility through the hydrologic hindcast sets.

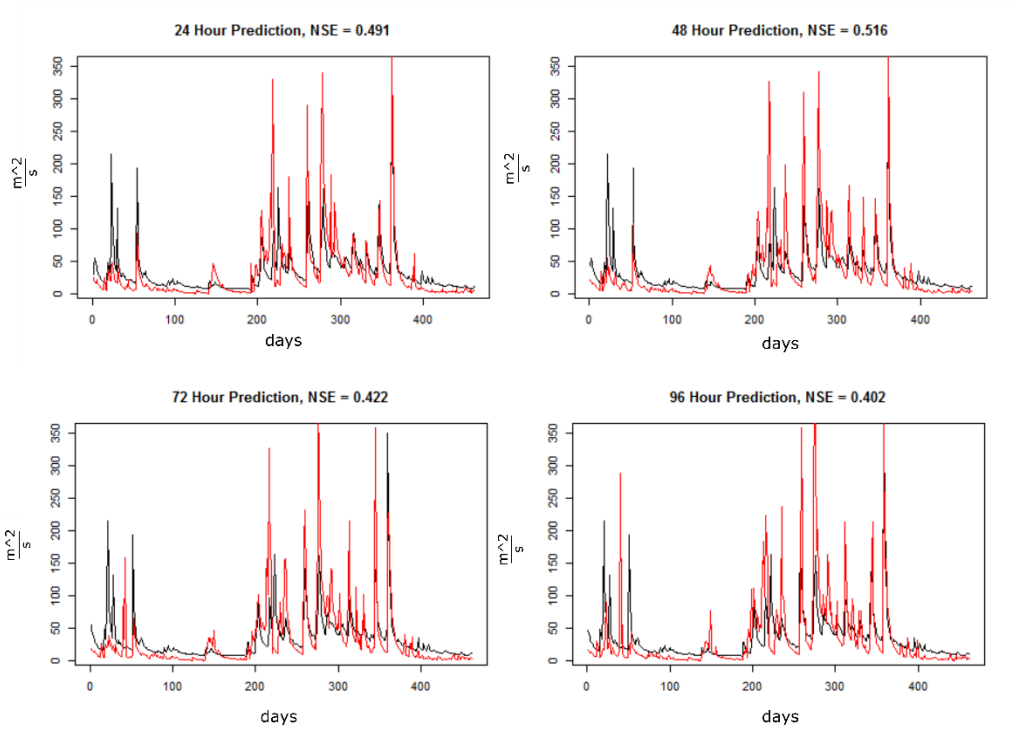


Figure 3.7. Flow estimates for forecast days 1 - 4 from the GFS hindcast forced model.

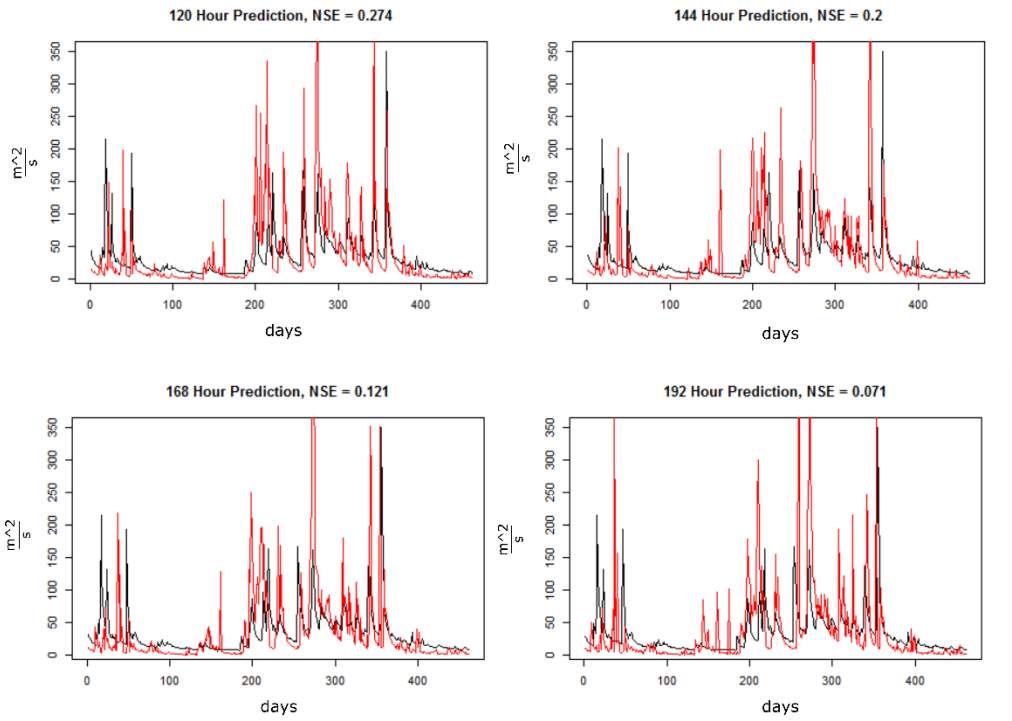


Figure 3.8. Flow Estimates for forecast days 5-8 from the GFS hindcast forced model.

3.3.4 Distributed Hydrologic Predictions: A summary of the distributed model predictions of soil moisture classification is shown in Fig. 3.9. These predictions were selected to correspond to field measurements made in the watershed on December 19, and December 26 2015. The soil moisture prediction is a binary classification 3 m resolution raster, and the boundaries are from data collected in the field and define areas mapped as fully saturated or unsaturated. Figure 3.9 presents the predicted soil moisture classifications and a selected number of measured saturated area boundaries. The probability that each of these areas is saturated is assumed equal to the ratio of raster cells classified by the model as saturated to total classifications contained in each measured boundary. In these example cases, all areas were predicted as saturated using a 50% probability threshold.

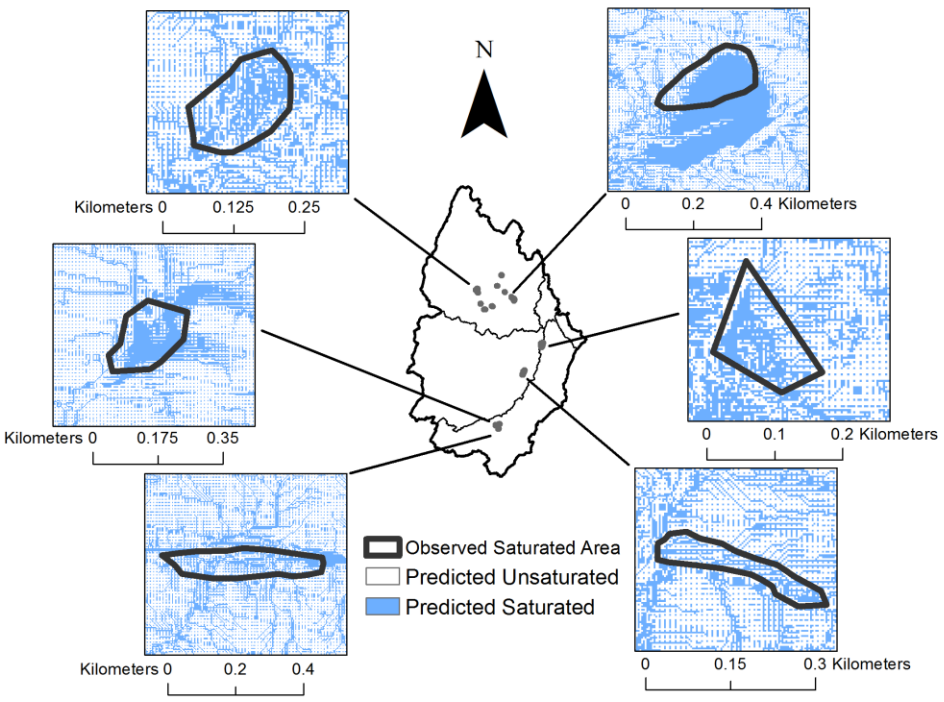


Figure 3.9. Example model output soil moisture classification forecast and measured saturated areas.

To assess the model's ability to predict distributed soil saturation, a receiver operating characteristic (ROC) curve was constructed with the predicted saturation probabilities of all areas and their measured classifications (Fig. 3.10). In all, 24 areas for December 19 and 25 areas for December 26 are included with 37 total saturated areas and 12 total unsaturated areas. The ROC curve shows the relationship between the true positive and false positive rate (TPR/FPR) for different classification thresholds and compares the model's ability to predict a binary classification compared to random selection. The ROC curve shows a distinctive bow towards the top left corner of Fig. 3.10, indicating the model performs well in comparison to random selection (represented by the diagonal). The classification ratio at a 50% probability with equal importance given to false positives and false negatives is indicated by the dashed lines and

highlights a TPR/FPR ratio of 0.88/0.25. The area under the ROC curve (AUC) is a common metric used to quantify the ROC analysis into one number. An AUC of 0.5 is interpreted as the model is equally good at predicting a binary outcome as would be a random selection, and a model with an AUC of 1 predicts all model outcomes correctly. The model's AUC from the ROC curve in Fig. 3.10 is 0.86, indicating good performance in spatial soil moisture prediction.

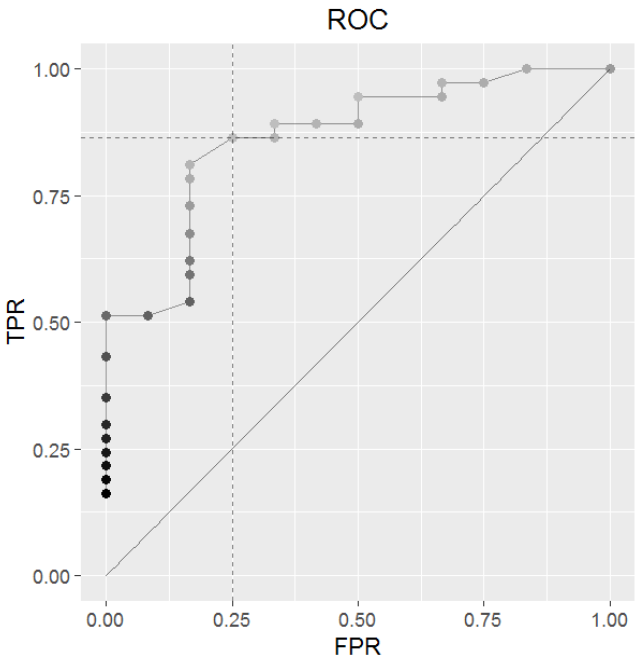


Figure 3.10. SWAT-VSA coupled with GFS hindcast performance. sDashed lines show an optimum threshold selection along a pareto-optimal front.

Using the 50% probability threshold, all predictions for the defined areas could be compared to measured data. Figure 3.11 below summarizes the results. Of the 49 areas measured, two were incorrectly classified as saturated when they were unsaturated (false positives) and seven were incorrectly classified as unsaturated when they were saturated (false negatives). Thirty true positives (areas correctly classified as saturated) and 10 true negatives

(areas correctly classified as unsaturated) were predicted. In Fig. 3.11 the "spinning top" shape distribution is due to majority of probabilities between 0.5 and 0.7, and a group of probabilities around 1.0. The group of high predictions around 1.0 is due to instances where all saturation classification raster cells contained within an area boundary are correctly labeled as saturated. Since the model derives the probabilities from this spatial relationship predicted probabilities of 1.0 are possible.

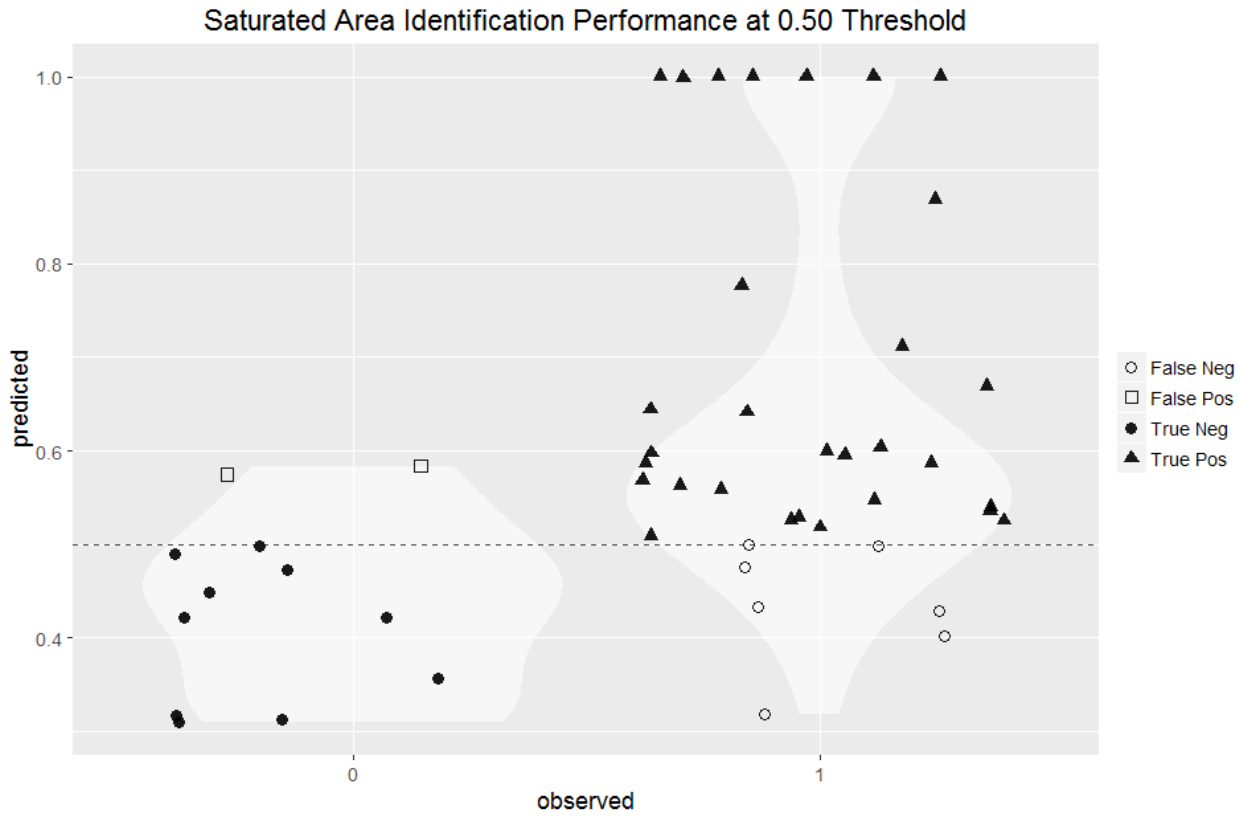


Figure 3.11. Saturated Area identification with 0.5 probability threshold.

3.4 Discussion

The development of the SWAT-VSA model coupled with the GFS identifies not only the locations of areas prone to saturation or runoff, but is also capable of predicting the probability of particular areas representing an aggregate of the spatial output to be saturated or unsaturated.

This information is critical to characterize field-specific saturation dynamics, which are helpful to formulate landscape management strategies. The daily updated probability maps provide scientifically grounded means for planners and producers to increase efforts that focus on the day-to-day protection of areas with high transport potential.

The model performed well spatially, providing single day probability saturated area forecasts (0.86 AUC), and it was able to forecast stream flow up to 96 hrs in advance with reasonable accuracy. A primary reason for lower performance on stream flow forecasts after 96 hrs was a volatile time series, under predicting low flow events, and over predicting high flow events. Interestingly, and based on the probability distributions of the archived forecasts, the meteorological forcings do not likely fully explain this volatility. The flow forecasts become increasingly unstable through the hindcast days, in part due to error resulting from the multiple model initializations. For instance, the 8-day hydrologic hindcast includes data from 8 different SWAT-VSA model initializations run over different time periods with the same calibration parameters. The error from these different model initializations propagates through the forecasts, limiting the satisfactory forecast of the framework to 96 hrs. Even with this limitation, the SWAT-VSA GFS coupling performed well in predicting the classification of saturated areas. The spatial classification method is a clear way to communicate the risk of agricultural managements applied to particular areas.

Agricultural field management operations could benefit from using the SWAT-VSA outputs to identify HSAs. The saturation classification raster overlaid on aerial or satellite imagery would help identify HSAs several days in advance. A producer could then employ this framework to avoid areas in the field where a high future probability of saturation is predicted, or the producer could elect to move the operations to other fields where probabilities are lower.

This could inform agricultural practices in a way that more simplified watershed models that rely on seasonal or annual load estimates fall short. A web-based forecast framework that displays the relevant forecast model outputs could provide the necessary information to drive field level management. The weather forecast data processing described here can be applied to real-time GFS forecasts, allowing distributed hydrologic outputs to be used in accessible forecasts designed as decision making tools for land managers and watershed stakeholders

3.5 Conclusions: Distributed hydrologic models like SWAT-VSA can be used to provide surface hydrology forecasts at a daily time scale. In this study, a framework consisting of the SWAT-VSA model and the CFSR and GFS gridded weather models was developed to provide daily volumetric flow forecasts and distributed saturation probability predictions for the South Fork of the Shenandoah watershed in north-central Virginia. A unique data set of archived GFS weather forecasts was used to perform a hindcast corroboration of the model predictions. The archived GFS forecast data set was preprocessed in a manner repeatable with real-time weather forecasts from the GFS model. The hindcast procedure provided satisfactory volumetric flow forecasts at the watershed outlet up to 96 hrs into the future, with NSEs 0.400-0.500 for daily model runs for a seven-month period.

The ability of the SWAT-VSA model to provide distributed hydrologic outputs based on TI/landuse defined HRUs allows the forecast system to make distributed hydrology forecasts in addition to hydrologic flow predictions. Distributed forecasts could be used to inform operational management practices in real-time on agricultural land to help protect surface runoff from pollution, especially practices like spreading manure without soil incorporation. Modifications in

practices influenced by these forecasts could reduce the impact of current agricultural practices on surface water quality. As agricultural producers continue to adopt new technologies and to use precision agriculture tools, short-term hydrologic forecasts can assist in the effort to better manage agricultural land and reduce NPS pollution.

3.6 Acknowledgements: We would like to acknowledge high-performance computing support from Yellowstone (http://n2t.net/ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation under project code 1360415 and funding support from the USDA under project number 2012-67019-19434

3.7 References

Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2014. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. J. Hydrol. 519, 2692–2706. doi:10.1016/j.jhydrol.2014.08.038

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Griensven, a. Van, VanLiew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: Model Use, Calibration, and Validation. Asabe 55, 1491–1508.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large Area Hydrologic Modeling And Assessment Part I : Model Development J. Am. Assoc. Am. Water Resour. Assoc. 34, 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth Syst. Sci. 5, 1–12. doi:10.5194/hess-5-1-2001

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. JAWRA J. Am. Water

Resour. Assoc. n/a–n/a. doi:10.1111/1752-1688.12296

Collick, A.S., Fuka, D.R., Kleinman, P.J. a, Buda, A.R., Weld, J.L., White, M.J., Veith, T.L., Bryant, R.B., Bolster, C.H., Easton, Z.M., 2014. Predicting phosphorus dynamics in complex terrains using a variable source area hydrology model. Hydrol. Process. 601, 588–601. doi:10.1002/hyp.10178

Dahlke, H., Easton, Z., Fuka, D., Walter, M., Steenhuis, T., 2013. Real-Time Forecast of Hydrologically Sensitive Areas in the Salmon Creek Watershed, New York State, Using an Online Prediction Tool. Water 5, 917–944. doi:10.3390/w5030917

Dahlke, H.E., Easton, Z.M., Lyon, S.W., Todd Walter, M., Destouni, G., Steenhuis, T.S., 2012. Dissecting the variable source area concept - Subsurface flow pathways and water mixing processes in a hillslope. J. Hydrol. 420-421, 125–141. doi:10.1016/j.jhydrol.2011.11.052

de Alwis, D. a, Easton, Z.M., Dahlke, H.E., Philpot, W.D., Steenhuis, T.S., 2007. Unsupervised classification of saturated areas using a time series of remotely sensed images. Hydrol. Earth Syst. Sci. 11, 1609–1620. doi:10.5194/hess-11-1609-2007

Douglas-Mankin, K.R., Srinivasan, R., Arnold, J.G., 2010. Soil and Water Assessment Tool (SWAT) Model : Current Developments and Applications. Am. Soc. Agric. Biol. Eng. ISSN 2151-0032. 53, 1423–1431.

Dunne, T., Black, R.D., 1970. to Storm Runo in a Small NeW , England Watershed Dense , till. Water Resour. Res. 6, 1296–1311.

Dunne, T., Leopold, L.B., 1978. Water in Environmental Planning. W. H. Freeman and Company, New York.

Dutta, D., Welsh, W.D., Vaze, J., Kim, S.S.H., Nicholls, D., 2012. A Comparative Evaluation of Short-Term Streamflow Forecasting Using Time Series Analysis and Rainfall-Runoff Models in eWater Source. Water Resour. Manag. 26, 4397–4415. doi:10.1007/s11269-012-0151-9

Easton, Z., Fuka, D., Walter, M., Cowan, D., Schneiderman, E., Steenhuis, T., 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. J. Hydrol. 348, 279–291. doi:10.1016/j.jhydrol.2007.10.008

Easton, Z.M., Fuka, D.R., White, E.D., Collick, a. S., Biruk Ashagre, B., McCartney, M., Awulachew, S.B., Ahmed, a. a., Steenhuis, T.S., 2010. A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia. Hydrol. Earth Syst. Sci. 14, 1827–1841. doi:10.5194/hess-14-1827-2010

Easton, Z.M., Walter, M.T., Fuka, D.R., White, E.D., Steenhuis, T.S., 2011. A simple concept for calibrating runoff thresholds in quasi-distributed variable source area watershed models. Hydrol. Process. 25, 3131–3143. doi:10.1002/hyp.8032

Fuka, D.R., Walter, M.T., Macalister, C., Degaetano, A.T., Steenhuis, T.S., Easton, Z.M., 2013. Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrol. Process. 5623, 5613–5623. doi:10.1002/hyp.10073

Fuka, D.R., Walter, M.T., Macalister, C., Steenhuis, T.S., Easton, Z.M., 2014. SWATmodel: A multi-operating system, multi-platform SWAT model package in R. J. Am. Water Resour. Assoc. 50, 1349–1353. doi:10.1111/jawr.12170

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: Historical development, applications, and future research directions. Trans. Asabe 50, 1211–1250. doi:10.1.1.88.6554

Gassman, P.W., Sadeghi, A.M., Srinivasan, R., 2014. Applications of the SWAT Model Special Section: Overview and Insights. J. Environ. Qual. 43, 1–8. doi:10.2134/jeq2013.11.0466

GFS, 2015. Global Forecast System (GFS) | National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC) [WWW Document]. URL

https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs (accessed 7.24.15).

Girvetz, E.H., Maurer, E., Duffy, P., Ruesch, A., Thrasher, B., Zganjar, C., 2013. Making Climate Data Relevant to Decision Making: The important details of Spatial and Temporal Downscaling 43.

Hanrahan, L.P., Jokela, W.E., Knapp, J.R., 2004. Dairy diet phosphorus and rainfall timing effects on runoff phosphorus from land-applied manure. J. Environ. Qual. 38, 212–217. doi:10.2134/jeq2007.0672

Hewlett, J.., Hibbert, A.R., n.d. Factors affecting the response of small watersheds to precipitation in humid regions. For. Hydrol. 275–290.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sensing 81, 345–354.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle, in: Transactions of the American Geophysics Union. pp. 446–460.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting : the forecast package for R Automatic time series forecasting : the forecast package for R. J. Stat. Softw. 27, 1–22.

IUSS, W.G.W., 2007. World Reference Base for Soil Resources 2006, first update 2007 (No. 103), World Soil Resources Reports. Rome.

Marjerison, R.D., Dahlke, H., Easton, Z.M., Seifert, S., Walter, M.T., 2011. A Phosphorus Index transport factor based on variable source area hydrology for New York State. J. Soil Water Conserv. 66, 149–157. doi:10.2489/jswc.66.3.149

Mohamoud, Y., 2004. Comparison of hydrologic responses at different watershed scales 1–81.

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

NCEP, 2003. The GFS Atmospheric Model, Note 442. 14.

Needelman, B. a., Gburek, W.J., Petersen, G.W., Sharpley, A.N., Kleinman, P.J. a., 2004. Surface Runoff along Two Agricultural Hillslopes with Contrasting Soils. Soil Sci. Soc. Am. J. 68, 914. doi:10.2136/sssaj2004.0914

Page, T., Haygarth, P.M., Beven, K.J., Joynes, A., Butler, T., Keeler, C., Freer, J., Owens, P.N., Wood, G. a, 2005. Spatial variability of soil phosphorus in relation to the topographic index and critical source areas: sampling for assessing risk to water quality. J. Environ. Qual. 34, 2263–2277. doi:10.2134/jeq2004.0398

Puckett, L.J., 1995. Identifying the major sources of nutrient water pollution. Environ. Sci. Technol. 29, 408A–414A. doi:10.1021/es00009a001

Schneiderman, E.M., Steenhuis, T.S., Thongs, D.J., Easton, Z.M., Zion, M.S., Neal, A.L., Mendoza, G.F., Walter, M.T., 2007. Incorporating variable source area hydrology into acurve-number-based watershed model. Hydrol. Process. 21, 3420–3430.

Sela, J., 1982. The NMC Spectral Model, , NOAA Technical Report NWS-30.

Sela, J., 1979. Spectral modeling at the National Meteorological Center. Mon. Weather Rev. 108, 1279–1292.

Sharma, S.K., Mohanty, B.P., Zhu, J., 2006. Including Topography and Vegetation Attributes for Developing Pedotransfer Functions. Soil Sci. Soc. Am. J. 70, 1430. doi:10.2136/sssaj2005.0087

Shortle, J.S., Ribaudo, M., Horan, R.D., Blandford, D., 2012. Reforming agricultural nonpoint pollution policy in an increasingly budget-constrained environment. Environ. Sci. Technol. 46, 1316–1325.

doi:10.1021/es2020499

Smith, D.R., Owens, P.R., Leytem, a. B., Warnemuende, E. a., 2007. Nutrient losses from manure and fertilizer applications as impacted by time to first runoff event. Environ. Pollut. 147, 131–137. doi:10.1016/j.envpol.2006.08.021

Sommerlot, A.R., Pouyan Nejadhashemi, A., Woznicki, S.A., Prohaska, M.D., 2013. Evaluating the impact of field-scale management strategies on sediment transport to the watershed outlet. J. Environ. Manage. 128, 735–748. doi:10.1016/j.jenvman.2013.06.019

Thompson, J. a., Pena-Yewtukhiw, E.M., Grove, J.H., 2006. Soil-landscape modeling across a physiographic region: Topographic patterns and model transportability. Geoderma 133, 57–70. doi:10.1016/j.geoderma.2006.03.037

Tsai, M.J., Abrahart, R.J., Mount, N.J., Chang, F.J., 2014. Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. Hydrol. Process. 28, 1055–1070. doi:10.1002/hyp.9559

Vadas, P. a., Jokela, W.E., Franklin, D.H., Endale, D.M., 2011. The effect of rain and runoff when assessing timing of manure application and dissolved phosphorus loss in runoff. J. Am. Water Resour. Assoc. 47, 877–886. doi:10.1111/j.1752-1688.2011.00561.x

Walter, M.T., Walter, M.F., 1999. The New York City Watershed Agricultural Program (WAP): A model for comprehensive planning for water quality and agricultural economic viability. Water Resour. Impact 1, 5–8.

White, E.D., Easton, Z.M., Fuka, D.R., Collick, A.S., Adgo, E., McCartney, M., Awulachew, S.B., Selassie, Y.G., Steenhuis, T.S., 2011. Development and application of a physically based landscape water balance in the SWAT model. Hydrol. Process. 25, 915–925. doi:10.1002/hyp.7876

Woodbury, J.D., Shoemaker, C.A., Easton, Z.M., Cowan, D.M., 2014. Application of SWAT with and

without Variable Source Area Hydrology to a Large Watershed. J. Am. Water Resour. Assoc. 50, 42–56. doi:10.1111/jawr.12116

Wuertz, D., Miklovic, M., Boudt, C., Chausse, P. 2009. fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling.

Zambrano-Bigiarini, M., Rojas, R., 2013. A model-independent Particle Swarm Optimisation software for model calibration. Environ. Model. Softw. 43, 5–25. doi:10.1016/j.envsoft.2013.01.004

CHAPTER 4. Development of a Free and Open Source Web Based Interface for Distributed Short-Term Hydrologic Forecasts*

***Citation:** Sommerlot. A.R.* and Z.M. Easton. 2017. Development of a free and open source web based interface for distributed short-term hydrologic forecasts. Water. 2017, 9, 604; doi:10.3390/w9080604

Key Words: Forecast soil moisture, Free and open source code, Online tool, Short-term risk, SWAT, SWAT-VSA, Web based decision support system.

4.0 Abstract

Non-point source (NPS) pollution from agricultural activity is a major source of surface water quality impairment in the Chesapeake Bay region, and demonstrates widely varying spatial and temporal pollution potential. Many efforts to protect waterways in this region are based on seasonal and annual estimates of pollutant loss potential (e.g., NRCS 590 nutrient management standard, P-Index) that may not adequately address the underlying hydrologic processes driving NPS pollution. One barrier to adopting practices that address these variable source areas (VSAs) of NPS pollution is a lack of tools that are designed to transfer information at sufficient spatial and temporal resolution so that end-users can make informed decisions. This study introduces a web-based interface displaying distributed hydrologic forecasts using free and open source software. The forecast system consists of three primary components; 1) a hydrology model that predicts the size, location and timing of VSAs as short-term distributed forecasts; 2) a data structure component that is capable of re-structuring large, high-resolution rasters for web display, and; 3) a user interface, that employs adaptive map viewing technology to create a web page that allows users to dynamically interact with the data. This system provides easily accessible, real-time risk forecast maps designed to help users avoid high-risk areas when planning agricultural practices.

4.1 Introduction

Non-point source (NPS) pollution from agricultural lands is a major source of impairment to surface waters in the United States (Shortle et al., 2012). Various government agencies have introduced standards and tools (e.g., NRCS 590 Standard, P-Index) in an effort reduce NPS pollution; however, these standards and tools have mixed success in reducing NPS pollution (Puckett, 1995). While there are many causes for this, one major limitation is a lack of success in communicating information pertinent to planning land management activities in a manner, which can both reduce NPS pollution and increase agricultural management flexibility (Easton et al., 2017). For tools, such as the P-Index which are annual and developed at a field level resolution, the spatial and temporal scales used to identify areas at a high risk of contributing to NPS pollution often fail to provide information at sufficient spatial and temporal resolution to consistently allow targeted landscape management (Easton et al., 2008; Marjerison et al. 2011; Dahlke et al., 2012), and effectively communicating the results remains a challenge (Easton et al., 2017).

As an example, the P Index is widely used to direct manure and fertilizer management but may fail to account for short-duration high-risk periods, such as manure application to saturated areas, so called variable source areas (VSAs). Effectively communicating when and where VSAs form could be used to further optimize agricultural nutrient management operations (Hanrahan et al., 2004; Smith et al., 2007; Vadas et al., 2011). While the capability exists to model VSAs, the results are not immediately available to users or in a format conducive to decision making. Agricultural producers (end users) need a model to run and update in real-time with necessary spatial and temporal resolution imagery to provide insight into the nature of near-future conditions on the ground. In the surface hydrology modeling and land management

communities there is a gap between the predictive ability of the modeling systems and how that information is delivered to end users (Jones et al., 2016).

Real time decision support systems (DSS) are an effective approach used to disseminate information by leveraging internet connected devices to directly deliver relevant information to end users (Bharati and Chaudhury, 2004; Durmuşoğlu and Barczak, 2011; Power and Phillips-Wren, 2011). In order to provide VSA forecasts to a large audience of end users and tools to interested scientist developing solutions to similar problems, this study introduces a DSS that provides real-time and spatially-detailed, short-term forecasts in a web-based interface built on free and open source software (FOSS). The DSS includes an interface and data management structure component designed to disseminate high-resolution spatial forecasts in an effort to bridge the gap between modeling capabilities and application to land management decisions made by end users. The method described here can display any georeferened raster model output at very fine resolutions as an interactive map. Although DSS platforms with similar goals have been introduced (e.g., predicting streamflow or reservoir status (Choi et al; 2005; Dutta et al., 2012; Abaza et al., 2014; Tsai et al., 2014) and software tools have been designed to communicate geospatial information including real time data and forecasts from earth systems models (Kulawiak et al., 2010; Swain et al., 2015; Snow et al, 2016; Tayyebi et al., 2016) we describe a system that is tailor-made for the particular problem of rapid refresh, high resolution, scalable extent native forecast products. We postulate that a web based interface built entirely with open source software can meet the end user requirements of the target audience while being flexible and scalable enough to provide scientists working on similar problems a method to communicate real time, spatially detailed model results to land managers and decision makers.

4.2 Methods

4.2.1 Design Requirements

This system is intended to be used by agricultural producers, conservation personnel, or agricultural extension personnel to better plan potentially polluting farm activities, such as manure or fertilizer application prior to a precipitation event or on saturated soils (VSAs). The ability to identify when and where these VSAs form allows the potential to still apply fertilizers and manures to specific areas of a landscape that are less prone to soil saturation and runoff generation. A producer using the DSS could reschedule the time or location of an application based the near future forecast to a day or area where less of the field is saturated, thus avoiding water quality degradation. In order to effectively inform this type of decision making, both the spatial and temporal resolution of the forecasts must be adequate. Users of this system need to make sub-field level decisions on a day-to-day basis, thus, outputs of the system need to be "sub-field-scale" spatial resolution and update sub-daily to allow flexibility in timing. A number of studies aimed at describing watershed processes mention a raster cell size between 2.5 and 25 m as sub-field-scale or high resolution, with more recent studies tending towards 10 m or less (Cambardella et al; 1994; Jetten et al., 1999; Famiglietti, et al., 2008; Chaney, et al., 2015). Review of these studies leads us to set the spatial requirement of our system to the current interpretation of sub-field scale at 10 m or higher resolution. The temporal resolution is more straight forward; as day-to-day management decisions require at minimum daily updated information. Ideally, the system would be updated sub-daily as weather forecasts are released by the National Weather Service (NWS), providing the user with the most current information possible. Thus, we set our temporal resolution requirements to sub-daily. Remaining user

requirements are focused on usability and accessibility. We propose a system that anyone can access with an internet connected device. We define usability as full functionality, i.e., all designed utilities of the interface work on a wide variety of devices and that user experience is positive. The Google PageSpeed[1] Insights evaluation tool was used to evaluate the interface for implementation, speed, and usability. PageSpeed Insights evaluates a web page for both desktop and mobile performance and provides scores ranging from 0 to 100, along with some reasoning for mobile speed, mobile user experience, and desktop speed (About PageSpeed Insights, 2015). The interface scores are summarized in Table 4.1.

The design of the system was driven by the need for data structures and display technology to be easily modified by other scientists or developers trying to communicate similar types of model results. In order to be accessible to researchers all over the world, including areas that may have trouble obtaining licensed software due to prohibitive costs or distribution limitations, the DSS was developed on free and open source software. In summary, we propose a system with two sets of design requirements, one set meets criteria designed to plan agricultural management practices based on modeled forecasts, and another set that allows researchers with similar challenges a free, and open source method to communicate model outputs with similar spatial and temporal resolution needs. Table 4.1 summaries the design requirements for the system.

Table 4.1: Design Requirements of the DSS.

End-User Requirements	**Spatial Resolution**	10 m or higher

[1] https://developers.google.com/speed/pagespeed/insights/

	Temporal Resolution	sub-daily
	Accessibility	pc and mobile internet connected devices
	Usability	80 point score on Page Speed Insights
Research Requirements	**License**	open source
	Cost	free
	Applicability	world wide
	Flexibility	Able to display a wide range of rasters

4.2.2 System Components

The DSS is composed of three major components (Fig. 4.1); 1) the model, which acquires, preprocesses the necessary hydrologic forcings, runs the watershed model, and outputs spatially distributed VSA forecasts; 2) a data management structure component which converts high resolution rasters into overlay web map tiles, and the user interface hosting, which publishes both the overlay tiles, and; 3) the user interface component, a web page that allows the user to interact with the processed output (Fig. 4.1). The first two components of the system must be automated, as forecasts are updated regularly (four times per day in this case). The final component, the interface hosting, only needs to be put in place once. Locations of the tile overlays do not change, but rather the files that populate the tiles with data are updated with each model run, so the server always looks in the same place for the forecast information. The base map does not need to be updated regularly, as it is a worldwide map sourced from satellite imagery, and serves to show users what spatial extent the overlay forecast tiles refer to. This allows the system to seamlessly host the overlay tiles even while they are being updated.

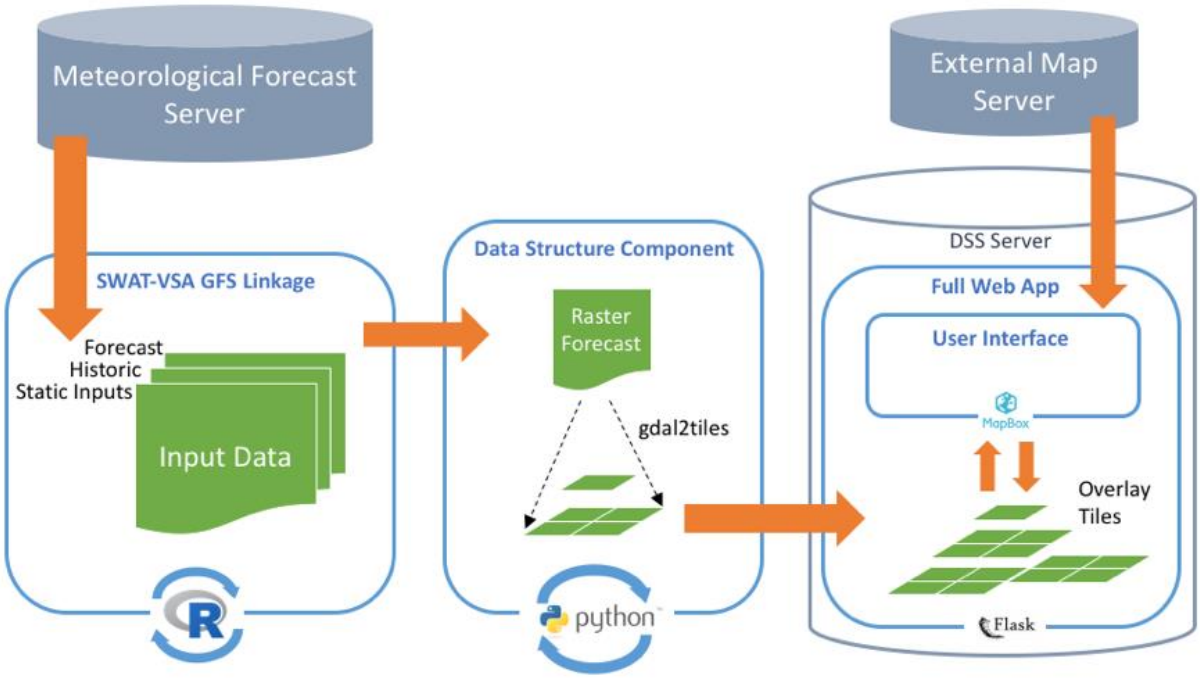


Figure 4.1. Overview of the DSS including the weather forecast (GFS-MOS), the model (SWAT-VSA GFS linkage) component, the data structure component, and the user interface.

4.2.3 SWAT-VSA Model Component

The model component, which produces the forecasts must satisfy the resolution requirements of the system in its forecast outputs. The model framework and development is described in Sommerlot et al. (2016a). Briefly the model framework links a VSA watershed model (SWAT-VSA) with the National Oceanic and Atmospheric Association (NOAA)'s Global Forecast System-Model Output Statistics (GFS-MOS) forecast, and sufficiently meets these standards. SWAT-VSA is a modification of the Soil and Water Assessment Tool (SWAT), which has been widely applied in surface water modeling applications in various climates and

regions. SWAT is a process-based, semi-distributed watershed model, developed by the US Department of Agriculture-Agricultural Research Service (USAD-ARS) to estimate the effects of agricultural landuse and management on water quantity and quality (Gassman et al., 2007; Arnold et al., 1998, 2012). Inputs required by the model include weather, soil, land cover, and land management data to simulate surface and subsurface hydrology and various chemical and sediment fluxes. Elevation data, spatial and tabular landuse, and spatial and tabular soil attributes are required for model initialization. SWAT lumps unique combinations of soil type, landuse, and if required, slope into hydrologic response units (HRUs) during the initialization process. Although the HRU formulation reduces computational requirements of the model, it often fails to correctly predict VSAs, which are common in humid temperate climates. In SWAT, unique HRU classifications may define multiple locations in the watershed that all have identical properties and normalized outputs. However, runoff from HRUs in watersheds dominated by VSA hydrology may vary based on topographic position. In this case, the SWAT model may oversimplify the hydrologic response of the watershed and, as a result may fail to account for critical runoff source areas (Easton et al., 2008; Douglas-Mankin et al., 2010; Marjerison et al., 2011; Tuppad et al., 2011; Collick et al., 2014; Gassman et al., 2014). In order to more effectively account for VSAs Easton et al. (2008) introduced the SWAT-VSA model, replacing the soil HRU layer with a Topographic Wetness Index (TI) / soil hybrid layer (Fuka et al., 2017). In SWAT-VSA the TI quantifies the saturation potential of areas within a watershed and ranges from low to high (1 to n), 1 being least likely and n being most likely to saturate (Easton et al., 2008). Spatial parameters and model outputs are therefore dependent on TI distribution throughout the watershed. Runoff and soil moisture are calculated from each HRU based on a modification of the Curve Number equation to define runoff through saturation, rather than

infiltration excess processes (Easton et al., 2008, 2010, 2011; Collick et al., 2014; Fuka et al., 2017).

Input data used in SWAT-VSA / GFS linkage consists of both dynamic and static data. The spatial data required for the initialization of the hydrology prediction model are all static inputs (soil, landuse, and elevation data). SWAT-VSA is forced with dynamic meteorological forecasts to produce a 24-96 hr distributed hydrologic forecast described in Sommerlot et al. (2016a). The DSS downloads and parses meteorological data from the GFS-MOS forecast dataset every six hrs. The GFS-MOS time series is obtained from an automatic implementation of the SWATmodel (Fuka et al., 2014) and getMet (Sommerlot et al., 2016b) R packages. The SWAT-VSA model is then re-initialized with the dynamic meteorological inputs for each model run to produce predictions for each model forecast day.

4.2.4 Data Structure Component

The output of the above process is a 3 x 3m resolution raster of soil moisture forecasts, which, in this application is multiple GB in size (see proof of concept application section, below), and is too large to be conveniently used directly as an overlay in an online mapping interface. In an effort to meet the design requirements, we employed the Open Street Map (OSM) Slippy Map structure (Slippy Map, 2016). This solution was chosen as it provides a means to display overlay rasters at the field-scale and higher resolutions with no specialized software other than common web browsers. The web map tiles are many 256x256 pixels files that together make up a large, high-resolution map at various zoom levels. Slippy Map works as an Asynchronous JavaScript and XML (AJAX) component, which dynamically requests map

tiles from a server without reloading the entire web page (Slippy Map, 2016). The number of tiles necessary to represent a geographic extent depends on the maximum zoom level rendered:

$$n_t = 2^{m_z} * 2^{m_z} \quad (4.1)$$

where n_t is the total number of tiles and m_z is the maximum zoom. For most applications, m_z is equal to 18 (Slippy Map, 2016). Therefore, to fully represent a worldwide map in a Slippy Map framework, 68,719,476,736 tiles are necessary. If the tile rendering takes too long, increasing hardware capabilities of the server or tile-rendering on-demand are options to reduce computation requirements to the desired time, but come at a cost of reduced end user performance (Quinn et al., 2010; Slippy Map, 2016). To create these map tiles with necessary names and file locations, a Python implementation of GDAL2Tiles (Fig. 4.2) capable of running in parallel creates a layered structure of map tiles that represents a georeferenced image at various zoom levels as a group of small, individually transferable images (GDAL2Tiles, 2015). Tiles are created from a georeferenced image by separating the image into a file tree of named map tiles in a particular structure. At each zoom level, the latitude and longitude of the top left corner of the tile is converted to a z-x-y naming and file-storing scheme (Fig. 4.2). The tiles are named and stored following Eqs. 4.3 and 4.4 (Slippy Map Tilenames, 2015):

$$x = \left[\frac{lon+180}{360}\right] * 2^z \quad (4.2)$$

where x is the resulting piece of the z-x-y naming scheme necessary for the OSM implementation of Slippy Maps, *lon* is the longitude (degrees) of the top left corner of the tile to be created, and *z* is the zoom level, which is also the z of the naming scheme. The *y* naming scheme follow Eq 4:

$$y = \left[\left(1 - \frac{\ln\left(\tan\left(lat * \frac{\pi}{180}\right) + \frac{1}{\cos\left(lat * \frac{\pi}{180}\right)}\right)}{\pi}\right) * 2^{z-1}\right] \quad (4.3)$$

where y is defined as the y component of the z-x-y naming scheme, lat is the latitude (degrees) of the top left corner of the tile to be created, and z is the zoom level the tile will be created at, which is also equal to the z component of the *z-x-y* naming scheme. Starting with the base file location, the *z* component defines a folder containing corresponding *x* folders which contain the actual map tile files named with the *y* component in portable network graphics (.PNG) format. All tiles at every necessary zoom level are made and structured in this way. The resulting file tree of tiles was stored on a beta server, which also hosted the user interface of the DSS. The overlay map in this structure can be queried by the user interface over regular HTTP. This process is readily automated, which is crucial to communicating outputs that must be updated often. In the case of our study area, the process of creating map tiles must be repeated four times daily in order to represent the most up-to-date information from the GFS-MOS.

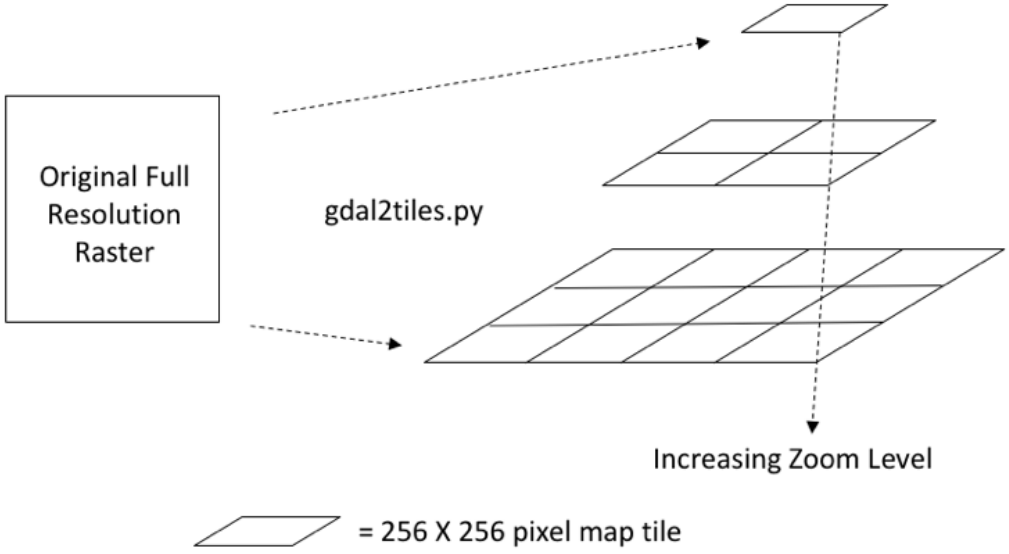


Figure 4.2. Conversion of raster file to map tile data structure via gdal2tiles.py showing a representation of increasing zoom levels. Maps in this data structure can be dynamically requested by the user interface.

4.2.5 User Interface

The public-facing portion of the web-based DSS is the user interface. The interface was built using the free Mapbox, JavaScript (JS)-based Application Programming Interface (API), Hyper Text Markup Language (HTML), and Cascading Style Sheets (CSS), bundled into an application with Flask, a light-weight web development platform implemented in Python (Ronacher, 2015). The HTML-CSS-JS combination is standard for building web applications, and Flask was chosen for its simplicity, as no communication between the user interface and a database is necessary. A critical component of the user interface is loading the base map. The base map, which is a world-wide hybrid transportation and satellite map is hosted by MapBox,

provides users with a high-quality spatial reference for the forecast overlays, updated regularly as the base map is improved. The base map is necessary for users to frame the forecast outputs in a usable way, without it, interpreting what the forecast represents on the ground would be very difficult. Other maps can be used as base maps, and it is worth noting that the MapBox hybrid map employed here is free to use only up to a certain number of user requests per month, though there are many fully free options which can replace this map with minor code edits. Researchers employing this method could create a custom base map and host it on their own server, again integrating it with minor edits. The DSS interface includes an address search box, legend, zoom bar, basic map manipulation controls, selectable forecast day, and a transparency slider for the overlay tiles. All functionalities of the user interface were designed to be simple and to translate well across devices.

The web based DSS (available at: http://zachary.bse.vt.edu/beta) was deployed on a server running the CENT-OS distribution of Linux. Under this deployment, the interface is capable of updating the distributed soil saturation forecast within the 6-hr intervals between the releases of new GFS-MOS forecasts. The elapsed time between the first step of triggering the meteorological forecast download to a complete map-tile update is roughly 4-hrs, even with the rather humble computational capabilities (8 core, 4 GHz Intel processor, and 32 GB ram) of the beta server. The full run time depends directly on the hardware capabilities of the machine on which the system is deployed and could be reduced with a more powerful processor, greater ram, or, most importantly for speed in this case, a larger number of logical processors available for parallel computing. The use of parallel computing in generating map tiles is very important to obtain a reasonable run time. Generating map tiles is the computational bottleneck of this framework, but provides the user with fast-loading overlay views of large raster data over a

standard HTTP internet connection. Other options exist, such as tile caching, but come at the expense of rendering speed to the end user, and when the server system is capable, using full tile generation ensures the highest performance (Qinn et al., 2010).

4.2.6 Proof of Concept Application

As a proof of concept, the model component was deployed in the South Fork of the Shenandoah River watershed (HUC 02070005) in the Shenandoah Valley of Virginia (Fig. 4.3). The outlet was defined at the location of the USGS gaging station 01628500. The watershed's 2600 km^2 area is forest (50%), agricultural land (38%), urban land (11%), and water (1%) (Homer et al., 2015). Silt loams and silty clay loams dominate the soil distribution and elevation ranges from 310 m to 1336 meters with slopes up to 60% (Fig. 4.3). The climate is classified as humid continental and yearly average rainfall in the watershed is 1057 mm (Mohamoud, 2004). This location was chosen as a proof of concept test bed based on the success of the short-term hydrologic forecast framework recently tested here (Sommerlot et al., 2016a). Geomorphic features in the watershed are typical of the Appalachian Mountain region and result in spatially and temporally discrete runoff generation mechanisms (Sommerlot et al., 2016a). Saturation excess runoff generation from soils with shallow restrictive layers at lower landscape positions (e.g., VSAs) accounts for the majority of surface runoff. Surface runoff from upslope soils is comparatively infrequent and of lower volume.

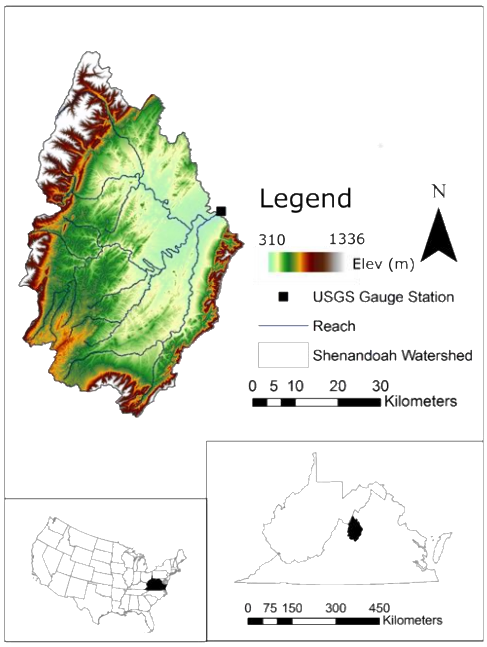


Figure 4.3. Study Area: South Fork of the Shenandoah river watershed in north central Virginia.

4.2.7 Model Setup and Calibration

In order to deploy the forecast component, the SWAT-VSA framework had to be set up and calibrated. Setup and calibration procedures are described in Sommerlot et al. (2016a). SWAT-VSA was initialized with distributed soil properties defined from the United Nations Food and Agriculture Organization (FAO) world wide database (IUSS, 2007), and spatially downscaled following the methods of Moore et al. (1993), Saxton and Rawls (2006), and Fuka et al. (2017). Landuse data was obtained from the National Land Cover Database (NLCD), which is available for the continental United States at 30 m resolution (Homer et al., 2015). The National Elevation Dataset provided the topography at varying resolutions, and the Digital Elevation Model (DEM) used was a combination of the highest resolution available for the area, combining 3 m and 10 m DEMs resampled to 3 m full coverage of the study area.

Meteorological and hydrologic time series data were also needed for calibration and corroboration of the SWAT-VSA model. Historic meteorological data, used to calibrate the model, included Climate Forecast System Reanalysis (CFSR) model data and observed weather data from the Global Historical Climate Network (GHCN) weather stations. The CFSR model was chosen as it has global coverage and contains all necessary meteorological parameters without missing data at a daily time step to force the watershed model: precipitation (mm), minimum and maximum temperature (deg C), percent relative humidity, wind speed (m/s), and solar radiation (MJ/m^2) (Fuka et al., 2013). Rainfall data sourced from the National Centers for Environmental Information Climate Data Online tool from five GHCN stations (UCS00448941, USC00442208, USC00448062, USC0044322 and USC00445096) were used to bias-correct the CFSR data, as bias was found in CFSR precipitation over the watershed area. Bias correction of CFSR precipitation was performed by modifying functions from the QMAP R Package using methods suggested by Gudmundsson et al. (2012) and Looper and Vieux (2012). The bias correction was applied by separating each precipitation data set into unique months, and creating a Cumulative Distribution Function (CDF) for each month. Next, a CDF mapping procedure was applied following Girvetz et al. (2013). Bias corrected CFSR data were verified to suitably map to the GHCN data by employing the Equiratio Cumulative Distribution Function matching method (Wang et al., 2010).

The resulting model was found to satisfactorily forecast the streamflow response and distributed runoff risk in the watershed during a hindcast test designed to reproduce actual past forecasts for evaluation and a 24-96 hr forecasts of streamflow estimates and binary spatial forecasts of saturated or unsaturated classifications with an accuracy of 86 percent when compared with measured data (Moriasi et al., 2007; Sommerlot et al., 2016a).

4.3 Results

The output raster in this application with the spatial extent of the study area, (roughly 2600 km^2 in web Mercator projection, and at a max zoom of 18) required the server to generate 350,266 new tiles four times per day for each forecast (24, 48 h etc.). With the output at this size, the system hardware was able to recreate the tile set within the time between every update to the distributed hydrology forecast raster while meeting our temporal resolution requirement. With the zoom at 18, each square pixel of a tile represents an area roughly 0.6 m across, which is capable of recreating the 3 x 3 m SWAT-VSA forecast output, more than fulfilling our spatial resolution requirement. Thus, both the forecast component and data structure component were able to meet the end-user requirements for spatial and temporal resolution.

As the interface is a simple web page it can be accessed by any device with a web browser. The user interface employing the MapBox API to display the map performed satisfactorily with 381 lines of JS, HTML, and CSS code needed to produce the output. Although full functionality was realized on the popular browser Chrome and Firefox, Internet Explorer was not able to properly render the maps consistently across versions due to limitations with the MapBox API. PageSpeed Insights provided estimations of the user interface performance (Table 4.2). The desktop platform performed better than the mobile in the speed category by 24 points, receiving a score of 82, while receiving a score of 68 in the mobile category. The reason for this score, and for speed reductions in the desktop category were both the need for external JavaScript Resources. User experience received a mobile friendly score, and the user experience for the desktop platform is not provided by PageSpeed Insights.

Table 4.2. Google PageSpeed Insights Summary.

Platform	Speed	User Experience	High Score Component	Low Score Component
Desktop	82	N/A	Reduced Server Response Time	External JavaScript Resource
Mobile	68	Mobile Friendly	Simple Design	External JavaScript Resource

Since all the software used was free and open source, and available to download the first two of our research design requirements, license and cost, were met. The third requirement, worldwide applicability is met with the caveat that each new area has a single time automated installation pre-process. This pre-process allows a different model output raster in a different area of the world to be easily displayed with the same data structure and user interface components described with only minor edits to the code by redefining the size, location, and resolution of the new model output. A range of raster types can be preprocessed according to specification in the gdal2tiles documentation (Gdal2tiles, 2015). The base map, which is a hybrid satellite and transportation label map, is already worldwide, thus, all that is needed is the new overlay raster in the correct size and location to change the display of the interface. The legend can be edited easily by anyone with basic knowledge of HTML and JS. A screen shot of the interface is shown in Figure 4.44, in which the DSS is predicting saturated areas during a precipitation event 48-hrs from the current conditions. As a further example of the DSS display, Figure 4.5 shows a large saturated area becoming less and less saturated over the 4-day forecast period. These results illustrate a case where land management operations, such as the surface

application of manure could be spatially and temporally adjusted around the short-term existence of a VSA.

Figure 4.4. DSS showing a 48-hr distributed soil saturation forecast over farm fields a few miles west of Harrisonburg, Virginia. The shapes of flow paths and VSAs can be seen. Top left includes zoom controls and a transparency slider for the forecast overlay, bottom left includes an address search box and semi-transparent scale that adjusts with zoom level, bottom right contains the legend and citations, and located at the top right is the forecast hr selector which changes the overlay display according to the selection.

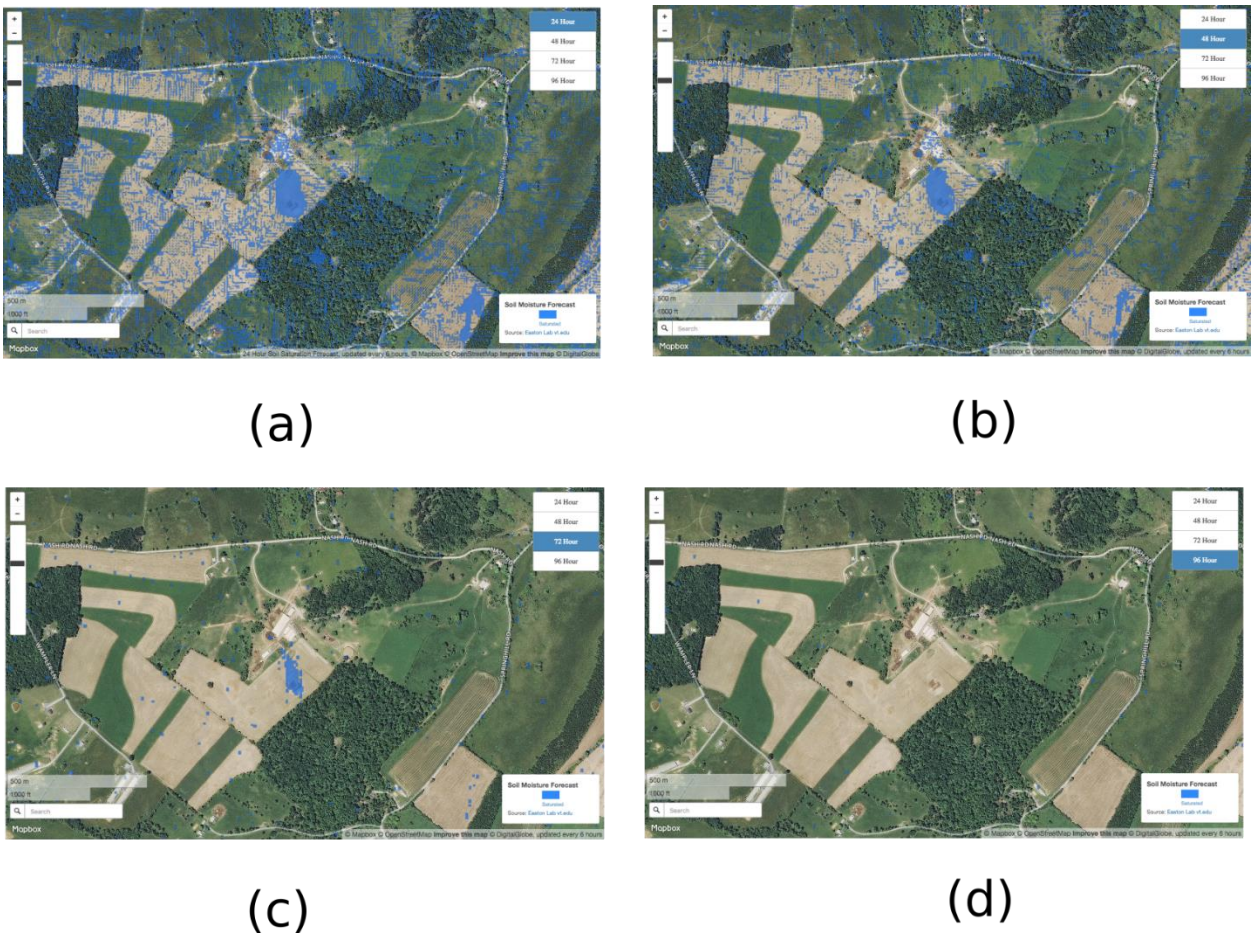


Figure 4.5. DSS showing a diminishing soil saturation forecast for 24 (a), 48 (b), 72 (c), and 96 (d) over farm fields west of Harrisonburg, Virginia. At the center of each screen shot, a large, concentrated area of soil saturation can be seen. Over the forecasts the extent of the area decreases, indicative of a forecasted precipitation then drying period. This illustrates the capability of the DSS to provide daily, sub-field resolution forecasts that communicate the variable nature of NPS pollution in the study area region.

Figure 5a highlights the sub-field-scale prediction abilities, showing a saturated area in a field (center blue shaded portion) following a 50 mm precipitation event. The DSS forecasts that the area will decrease in size over the next few days, but persist compared to surrounding areas.

If these forecasts were used to plan a surface manure spreading operation on the surrounding fields, assuming the operation must be completed in the 24 to 48-hr period, the field containing this persistent saturated area could be avoided and the application moved to adjacent, less saturated areas. Another option that the DSS presents is to modify the timing of the operation; a producer using the DSS might decide to wait a few days to apply manure and then the entire region in Fig. 4.5(d) is much lower risk. It is envisioned that this DSS can provide a measure of risk avoidance, particularly for agricultural producers who land apply manure with the simple communication format of a forecast labeling areas as saturated or not saturated.

4.4 Discussion

The DSS performed well in meeting the design requirements and was capable of communicating SWAT-VSA forecasts up to four days in advance at 3 x 3m resolution updated multiple times per day. One notable limitation of this system is the computational power required to render map tiles. Although rendering tiles regularly for entire overlay maps process makes the interface fast and light weight for the user, it does require at least a multi core server which will be routinely busy with tile creation. However, this paper demonstrates that reasonable performance can be had through the use of an entry-level server, and that data structure component can run in parallel, and thus is scalable to deployments with more powerful hardware. This result improves upon similar systems, such as SmartScape, an ecological decision making tool built with the GoogleMaps API displaying 30 m resolution rasters, who's designers list the static nature of displayed model results as a major limitation (Tayyebi et al., 2016).
Through the responsive Mapbox API, the interface was accessible on both pc and mobile devices with a number of common browsers, with the notable exception of Internet Explorer. Although

Internet Explorer is a common browser, the only version that will be supported by early 2017 holds, as of November 2016, an estimated 12% of the market share compared with Chrome's 54%, thus it was not considered a fatal flaw (Net Market Share, 2016). In addition, new versions of Windows OS are phasing out Internet Explorer in favor of a new browser, Edge, which presumably will be able to provide the requisite functionality.

The PageSpeed Insights scores (68/100 for mobile and 82/100 for desktop) and mobile friendly designation from were primarily due to the simple web page design, as most of the difficult programming and computationally expensive processes are encountered when developing the map tiles, leaving the final interface with simple requirements for viewing, translating well across different hardware and software platforms. In both the pc and mobile categories, the same major cause of speed bottlenecks was the requirement of an external JavaScript script, in this case the MapBox JS API. This may reduce the speed of mobile platforms more as they are generally running with slower internet or data connections. This could be addressed by storing the MapBox JS API on the server, rather than sourcing it from MapBox servers and rewriting some of the JS code in the user interface, though that would complicate future development and application efforts. To evaluate the DSS comparatively, similar systems reviewed in Easton et al., (2017) designed to display large scale spatial data were tested with the PageSpeed Tool (Table 4.3). The DSS was one of less than half of the tools that were categorized as mobile friendly. The DSS was out-performed in only one category by the Missouri Design Storm maps, which does not display high resolution rasters. It received the highest score for Desktop speed and exceeded the average speed of the other tools by 12 points in the mobile category and 24 points in the desktop category. A major reason many of these tools received low speed scores was non-optimized images on the page, and most of the tools were

designated as not mobile friendly due to the presence of small text and interaction points on the web pages. The DSS described in this study stands out due to simple design, allowing for higher cross device rendering speed while retaining the ability to display rapid update, high resolution outputs. Overall, the PageSpeed analysis was intended to provide a simple and semi-formal means to quantify the usability of the interface, point out areas of improvement in the prototype version, compare the interface with other tools, and to prepare for a higher volume of server requests.

Table 4.3 Comparative PageSpeed Scores

DSS Name	**PageSpeed Score**		**Mobile Friendly**	**URL**
	Mobile	**Desktop**		
Fusarium Head Blight Risk	59	31	No	http://www.wheatscab.psu.edu/
Climate Normals North East	61	75	No	http://www.nrcc.cornell.edu/regional/climatenorms/climatenorms.html
Climate Patterns Viewer	36	34	No	https://mygeohub.org/groups/u2u/cpv
CornSplit Nitrogen Application	36	34	No	https://mygeohub.org/groups/u2u/splitn
Wisconsin Manure Runoff Risk	55	75	No	http://www.manureadvisorysystem.wi.gov/app/runoffrisk

Missouri's Design Storm Maps	76	59	Yes	http://ag3.agebb.missouri.edu/design_storm/
Manure Spreading Advisory	61	77	Yes	https://maps.whatcomcd.org/whatcom_msa
Soil Moisture Forecast*	68	82	Yes	http://zachary.bse.vt.edu/beta

* The DSS described in this study

It is also worth noting that the VSA model in this deployment is capable of outputting a number of distributed forecasts that could be used in a similar way as soil moisture is used in this study. Any output that SWAT-VSA creates in the spatial output file can be easily mapped to the user interface, including runoff, precipitation, evapotranspiration, nitrogen, phosphorus, and loading to the river, percolation, and biomass and plant growth estimations such as crop yields. Although the data structure and interface components can be used to communicate any high resolution geo-referenced raster output, researchers in agricultural sciences could use the same SWAT-VSA modeling framework towards different goals, as the outputs are readily available.

Overall, the system presented in this paper met the design requirements. More work is needed however, to increase the speed of the user interface when accessed with mobile devices. Although the interface, which is designed generally to be accessed by a wide range of devices may be difficult to optimize, it has the advantage of being usable and accessible right away upon deployment with very little ongoing maintenance required. For some researchers, this may be of great importance, as more complicated web systems which require in-house database management or the purchase cloud server accounts may be out of reach for prototyping projects. The focus of the system described in this paper is not to provide a platform or a system that

allows users to edit data through communication with a database, but rather to communicate high resolution model output at sub-daily timescales in a simple way. In doing so, it remains simple enough for researchers to deploy with minimal continuing management.

4.5 Conclusions

This study introduces a web based DSS framework capable of providing large, high resolution rasters representing distributed soil saturation forecasts. The data structure and user interface components are capable of reproducing model outputs at the high spatial and temporal resolutions required by end-users. The data structure component employed for displaying the overlay maps is scalable and applicable worldwide limited by hardware capabilities. The interface successfully demonstrated the delivery of large, distributed VSA model outputs as an interactive overlay in a simple web-map application that can be used by many people with an internet or data connected device. Most of the design requirements were met, with minimal maintenance work required to confirm cross-device functionally and speed. The accessibility to users, and the flexibility for developers make this framework a useful and powerful tool in supplying important distributed hydrologic forecasts to land managers and decision makers. This work will strengthen the tools available of informing short-term management decisions, ultimately reducing NPS pollution in VSA dominated watersheds.

4.6 Acknowledgement

We would like to acknowledge high-performance computing support from Yellowstone (http://n2t.net/ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information

Systems Laboratory, sponsored by the National Science Foundation under award numbers 1360415 and 1343802, and funding support from the USDA under project number 2012-67019-19434.

4.7 References

Abaza, M., Anctil, F., Fortin, V., Turcotte, R., 2014. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. J. Hydrol. 519, 2692–2706. doi:10.1016/j.jhydrol.2014.08.038

About PageSpeed Insights [WWW Document], n.d. . 2015. URL https://developers.google.com/speed/docs/insights/about (accessed 2.10.16).

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Griensven, a. Van, VanLiew, M.W., Kannan, N., Jha, M.K., 2012. SWAT: model use, calibration, and validation. Asabe 55, 1491–1508.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: Model development of a basin scale model called SWAT (Soil and Water Assessment Tool. J. Am. Water Resour. Assoc. 34, 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth Syst. Sci. 5, 1–12. doi:10.5194/hess-5-1-2001

Bharati, P., Chaudhury, A., 2004. An empirical investigation of decision-making satisfaction in web-based decision support systems. Decision support systems, 37(2), 187-197.

Cambardella, C. A., T. B. Moorman, T. B. Parkin, D. L. Karlen, J. M. Novak, R. F. Turco, and A. E. Konopka., 1994. Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Sci. Soc. Am. J. 58, 1501-1511. doi:10.2136/sssaj1994.03615995005800050033x

Chaney, N. W., J. K. Roundy, J. E. Herrera-Estrada, and E. F. Wood., 2015. High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design, Water Resour. Res., 51, 619–638. doi:10.1002/2013WR014964.

Choi, Jin-Yong, Bernard A. Engel, and Richard L. Farnsworth., 2005. Web-based GIS and spatial decision support system for watershed management. J Hydroinform. 7.3, 165-174.

Collick, A.S., Fuka, D.R., Kleinman, P.J. A, Buda, A.R., Weld, J.L., White, M.J., Veith, T.L., Bryant, R.B., Bolster, C.H., Easton, Z.M., 2014. Predicting phosphorus dynamics in complex terrains using a variable source area hydrology model. Hydrol. Process. 601, 588–601. doi:10.1002/hyp.10178

Dahlke, H., Easton, Z.M., Fuka, D., Walter, M., Steenhuis, T., 2013. Real-Time Forecast of Hydrologically sensitive areas in the Salmon Creek watershed, New York State, using an online prediction tool. Water 5, 917–944. doi:10.3390/w5030917

Douglas-Mankin, K.R., Srinivasan, R., Arnold, J.G., 2010. Soil and Water Assessment Tool (SWAT) Model: Current developments and applications. Am. Soc. Agric. Biol. Eng. ISSN 2151-0032. 53, 1423–1431.

Dunne, T., and R.D. Black. 1970. Partial area contributions to storm runoff in a small New England watershed. Water Resour. Res. 6:1296-1311.

Durmuşoğlu, S. S., Barczak, G., 2011. The use of information technology tools in new product development phases: Analysis of effects on new product innovativeness, quality, and market performance. Industrial Marketing Management, 40(2), 321-330.

Dutta, D., Welsh, W.D., Vaze, J., Kim, S.S.H., Nicholls, D., 2012. A Comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in eWater Source. Water Resour. Manag. 26, 4397–4415. doi:10.1007/s11269-012-0151-9

Easton, Z.M., Fuka, D., Walter, M., Cowan, D., Schneiderman, E., Steenhuis, T., 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. J. Hydrol. 348, 279–291. doi:10.1016/j.jhydrol.2007.10.008

Easton, Z.M., Fuka, D.R., White, E.D., Collick, a. S., Biruk Ashagre, B., McCartney, M., Awulachew, S.B., Ahmed, a. a., Steenhuis, T.S., 2010. A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia. Hydrol. Earth Syst. Sci. 14, 1827–1841. doi:10.5194/hess-14-1827-2010

Easton, Z.M., Walter, M.T., Fuka, D.R., White, E.D., Steenhuis, T.S., 2011. A simple concept for calibrating runoff thresholds in quasi-distributed variable source area watershed models. Hydrol. Process. 25, 3131–3143. doi:10.1002/hyp.8032

Easton, Z.M., Kleinman, P.J.A, Buda, A.R., Goering, D., Emberston, N., Reed, S., Drohan, P.J., Walter, T.M., Guinan, P., Lory, J.A., Sommerlot, A.R., Sharpley, A., 2017. Short-term forecasting tools for agricultural nutrient management. J.Environ.Qual. doi:10.2134/jeq2016.09.0377.

Famiglietti, J. S., D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, 2008. Field observations of soil moisture variability across scales. Water Resour. Res. 44, W01423. doi:10.1029/2006WR005804

Fuka, D.R., Collick, A.S., Kleinman, P.J.A., Auerbach, D., Harmel, D., Easton, Z.M., 2016. Improving the spatial representation of soil properties and hydrology using topographically

derived initialization processes in the SWAT model. Hydrol. Process. doi:10.1002/hyp.10899

Fuka, D.R., Walter, M.T., Macalister, C., Degaetano, A.T., Steenhuis, T.S., Easton, Z.M., 2013. Using the Climate Forecast System Reanalysis as weather input data for watershed models. Hydrol. Process. 5623, 5613–5623. doi:10.1002/hyp.10073

Fuka, D.R., Walter, M.T., Macalister, C., Steenhuis, T.S., Easton, Z.M., 2014. SWATmodel: A multi-operating system, multi-platform SWAT model package in R. J. Am. Water Resour. Assoc. 50, 1349–1353. doi:10.1111/jawr.12170

Fuka, D.R., D. Auerbach, A.S. Collick, And Z.M. Easton. 2017. The TopoSWAT toolbox: Enhanced basin characterization in SWAT initializations. Envron. Model. Software. (In Review).

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The Soil and Water Assessment Tool: Historical development, applications, and future research directions. Trans. Asabe 50, 1211–1250. doi:10.1.1.88.6554

Gassman, P.W., Sadeghi, A.M., Srinivasan, R., 2014. Applications of the SWAT model special section: Overview and insights. J. Environ. Qual. 43, 1–8. doi:10.2134/jeq2013.11.0466

GDAL2Tiles [WWW Document], 2015. URL http://wiki.openstreetmap.org/wiki/GDAL2Tiles (accessed 2.10.16).

Girvetz, E.H., Maurer, E., Duffy, P., Ruesch, A., Thrasher, B., Zganjar, C., 2013. Making climate data relevant to decision making: The important details of spatial and temporal downscaling The World Bank, March 27, 2013.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations: A

comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. doi:10.5194/hess-16-3383-2012

Hanrahan, L.P., Jokela, W.E., Knapp, J.R., 2004. Dairy diet phosphorus and rainfall timing effects on runoff phosphorus from land-applied manure. J. Environ. Qual. 38, 212–217. doi:10.2134/jeq2007.0672

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sensing 81, 345–354.

IUSS, W.G.W., 2007. World Reference Base for Soil Resources 2006, first update 2007 (No. 103), World Soil Resources Reports. Rome.

Jetten, V., de Roo, A., Favis-Mortlock, D., 1999. Evaluation of field-scale and catchment-scale soil erosion models. Catena. 37, 521-541

Kulawiak, M., Prospathopoulos, A., Perivoliotis, L., Kioroglou, S., & Stepnowski, A., 2010. Interactive visualization of marine pollution monitoring and forecasting data via a Web-based GIS. Computers & Geosciences, 36(8), 1069-1080.

Looper, J.P., Vieux, B.E., 2012. An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model. J. Hydrol., Hydrology Conference 2010 412-413, 114–132. doi:10.1016/j.jhydrol.2011.05.046

Marjerison, R.D., Dahlke, H., Easton, Z.M., Seifert, S., Walter, M.T., 2011. A Phosphorus Index transport factor based on variable source area hydrology for New York State. J. Soil Water Conserv. 66, 149–157. doi:10.2489/jswc.66.3.149

Mohamoud, Y., 2004. Comparison of hydrologic responses at different watershed scales.

National Exposure Research Laboratory Office of Research and Development U.S. Environmental Protection Agency. EPA/600/R-04/103. 1–81.

Moore, I.D., Gessler, P., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443–452. doi:10.2136/sssaj1993.572NPb

Moriasi, D., Arnold, J., 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. Trans. ASABE 50, 885–900. doi:10.13031/2013.23153

Needelman, B. a., Gburek, W.J., Petersen, G.W., Sharpley, A.N., Kleinman, P.J.A., 2004. Surface Runoff along two agricultural hillslopes with contrasting soils. Soil Sci. Soc. Am. J. 68, 914. doi:10.2136/sssaj2004.0914

Net Market Share. [WWW Document] 2016. URL http://www.netmarketshare.com/browser-market-share.aspx?qprid=2&qpcustomd=020 (accessed: 12.20.2016)

Power, D. J., Phillips-Wren, G., 2011. Impact of social media and Web 2.0 on decision-making. Journal of Decision Systems, 20(3), 249-261.

Puckett, L.J., 1995. Identifying the major sources of nutrient water pollution. Environ. Sci. Technol. 29, 408A–414A. doi:10.1021/es00009a001

Quinn, S., Gahegan, M., 2010. A predictive model for frequently viewed tiles in a web map. Transactions in GIS, 14(2), 193-216.

Ronacher, A., 2015. Welcome to Flask [WWW Document]. URL http://flask.pocoo.org/docs/0.11/

Saxton, K., Rawls, W., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Sci. Soc. Am. J. 70, 1569–1578. doi:10.2136/sssaj2005.0117

Shortle, J.S., Ribaudo, M., Horan, R.D., Blandford, D., 2012. Reforming agricultural nonpoint

pollution policy in an increasingly budget-constrained environment. Environ. Sci. Technol. 46, 1316–1325. doi:10.1021/es2020499

Slippy Map [WWW Document], 2016. URL http://wiki.openstreetmap.org/wiki/Slippy_Map (accessed 1.10.16).

Slippy map tilenames [WWW Document], 2015. URL http://wiki.openstreetmap.org/wiki/Slippy_map_tilenames#Tile_numbers_to_lon..2Flat (accessed 1.10.16).

Smith, D.R., Owens, P.R., Leytem, a. B., Warnemuende, E. a., 2007. Nutrient losses from manure and fertilizer applications as impacted by time to first runoff event. Environ. Pollut. 147, 131–137. doi:10.1016/j.envpol.2006.08.021

Snow, A.D., Christensen, S.D., Swain, N.R., Nelson E.J., Ames, D.P, Jones N.L., Ding, D., Noman, N.S., David C.H., Pappenberger, F., Zsoter, E., 2016. A High-Resolution National-Scale Hydrologic Forecast System from a Global Ensemble Land Surface Model. J. Am. Water Resour Assoc. 52, 950–964. doi: 10.1111/1752-1688.12434

Sommerlot, A.R., Berbero, M., Fuka, D.R., and Easton, Z.M., 2016a. Coupling the short-term Global Forecast System weather data with a variable source area hydrologic model. Environ. Model. Softw. 86, 68-80. http://dx.doi.org/10.1016/j.envsoft.2016.09.0081364-8152.

Sommerlot, A.R., Fuka, D.R., and Easton Z.M., 2016b. getMet: Get Meteorological Data for Hydrologic Models. R package version 0.3.2. https://CRAN.R-project.org/package=getMet

Swain, N.R., 2015. Tethys Platform: A Development and Hosting Platform for Water Resources Web Apps. All Theses and Dissertations. 5832. http://scholarsarchive.byu.edu/etd/5832

Tayyebi, A., Meehan, T. D., Dischler, J., Radloff, G., Ferris, M., & Gratton, C., 2016. SmartScape™: a web-based decision support system for assessing the tradeoffs among multiple ecosystem services under crop-change scenarios. Computers and Electronics in Agriculture, 121, 108-121.

Tsai, M.J., Abrahart, R.J., Mount, N.J., Chang, F.J., 2014. Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. Hydrol. Process. 28, 1055–1070. doi:10.1002/hyp.9559

Tuppad, P., Mankin, K.R.D., Lee, T., Srinivasan, R., Arnold, J.G., 2011. Soil and Water Assessment Tool (SWAT) hydrologic/water quality model: Extended capability and wider adoption. ASABE 54, 1677–1684.

Vadas, P. A., Jokela, W.E., Franklin, D.H., Endale, D.M., 2011. The effect of rain and runoff when assessing timing of manure application and dissolved phosphorus loss in runoff. J. Am. Water Resour. Assoc. 47, 877–886. doi:10.1111/j.1752-1688.2011.00561.x

Wang, X.L., Chen, H., Wu, Y., Feng, Y., Pu, Q., 2010. New techniques for the detection and adjustment of shifts in daily precipitation data series. J. Appl. Meteorol. Climatol. 49, 2416–2436. doi:10.1175/2010JAMC2376.1

CHAPTER 5. Deep Learning for High Resolution, Rapid Update Soil Moisture Predictions at Large-Scales: Applying and Interpreting Artificial Intelligence in Water Resources*

*Under Review in Environmental Modelling & Software, Citation: Sommerlot, A.R., and Z.M. Easton. 2017.

Keywords: Soil Moisture Forecast, Short-term, machine learning,

5.0 Abstract

Great strides have recently been made in machine learning; however, gaps still exist in our ability to interpret and fully exploit results of this technique. This study attempts to harness the advantages of both stochastic and process based modeling, providing a method capable of consistently estimating sub-daily, high resolution (3 m) soil moisture forecasts. A deep learning (DL) model is trained using geomorphic data, climate data, and results from a process based model. The DL model resulted in a 9% increase in predictive power compared to the process based model and a ten-fold decrease in run time. Additionally, DL interpretation methods applicable beyond this study are described including hidden layer visualization and equation extractions describing a quantifiable amount of variance in hidden layer values, providing a means to interpret processes learned by the DL model, and an area under the curve (AUC) metric of 0.9 when predicting on corroboration data. A DL model was successfully trained to make soil moisture predictions superior to a process based model. Additionally, some indications exist that the model learned relationships which behave similar to known hydrologic process.

5.1 Introduction

Hydrologic models are often classified into two groups: stochastic or data driven models, which are based on mathematic or statistical data representations, such as regression or neural networks, and process-based models, which represent the physical process occurring, such as rainfall-runoff responses or channel flow (Daniel et al., 2011). Processed-based models are routinely used in agricultural science and water resources fields (Daniel et al., 2011). An emerging application of these models in water resources is rapid update, spatial forecasts of distributed hydrologic parameters, such as soil moisture levels or runoff response (Easton et al., 2017). For these applications, various spatial discretization schemes exist, broadly falling into fully distributed, semi-distributed, and lumped categories (Daniel et al., 2011). Although benefits and costs exist with each method, the computational efficiency and scientific appeal due to the widely applicable characterization schemes of semi-distributed models have led to research and applications of these methods to predict distributed hydrology (Easton et al., 2008; Dalhke et al., 2011; Easton et al., 2017).

The basic structure of Neural Networks (NN), appearing in literature as Artificial Neural Networks (ANN), Multi-Layer Perceptions (MLP), and Deep Learning (DL), have been previously used in hydrology (Hsu, et al., 1995; Dawson and Wilby, 1998; Sajikumar and Thandaveswara, 1999; Abrahart and See, 2000; Luk et al., 2000; Dawson and Wilby 2001; Nagy et al., 2002; Rajurkar et al., 2004; Chen et al., 2012; Valipour et al., 2013; Yuan and Jia, 2016). "Deep learning" does not yet have a consensus-supported definition; however Schmidhuber (2015) suggests DL can be defined as a NN structure with three or more hidden layers. Applications of NN include reservoir inflow forecasting (Bai et al., 2016), rainfall-runoff response modeling (Wilby et al., 2003), groundwater dynamics (Nayak et al., 2006), soil moisture estimation (Jiang et al., 2004) and forecasting extreme events (Coulibaly et al., 2001). Although these methods show promise, DL models are used less often than process-based models to drive decisions in water resource management (Fatichi et al., 2016). An often-cited reason is that hydrologic models for water resource management must include attempts to describe physical processes not dependent on input data in order to create a "virtual laboratory" where scenarios can be assessed and processes understood (Fatichi et al., 2016). Though long-term scenario modeling is important for water resource management, there has been a recent interest in real-time and short-term forecasting and modeling as components of decision support systems (DSS) designed for water resource management (Easton et al., 2017). In these types of applications, it more important to provide accurate forecasts than to have the ability to model scenarios or fully understand the underlying processes. These short-term forecasting applications present an opportunity to benefit from the predictive power of DL. Ideally, in addition to providing high performance estimations, DL could be used to develop new rules and equations that describe physical processes (Jain et al., 2004). Deep Learning has a proven track record of

high performance in a variety of fields, and with recent improvements in computation and software structure, these methods are being widely adopted to solve difficult and complex problems (LeCun et al., 2015). Additional improvements in DL platforms have made implementation, training, and in-depth exploration of the final model relatively straightforward for researchers and practitioners (Candel et al., 2016; LeCun et al., 2015; Daniel et al., 2011).

This study demonstrates an implementation of DL designed for enterprise-like applications in providing rapid update, high resolution soil moisture forecasts. A data driven DL model is trained with an input dataset containing parameters that relate to hydrologic processes. The input dataset required to train the DL model is built by applying scientific principles of hydrology in a preprocessing step, guided by what is known and assumed about hydrology in the region of application. In this way, decades of scientific knowledge and study are used to create an input dataset which provides a physical context built with known processes for training the DL model. Based on this input data and the dependent variable (soil moisture), the DL model is then free to learn while optimizing a loss function during the training process. The DL model learns both major driving relationships that may be similar to process-based models, and additional intricacies represented by data driven relationships. Additionally, the DL model can be trained on higher discretization resolution where initialization of the process based model fails or becomes too cumbersome for calibration or rapid-update implementation.

5.2 Methods

5.2.1 Overview

The DL model was developed by creating a multivariate dataset with each row representing data describing a unique combination of static geomorphic variables for each unique

area in the study domain (Table 5.1). Additionally, a binary dependent variable consisting of soil moisture status (saturated and unsaturated) was produced for each row of static inputs over a range of dynamic meteorological conditions. The soil moisture estimates, over a range of wet and dry days in 2016, for the training dataset were made by a process based model (SWAT-VSA, Sommerlot et al., 2016) which has been extensively evaluated against measured soil moisture levels. The main difference between the input variables for the process based model and the DL model in this study was the input set built here had roughly a 50 times higher spatial resolution. This setup conceivably provided an opportunity for the DL model to learn processes and intricacies that the process based model could not capture.

5.2.2 Study Area

The South Fork of the Shenandoah (SFS) watershed (the training area) hydrologic unit code (HUC) 02070005 is located in the Shenandoah Valley in the state of Virginia, United States (US). Maximum and minimum elevation in the watershed are 1057 m and 310 m, respectively (Fig. 5.1). Soils are primarily silty clay loams and silt loams, and the watershed is covered primarily by forest (50%) with 38% agricultural land, 11% urban, and 1% of the 2600 km^2area covered by water (Homer et al., 2015). Annual average rainfall is 1057 mm, and climate is defined as humid continental (Mahomoud, 2004). This location was chosen as a proof of concept test bed based on the success of the short-term hydrologic forecast framework recently tested here (Sommerlot et al., 2016a). Geomorphic features in the watershed are typical of the Appalachian Mountain region and result in spatially and temporally discrete runoff generation mechanisms (Sommerlot et al., 2016). Saturation excess runoff generation from soils with shallow restrictive layers at lower landscape positions (e.g., VSAs) accounts for the majority of

surface runoff. Surface runoff from upslope soils is comparatively infrequent and of lower volume.

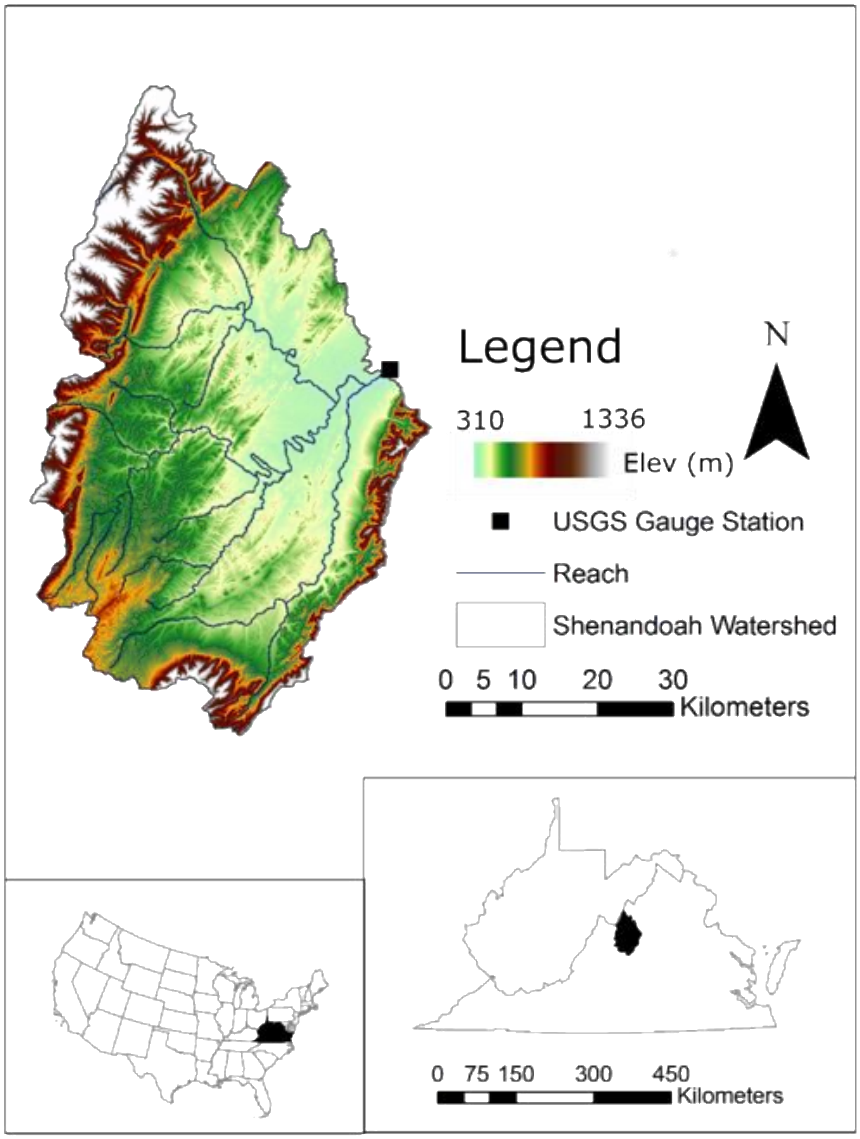


Figure 5.1. Location of the study area from Sommerlot et al. (2016) showing watershed elevation and relative location.

5.2.3 Process Based Model Description

Sommerlot et al. (2016) developed a coupling of live weather forecasts with a variant of the Soil and Water Assessment Tool (SWAT) model modified specifically for distributed simulations based on non-Hortonian, saturation-excess flow, known as variable source are (VSAs) hydrology (Easton et al., 2008; Arnold et al., 2012). Sub-daily weather forecast from the Global Forecast System (GFS) (NCEP, 2003) in the SFS river watershed were used to force short-term estimations of distributed soil moisture status with the SWAT-VSA model. The process-based model linkage was designed as a binary classification (saturated or not saturated) of daily, distributed soil moisture at 3 m resolution. The model predicted soil moisture states well, providing an area under curve (AUC) metric of 0.86 compared to observed areas (Sommerlot et al., 2016). The soil moisture classification data developed during the SWAT-VSA initialization procedure was used to develop the training data for the DL model, as the process based soil moisture forecasts were previously corroborated.

5.2.4 Deep Learning Model Setup

The DL model developed in this study is a deep feed forward neural network; deep feed forward neural networks are composed of an input layer, hidden layers, and an output layer (Schmidhuber, 2015). Layers are made up of nodes called neurons that produce real-valued activations from a number of weighed connections to neurons from the previous layer (Schmidhuber, 2015). Input variables are feed into the network through the input neurons, with one neuron per variable. The signal from each neuron is fed into the neurons in the next layer with a unique weight parameter; neurons in the next layer calculate activations based on the weighted sum resulting from the addition of the signal from these connections as an input to an activation function (Schmidhuber, 2015). This process is repeated through all layers and nodes and a final computation is done to calculate the values for the dependent output layer (Schmidhuber, 2015). Training the DL model involves incrementally adjusting the weights assigned to each connection between nodes to optimize a loss function (Schmidhuber, 2015). The backpropagation

algorithm is a comparatively fast and efficient training algorithm and is currently very common among modern DL code libraries (LeCun et al., 2015) and is implemented in this study.

For this study, the H2O platform (Candel et al., 2016) was selected for implementing the DL model due to the many built-in features available, its default ability to run in parallel, and its R interface. The DL model was setup with the goals of training a model using a dataset obtained from a calibrated, corroborated process-based model and geomorphic/climate data from the watershed and to provide an opportunity to interpret the final trained model to gain insight into hydrologic processes (Lipton, 2016). In an attempt to increase interpretability, the number of neurons in hidden layers were chosen to "tail-off" at the end of the feed forward structure, with five hidden layers containing, 64, 128, 32, 16, and 2 neurons, respectively. It is notable that this was not the best performing architecture tested, where models with a higher number of minimum neurons in hidden layers performed better, however the 16 and two-neuron hidden layer model was chosen to increase the interpretability by extracting the neuron values in the final layer with respect to the training dataset, allowing the development of a relationship between the hidden layer values, inputs, and outputs. Other important DL model settings included maxout selected as the activation function, a recently popularized modification of the common rectified linear activation (Schmidhuber, 2015; Candel et al., 2016), learning rate of 1e-6, an objective function set to maximize AUC, and neuron dropout (Candel et al., 2016) of 20% in the first three hidden layers, and 0% in the remaining two.

5.2.5 DL Model Input Data

An input data set consisting of a number of independent variables and a dependent variable was constructed for training the DL model (Table 5.1). Independent variables included both dynamic (weather) and static (geomorphic) data similar to the structure described by Song et al. (2016). The static dataset was built from an aggregation of data created during the SWAT-VSA initialization process using ArcSWAT2012, TopoSWAT (available from http://ww2.bse.vt.edu/eastonlab/) and from geomorphic data

from the SFS watershed. TopoSWAT links external databases including the United Nations Food and Agriculture Organization (FAO) soils database (IUSS, 2007), and local inputs in ArcMap to create model initialization data based on a topographic index linkage. Additional spatial inputs include landuse from the National Land Cover database (Homer et al., 2015), and a DEM created through a resampling of 3 m and 10 m resolution United States Geologic Survey (GHCN) National Elevation Datasets (NED). Data was prepared as if it would be used in a SWAT-VSA model, only considerations for database size, model run time, and success of SWAT-VSA initialization (all of which may present problems when building large input database tables) were not important. The model initialized by Sommerlot et al. (2016) was limited to 859 Hydrologic Response Units (HRU-spatial intersection of landuse, soil type and topographic index) due to computational limitations in ArcSWAT, while the data for the DL model was not limited in this manner and included 54,789 HRUs over the same study area. Database tables describing the HRUs, soil inputs, landcover, and topography were used to create a static input data set of independent variables considering each HRU a separate observation (Table 5.1).

Additionally, a process which gathers dynamic inputs was designed to connect with the static inputs providing the final inputs (Table 5.1). Dynamic data was acquired from Weather Underground's Application Programming Interface (API; https://www.wunderground.com/weather /api/), which has been used in a number of applications, including forecasting and real-time weather warning applications (Wang et al., 2006; Huang et al., 2008; David et al., 2013; Black and Stephen, 2014; Retchless et al., 2014; Samaranayake et al., 2014; Wong and Kerkez, 2016). This study uses both historical and forecast data from the API to create a dynamic input dataset consisting of current conditions, three days previous weather, and a 24 hr forecast. These inputs were queried from the API by latitude and longitude defined by each watershed subbasin centroid from the initialization discretization in Sommerlot et al. (2106), and applied to all the HRUs which were contained in each subbasin. All variables and date combinations were

treated as unique independent variables. Weather variables (Table 5.1) combined to create a dynamic input dataset consisting of 20 variables, five for each day queried.

Table 5.1. Static and dynamic input data used in the Deep Learning model, source is in parenthesis.

Variable Description (Source)	**Data Type**
Area of Hydrologic Response Unit (HRU) m^2 (SWAT-VSA)*	Continuous
Mean Slope Steepness in HRU m m^{-1} (SWAT-VSA)	Continuous
Mean Slope Length m (SWAT-VSA)	Continuous
Mean Slope in HRU m m^{-1} (SWAT-VSA)	Continuous
Max Depth of Soil in HRU mm (SWAT-VSA)	Continuous
Soil Hydraulic Conductivity mm hr^{-1} (SWAT-VSA)	Continuous
Percent Soil Weight Particles < 0.002 mm (SWAT-VSA)	Continuous
Percent Soil Weight Particles between 2.0 and 0.05 mm (SWAT-VSA)	Continuous
Percent Soil Weight Particles between > 2.0 mm (SWAT-VSA)	Continuous
Available Water Content mm water / mm soil depth (SWAT-VSA)	Continuous
Soil Hydrologic Group (SWAT-VSA)	Categorical
Landuse Category (SWAT-VSA)	Categorical
Soil Aspect Flow Direction (SWAT-VSA)	Categorical
Low Resolution Soil Name (FAO)	Categorical
Topographic Index Class (SWAT-VSA)	Categorical
Cumulative Rainfall in (wunderground)[1]	Continuous
dewpoint °F (wunderground)[1]	Continuous
Windspeed m s^{-1} (wunderground)[1]	Continuous

Maximum Temperature °F (wunderground)[1]	Continuous
Minimum Temperature °F (wunderground)[1]	Continuous

*Shaded rows are dynamic inputs, unshaded are static

[1]includes data from three previous days, current day, and the 24 h forecast

In order to further refine the training data, and to ensure the dataset was representative of a variety of soil moisture conditions, a range of wet and dry days for the watershed were selected: March 17, 2016, April 28, 2016, May 8 2016, June 5, 2016, August 9, 2016, September 16, October 28, 2016, and November 27, 2016. After selecting the wet and dry days, the SWAT-VSA framework was run on each of these days, producing predictions of saturated or unsaturated conditions in the watershed at a 3 m resolution. Each output raster from this process was used in a spatial overlay operation to label the 54,768 HRU raster designed for the DL model for each of the training days as saturated or unsaturated. Through the matching of HRU identifiers, the saturation labels of the resulting 54,768 HRU raster could be related to the selected independent variables in Table 5.1. The weather data (three history days, one current, and one forecast) were kept as separate variables, thus every row of the resulting table from this process represented a unique observation in time over a uniquely defined space and could then be combined in a single training dataset. In all, the training data totaled 46 (number of variables greater than Table 5.1 due to expansion of categorical variables) independent variables with 438,072 observations including labeled saturation classifications spanning a range of watershed moisture conditions.

5.2.6 Training and Corroboration

The DL model was trained using the dataset described above through backpropagation, which is the included training algorithm in H20 (Candel et al., 2016). An AUC metric, commonly used in training DL models and other machine learning algorithms, was selected as a loss function (Candel et al., 2016). The input dataset was split into a training and corroboration set. The training set was used to adjust the weights of the DL model through backpropagation, and the corroboration set was held out and the loss

function evaluated during every training iteration and the final model. This holdout corroboration process provides an opportunity to observe how the model performs on data which it was not trained on, a process analogous to the calibration and corroboration process of a watershed model with time series data. In all, 394,246 samples were used for training, about 90% of the input data, and 43,807 were used for corroboration, about 10%. Due to class imbalance, (there were far more unsaturated instances compared to saturated) the training dataset was balanced by using the built-in class balance option in H20, which increased the size of the effective training dataset by a factor of roughly two (Wei et al., 2013). Additionally, following the method described in Sommerlot et al. (2016), the DL model output was compared with observed data defining the extent of saturated areas in various locations of the SFS watershed. This process was completed for 25 separate saturated areas over one day, December 26, 2015; the data used by Sommerlot et al. (2016) in corroboration of their framework. In this way, a two-step corroboration was completed, one to provide evidence that the DL model compares well to a well-behaved, process based model (SWAT-VSA) in the watershed and a second, to provide evidence that the DL model predicts the extent of observed saturated areas, the manner in which it would be used in a live application over a single day without class balancing.

5.2.7 Deep Learning Interpretation

The DL model was subjected to a number of procedures after corroboration aimed at increasing the understanding of the learned processes by using a function in H2O which extracts the values of hidden layers in the DL model. Applying this function to the model and training data, the final hidden layer of the DL model consisting of only two neurons was used to explore how different inputs are related to the resulting probability of soil saturation estimates. Hidden layer number four, or the second to last hidden layer was also extracted to observe class separations. Due to having 16 neurons, a dimensionality reduction technique called t-tailed

stochastic nearest neighbor estimation (tSNE) was used to visualize the hidden layers to assist in interpreting the model (Maaten and Hinton, 2008; Mohamed et al., 2014). Furthermore, elastic net linear regression (Friedman et al., 2010) and non-linear regression were performed on the final hidden layer with the DL model saturated class probability outputs as dependent variables, and real valued inputs as independent variables. By using the elastic net linear regression to estimate values from a hidden layer neuron, and using the result as inputs to a second regression equation relating the final DL hidden layer to saturated class probabilities, an explanatory linkage equation was extracted and can be interpreted as a simple summary of the DL model. The linkage equation is composed of simple coefficients that correspond to real-valued inputs and outputs and can be interpreted with relative ease. The extracted equation was used on the corroboration dataset to evaluate its predictive capability. This equation can be used both as evidence for interpreting the processes learned by the DL model and for making explicit predictions on its own, given suitable performance on corroboration data. This is a key step in the data-driven modeling approach, as it attempts to extract a usable equation from a DL model, a task that has remained difficult for researchers and practitioners (Jain et al., 2004). Although the equation does not fully describe the DL model, it can describe a quantifiable amount of variance in parts of the hidden layers and be interpreted as a simple summary of the "knowledge" contained in the model after training over the input data.

5.3 Results

5.3.1 Accuracy Assessment

The training procedure resulted in a satisfactory DL model reproducing the process based model results. After 700 learning iterations, the DL model had an AUC of 0.99 for the training

dataset and 0.98 for the corroboration dataset compared to the process based model. The calibration and corroboration results produced similar AUC results, indicating the model is capable of generalizing to unseen, new data (Fig. 5.2).

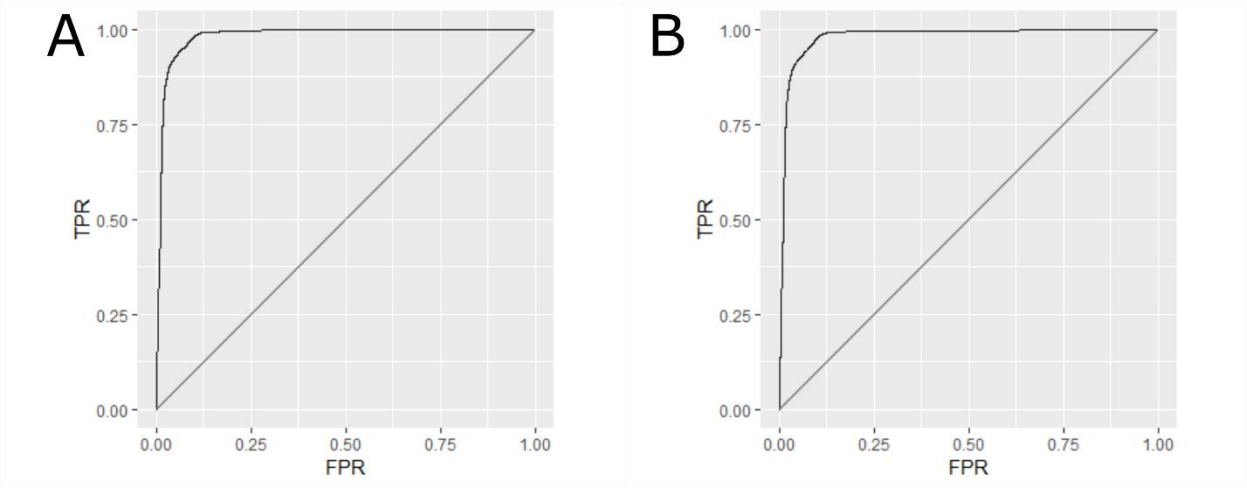


Figure 5.2. Training (A) and corroboration (B) AUC curves. Both cures exhibit similar shapes and values, suggesting a model which may be effective in generalizing to new data. On both graphs, the y-axis is True Positive Rate (TPR), and the x-axis is False Positive Rate (FPR).

The confusion matrix of the DL model shows high overall accuracy in both the training and corroboration sets (Table 5.2). The DL model performed slightly worse in estimating unsaturated classes (91-94% accurate) than saturated classes for both the training and corroborations sets (Table 5.2). Saturated areas were generally well predicted, with user's accuracy of 96-98%, and a producer's accuracy of 90-94%. Training and corroboration performance were similar, indicating a model which is able to generalize to new data.

Table 5.2. Confusion matrices for the training and corroboration datasets, comparing the Deep Learning model to the process based model.

Training	**Process Based Model**			
Deep Learning	Unsaturated	Saturated	Row Total	Producers Accuracy
Unsaturated	375220	14277	389497	**96.1**
Saturated	22382	367115	389497	**94.0**
Column Total	397602	381392		Overall Accuracy
Users Accuracy	**94.3**	**96.2**		**95.2**
Corroboration	**Process Based Model**			
Deep Learning	Unsaturated	Saturated	Row Total	Producers Accuracy
Unsaturated	21461	393	21854	**98.1**
Saturated	2108	19746	22247	**90.2**
Column Total	23569	20139		Overall Accuracy
Users Accuracy	**91.1**	**98.0**		**94.2**

Class probability distributions for the DL model compared to the process based model from the corroboration dataset are shown in Fig. 5.3 In both true positive and true negative cases, areas of high point density are seen near the top and bottom of the plot, indicating highly confident class separations. The DL model performed well when evaluated against observed saturated areas from Sommerlot et al. (2016), in which the fraction of the observed area that is classified as saturated by the model is interpreted as the probability that the whole area is saturated. Additionally, an observed saturated area that is covered by a 0.5 or greater fraction of predicted saturated labels is considered fully saturated. The results of this labeling process are shown in Fig. 5.4. All non-saturated areas were classified correctly, and four saturated areas were

classified incorrectly, there were no false positive predictions. Distribution shapes are shaded in light gray in the background, their widths corresponding to the frequency of the probability distribution in each data set.

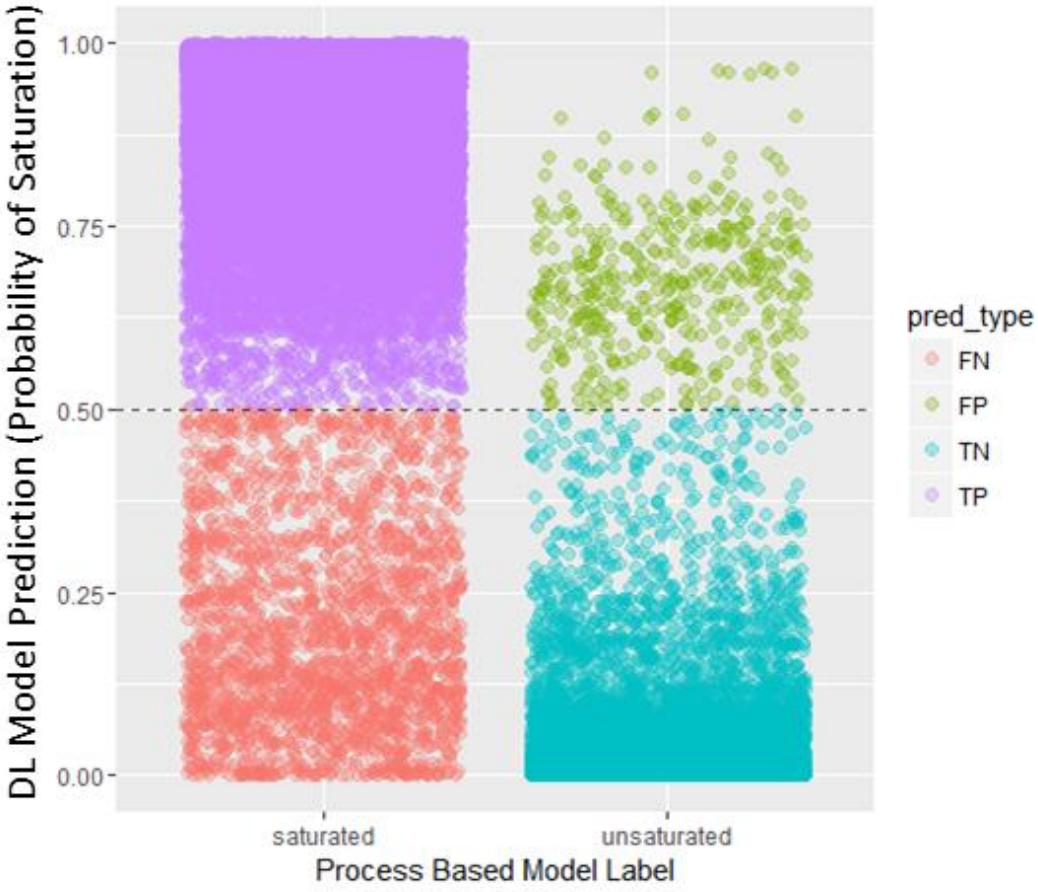


Figure 5.3. Deep Learning model output probability of saturated class compared with the process based model label for the corroboration dataset with 0 as unsaturated and 1 as saturated. The y-axis is the output of the DL model, which are predicted probabilities of saturation, the x-axis is the labeled class, a binary variable. The classification threshold line is drawn at 0.5. FN=False Negative, FP=False Positive, TN=True Negative, TP=True Positive,

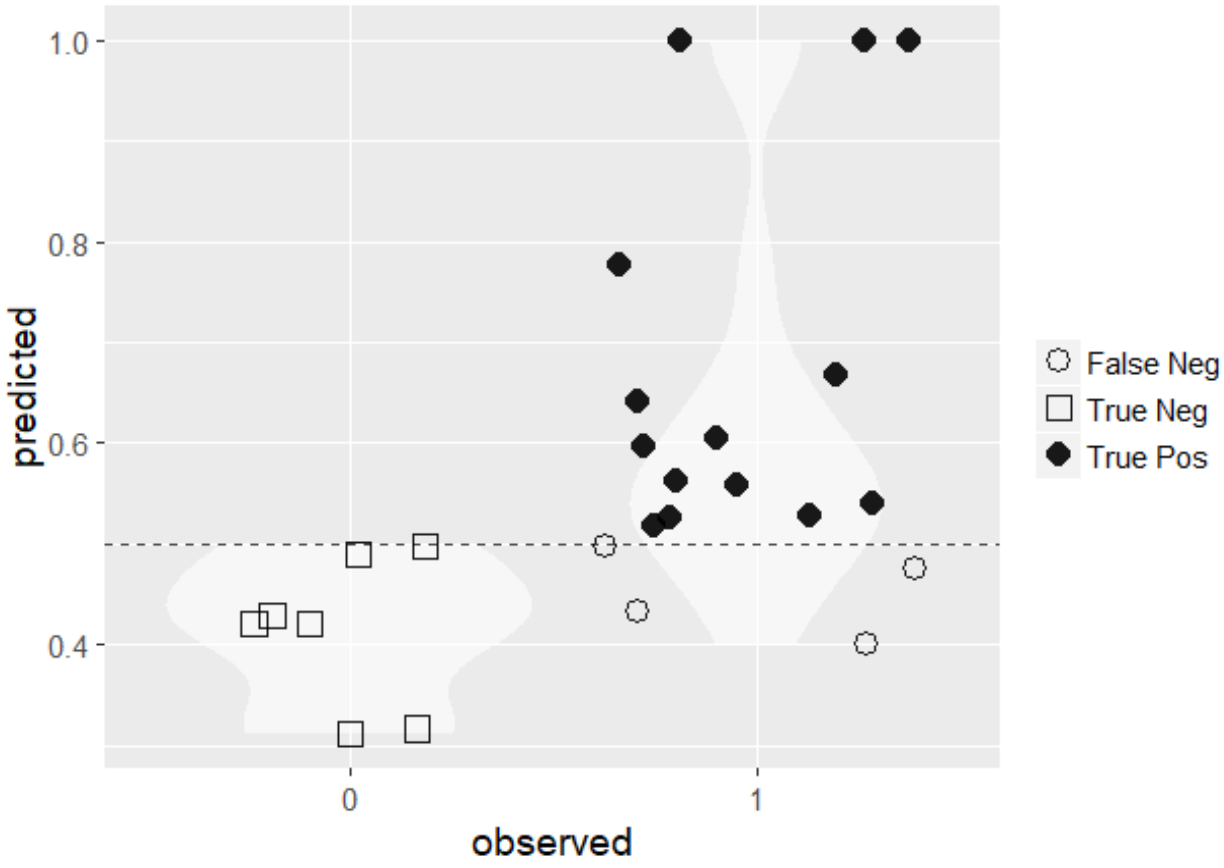
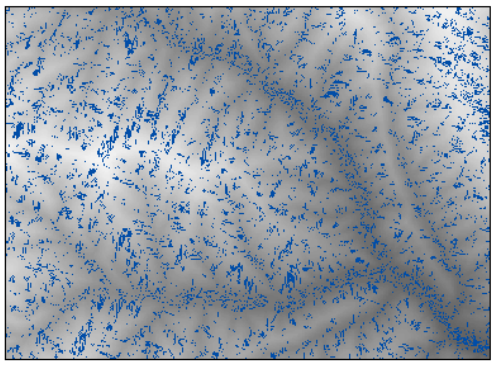


Figure 5.4. Class label distributions for the Deep Learning model compared to observed data following the process from Sommerlot et al. (2016). A 0 indicates an unsaturated area and a 1 a saturated area. The fraction of area that is saturated is interpreted as the probability that the whole area is saturated is represented on the y axis. The dashed line shows 0.5 probability threshold under which all points are predicted unsaturated and above which all points are predicted saturated.

5.4 Model Interpretation and Discussion

The DL model performed well, replicating the process based model with an overall accuracy of 98%. Thus, in an effort to better understand the processes represented in the DL model, several exploratory techniques were employed. First, incremental stages of the training process were observed and compared to the training labels from the process based model. The results of this process are shown in Fig. 5.5, which shows topography is a very important component for learning where saturated areas located, as even from an early learning epoch, major flow paths draining upland areas are visible. Beginning at 10 epochs, the DL model has identified major flow paths, but fails to make an effective differentiation between unsaturated and saturated classes, as is seen by the large saturated area predicted at the top of a high elevation area on the left side of Fig. 5.5. After 40 more epochs, an effective separation of classes is observed, but overall there are instances of over prediction, where some areas are more densely saturated than the process based model indicates. At 200 epochs, the shape and prevalence of flow paths are well defined, though the DL model still over predicts the density of these paths. Finally, at 700 epochs the shape, location, and density of the saturated areas are very similar to those predicted by the process based model. Overall, the DL model produces similar, but not identical results when compared with the process based training data, is roughly 50 times more spatially complex, and takes about several hundred epochs before behaving as if it understands the hydrology at a similar level as the process based model.



Process Based Model

Model Estimate

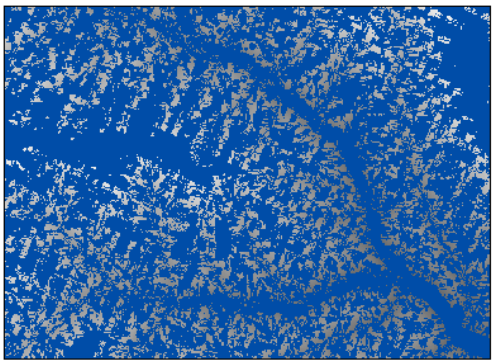
Unsaturated

Saturated

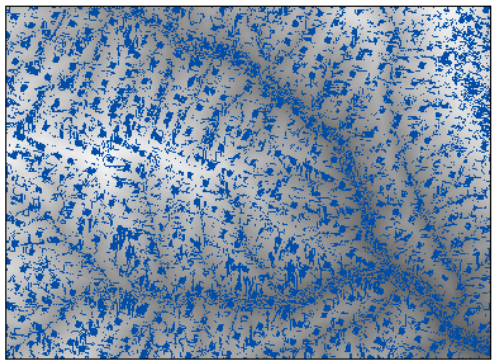
Elevation (m)

1100

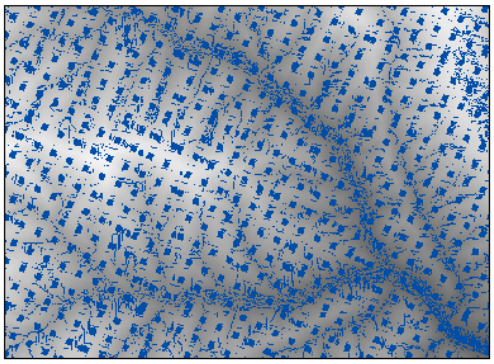
100



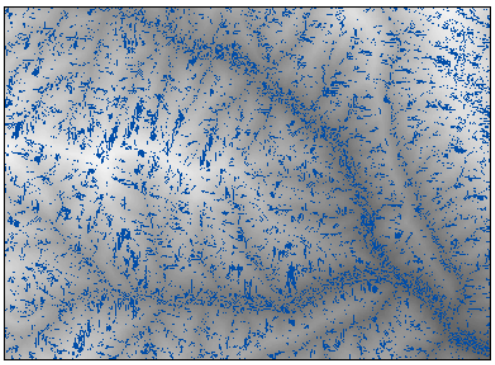
Deep Learning 10 Epochs



Deep Learning 50 Epochs



Deep Learning 200 Epochs



Deep Learning 700 Epochs

Figure 5.5. Saturation classifications from the Deep Learning model at various points in the training process. An epoch refers to a completed pass of a backpropagation iteration. The uppermost inset, shows the final calibrated/corroborated predictions from the Sommerlot et al (2016) process based model.

Using the deep layer extraction function in H20, values from neurons in the hidden layers corresponding to input values can be extracted. Yosinski et al. (2015) used deep layer extraction to assist in interpreting deep convolutional neural networks and asserted that studying these interim values allows researchers and practitioners to gain insight into how the input space is being altered in the trained DL model, and possibly gain insight into model behavior. The extracted values of hidden layer four contained 16 neurons, and thus 16 dimensions after extraction, was reduced to two dimensions using tSNE (Mohamed et al., 2014). The resulting two-dimensional scatter plot (Fig. 5.6) was then labeled with saturation (Fig. 5.6A), landuse (Fig. 5.6B), TI class (Fig. 5.6C), and forecast day precipitation (Fig. 5.6D) from the training dataset corresponding to each data point. The forecast day precipitation was classified into "0", "LOW", "MED", "M-HIGH" (medium-high), and "HIGH" classes to assist in interpreting the plot. An example of clear class separation is the small cluster of data points labeled "saturated" at the top of Fig. 5.6A. In Fig. 5.6B mixture of landuses exist all across the scatter plot indicating no single landuse dominates this saturated/unsaturated class separation. Figure 5.6C shows a small, top-most cluster of high TI class predictions, "08." Additionally, other clusters of saturated cases in Fig, 5.6A correspond to a group of TI classes in Fig. 5.6C consisting primarily of "10," "09," "08," and "07," all high (or saturated) classes. In Fig. 5.6C the same points are labeled with "HIGH" to "MED" precipitation classes thus higher precipitation dominates where saturated

classes are clearly separate from the unsaturated classes. Finally, in Fig. 5.6D the DL model understands that different precipitation levels form clearly different groups. Overall, Fig. 5.6 shows that landuses can be saturated or unsaturated (and do not really contribute to the model skill), that high TI classes are dominantly correlated to saturated cases compared to low TI classes, and that higher precipitation is correlated to saturation.

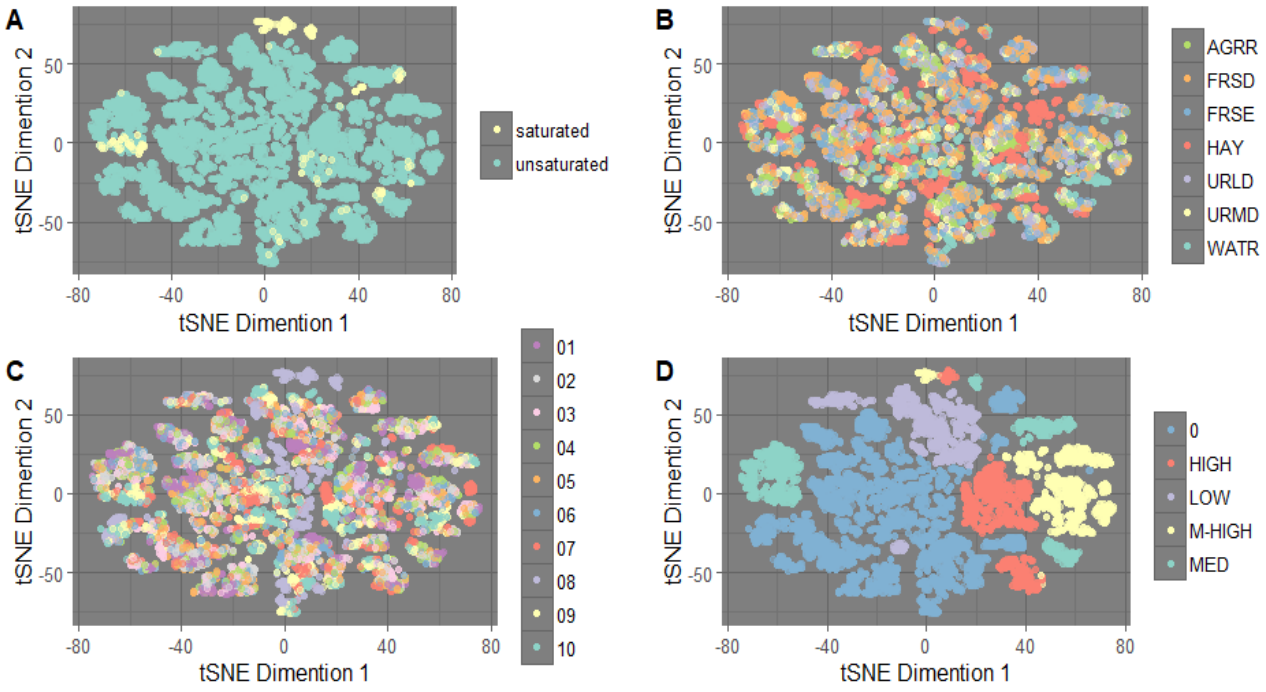


Figure 5.6: Interim training dataset values produced by extraction of values from the neurons of hidden layer four and tSNE dimensionality reduction to a 2D dataset. Both axes are unitless. A number of categorical variables provide the color scheme for each sub-figure and illustrate how the Deep Learning (DL) model is making discretization after learning from the training dataset. Precipitation is causing distinct groups to form in the DL model's hidden layers (D), TI class is correlated to saturation in isolated clusters (C), landuse remains mixed as almost any can be saturated (B), and the saturation classifications are dispersed in a way qualitatively consistent with VSA hydrology assumptions (A). The DL model is exhibiting types of behaviors that would be expected in a process based model.

Training dataset values from the two-neuron final layer (layer five) were extracted, plotted against the DL model output saturation probabilities, and labeled with the process based model saturated categories (Fig. 5.7). As expected, the points labeled with the saturated class cluster near the top and the unsaturated classes cluster near the bottom. Instances of error exist where a particular class has received a lower or higher probability than it should have based on its label, for example, some saturated labels are very near the 0 on the Y axis, and a number of unsaturated labels exist at relatively high values near 1 on the Y axis. Overall, many of the process based class labels are well represented by their corresponding DL model predicted probability. Additionally, clear shapes are exhibited by values from both neurons, with the head-and-tail curves due to the softmax function used to calculate the output probabilities in the DL model (Candel et al., 2016).

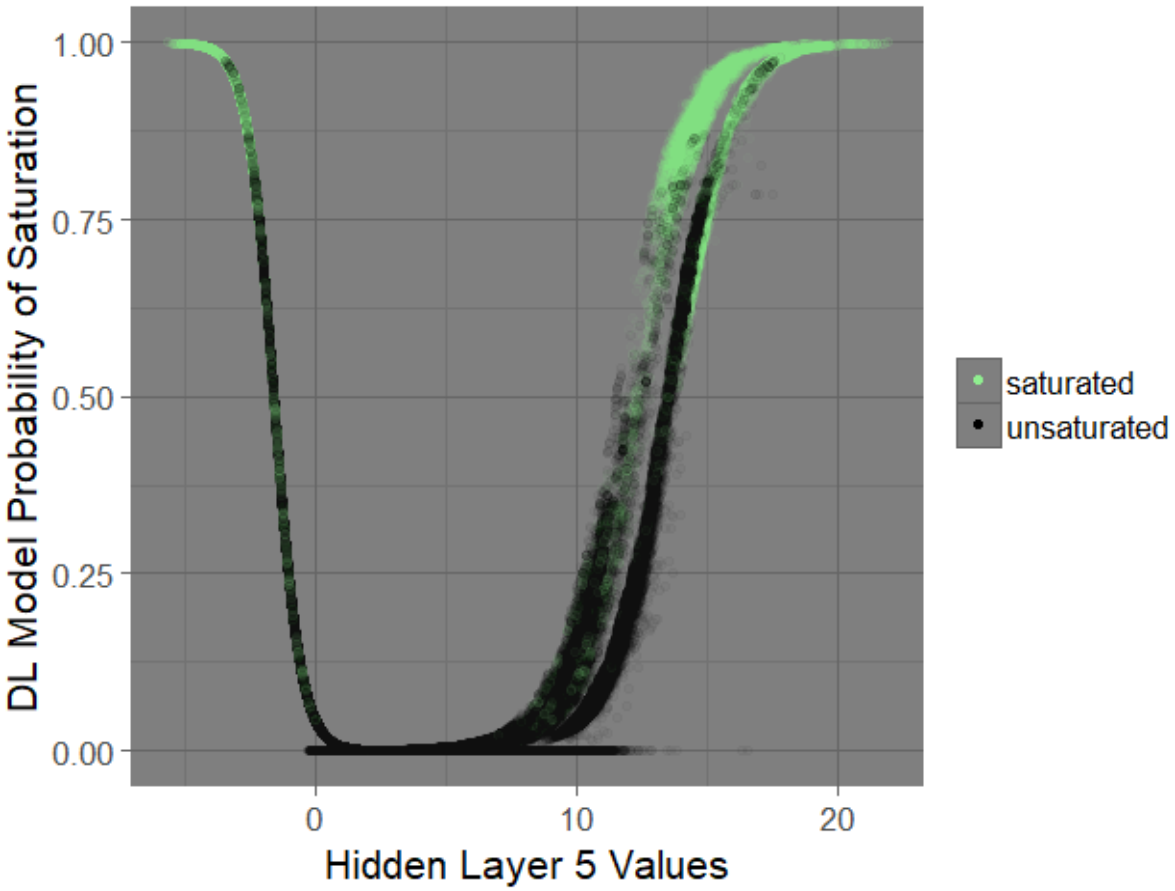

1.00
0.75
0.50
0.25
0.00
DL Model Probability of Saturation
0
10
20
Hidden Layer 5 Values
saturated
unsaturated

Figure 5.7. Interim values from both final hidden layer nodes (separated left/right). They are both similar, suggesting the possibility that the full dimensions are well represented. The corresponding processed based labels color the points, and the Deep Learning (DL) model output probability is on the Y axis. Most instances of saturated examples are seen at higher DL model probabilities, and the opposite is true for unsaturated examples. Near the bottom, there is a flat line, representative of areas which almost never saturate.

Proceeding further into interpretative efforts, a simple mathematical linkage attempting to represent the DL model was constructed. The following considers a multiple linear regression fit to interim hidden layer values linked to a non-linear regression describing the relationship between hidden layer values and the DL model probability estimates would constitute an equation extraction, representing a simplification of the DL model. Interim values from the second neuron of hidden layer five (Fig. 5.7) appear to have the greatest correlation to the DL probability estimates and were chosen for the extraction. In binary classification, H2O uses a softmax function to map hidden layer values to class estimates in the output layer (Candel et al., 2016). Given that softmax is a generalization of the logistic function, logistic growth was chosen as a function structure for the non-linear least squares regression relating DL saturation probability estimates to interim deep values from hidden layer five, neuron two:

$$p_{sat} = \frac{1}{1+ \alpha\, e^{k\, L_{5,2}}} \qquad \text{Eq. 5.1}$$

In Eq. 5.1, p_{sat} is the probability of the saturated class label, with possible values ranging from 0 to 1, and $L_{5,2}$ represents the interim values from the final hidden layer (layer five) produced by neuron two with the training dataset. After fitting, α and k, the regression parameters were found

to be 21.59 and 1.947, respectively. The line produced a fit with a 0.98 r-squared metric and normally distributed residuals.

The glmnet R package was used to estimate variable $L_{5,2}$ using an elastic net regression to from new data (Friedman et al., 2010). Real-valued inputs variables were used and automatically dropped with least absolute shrinkage and selection operator (LASSO), with the alpha parameter set at 0.5 (alpha ranges from 0, or ridge-like to 1, or LASSO-like), which is a balance between Ridge-like and LASSO-like parameter penalties (Friedman et al., 2010). Categorical variables were expanded into binary numerical variables and one class was left out as a base line, similar to logistic regression. The resulting elastic net linear regression produced an r-squared metric of 0.82 and displayed normally distributed residuals. Taking the equation obtained from this regression, the $L_{5,2}$ values can be estimated with coefficients (Table 5.3), and interpreted with a manner somewhat similar to the odds ratio of ordinary logistic regression:

$$OR_v = e^{-k\,C_v} \qquad \text{Eq. 5.2}$$

In Eq. 5.2, k is the regression parameter from Eq. 5.1 (1.947), C_v is the coefficient in Table 5.3 corresponding to variable v, and OR_v is the resulting ratio, similar to the odds ratio for variable v. In Table 5.3, all categorical coefficients are evaluated with respect to a base case, for landuse it was set to agriculture, and for TI class it was set to TI class 1. Precipitation odds ratios can be interpreted as increase in odds of saturation of an HRU per unit precipitation (in) it receives. Table 5.3 clearly shows that the forecast day precipitation is very important; the odds ratio corresponding to the antecedent day precipitation is still large, but much smaller than the forecast day. The landuse odds ratios are all less than one, meaning the odds of saturation decrease for every other landuse compared to agriculture, with hay and forest decreasing more than urban

landuses. TI class 10 increases the odds of saturation 366 compared to TI class 1, and the odds of saturation generally decrease with deciding TI class consistent with observations in the watershed (Sommerlot et al.,2016). These behaviors are sensible when considering the physical process that govern soil saturation, and they provide a meaningful interpretation through a simplification, which describes 82% of the variance of final hidden layer values of one neuron highly descriptive of the output probability.

Table 5.3. Coefficients from elastic net regression relating real valued inputs to $L_{5,2}$ and a metric analogous to an odds ratio. All categorical coefficients are evaluated with respect to a base case, for landuse it is agriculture, and for TI class it is TI class 1

Variable	**Coefficient to Estimate $L_{5,2}$**	**Odds Ratio for p_{sat}**
Intercept	5.05	NA
landuse Deciduous Forest	0.80	0.21
landuse Evergreen Forest	0.77	0.22
landuse Hay	0.90	0.17
landuse Developed Low Density	0.59	0.31
landuse Developed Medium Density	0.27	0.59

TI class 2	-0.48	2.57
TI class 3	-0.25	1.65
TI class 4	-0.51	2.71
TI class 5	-0.42	2.27
TI class 6	-0.34	1.95
TI class 7	-1.31	12.96
TI class 8	-1.63	24.12
TI class 9	-2.22	76.13
TI class 10	-3.03	366.56
Forecast Day Precipitation (in)	-5.29	30137.46
Forecast - 1 Day Precipitation (in)	-2.66	178.89

To create a complete linkage, $L_{5,2}$ from Eq. 5.1 is substituted with the regression equation defined by the coefficients in Table 5.3, yielding an equation for estimating the probability of saturation linked to real-valued inputs:

$$p_{sat} = \frac{1}{1+ \alpha\, e^{k \sum_{i=1}^{n} \beta_i V_i + C}} \qquad \text{Eq. 5.3}$$

where p_{sat} is the estimated probability of saturation, α is a regression parameter from Eq. 1 equal to 21.59, k is a regression parameter from Eq. 5.1 equal to 1.947, n is the number of variables in the elastic net regression (Table 5.3), i refers to the ith parameter and coefficient (Table 5.3), which are defined by β_i and V_i, respectively, and C is the intercept from Table 5.3. This equation is similar to the logit function in logistic regression, and in this application, serves the same purpose, to map an input variable ($L_{5,2}$) to a 0-to-1 range representing the probability of a

positive (in this case saturated) class (Glas et al., 2003; Pepe et al., 2004). In order to test this relationship, the equation was used to predict the probability of saturation and compared with the DL model estimates from the corroboration dataset. Comparing the predicted probabilities from the equation with the corroboration dataset labels yielded an AUC of 0.97 (Fig. 5.8). This is a surprisingly good performance, given the simplicity of the series of equations compared to the DL model. For reference, an ordinary logistic regression performed on the training dataset with the same variables provides an AUC of 0.42 when used to predict saturation probabilities for the corroboration dataset. Equation 5.3 can be interpreted as a custom logistic regression in which some of the "knowledge" contained within the DL model from hierarchal information extraction in the deep layers is summarized in the coefficients (β_i, α, and k) making it perform far better than a logistic regression alone.

Important limitations of the equations include applications containing new data outside the range of the training data. It is very unlikely that the equations would produce satisfactory results, for example, on data from a watershed unlike the study area or where weather patterns are different. However, this study provides evidence that hidden layer summary equation extraction can perform well over a relatively large spatial area for estimating the probability of a saturated HRU and additionally, a means to explore the relationships learned by the DL model. It is possible the equation extraction process described here would be useful to explore estimations over larger areas with different variables, or to predict outputs from different DL implementations, such as regression models. The result would be a simple to use equation that relates real-valued inputs to a desired outcome, provides evidence for interpreting the DL model it represents, explains a known amount of the variance in values from DL model hidden layer neuron(s), and has quantifiable performance on new data. This equation extraction process could

potentially be used in any discipline where DL models are employed, to assist in interpretation and to create equations which may assist in scientific exploration.

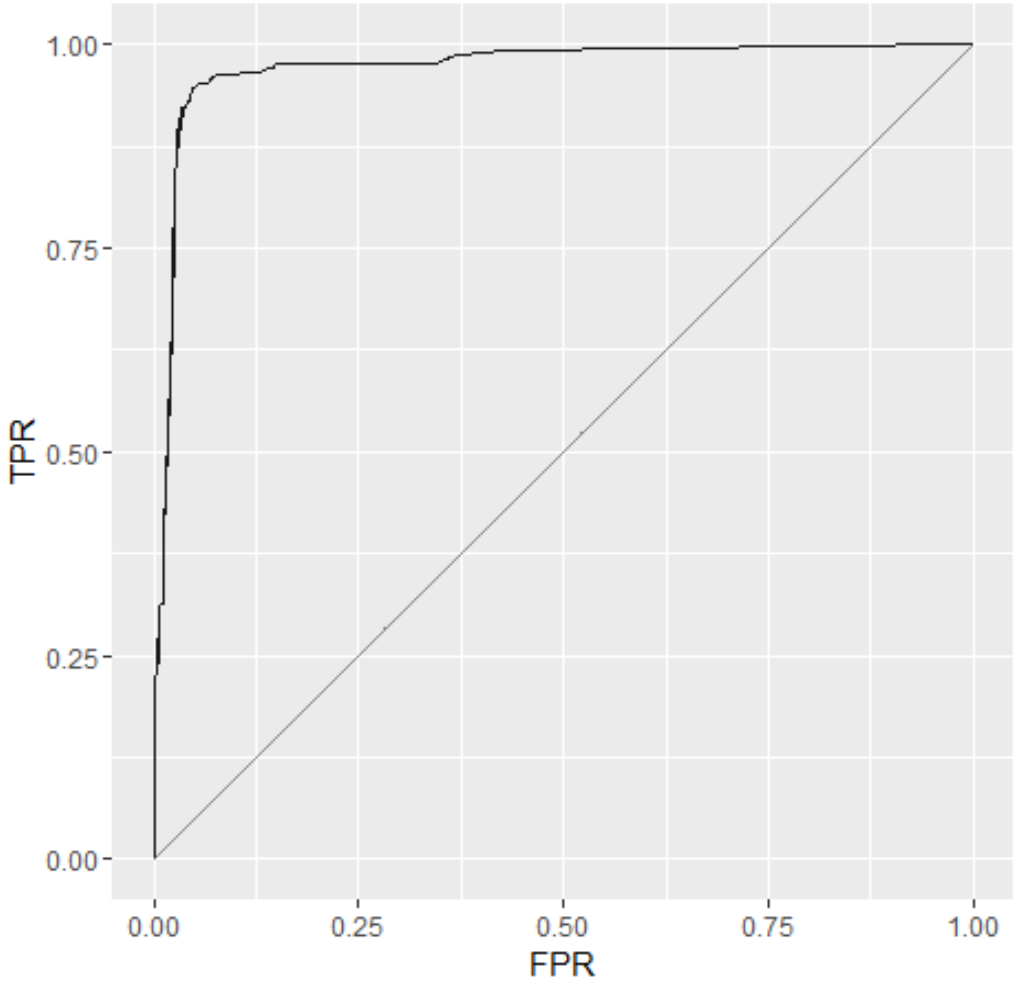


Figure 5.8. Area Under the receiver operating Curve (AUC) of the linkage equation on the corroboration dataset. The Y axis is True Positive Rate (TPF), and X axis is False Positive Rate (FPR). The AUC metric was 0.97, only 0.01 less than that of the Deep Learning model.

Conclusion

This study demonstrates training a DL model to reproduce calibrated and corroborated results from a process based model and provides acceptable predictions of saturated areas. The DL model can learn from the data and be used to explain underlying processes similar to a process based model. The DL model was able to extricate intricacies based on highly discretized input data and a learning process with high predictive power governed by comparatively few assumptions. The DL model provided satisfactory performance both predictively and functionally achieving a 9% increase in AUC score on observed saturated areas and a nearly 10-fold decrease in run time when compared with the process based model. Building, training, and iterating with various model structures and settings is much improved in modern implementations of DL models, many of which are free and open source including the one used in this study. The DL model was not immune to common problems in machine learning and required regularization in order not to overfit to the training data. Inferences into the relationships within the DL model can be made through extracting interim values from neurons in hidden layers and plotting them in low dimensional space. Additionally, a series of equations was extracted with linear elastic net and non-linear regressions relating real-valued inputs to estimated probability of saturation. This series of equations performed well making prediction on corroboration data, with an AUC metric of 0.97, only 0.01 less than that of the DL model.

Though process based models have been a cornerstone of hydrology for many years and will continue to serve in that capacity, recent improvements in modern machine learning, including new implementations of DL model architectures can demonstrate an impressive increase in functionality and predictive power when compared with process based models. Despite having a noticeable footprint in the literature, DL models have not been widely adopted

into routine hydrologic research due to a prevailing opinion that they are “black box.” The results in this study demonstrate that not only are DL models not entirely black box, but they provide a rich opportunity for interpreting relationships, a capability of learning processes similar to those observed in the physical world, and perhaps most importantly, they give researchers a tool with which to explore the unknown through the extraction of equations. As the depth, breadth, and sheer mass of data increase, it is of utmost importance for researchers and practitioners to keep an open mind to data driven modeling methodologies as we learn to apply them with ever-improving effectiveness, solving problems which increase our knowledge of known processes and contribute to the exploration of relationships yet to be defined.

5.5 Acknowledgements

We would like to acknowledge high-performance computing support from Yellowstone (http://n2t.net/ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, support from the National Science Foundation under award numbers 1360415 and 1343802, and funding support from the USDA under project number 2012-67019-19434.

5.6 References

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H. and Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology, 524, pp.733-752.

Abrahart, R.J. and See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrological processes, 14(11-12), pp.2157-2172.

Arnold, J.G., Moriasi, D.N., Gassman, P.W., Abbaspour, K.C., White, M.J., Srinivasan, R., Santhi, C., Harmel, R.D., Van Griensven, A., Van Liew, M.W. and Kannan, N., 2012. SWAT: Model use, calibration, and validation. Transactions of the ASABE, 55(4), pp.1491-1508.

Bai, Y., Chen, Z., Xie, J. and Li, C., 2016. Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. Journal of Hydrology, 532, pp.193-206.

Black, A. and Stephen, H., 2014. Relating temperature trends to the normalized difference vegetation index in Las Vegas. GIScience & Remote Sensing, 51(4), pp.468-482.Ganguly, N.D., 2009. Variation in atmospheric ozone concentration following strong earthquakes. International Journal of Remote Sensing, 30(2), pp.349-356.

Candel, A., Parmar, V., LeDell, E. and Arora, A., 2016. Deep Learning with H2O. H2O. ai Inc.

Chen, J., Jin, Q. and Chao, J., 2012. Design of deep belief networks for short-term prediction of drought index using data in the Huaihe river basin. Mathematical Problems in Engineering, 2012.

Coulibaly, P., Bobée, B. and Anctil, F., 2001. Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. Hydrological Processes, 15(8), pp.1533-1536.

Daniel, E.B., Camp, J.V., LeBoeuf, E.J., Penrod, J.R., Dobbins, J.P. and Abkowitz, M.D., 2011. Watershed modeling and its applications: A state-of-the-art review. The Open Hydrology Journal, 5(1).

Dawson, C.W. and Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. Progress in physical Geography, 25(1), pp.80-108.

David, L.M., Oliveira, A., Rodrigues, M., Jesus, G., Póvoa, P., David, C., Costa, R., Fortunato, A., Menaia, J., Frazão, M. and Matos, R., 2013. Development of an integrated system for early warning of recreational waters contamination. Novatech 2013.

Dawson, C.W. and Wilby, R., 1998. An artificial neural network approach to rainfall-runoff modelling. Hydrological Sciences Journal, 43(1), pp.47-66.

Easton, Z.M., D.R. Fuka, M.T. Walter, D.M. Cowan, E.M. Schneiderman, and T.S. Steenhuis. 2008. Re-Conceptualizing the Soil and Water Assessment Tool (SWAT) model to predict runoff from variable source areas. J. Hydrol. 348: 279-291.

Easton, Z.M., P.J. Kleinman, A.R. Buda, D. Goering, N. Emberston, S. Reed, P.J. Drohan, M.T. Walter, P. Guinan, J.A. Lory, A.R. Sommerlot, A. Sharpley. 2017. Short-term forecasting tools for agricultural nutrient management. J. Environ. Qual. doi:10.2134/jeq2016.09.0377.

Friedman, J., Hastie, T. and Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1).

Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B. and Jones, N., 2016. An overview of current

applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology, 537, pp.45-60.

Glas, A.S., Lijmer, J.G., Prins, M.H., Bonsel, G.J. and Bossuyt, P.M., 2003. The diagnostic odds ratio: a single indicator of test performance. Journal of clinical epidemiology, 56(11), pp.1129-1135.

Homer, C.G. Dewitz, J.A., L. Yang, L. Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sens., 81. pp. 345–354

Hsu, K.L., Gupta, H.V. and Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Research, 31(10), pp.2517-2530.

Huang, K., Zhuang, G., Xu, C., Wang, Y. and Tang, A., 2008. The chemistry of the severe acidic precipitation in Shanghai, China. Atmospheric Research, 89(1), pp.149-160.

IUSS W.G.W., World Reference Base for Soil Resources 2006, First Update 2007. 2007. (No. 103). World Soil Resources Reports. Rome

Jain, A., Sudheer, K.P. and Srinivasulu, S., 2004. Identification of physical processes inherent in artificial neural network rainfall runoff models. Hydrological Processes, 18(3), pp.571-581.

Jiang, H. and Cotton, W.R., 2004. Soil moisture estimation using an artificial neural network: a feasibility study. Canadian Journal of Remote Sensing, 30(5), pp.827-839.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553), pp.436-444.

Luk, K.C., Ball, J.E. and Sharma, A., 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. Journal of Hydrology, 227(1), pp.56-65.

Mohamed, A.R., Hinton, G. and Penn, G., 2012, March. Understanding how deep belief networks perform acoustic modelling. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on (pp. 4273-4276). IEEE.

Mohamoud, Y., 2004. Comparison of hydrologic responses at different watershed scales (pp. 1-81). Office of Research and Development, United States Environmental Protection Agency.Nagy, H.M., Watanabe, K.A.N.D. and Hirano, M., 2002. Prediction of sediment load concentration in rivers using artificial neural network model. Journal of Hydraulic Engineering, 128(6), pp.588-595.

Nayak, P.C., Rao, Y.S. and Sudheer, K.P., 2006. Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resources Management, 20(1), pp.77-90.

NCEP, 2003. Note: The GFS Atmospheric Model, vol. 442, pp. 14

Pepe, M.S., Janes, H., Longton, G., Leisenring, W. and Newcomb, P., 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. American Journal of Epidemiology, 159(9), pp.882-890.

Glas, A.S., Lijmer, J.G., Prins, M.H., Bonsel, G.J. and Bossuyt, P.M., 2003. The diagnostic odds ratio: a single indicator of test performance. Journal of Clinical Epidemiology, 56(11), pp.1129-1135.

Rajurkar, M.P., Kothyari, U.C. and Chaube, U.C., 2004. Modeling of the daily rainfall-runoff relationship with artificial neural network. Journal of Hydrology, 285(1), pp.96-113.

Retchless, D., Frey, N., Wang, C., Hung, L.S. and Yarnal, B., 2014. Climate extremes in the United States: recent research by physical geographers. Physical Geography, 35(1), pp.3-21.

Samaranayake, S., Glaser, S., Holstius, D., Monteil, J., Tracton, K., Seto, E. and Bayen, A., 2014. Real-Time Estimation of Pollution Emissions and Dispersion from Highway Traffic. Computer-Aided Civil and Infrastructure Engineering, 29(7), pp.546-558.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural etworks, 61, pp.85-117.

Sommerlot, A.R., M.B. Wagena, D.R. Fuka and Z.M. Easton. 2016. Coupling the short-term Global Forecast System weather data with a variable source area hydrologic model. Environ. Model. Software. http://dx.doi.org/10.1016/j.envsoft.2016.09.0081364-8152.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), pp.1929-1958.

Song, X., Zhang, G., Liu, F., Li, D., Zhao, Y. and Yang, J., 2016. Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. Journal of Arid Land, 8(5), pp.734-748.

Sajikumar, N. and Thandaveswara, B.S., 1999. A non-linear rainfall–runoff model using an artificial neural network. Journal of Hydrology, 216(1), pp.32-55.

Valipour, M., Banihabib, M.E. and Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, pp.433-441.

Wang, Y., Zhuang, G., Sun, Y. and An, Z., 2006. The variation of characteristics and formation mechanisms of aerosols in dust, haze, and clear days in Beijing. Atmospheric Environment, 40(34), pp.6579-6591.

Wei, Q. and Dunbrack Jr, R.L., 2013. The role of balanced training and testing data sets for binary classifiers in bioinformatics. PloSOne, 8(7), p.e67863.

Wilby, R.L., Abrahart, R.J. and Dawson, C.W., 2003. Detection of conceptual model rainfall—runoff processes inside an artificial neural network. Hydrological Sciences Journal, 48(2), pp.163-181.

Wong, B.P. and Kerkez, B., 2016. Real-time environmental sensor data: An application to water quality using web services. Environmental Modelling & Software, 84, pp.505-517.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, H., 2015. Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579.

Yuan, Y. and Jia, K.B., 2016. A Semi-supervised Approach for Water Quality Detection based on IoT Network. J. Inf. Hiding Multimed. Signal Process, 7, pp.858-866.

CHAPTER 6. A Bayesian Approach to Estimating Long Term Annual Agricultural Soil Phosphorus Concentrations Over the Chesapeake Bay Watershed*

*Under USGS Internal Review, Citation: A.R. Sommerlot, G. Shenk, G. Yacatyow, G. Bhatt, Z.M. Easton. 2017.

Keywords: APLE, Uncertainty, soil phosphorus modeling, long-term

6.0 Abstract

Long-term phosphorus (P) applications to agricultural soils in the Chesapeake Bay Watershed have resulted in many soils saturated with P. These P saturated soils are a major legacy source for P that are difficult to control. Most current control efforts have been aimed at reducing P input to the soils; however, these control efforts are exceedingly difficult to quantify, usually necessitating many years of high spatial and temporal frequency sampling, which unfortunately, does not exist in the Chesapeake Bay watershed. Furthermore, privacy issues (landowners not wanting to make sensitive information about their property known) surrounding data collection make interpreting soil test P at a scale necessary to drive management decision exceedingly difficult. This study presents a method to estimate annual soil P concentration histories for 13 agricultural landuses across the entire Chesapeake Bay Watershed at the county level by developing a Bayesian inference model, incorporating two sources of soil P information, each with its own uncertainty; 1) a combinatory census of multiple soil P concentration data sets from multiple watershed sources, and 2) estimates from the Annual Phosphorus Loss Estimator (APLE), the NRCS approved soil P estimation method. These two sources were used to provide estimates of soil P concentrations over a 31-yr period and the associated uncertainty in those estimates. The final estimates in mg per kg covered all years and landuses ranging from ~6 to ~975 with a median of ~84 and a mean of ~96. Uncertainty estimates of soil P concentrations were quantifiable via this framework, and tended to decline in areas of the watershed with greater sampling density, or a longer history of soil test P sampling. These results can be can be used by policymakers and land managers to prioritize landscape level soil test P sampling intensification and to drive management aimed at reducing or controlling P loss from critical P

source areas, or as input to the Chesapeake Bay Watershed model to quantify the potential P reductions from attenuation measures.

6.1 Introduction

Phosphorus (P) loss from agricultural lands is a major threat to surface water quality in the United States (Sharpley et al., 2015; Dubrovsk et al., 2010). This problem is especially acute in the Chesapeake Bay region in the Eastern United States, where high population densities and legacy effects all contribute to high levels of P in many soils of the region (USEPA, 2010). Excess P can trigger algal blooms in waterways causing fish kills, reduction in water quality and clarity, and ultimately harm local industries including fisheries and recreation, which are valuated at billions of dollars per year in the region (Phillips et al., 2016).

Soil P concentrations in agricultural soils are a major legacy source for potentially polluting species of P, and monitoring and/or modeling this concentration provides important metrics for policy makers in designing efforts to keep excess P out of the Bay waters (Sharpley et al., 2015). Monitoring the soil P levels consistently and completely across the Bay has, and will continue to be a challenge, as multiple states and jurisdictions are included in the monitoring area, protection of land owner privacy is important, and obtaining data from many locations is dependent on voluntary participation (Bishop et al., 2005). Even though many soil P tests have been conducted over the years, no large scale, long term effort has been made to evaluate these data in a comprehensive manner at the spatial and temporal scales needed to better inform

management practices, and potentially estimate the future impact of current P control measures (Skown et al., 2017; Yen et al., 2016).

This study presents a method to estimate annual soil P concentration histories, and the associated uncertainty in those estimates, for 13 agricultural landuses over a 31-yr period across the entire Chesapeake Bay watershed at the county level. Two sources of soil P information are used in this work: 1) a combinatory census of multiple soil P concentration data sets, and 2) estimates from the Annual Phosphorus Loss Estimator (APLE), the NRCS approved soil P estimation method was used to provide estimates of soil P concentrations.

A hierarchal Bayesian framework was constructed in an effort to address the shortcomings of both soil test aggregation and APLE model estimates in describing the soil P concentrations. The Bayesian method was designed to incorporate the strengths of each of the data sets into annual estimate of soil P, account for uncertainties in these estimates, providing important assumptions and information leading to credible estimation intervals for each soil P estimate. Methods incorporating Bayesian methods in fusion of multiple sources of estimations have been successfully used to fill temporal gaps in data or correct for bias, and have the advantage of being flexible enough to use with varying quality and types of data in the same statistical process (Qu and Wang, 2012; Pierdicca et al., 2010).

6.2 Methods

6.2.1 Soil P Tests

Soil tests P concentrations from six sources (Table 6.1) were aggregated through a four-step process: 1) each data point was converted from the base soil test P measurement method to Mehlich III soil P (Mehlich, 1984), 2) each data point was assigned an agricultural landuse (grain

with manure, legume hay, agronomic crops, other hay, pasture, specialty crop high, specialty crop low, small grains and grains, full season soy, silage with manure), 3) land segment locations within the watershed were assigned to each data point, and 4) for each landuse/land segment combination data distribution shape and uncertainty level were quantified.

Uncertainty was quantified on two levels: 1) by source, where data derived from the different providers in Table 6.1 were assigned uncertainties based on a combination of expert opinion, and the level of interannual variation between pairs of data points collected within the same field (recognizing that intra-field interannual variations in soil test P of more than 10% are exceedingly unlikely, (Sharpley et al., 2015), and 2) at the land segment/landuse combination level, where variance was added through scaling the uncertainty of individual data points based on their distance from the central tendency of the discretized set. The final aggregated dataset was a collection of normal distributions centered about each soil P data point with a standard deviation defined by the two-tier estimation of uncertainty, described above. The resulting dataset represents a large number of soil tests P concentrations converted to a common measurement unit and labeling scheme, covering a large spatial and temporal range, and a guide for interpreting these data based on an estimate of their uncertainty defined through distributions.

Table 6.1: Soil Test Data Sources

SOURCE	YEARS	LOCATION	UNITS	TYPE
AgriAnalysis*	2003 - 2014	DE, MD, NY, PA, VA, WV	lbs P/ac	by county & zip code
Penn State [1]	2001 - 2014	PA	Mehlich III soil P (ppm)	by county and by crop
Virginia Tech[1]	2012	VA	Mehlich III soil P (ppm)	by county and by crop
U Maryland 1*	1954 - 2002	MD	Percent of samples in Mehlich I range	by county
U Maryland 2	1992	DE, MD, NY, PA, VA, WV	Mehlich III soil P (ppm)	by county
U Delaware	1992 - 2015	DE	P-FIV, Equal to 0.5 X Mehlich III lbs/ac	by county

*Required conversion to Mehlilch III

[1] source includes landuse metadata

6.2.2 Mehlich III Conversions

Separate conversion methods were used to convert base soil test P methods (Table 6.1) into in Mehlich III soil P. Mehlich III is a soil P quantification method and is cited as a standard metric for quantifying P concentrations in agricultural soils (Donohue, 1992). Soil test P sources that needed conversion were the AgriAnalysis test, the University of Maryland 1 test, the Virginia Tech test, and the University of Delaware test. Units of pounds (lbs) per acre (ac) were reported by AgriAnalysis and were converted to Mehlich III mg per kg using equation 6.1, approved by the state of Virginia (Virginia Soil and Water Conservation Board, 2014):

$$\boldsymbol{C_{m3} = \frac{1}{2} * M_{m3}} \qquad \textit{Equation (6.1)}$$

where $\boldsymbol{C_{m3}}$ is the Mehlich III P concentration in the soil (ppm) and $\boldsymbol{M_{m3}}$ is the Mehlich III mass of P (lbs) per unit area, (ac) of soil. Although reported in Mehlich III , the Virginia Tech data contained documentation indicating the original values were collected in Mehlich I and converted to Mehlich III using one of two equations (Equation 6.2), depending on the value of Mehlich I being converted (Commonwealth of Virginia, 2005).

$$\boldsymbol{C_{m3} = \begin{cases} \frac{(C_{m1}+3.26)}{0.458} & C_{m1} < 90.63 \\ \frac{(C_{m1}+103.5)}{0.945} & C_{m1} \geq 90.63 \end{cases}} \qquad \textit{Equation (6.2)}$$

Equation 2 converts $\boldsymbol{C_{m1}}$, Mehlich I soil concentration (ppm) to $\boldsymbol{C_{m3}}$, Mehlich III soil concentration (ppm). Data from University of Maryland 1 test were also converted from the

original reporting units to Mehlich III. These data were reported as percent of samples in three Mehlich III ranges, low, 0-25 ppm, medium 26-50 ppm, optimum, 51-100 ppm, and three separate excessive ranges, 101-150, 151-250, and greater than 250 ppm. Additionally, the data spanned a change in category convention in the source, and samples summarized from years earlier than 1998 were only categorized as high, medium, and low, qualitative labels without listed ranges. The low, medium, and first excessive range from the newer data in this set were used to define the values of the older test summaries. A weighted average (Equation 6.3) using the percent of samples and the center value of the defined ranges to assign a single value estimate for each summary of observations was employed:

$$\boldsymbol{C_{m3} = (P_1)C_{r1} + (P_2)C_{r2} + \ldots (P_x)C_{rx}} \qquad \textit{Equation (6.3)}$$

where $\boldsymbol{C_{m3}}$ is soil Mehlich III concentration (ppm), $\boldsymbol{P}$ represents reported percentage of soil tests the fell within their corresponding Mehlich III range $\boldsymbol{C_r}$ (ppm) as defined above, and *x* is the total number of ranges in the reported, equal to six for reports from 1998 to 2002, and three for data from 1997 and older. Summing all ***P's*** for a single calculation always equals 100%. Data in the University of Delaware set were reported in P-FIV, a fertility index value equal two times the Mehlich III P measurement in lbs per acre. Thus, the data were converted to Mehlich III ppm with Equations 6.4 and 6.5:

$$\boldsymbol{M_{m3} = 2\,PFIV} \qquad \textit{Equation (6.4)}$$

$$\boldsymbol{C_{m3} = * \frac{1}{2} M_{m3}} \qquad \textit{Equation (6.5)}$$

where $\boldsymbol{C_{m3}}$ is the Mehich III soil P concentration (ppm), $\boldsymbol{M_{m3}}$ is the Mehlich III mass in lbs per ac, and $PFIV$ is the soil test report value in University of Delaware P fertility index value units.

6.2.3 Landuse Categorization

The landuse associated with each data point was categorized based on crop types reported in the soil test P data sources and included agricultural classifications such as harvested grains, soybeans, vegetables, or grasses. In order to assign landuse categories, crop types were related to landuse labels. This was accomplished by translating landuse definitions adopted from the Chesapeake Bay Model's definitions of crop type to landuse based on similarities between crop types (Figure 6.1). For example, in the adopted definitions, the crop type "Alfalfa Harvested Area" is defined as landuse "legume hay," in the soil P test sets, this relationship was used to label crop types from the Penn State and Virginia Tech sources "Planting Alfalfa" and "Alfalfa, Alf-Grass – Estab" as landuse "legume hay." The resulting categorizations are shown Tables 6.2 and 6.3.

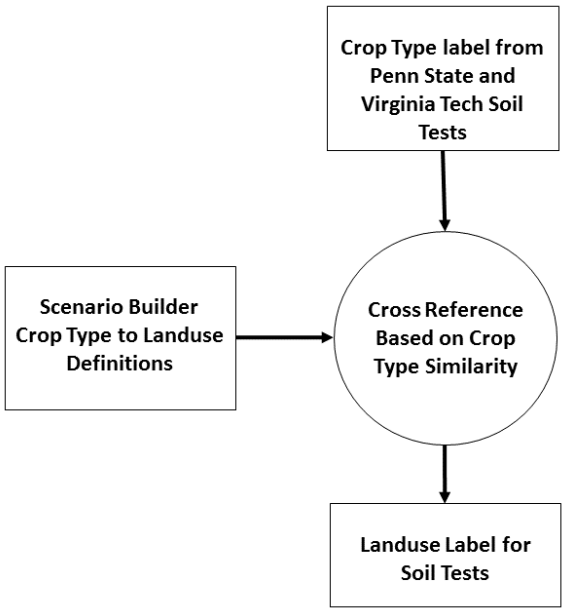


Figure 6.1: Illustration of decision process used to categorize crop types into landuses.

Categorized in this way, data already containing crop type labels could be compared across landuses and data sources. However, since using this method required some subjective rules, statistical tests were developed to assess the efficacy of the labeling methodology. The Penn State University source included data in the state of Pennsylvania (PA), and the Virginia Tech source contained data for the state of Virginia (VA). A test was designed to leverage this source difference as a means to assess the landuse categorization process. Under the supposition that soil test P concentration data labeled with landuse categories reflecting reality would display significant differences between data from single landuse categories compared with unlabeled data, the concentration data were tested for significant differences. Distributions were non-normal and contained different numbers of samples, thus the Mann-Whitney U test was chosen

to assess the differences between the data sets (Mann and Whitney, 1947). The Mann-Whitney U test is similar to the t-test, but is non-parametric and appropriate to use when the data do not meet the required normality assumptions (McKnight and Najab, 2010). Test of hypothesis 6.1 was performed for the 10 landuses present in the PA and VA data sets after the labeling process, and the hypotheses were formulated as follows:

Null Hypothesis 6.1 – Data sharing a common landuse label have the same distribution of soil test P as unlabeled data

Alternate Hypothesis 6.1 – Data sharing a common landuse label do not have the same distribution of soil test P as unlabeled data

Test of hypothesis 6.1 was performed separately for all landuses in both PA and VA, always comparing against the overall dataset as if the data were unlabeled. Test of hypothesis 6.2 was used to check for the presence of factors affecting the data that would make inter-state interpretation of the results unadvisable. Test of hypothesis 6.2 was performed for the 10 landuses present in the PA and VA data sets after the labeling process, and the hypotheses for test 6.2 were formulated as follows:

Null Hypothesis 6.2 – Data from VA and PA sharing a common landuse label have the same soil test P distribution

Alternate Hypothesis 6.2 – Data from VA and PA sharing a common landuse label do not have the same soil test P distribution

6.2.4 Spatial Discretization

Most of the soil test P data were reported at the county level excluding the AgriAnalysis data, which reported data by zip code. In order to maintain a consistent spatial extent among data sources the AgriAnalysis data were aggregated by mean to the county level by defining the zip code to Federal Information Processing Standard codes from the Untitled States Census, which correspond to counties (US Census, 2010). In order to obtain one value for each year in every landuse/county combination, data were averaged across source when there were overlapping values. Thus, each landuse county combination can have a combination of averages across sources and/or various data from single sources depending on overlap. This method aggregated soil test P from all sources into a single data set representing annual county average Mehlich III P concentrations for the 10 labeled landuses.

6.2.5 Uncertainty Definitions and Distributions

Uncertainties were defined at two levels, one through soil test P source, considering the reporting structures, goals of the data collection, and if/how the results were converted to Mehlich III values. These uncertainties were based on how much preprocessing the data from each source were subject to before being represented as real valued Mehlch III values. Data points were then assumed (assumption 6.2) to be the medians of normal distributions with a standard deviation defined by data source and state. From these general guidelines (Table 6.2), standard deviations were developed to represent reasonable variations following the guidelines discussed above for each distribution (Table 6.3).

Table 6.2: Uncertainty guidelines for each data source

SOURCE	TIME RANGE	STATE	RELATIVE UNCERTAINTY
AgriAnalysis	2003 - 2014	DE, MD, NY, PA, VA, WV	Medium
Penn State University	2001 - 2014	PA	Medium Low
Virginia Tech	2012	VA	Medium High
University of Maryland 1	1954 - 2002	MD	High
University of Maryland 2	1992	DE, MD, NY, PA, VA, WV	Medium to High
University of Delaware	1992 - 2015	DE	Medium

Table 6.3: Standard deviations assigned by source and state

SOURCE	STATE	STANDARD DEVIATION
AgriAnalysis	DE	25
AgriAnalysis	MD	25
AgriAnalysis	NY	25
AgriAnalysis	PA	25
AgriAnalysis	VA	25
AgriAnalysis	WV	25
Penn State University	PA	15

Virginia Tech	VA	30
University of Maryland 1	MD	50
University of Maryland 2	DE	25
University of Maryland 2	MD	40
University of Maryland 2	NY	50
University of Maryland 2	PA	15
University of Maryland 2	VA	50
University of Maryland 2	WV	20
University of Delaware	DE	20

Assumption 6.1 – The true state of annual soil P concentration for a particular land segment/landuse combination exists within a normal distribution centered about the soil test P data point that exists for that year.

Assumption 6.2 – The true state of annual soil P concentration for a particular land segment/landuse combination exists within a uniform distribution bounded by 0 and 2000 when no soil test data point exists for that year.

From the first level of uncertainty, each individual data point received an uncertainty scaling factor based on the central tendency of the final discretization in which it was assigned. Thus, for every land segment/landuse combination, the first level uncertainties were scaled with equation 6.6 to attenuate the influence of outliers.

$$\boldsymbol{\sigma_{scaled1}} = \boldsymbol{\sigma_{unscaled}} \left(\frac{\mathbf{1.01}}{\mathbf{1}}\right)^{\boldsymbol{abs(mean(S)-s_1)}} \qquad \textit{Equation (6.6)}$$

Equation 6.6 is defined by the final, annual scaled standard deviation ($\boldsymbol{\sigma_{scaled1}}$), which is unique for each year, the original standard deviation from the level 1 definition ($\boldsymbol{\sigma_{unscaled}}$), which is not unique for each year, the set including all annual soil test values for a single landuse/land segment combination after the complete aggregation process ($\boldsymbol{S}$), and the single soil test P value corresponding to the final unique annual scaled standard deviation. In this way, uncertainties were adjusted based on the distributions of the individual land segment/landuse combination soil test P sets resulting from the aggregation process. The scaling factor, $\left(\frac{\mathbf{1.01}}{\mathbf{1}}\right)$ and equation shape were chosen empirically based on sensible reductions of the influence of outliers.

Using the process described above and assumptions 6.1 and 6.2, a collection of distributions containing the true state of soil P concentrations over the entire time series for every land segment/landuse combination for the Chesapeake Bay were defined. Although this collection contains some uninformative distributions where there are missing data, it describes the probabilities of the soil test P values in a useful manner. When coupled with a quantification of the probability of inter-annual change in the state of the soil P concentrations, a new estimate can be produced that presents a reasonable combination of the strengths of the APLE model estimates and the aggregate data set attenuated by the defined uncertainties. Figures 6.2 and 6.3 summarize the aggregation process converting a group of soil test P concentrations from different sources into a single set of distributions.

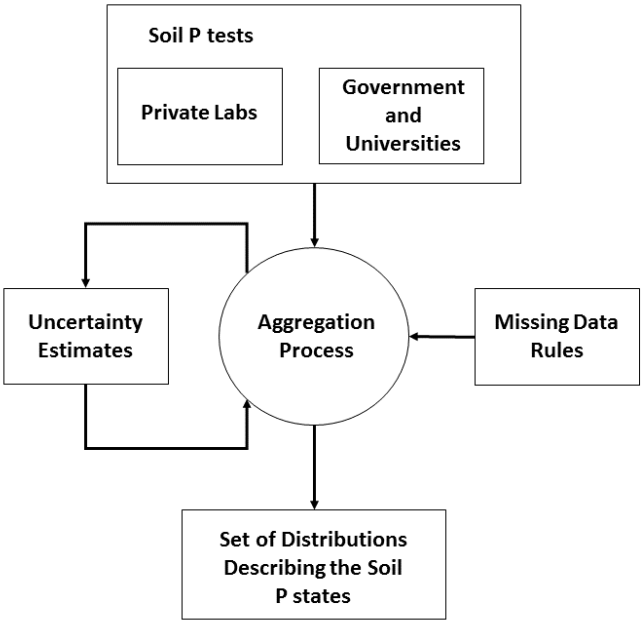


Figure 6.2: Summary of the overall process of combining evidence into soil P distributions.

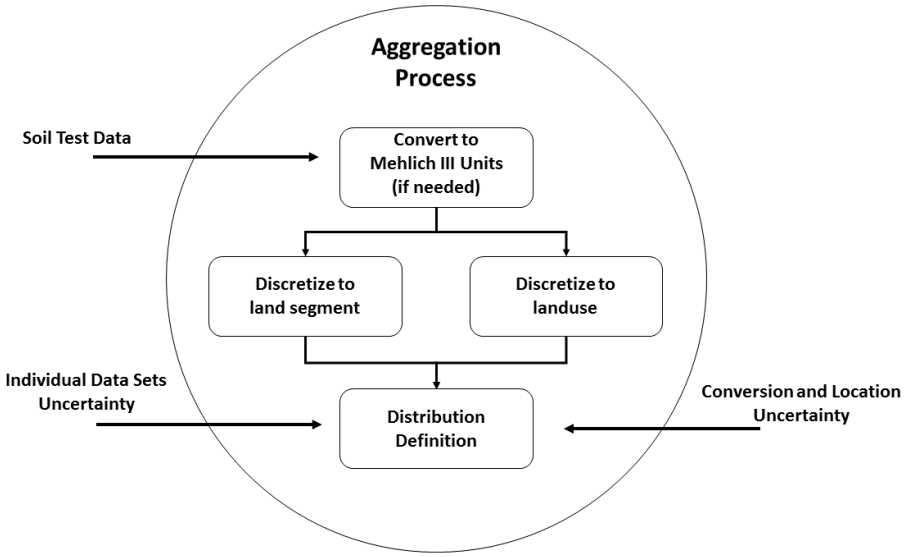


Figure 6.3: Summary of the aggregation process and conversion of soil test data sets into a complete set of distributions.

6.2.5 APLE Model Estimates

The APLE model was used to provide soil P estimates for every landuse and county combination. APLE is a mass balance based model that provides annual P loss estimates from dissolved fertilizer P, dissolved manure P, and P loss from soil bound nutrients along with soil P concentration estimates (Vadas et al, 2009). Recent applications of this model include the evaluation of state P indices and farm-scale P loss estimations (Bolster, 2011; Nelson and Shober, 2012; Vadas et al., 2015). APLE estimates P soil storage with the following equation (Vadas et al., 2009):

$$\boldsymbol{P_{soil} = ((27 * OM_{soil}) + (0.06 * LP_{soil}))^2} \qquad \textit{Equation (6.7)}$$

where $\mathbf{P_{soil}}$ is the total P content of surface soil (mg per kg), $\mathbf{OM_{soil}}$ is the organic matter content of the surface soil (g per kg), and $\mathbf{LP_{soil}}$ is soil labile P content of the surface soil (mg per kg). Inputs to APLE include annual manure and fertilizer applications, rainfall and runoff data, soil properties, plant/crop nutrient uptake rates, and tillage and mixing coefficients (Vadas et al., 2009). Input data were adopted from the Chesapeake Bay program's scenario builder data archive (Chesapeake Bay Scenario Builder, 2016). These data were compiled during the Chesapeake Bay Program's construction of the Phase 6 model in 2016 and represent a comprehensive summary of annual P applications differentiated by landuse and county over the 31-year range, and are publically available (Chesapeake Bay Scenario Builder, 2016). At the county level, the total mass of manure is estimated based on assumptions of manure nutrient content and annual animal production separately for each animal type; P applications are then

spread out across landuses in the county based on information provided by the jurisdictions (Chesapeake Bay Program, 2016). Fertilizer P applications are estimated based on commercial fertilizer sales information from the Association of American Plant Food Control Officials (AAPFCO) (AAPFCO, 2012). Documentation including detailed descriptions of this input data set are available through the Chesapeake Bay Program (Chesapeake Bay Program, 2016).

6.2.6 APLE Model Optimization

An optimization procedure was performed to calibrate the APLE estimates with respect to the available soil test P data. Soil test P data were not available at every time step estimated by the APLE model, thus an aggregate calibration scheme was developed and implemented across all landuse/land segment combinations, totaling 3,068 separate optimization runs. APLE soil P concentration estimates were adjusted by modifying the initial soil Mehlich III value for each landuse/land segment combination based on the mean error defined by the difference between all existing soil test data points and their corresponding APLE estimates (Equation 6.8):

$$\boldsymbol{ME} = \frac{\sum_{i=1}^{n} \boldsymbol{e_i} - \boldsymbol{o_{ti}}}{\boldsymbol{n}} \qquad \textit{Equation (6.8)}$$

where **ME** is the mean error of the APLE model for a single landuse/land segment combination (ppm), $\mathbf{e_i}$ is the i^{th} APLE annual soil P concentration estimate (ppm) corresponding to $\boldsymbol{o}_i$, the i^{th} soil test value, numbered 1 through n, the length of the set of existing soil test values for the unique landuse land segment combination. During the calibration procedure, for each landuse/land segment combination, the initial Mehlich III value was set equal to the first existing soil test P value in the dataset. APLE was then run and $\boldsymbol{ME}$ calculated; thus, an APLE estimation

providing a positive $\boldsymbol{ME}$, or a general over estimation drove the initial Mehlich III value down, and vis-versa. For land segments where no soil test P data existed for a particular landuse, the initial Mehlich III value was set equal to the mean concentration of soil tests in the same growing region as the missing data. This process was iterated separately for each of the 3,068 spatial and landuse discretization's until changes in $\boldsymbol{ME}$ approached zero.

6.2.7 Bayesian Estimates

The following assumptions were made about the datasets and used a guide development of the Bayesian estimates.

Assumption 6.3 - The soil test P dataset provides an estimate for absolute values of annual soil P concentrations but does not provide meaningful information about the inter-annual variability of soil P.

Assumption 6.4 - The APLE dataset provides estimates of the inter-annual change in the state of soil P concentrations, but does not provide meaningful information about the absolute value of annual soil P concentration.

In accordance with the assumptions above, the soil test P data were chosen to inform prior distributions for years where these data existed, and the APLE data set was used as a change model to update the current state of the model through time. Each soil test P data point was assumed to represent the mean of a normal distribution of possible values for that year. Standard deviations were assigned to all data points as listed in Table 6.3. When no data existed for a year,

the prior distribution was assumed to be uniform between two possible values. For each land segment landuse combination the model was formulated as:

$$m_t \sim \begin{cases} normal(s_t, e_1) & s_t \in S \\ uniform(a, b) & s_t \notin S \end{cases} \qquad \textit{Equation (6.9)}$$

where $\mathbf{m_t}$ is the estimate of the state of Mehlich III soil P level in year t, $\mathbf{s_t}$ is the soil test P value in year t, **S** is the full set of available annual soil test P data points for a particular land segment landuse subset, $\mathbf{e_1}$ is the estimated standard deviation, quantifying the uncertainty in soil test P data, **a** is the lower limit of possible Mehlich III soil P concentrations when no soil tests were available, and **b** is the upper bound under the same condition. These bounds were set at 0 and 1000 mg P kg^{-1}, defining a wide range of possible Mehlich III values informed by extreme values found in the literature to avoid over confident estimations. The interannual change in the state of the soil P concentrations was defined by adding the change defined by APLE to the current state and applying a standard deviation to account for the uncertainty in the inter-annual variability defined by APLE. This value was chosen based on expert opinion and estimation of uncertainties in the inputs required to produce APLE estimates and was used in the absence of parameterized APLE inputs, designed to estimate the resulting uncertainty in the estimated annual changes due to the inputs.

$$\Delta APLE_t = APLE_t - APLE_{t-1} \qquad \textit{Equation (6.10)}$$

$$m_t \sim normal(\square_{t-1}, + \Delta APLE_t, e_2) \qquad \textit{Equation (6.11)}$$

where $\mathbf{m_t}$ is the estimate of the Mehlich III soil P in year t through updating the previous year's state ($\mathbf{m_{t-1}}$) by the estimated change in concentration at year t ($\Delta\mathbf{APLE_t}$), $\mathbf{APLE_t}$ is the soil Mehlich III estimate from the APLE data set at year t, $\mathbf{APLE_{t-1}}$ is the soil Mehlich III estimate from the APLE data set on the previous year, and $\mathbf{e_2}$ is the estimated standard deviation of the normal distribution, representing the error in $\Delta\mathbf{APLE_t}$ estimates. This framework is visualized in Figure 4.

The Bayesian model was implemented in the compiled code language stan, and the final distribution of the Bayesian soil P concentrations estimated through the default stan implementation of Markov Chain Monte Carlo (MCMC) method, employing No-U-Turn-Sampling (NUTS) to adaptively fit the Hamiltonian Monte Carlo procedure (Hoffman, 2014; Stan Development Team, 2016b). This procedure was run separately for each landuse/land segment combination on four chains with 1000 samples each, 200 of which were used for warm up. Input data were prepared with an R script, and the stan model was run with Rstan, an R implementation of stan (Stan Development Team, 2016a).

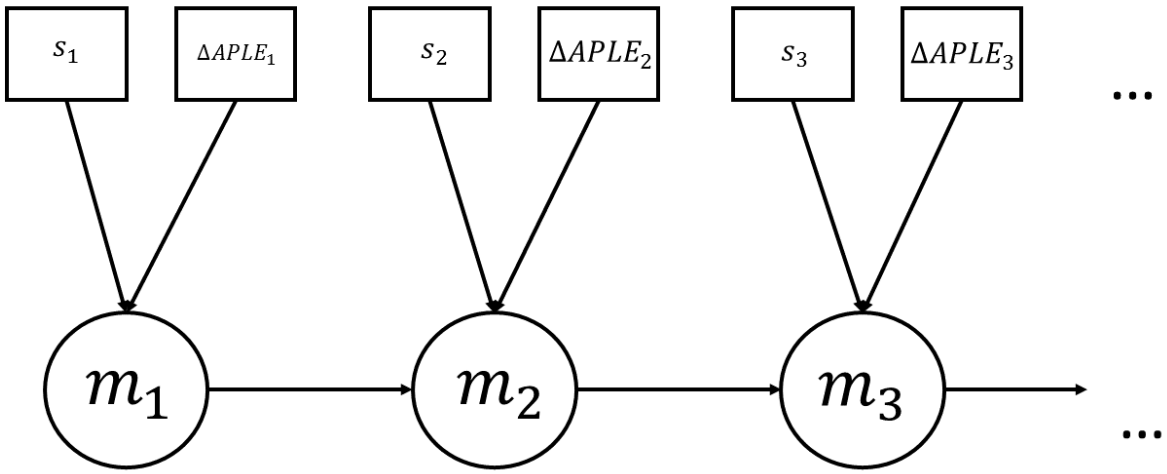


Figure 6.4: Diagram of the update of estimated soil phosphorus concentration states. Each new state contains information from the soil test P dataset, the APLE dataset, and the previous estimated state.

6.3 Results

6.3.1 Landuse Discretization Significance

The p-values for test 6.1 and test 6.2 concerning landuse discretization are summarized in Table 6.4. At a 0.05 significance level, test 6.1 provided evidence to reject null hypothesis 6.1, finding that data sharing a common landuse label between PA and VA do not have the same distribution as unlabeled data, for all but the full season soy landuse. Test 6.2 did not provide evidence to reject null hypothesis 6.2, in all but the silage with manure and specialty crops high landuses. Rejecting hypothesis 6.1 provides evidence that the landuse labels are indeed describing the data in a way that would be expected if the labels were true, and a general

inability to reject null hypothesis 6.2 provides that central tendencies of distributions with common landuse labels do not differ between PA and VA. In summary, test 6.1 and test 6.2 provided a means to assess both the discriminatory behavior of the landuse labels, and the spatial applicability of the central tendency of the soil test P concentrations for each landuse category. Each landuse category for both PA and VA are shown in Figure 6.5.

Table 6.4: P-values for Test 6.1 and Test 6.2 which test Null Hypothesis 6.1: Data sharing a common landuse label have the same distribution as unlabeled data, and Null Hypothesis 6.2: Data from VA and PA sharing a common landuse label have the same distribution, respectively. Values shaded in gray were used to reject the null hypotheses at a significance level of 0.05.

	Test 6.1		**Test 6.2**
Landuse	**PA**	**VA**	
Grain with Manure	1.51E-11	5.25E-01	0.8197
Legume Hay	5.38E-144	1.28E-02	0.4303
Other Agronomic Crops	7.97E-65	8.25E-06	0.9946
Other Hay	1.32E-113	7.50E-03	0.3011
Pasture	3.51E-214	1.17E-04	0.5941
Specialty Crop High	5.10E-11	6.88E-21	0.0216
Specialty Crop Low	1.88E-160	2.89E-12	0.2399
Small Grains and Grains	1.21E-62	7.15E-01	0.1312
Full Season Soy	1.27E-01	2.01E-01	0.1535
Silage with Manure	1.78E-10	1.59E-02	0.0348

Following the results of test 6.1 and test 6.2 we assumed that the soil test P from points having landuse labels are significantly different from those without labels. Following this assumption, a ratio method was developed to shift the mean of an unlabeled data set to a value that would describe the central tendency of a dataset with a common landuse label. Conjecture 6.1 was used to scale the remainder of the unlabeled soil P concentrations into landuse categorizations.

Conjecture 6.1 – Differences in central tendencies can be used to describe the relationship between soil test P concentrations with and without landuse labels

Following conjecture 6.1, mean ratios were calculated for each labeled dataset compared to the full set for both Penn State and Virginia Tech data (Table 6.5). These ratios were used to obtain landuse labeled estimates from all other remaining soil test P concentration data sources (Table 6.1) by applying assumption 6.1. All data were multiplied by each corresponding mean ratio as a scaling factor, resulting in soil test P concentrations with landuse labels. The resulting differentiation are illustrated in Figure 6.5, showing the soil P concentrations on natural log scale, split by landuse after the mean ratios below were applied. Although only 10 landuses were present after crop to landuse assignments in the PA and VA data sets described above, all 13 landuses are shown in Figure 6.5. For APLE estimates, different inputs are defined for all landuses, so the Bayesian model estimates, which depend on both soil test P concentrations and APLE estimates can differ between all landuses. However, no evidence was found to differentiate soil tests P concentrations from landuses with or without manure, therefore, the grain with manure (gwm) and grain without manure (gom) soil test P concentrations were set as

identical, and the silage without manure (swm), and silage with manure (som) sets were set as identical. Additionally, double cropping (dbl) was assumed identical to small grains and grains (sgg).

Table 6.5: Resulting ratios from the mean of all soil test P concentration data points from Penn State and Virginia Tech compared with unlabeled data from the remaining sources.

Landuse	**Mean Ratio**
Specialty Crop High	1.803
Specialty Crop Low	1.615
Other Agronomic Crops	1.404
Grain with Manure	1.026
Full Season Soybeans	0.974
Silage with Manure	0.919
Small Grains and Grains	0.883
Other Hay	0.812
Legume Hay	0.779
Pasture	0.735

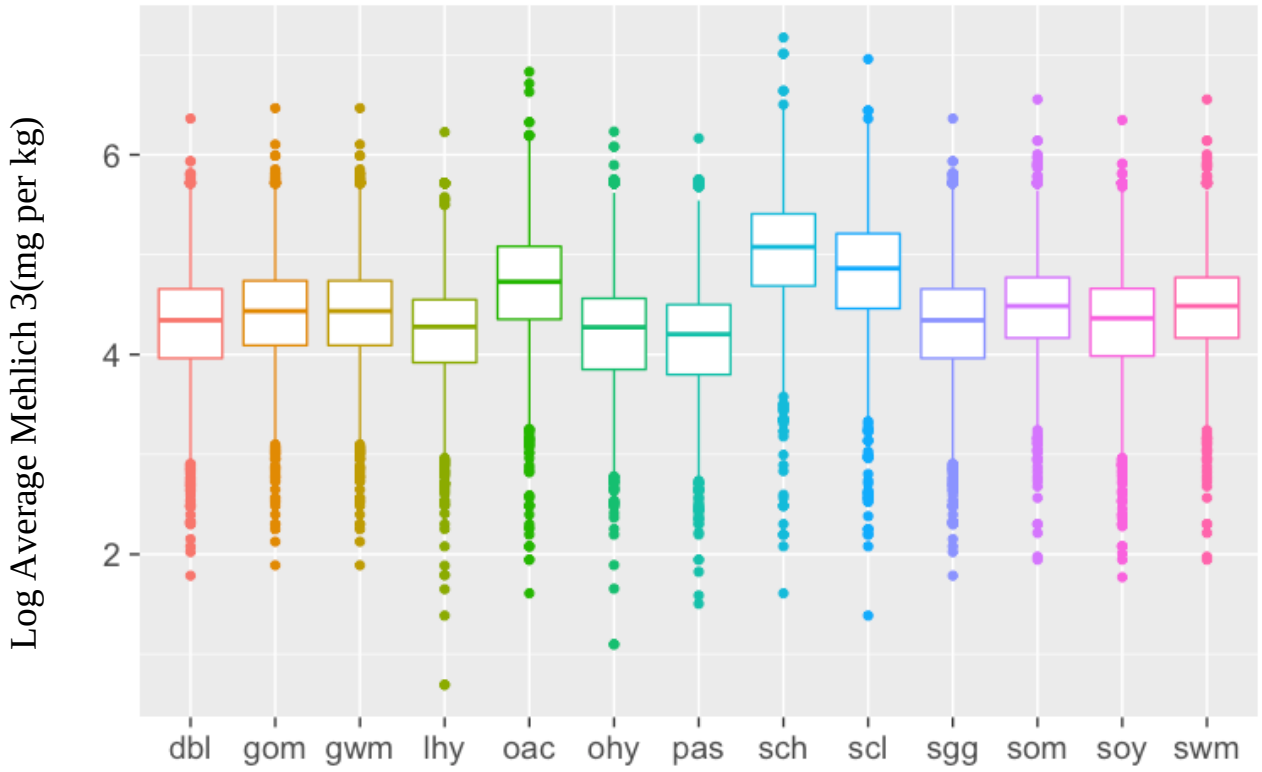


Figure 6.5: Soil tests after landuse mean scaling factors were applied. The three letter codes refer to the landuses as follows: dbl is double cropping, gom is grain without manure, gwm is grain with manure, lhy is legume hay, oac is other agronomic crops, ohy is other hay, pas is pastures, sch is specialty crops high, scl is specialty crops low, sgg is small grains and grains, som is silage without manure, soy is full season soy, and swm is silage with manure. More on these landuses can be found in the Chesapeake Bay Model Documentation.

6.3.2 APLE Model Optimization Results

The final distribution of APLE estimates compared with the soil test P concentrations distribution had a lower mean value (95.8 mg P kg^{-1}) than the measured soil tests (101.2 mg P kg^{-1}), however, the calibrated APLE estimates had a larger range, with a maximum value 600

ppm greater than the highest measured soil test value (Table 6.6). Both the distribution of the APLE and soil test P estimates were right skewed (Figure 6.6), with the measured soil test P values exhibiting variation in the right tail of the distribution with a notable increase in number of tests around 300 ppm, while the APLE distribution descended smoothly.

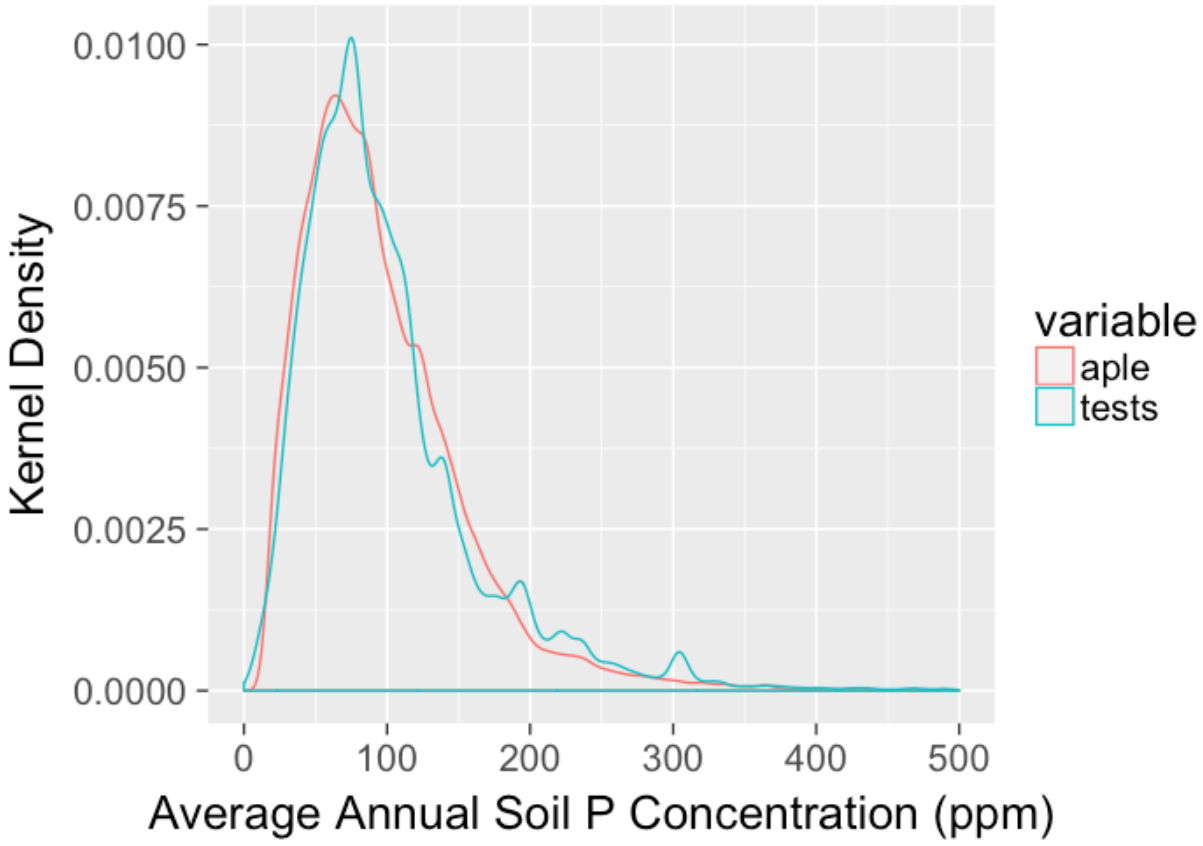


Figure 6.6: Distributions of calibrated APLE and soil test P concentration estimates. Post-calibration, APLE exhibits a similar shape and density to the measured soil test P concentrations. Kernel density smoothing was used to compare distributions over histograms due to the different number of observations in each set. X-axis limit threshold set at 500 to show detail.

6.3.3 Bayesian Estimates

Bayesian estimates of the soil test P concentrations had a lower mean than the measured soil tests P concentrations, similar to that of the APLE estimates. Notably, all methods had very similar median values. The range of the Bayesian estimates was roughly half of the APLE estimate range and almost 400 ppm less than that of the soil tests. When viewed in natural log scale (Figure 6.7), the set of Bayesian estimates exhibits a smaller range around its quartiles and a more balanced number of low and high values than the soil tests or the APLE estimates. The Bayesian method was also able to quantify the average uncertainty across the dataset, which at the 95% level was 54 ppm.

Table 6.6: Summary statistics from the three soil test P concentration estimation methods

Method	Min (ppm)	Max (ppm)	Mean (ppm)	Median (ppm)	95% Uncertainty (ppm)
Bayes	5	975	96	84	54
APLE	15	1925	96	84	NA
Soil Tests	2	1304	10	86	NA

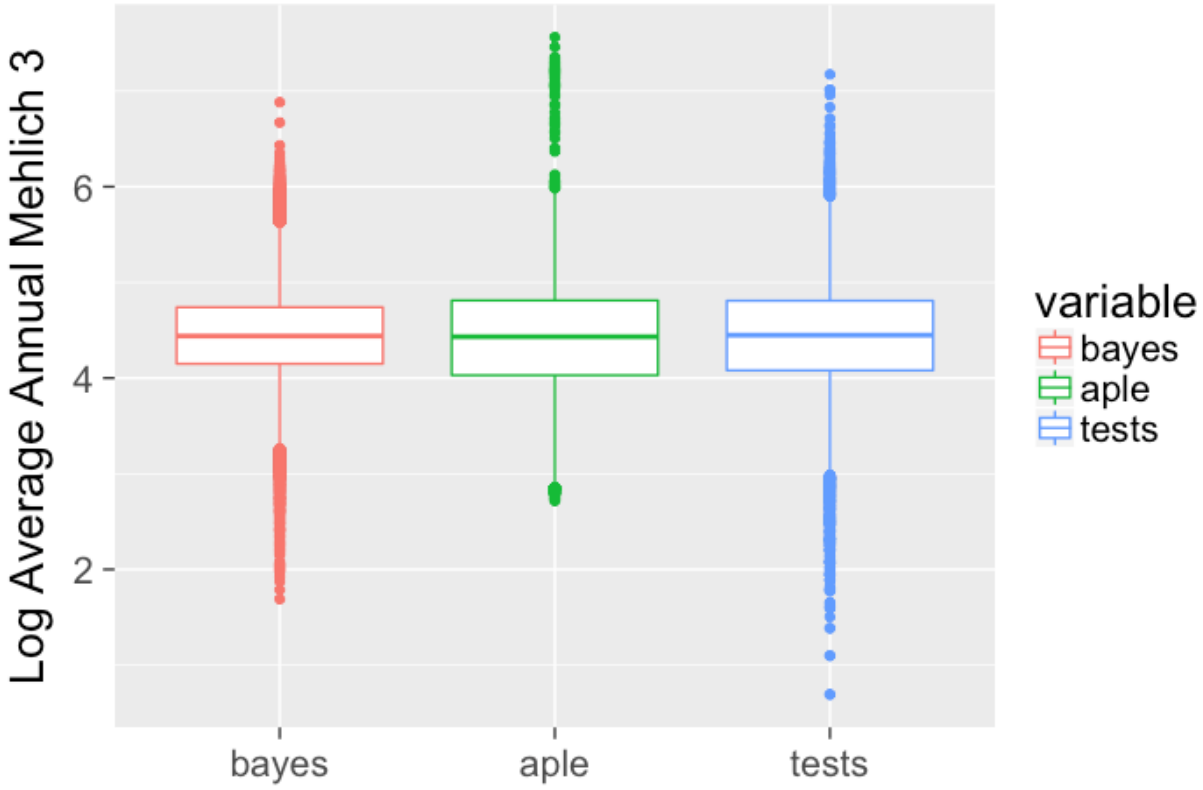


Figure 6.7: Natural log of soil test P concentration estimates from the three methods. The quartiles of the Bayesian method are less variable compared to the other sources. Additionally, low and high outliers are more evenly distributed in the Bayesian method than in the APLE model estimates or the observed estimates.

6.4 Discussion

During the 31-yr estimation period each method demonstrates variability (Figure 6.7), with the APLE and soil test P methods displaying approximately equal quartile distributions. The log annual soil test P estimates are shown in Fig. 6.8, where APLE display relatively low variation. This is due to two primary factors, the initial model boundary conditions imposed during development, a function of available data in 1984, and the cumulative (multiplicative)

nature of the APLE mass balance model. The Bayesian estimates begin with wider ranges in early years and progress to lower variance with very fewer outliers. This is initially large range in values is due to the low temporal resolution of measured soil test P data, upon which the Bayesian model is developed, in the early years. As the Bayesian time series progresses the variation declines by approximately 200 ppm from 1994 to 2000 due to much denser coverage of soil test P measurements in later years, giving the Bayesian estimates more confidence (Fig 6.8). The variation in the soil test P measurements are a direct function of the measurement density, with the early period of the time series exhibiting little variation, due to very low sampling density, and the later period displaying higher variation (several orders of magnitude, in fact) due to much higher spatial and temporal sampling density (Fig. 6.8). Two years stand out in the measured soil test P data, 1992 which contain soil test P concentrations that are, on average, markedly lower than nearby years, and 2012, which contains soil tests that are markedly higher, on average, than surrounding years. These anomalies are due to the addition of new soil test P sources during these time frame that only cover one year. The soil test set University of Maryland 2, present in 1992 reduced that year's mean value by nearly 0.5 on the log scale, while the presence of the Virginia Tech data set in 2012 had the opposite effect raising the soil P concentration by 0.5 on the log scale. Although the Bayesian estimates are intended to be used at the county scale and these anomalies do no directly indicate any problems with this geographic abstraction of the data, they illustrate the importance of data coverage when summary statistics are of interests. In addition, given the paucity of measured soil test P data, particularly in the early period of the time series, excluding any given dataset without clear evidence of inappropriateness would render complete Chesapeake Bay watershed level coverage nearly impossible, thus the Bayesian estimates are intended to combined these data sources in as

equitable manner as possible. This can be seen in Fig 6.8, where the Bayesian estimates for 1992 and 2012 do not appear to be unduly influenced by the anomalous soil test P levels. However, near the end of the time period the influence of several years of high APLE and soil test P estimates begin to have an effect on the Bayesian estimates at higher soil P concentrations, pulling them upwards.

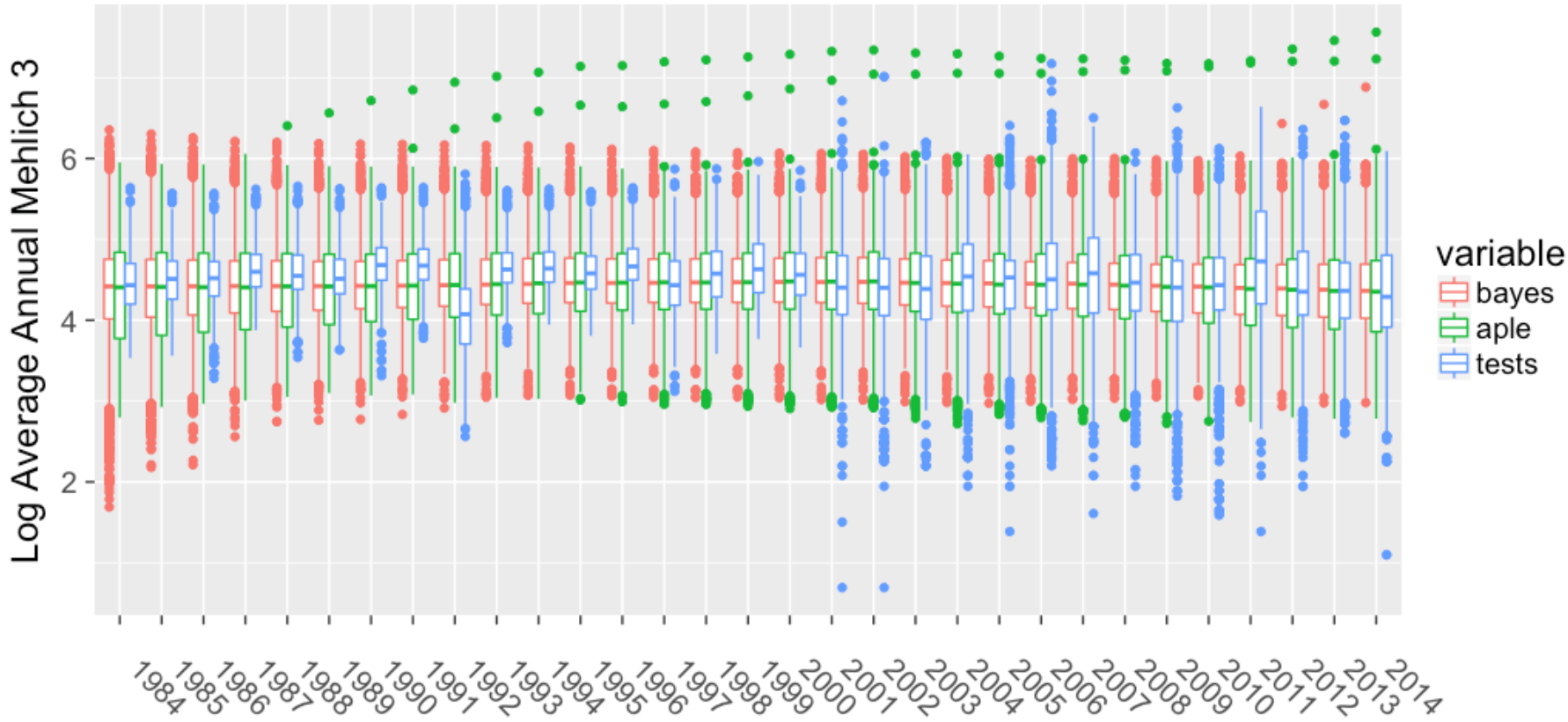


Figure 6.8: Boxplots of soil P concentration estimates from each estimate source through all 31 years. A small number of APLE estimates over 6 on the log scale can be seen beginning in the late 1980's, and continuing through 2014. The Bayesian estimates show wider ranges in early years, and soil test data show greater variability in later years as the number of tests increases in the year 2000.

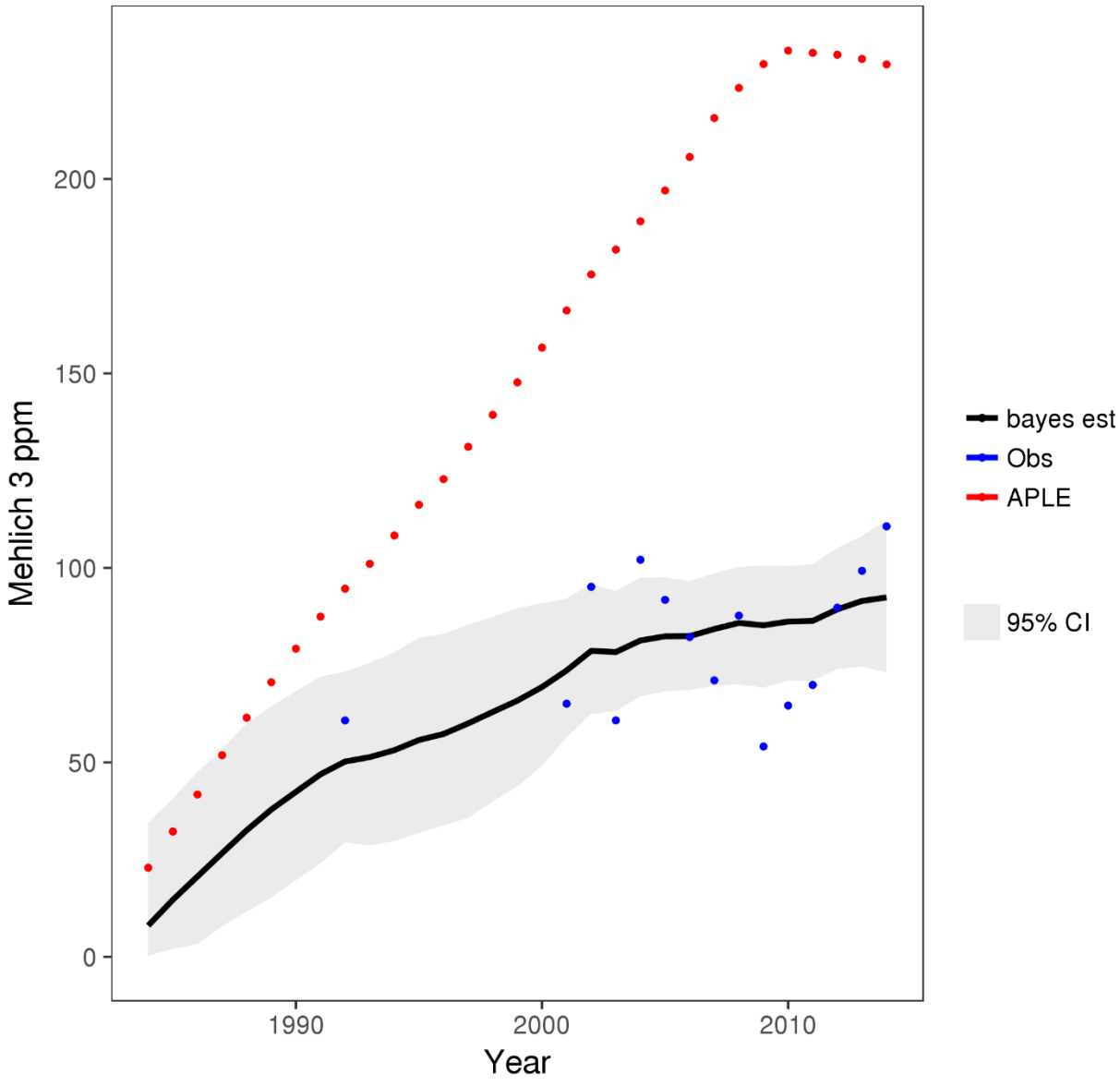


Figure 6.9: Bayesian, soil test, and APLE estimates for grain with manure landuse in Adams County Pennsylvania. An extreme APLE estimate raising far above the available soil tests appears unlikely. The Bayesian Method is able to smooth out this effect without completely losing the information APLE provides.

The advantages of the Bayesian estimates can be seen more clearly at the land segment level. Figure 6.9 shows all three methods estimated soil test P concentrations for the grain with manure landuse in Adams County, Pennsylvania. In this case, APLE estimates begin low (again due to the lack of parametrization data available in the early 1980's) and increase quickly and do not agree with the measured soil test P concentrations. The Bayesian method estimates, however, track the measured soil test P concentrations rather well, and provide uncertainty bounds around the point predictions, which can be seen to contract and expand as a function of the density of measured data, and can be used to quantify the relative confidence in model estimates. Although Bayesian estimates from 1995 forward seem reasonable for this county, early estimates appear low, due to a lack of soil test data combined with low APLE estimated soil P levels.

Figure 6.10 shows a selected result for the grain with manure landuse to illustrate the results from the Bayesian method at the county scale. Figure 6109a shows APLE predicting an upward trend in soil P, and while the APLE predictions fall within the 95% confidence interval of the Bayesian estimates, the shape of the APLE curve differs substantially from both the measured data and the Bayesian estimates. What is also clear from Fig. 6.10a is that there is a relatively high level of uncertainty in the estimated soil P concentrations, varying from approximately 40 ppm to almost 100 ppm, suggesting that a refined or intensified sample collection protocol be developed to better quantify soil P concentrations. Figures 6.10b, c, and d show instances of outliers in measured soil test P having no significant effects on the Bayesian estimates, primarily due to the standard deviation scaling method (Equation 6.6), which reduces the impact of outliers. Figures 6.10a and b elucidate the impact of large interannual variation between soil test P measurements in adjacent years; in landuse segments with high interannual

variability in soil test P, the Bayesian uncertainty is relatively high (200-300% of estimate). This is despite the generally good agreement if APLE with the measured soil test P data.

Figures 6.10c and d display the effect of method agreement on the resultant Bayesian uncertainty; in instances where both APLE and the measured soil test P data agree, both in magnitude and in interannual variation, the Bayesian estimates have much lower uncertainty, approximately 10% near the end of the periods in Fig 6.10c and d. Figure 6.9e depicts a typical time series of soil P estimates at the county level, rapidly increasing in the 1980s and 1990s, due to intensive expansion of agriculture in the region coupled with cheap fertilizers, followed by a decrease in the rate of increase in the late 2000s-2010s possibly due to increased fertilizer prices across the United States around that time (Cordell et al., 2009). The APLE estimates begin low, increase rapidly, leveling off around the year 2000, reflecting the increase in fertilizer prices. Soil test P data are sparse in early years, but increase in later years with denser coverage, generally agreeing with the APLE estimates. The Bayesian method, coupling these two time series provides smooth estimate influenced by both inputs, though may be over-influenced by rapidly increasing APLE estimates in early years where soil tests are less frequent. Figure 6.10f presents an ideal soil P response, decreasing over time. In this case both the soil test P measurements and APLE agree that soil P levels are declining This is reflected in the Bayesian estimates, which also decline, although the uncertainty bounds are still relatively large, due in part to the large year to year change in measured soil test P in the mid 2000s.

The estimates and uncertainty bounds from the Bayesian results in Fig. 6.10 could be used as inputs to other modeling frameworks (the Chesapeake Bay Watershed Model, for instance, or a process based model) to develop a plausible range of expected soil P concentrations, and their associated uncertainty. This could then be used to estimate the

probability that soil P reduction practices will result in the required water quality improvements needed to meet the Chesapeake Bay TMDL. Perhaps more directly, the uncertainty estimates provide direct evidence of locations (and perhaps times) where additional data collection would improve estimates of soil P concentrations. Additionally, the uncertainty bounds numerically quantify the credibility in the estimates across a known concentration range. Both of these interpretations of the credible interval can add important knowledge when model outcomes are designed to support management decisions, especially when margins of safety need consideration or when "worst case" scenario analysis would support objectives.

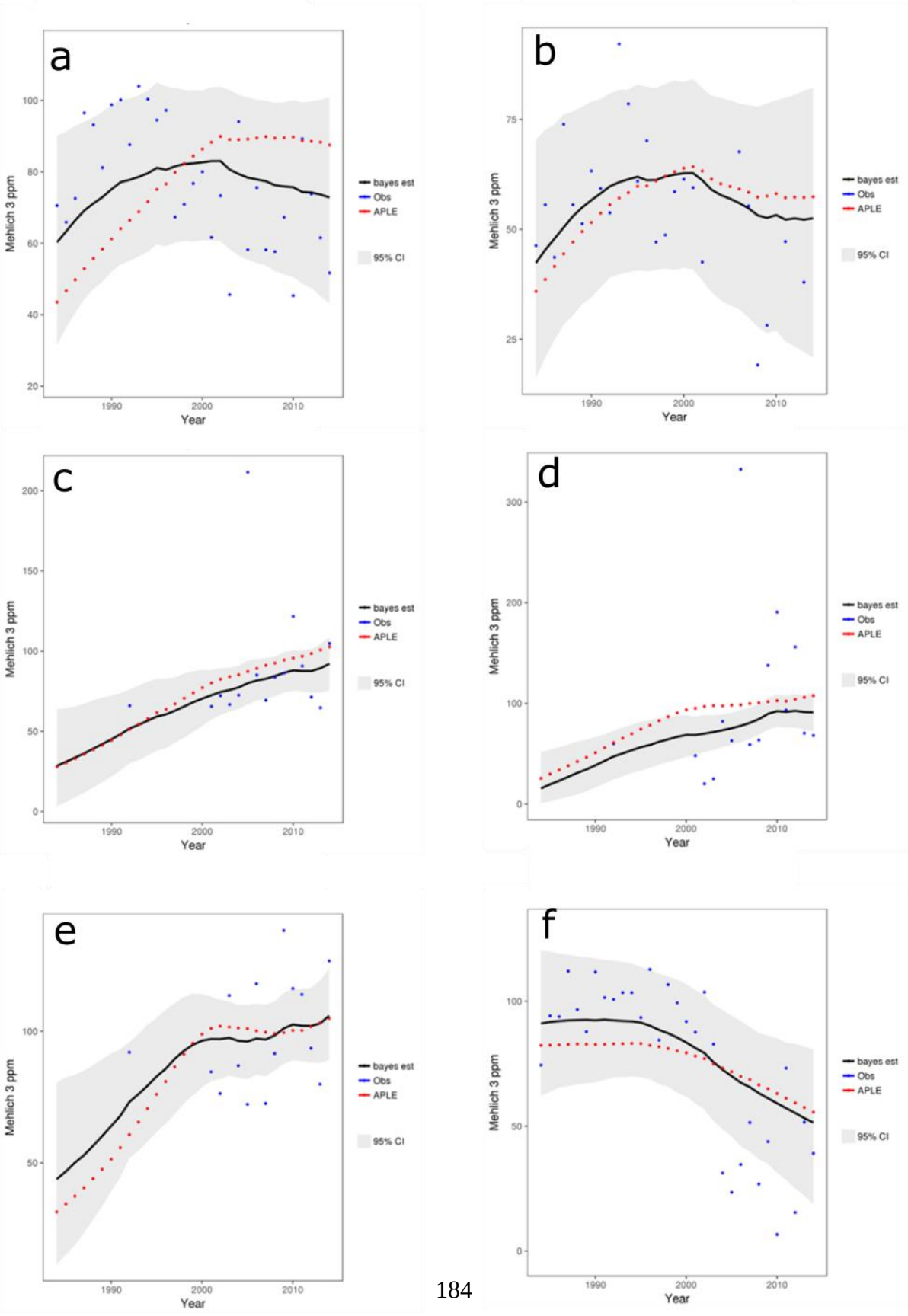

a
b
c
d
e
f
Mehlich 3 ppm
Year
bayes est
Obs
APLE
95% CI

Figure 6.10: Examples of Bayesian soil concentration estimates at the county scale from Maryland and Pennsylvania. (A) and (B) show examples of weak trends, (C), (d), and (E) show increasing trends, and (F) shows a decreasing trend.

6.5 Conclusion

Two datasets describing a 31-year history of soil test P concentrations in agricultural soils with varying quality, uncertainties, and coverages were used as inputs to a probabilistic Bayesian model designed to incorporate the desirable traits of each data set while quantifying the uncertainty associated with each data point. This methodology provides a single estimate of soil P for each year, as well as the resulting uncertainty in the estimate. The Bayesian model provided estimates with similar overall mean and median to the APLE estimates and had an attenuating effect compared with extreme APLE changes at the county and single landuse scale. These effects were most prevalent in cases with sparse soil test data, especially over a number of consecutive years with little or no soil test P measurements. The final Bayesian estimates include soil test P concentrations and 95% confidence intervals from the estimated posterior distribution for all 31 years across 3,068 unique landuse/land segment combinations. The data set encompassing the results of this study has been made publically available by the authors and is intended to support further research and study into soil P storage and provide stakeholders with valuable information to support decision making concerning management practices aimed at reducing the impact of P runoff at the county scale. Additionally, this study demonstrates the importance of future monitoring programs designed to collect high quality soil P concentration data and data sharing in order to avoid situations where working with highly uncertain data is necessary.

6.6 References

AAPFCO; The Association of American Plant Food Control Officials (AAPFCO), TFI; The Fertilizer Institute. 2012. Commercial Fertilizers. 2012 edition.

Bishop, P.L., Hively, W.D., Stedinger, J.R., Rafferty, M.R., Lojpersberger, J.L. and Bloomfield, J.A., 2005. Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. Journal of Environmental Quality, 34(3), pp.1087-1101.

Bolster, C.H., 2011. A critical evaluation of the Kentucky phosphorus index. Journal of the Kentucky Academy of Science, 72(1), pp.46-58.

Cordell, D., Drangert, J.O. and White, S., 2009. The story of phosphorus: global food security and food for thought. Global environmental change, 19(2), pp.292-305.

Chesapeake Bay Program. (2016). Draft Phase 6 Model Documentation: Terrestrial Inputs. 3(1 – 20). Accessed December 17 2016.

Chesapeake Bay Scenario Builder. (2016). Beta 3 APLE Inputs ftp://chesapkeabay.net/Modeling/Phase6/PhCalibration_Beta4/Inputs/APLE_Inputs/. Accessed December 20, 2016.

Commonwealth of Virginia. (2005). Virginia Nutrient Management Standards and Criteria. Virginia Department of Conservation and Recreation Division of Soil and Water Conservation

Donohue, S. J. 1992. Reference soil and media diagnostic procedures for the southern region of the United States.

Dubrovsky, N.M. and Hamilton, P.A., 2010. The quality of our nation's water: nutrients in the nation's streams and groundwater; national findings and implications. US Geological Survey, National Water-Quality Assessment Program.

Hoffman, M.D. and Gelman, A., 2014. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1), pp.1593-1623.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50-60.

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. Corsini Encyclopedia of Psychology.

Mehlich, A. 1984. Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. Comm. Soil Sci. Plant An. 15: 1409-1416.

Nelson, N.O. and Shober, A.L., 2012. Evaluation of phosphorus indices after twenty years of science and development. Journal of environmental quality, 41(6), pp.1703-1710.

Phillips, S. and McGee, B., 2016. Ecosystem Service Benefits of a Cleaner Chesapeake Bay. Coastal Management, 44(3), pp.241-258.

Scown, M.W., McManus, M.G., Carson, J.H. and Nietch, C.T., 2017. Improving Predictive Models of In-Stream Phosphorus Concentration Based on Nationally-Available Spatial Data

Coverages. JAWRA Journal of the American Water Resources Association, 53(4), pp.944-960.

Sharpley, A.N., Bergström, L., Aronsson, H., Bechmann, M., Bolster, C.H., Börling, K., Djodjic, F., Jarvie, H.P., Schoumans, O.F., Stamm, C. and Tonderski, K.S., 2015. Future agriculture with minimized phosphorus losses to waters: Research needs and direction. Ambio, 44(2), pp.163-179.

Stan Development Team. 2016a. RStan: the R interface to Stan. R package version 2.14.1. http://mc-stan.org

Stan Development Team. 2016b. The Stan Math Library, Version 2.14.0. http://mc-stan.org

US Census. (2010). 2010 FIPS Codes for Counties and County Equivalent Entities. https://www.census.gov/geo/reference/codes/cou.html. Accessed December 20, 2016.

Pierdicca, N., Pulvirenti, L. and Bignami, C., 2010. Soil moisture estimation over vegetated terrains using multitemporal remote sensing data. Remote Sensing of Environment, 114(2), pp.440-448.

Qu, Y., Zhang, Y. and Wang, J., 2012. A dynamic Bayesian network data fusion algorithm for estimating leaf area index using time-series data from in situ measurement to remote sensing observations. International journal of remote sensing, 33(4), pp.1106-1125.

Vadas, P.A., Good, L.W., Moore, P.A. and Widman, N., 2009. Estimating phosphorus loss in runoff from manure and fertilizer for a phosphorus loss quantification tool. Journal of environmental quality, 38(4), pp.1645-1653.

Vadas, P.A., Powell, J.M., Brink, G.E., Busch, D.L. and Good, L.W., 2015. Whole-farm phosphorus loss from grazing-based dairy farms. Agricultural Systems, 140, pp.40-47.

Virginia Soil and Water Conservation Board, 2014. Nutrient Management Training and Certification Regulations. 4VAC50-85

Yen, H., White, M.J., Ascough, J.C., Smith, D.R. and Arnold, J.G., 2016. Augmenting watershed model calibration with incorporation of ancillary data sources and qualitative soft data sources. JAWRA Journal of the American Water Resources Association, 52(3), pp.788-798.

Chapter 7: Conclusions and Future Work

Actions aimed at protecting water quality beyond course-scale temporal and spatial recommendations from currently available tools such as nutrient management plans or state sponsored nutrient indices is needed. The full potential of short-term forecasts to assist in filling this gap in agricultural management has yet to be realized. One pathway to utilizing short-term forecasts is creating accessible data delivery tools built on proven science that can inform end users in a valuable way. Challenges for academics seeking to build tools like these include a lack of free and open source frameworks to automatically run short-term watershed scale forecasts, a lack of data available for validating forecast methods, cumbersome models, and ineffective or untested communication strategies for the data intensive outputs. This collection of studies describes novel answers to these challenges and explores their efficacy. In Chapter 3, an automatable framework for running a scientifically proven watershed-scale model based on physical process was built. This framework was effectively used to produce short term soil moisture forecasts at the watershed scale as static reports which require some post-processing by a GIS trained user to create actionable demonstrations of the results.

Chapter 4 addressed the sharing and communication component of the short-term forecasting gap by creating a customizable, web ready front end user interface which was linked to the short term forecasting framework to display the output. The interface was designed with the goals of being reusable for researchers working on similar problems, dependent only on software available world-wide and completely free and open source, usable on a variety of internet connected devices, and to provide a positive user experience. Unlike similar tools in the past, the user-centric design goals of the interface were quantified and tested with widely

accepted? excepted user experience metrics. In addition to successfully meeting the technical criteria, the interface described in Chapter 4 stood out as a top performer in usability and accessibility. Analysis of the similar interfaces suggests user-centered design is a key area of improvement for short-term, public facing water quality tools in agricultural management. Although the interface performed in a satisfactory manner, the framework described in Chapter 4 which delivered the data for the displayed forecast had limitations. Over time, it became clear the forecast framework was difficult to run fully automatically on a simple server for long periods of time without errors causing the process to stop. If the framework stopped it may or may not require a small amount of trouble shooting before launching again. Therefore, some continuing human maintenance was needed for the forecast to reliably continue indefinitely.

The modeling schema described in Chapter 3 attempts to address problem of implementing scientific watershed scale models in a live application framework. This use case is beyond the design goals of scientific watershed models and presented many problems including awkward input data structures, necessary components which are not cross platform, inefficient use of computational resources, and the inability to scale to very large implementations. To address these short comings, Chapter 5 describes a method which takes advantage of the fact that scenario modeling is unnecessary to fulfil the goals of the short-term forecast tool and implements a fully scalable, enterprise ready, free and open source deep learning model. A reformulation of the modeling problem was necessary and the chapter describes the soil moisture forecast re-designed as a binary classification problem with input data distilled to a multi-variate grid through published VSA defined rules and spatial discretization schemes shown to improve spatial accuracy of the overland components of the hydrologic cycle compared to an unmodified semi-distributed watershed model. In formulating the problem this way, proven science was used

to distill the problem from a completely data driven approach to a data set with more real valued and process meaning. This was done to address the problem of interpretability in neural network architectures like the implemented deep learning. Training labels for the deep learning model were supplied by the process based model framework described in Chapter 3, thus, information describing a process based system was developed on both sides of the deep learning model, giving it the possibility to learn towards a more physically meaningful optimum representation. Furthermore, novel interpretation techniques were demonstrated and employed to find hierarchal linkage regression equations providing evidence of model relationships from real valued input data, through selected hidden layers, to output classification. Metrics analogous to odds ratios in regular logistic regression were calculated and used to interpret the results. The overall conclusion of the interpretation was the deep learning model was able to predict saturated areas well with a process that had evidence of being physically meaningful. Additionally, the linkage regressions described much of the variance in the model and could be used as a simple, standalone process that provided surprisingly accurate classifications, though they were not as effective as the full deep learning model.

Chapter 6 describes a weight of evidence approach with a Bayesian time series model in an attempt to quantify soil phosphorus levels at a large scale spanning 31 years. Although not a short-term forecast, this long-term information is beneficial to framing the use cases and applications of short term forecasts tools, like the one described in Chapter 4 which can be used in P management. Since much of P runoff enters the water ways of VSA dominated watersheds like the study area of Chapter 6, the Chesapeake Bay watershed, short term tools are needed to inform practices. However, in the absence of continuous data at this spatial scale, a longer term quantification of soil P levels can assist in the spatial framing and targeting of short term

planning approaches. A similar process could potentially be used to assess the effect of an implementation of a short-term planning tool on soil P levels in a particular area.

Future work in this area would benefit from a more complete study of the distillation of scientific knowledge into learning model parameters. The method of summarizing deep learning parameters and structure through simple regressing linkages could be used to increase confidence in predictions and interpret these models. Further investigation into the efficacy of using the linkage equations alone is needed, especially when used in a forecasting application. Additionally, easy to use and access public facing decision support tools can widen the approach to driving adoption of conservation practices. More work is needed which demonstrates actual user patterns, behavioral influence and especially long term use. Data shortages, such as soil P levels need to be addressed further, and the collection, curation, cataloging of much more data be made accessible. As the science of watershed management progresses, new work will benefit greatly from an increase in available data, open-minded approaches to advanced physical, statistical, and machine learning frameworks, and public facing demonstrations of results with the capability to influence real world decision making.

APPENDIX

The following are links to code published under open source licenses used to complete the work described in this dissertation. Code used in Chapter 3 can be found in an R package published under the General Public License version 3 (GPL 3) on the Comprehensive R Archive Network (CRAN), available at https://cran.r-project.org/web/packages/getMet/index.html. The package named getMet assists in the collection and formatting of meteorological historical data and forecast in addition to offering features for automatic formatting for SWAT model use. It has been downloaded over 3,700 times since its publication on CRAN. Key functions used in Chapters 3 and 4 can be found in a collection of unlicensed scripts available on github at https://github.com/andrewsommerlot/vsa-gfs-forecast. Code used in Chapter 5 has been compiled into a stand-alone R package available on github under GPL3 available at https://github.com/andrewsommerlot/startml. Though not yet on CRAN, startml (short for start machine learning) provides users with automatic model building, hyper parameter tuning, and ensembles for regression and binary classification problems. It is fully installable through github using the devtools package.