COMPARING TWO TOOLS FOR ECOSYSTEM SERVICE ASSESSMENTS REGARDING WATER RESOURCES DECISIONS

P. James Dennedy-Frank*, Rebecca Logsdon Muenichb,1, Indrajeet Chaubeyb, Guy Zivc

* Corresponding author:
Stanford University
Department of Earth System Science
Yang & Yamazaki (Y2E2) Building
473 Via Ortega, Room 140
Stanford, CA 94305-4216
USA
pjdf@stanford.edu

b Department of Agricultural and Biological Engineering
225 South University St.
Purdue University
West Lafayette, IN 47907
USA
rlogsdon@umich.edu
ichaubey@purdue.edu

c water@leeds & School of Geography
University of Leeds
Leeds LS2 9JT
UK
g.ziv@leeds.ac.uk

1 Present address:
Graham Sustainability Institute
University of Michigan
625 E. Liberty St., Suite 300
Ann Arbor, MI 48104
USA
1 ABSTRACT
We present a comparison of two ecohydrologic models commonly used for planning land management to assess the production of hydrologic ecosystem services: the Soil and Water Assessment Tool (SWAT) and the Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) annual water yield model. We compare these two models at two distinct sites in the US: the Wildcat Creek Watershed in Indiana and the Upper Upatoi Creek Watershed in Georgia. The InVEST and SWAT models provide similar estimates of the spatial distribution of water yield in Wildcat Creek, but very different estimates of the spatial distribution of water yield in Upper Upatoi Creek. The InVEST model may do a poor job estimating the spatial distribution of water yield in the Upper Upatoi Creek Watershed because baseflow provides a significant portion of the site’s total water yield, which means that storage dynamics which are not modeled by InVEST may be important. We also compare the ability of these two models, as well as one newly developed set of ecosystem service indices, to deliver useful guidance for land management decisions focused on providing hydrologic ecosystem services in three particular decision contexts: environmental flow ecosystem services, ecosystem services for potable water supply, and ecosystem services for rainfed irrigation. We present a simple framework for selecting models or indices to evaluate hydrologic ecosystem services as a way to formalize where models deliver useful guidance.

1.1 KEYWORDS:
ecosystem services, water yield, model comparison, SWAT, InVEST

2 INTRODUCTION
Hydrologic ecosystem services (HESs) are beginning to influence land management decisions through both regulations and investments targeted at protecting and improving water resources [Le Maitre et al., 2007; Gordon et al., 2010; Goldman-Benner et al., 2012]. HESs are the goods and services that ecosystems provide to people related to various uses of water, and include water availability for municipal, agricultural, and commercial use, the reduction of the magnitude and frequency of flow peaks to prevent floods, the reduction of sediment and nutrients in water, and the value of natural hydrologic systems for recreation [Brauman et al., 2007; Keeler et al., 2012; Brauman, 2015]. Land managers use HESs to support investments and meet regulatory requirements around water resources in both developing countries [Goldman, 2009] and developed countries [Lautenbach et al., 2010; Logsdon and Chaubey, 2013]. To support investments or meet regulations, scientists and managers assess HESs under expected future conditions based on some model of the landscape and ecosystem response; the landscapes are then managed to restore, retain, or optimize HESs and thus improve the status of the water resources. Scientists and managers assess HES provision in distinct decision contexts- combinations of investment and regulatory goals and institutional structures- which may decide both which models are applied and how model outputs are used to make decisions.
Scientists and managers often have different focuses when assessing HESs for inclusion in land management and water resource decisions. Scientists focus on formally linking ecosystem status and function to the goods and services they provide, while managers want models and results that can be easily included in the decision-making process. Clearly the best decisions can be made when both scientists’ and managers’ criteria are met; however, there is often a gap between advanced models, which are often more complex, and ease of inclusion in the decision-making process. Linking ecosystem function to services presents a challenge because models of ecosystem structure and function around water remain complex, and provide uncertain results [Kirchner, 2006; Beven, 2009]. The simplest analyses rely solely on correlation or lookup tables to link ecosystem structure and function to these goods and services [Costanza et al., 1997, 2014]. These simple approaches may be problematic in hydrologic systems where downstream interactions can be complex and nonlinear, especially on relatively short timescales that may be crucial for many water resource decisions. It is difficult to generalize which models can be included in the decision process because model results and their presentation can vary depending on the decision and institutional contexts. Various HES analyses have provided results at different levels of complexity and with different presentation strategies [Lautenbach et al., 2010], from overlay maps [Jackson et al., 2013] to scenario analysis [Van Liew et al., 2007] to explicit landscape optimization [Bekele and Nicklow, 2005; Lautenbach et al., 2010; Vogl et al., 2012; Cibin and Chaubey, 2015] and adaptive optimization [Kalcic et al., 2015].

Simple HES models have been developed by several groups to provide managers with estimates of HESs for land and water resource management decisions. These models attempt to link ecosystem structure and function to the water resource goods and services provided by the ecosystem. Models to assess HESs have been developed by groups including the Natural Capital Project [the InVEST model suite; Tallis and Polasky, 2009], the ARIES project [Bagstad et al., 2013], and LUCI [Jackson et al., 2013], while INFFER [Pannell et al., 2012] uses an expert elicitation approach to estimate the biophysical production of HESs. All of these groups work to increase the use of ecosystem service (ES) approaches in land management and water resources decisions by making it easy to include them in the decision process. Therefore, these groups’ models use simple approaches to ease the assessment of multiple ESs by non-experts in the developing world where data may be scarce. These groups hope that their models will be more widely used in decision-making than currently available hydrologic models and that such models will be used by people outside the hydrology research community. However, even among themselves, these models differ significantly in their simulation of biophysical complexity and the linkages of ecosystem structure to HESs. For example, InVEST uses relatively simple biophysical models developed from the hydrology literature to assess the landscape response under distinct scenarios, while INFFER uses expert opinion about landscape response.
Previous studies have not thoroughly assessed the accuracy of these simple HES models’ hydrologic estimates or how applicable their results are to relevant decisions. This is true despite the use of such models in making real decisions and investments [Goldman et al., 2010]. Vigerstol and Aukema [2011] compared the inputs and outputs of SWAT [Arnold et al., 1998], InVEST [Tallis and Polasky, 2009], the ARtificial Intelligence for Ecosystem Services (ARIES) model [Bagstad et al., 2013], and the Variable Infiltration Capacity (VIC) model [Liang et al., 1994] but did not compare model results. A recent review highlighted SWAT’s use and utility for ES assessments, especially for evaluating water yield [Francesconi et al., 2016]. InVEST’s annual water yield model has been compared with the SWAT model [Gassman et al., 2007] in a few studies [Goldman et al., 2010; Rocha, 2012], but none of them have been peer-reviewed. Hamel et al. [2015] did assess the InVEST annual water yield model in one North Carolina watershed against observed flows, but this does not evaluate the accuracy relative to other models, and further assessment is necessary to understand its value for application across many sites. Most of these comparisons have been performed at scales larger than those at which most management or investment decisions are made, making them less useful for real-world application. The ARIES hydrologic models have been assessed theoretically but have not been compared with more sophisticated models [Bagstad et al., 2013]. The LUCI models have been developed at smaller scales, and have been applied in the Pontbren Catchment in Wales [Jackson et al., 2013], but have not been applied broadly or in the developing world contexts where such investments may be the most important. These HES models are being used to make decisions [e.g., Goldman-Benner et al., 2012; Goldstein et al., 2012; Jackson et al., 2013], but the decision contexts under which they are useful have not been well-described. In addition, the accuracy of the model estimates will influence the land management contexts under which their use is justified.

Previous studies have used existing hydrologic models to assess HESs for land management and water resources decisions [e.g., Bekele and Nicklow, 2005; Van Liew et al., 2007; Lautenbach et al., 2012; Logsdon and Chaubey, 2013]; however, with few exceptions, while these models have been shown to provide useful hydrologic predictions in the hydrology literature, their application in ES decision contexts where there are many service tradeoffs have not been thoroughly described. For example, Logsdon and Chaubey [2013] developed a set of ES indices to describe the integrated services supplied by ecosystems. They tested their indices (hereafter the Logsdon ES indices) by examining simulation of several extreme landscape management scenarios built with the SWAT model, and found that the resulting ESs differed significantly across those scenarios. (Note that the Logsdon ES indices can be calculated based on observation of hydrologic and agricultural processes, in place of simulated outputs, to examine actual ES production rather than the potential of a given scenario). Prior to that, SWAT was used to examine ESs in agricultural areas as part of the Conservation Effects Assessment Project [Van Liew et al., 2007]. Other attempts have been made to quantitatively
assess HESs; however, as Logsdon and Chaubey [2013] point out, while ESs has become a significant buzzword in the hydrology literature, few quantitative assessment tools or measures for ESs have been developed.

This paper compares one of the simple HES models, the InVEST annual water yield model (hereafter the InVEST water model), against a very widely used ecohydrologic model, SWAT. Our unique contribution is to be the first to compare these models in the peer-reviewed literature, and to carefully consider the spatial and temporal scales at which these models should be selected for use in specific decision contexts. As shown in Figure 1, we address two major components, corresponding to the different rows in the figure: (1) the spatial distribution of water yield estimates from the InVEST model relative to the SWAT model; and (2) the usefulness for decision-makers of the INVEST annual water yield results versus the results from a SWAT model or ES indices developed by Logsdon and Chaubey [2013], both with and without a hydrologic model to apply the indices at a smaller scale. To address component (1), we ran both the InVEST water model and SWAT in two watersheds in the US with distinct climatic, physiographic, soil, and water use regimes. We investigated the spatial distribution of water yield estimates from InVEST by comparing the subbasin-scale results against those from the SWAT. Component (2) can be seen in the columns of Figure 1; to investigate this, we considered InVEST, SWAT, and the Logsdon ES index outputs and how each would address three distinct decision contexts about land management decisions to meet regulations and inform sustainable resource investments. We assessed each model’s usefulness for making decisions in each of these decision contexts. While the considered decision contexts are hypothetical, they are realistic both in terms of the types of tools that people are using to assess the ecosystem response for ES assessments, and for the types of regulations and investments ES approaches are being used to support.

3 METHODS

3.1 SITES

Our assessment includes two sites: Wildcat Creek Watershed, a predominantly agricultural watershed in Indiana; and the Upper Upatoi Watershed, a largely forested watershed on the Ft. Benning Army Base in Georgia (Figure 2).

Wildcat Creek Watershed is a 2083 km$^2$ watershed located in Central Indiana. The primary land use in the watershed is rainfed agriculture, which takes up about 70% of the watershed area. The soils have a high clay content and are poorly drained, thus many farmers rely on subsurface tile drainage to manage their croplands, and the climate includes a cold winter and aseasonal precipitation. Average annual temperature at the site is 10.7° C, with an average annual rainfall of 1027 mm [PRISM, 2012]. A full description of the Wildcat Creek watershed is supplied in Logsdon and Chaubey [2013].
The Upper Upatoi Creek watershed covers 876 km² in the Piedmont and Upper Coastal Plain physiographic provinces in Georgia, largely in the upstream portion of the Ft. Benning Army Base, and has an average annual temperature of 17.8° C and an average annual rainfall of 1231mm [PRISM, 2012]. The watershed consists of mostly level ridgetops and valleys separated by moderately steep to steep hillsides. The soils in the uplands are generally well-to-excessively drained, though there are also loam hills that have relatively high clay content; many of the lowland wetlands have much less-well-drained soils. Forests in the watershed consist of uneven-aged hardwood and hardwood/pine overstory assemblages, mostly deciduous in nature [Cavalcanti and Lockaby, 2005; Cohen et al., 2006]. Approximately 65% of the base is classified as forest, 17% as range, 6% each as agriculture and wetlands, and 4% as urban developed land [Fry et al., 2011]. Upatoi Creek flows generally SW towards the Chattahoochee River, which it joins near the edge of the Ft. Benning Army Base.

These two sites have distinct climates, physiographies, soils, and flow regimes, making them useful for comparing ES provision across a range of biophysical factors. These differences result in distinct hydrological responses and, as a result, large differences in hydrological ES provision.

These two sites also have significantly different water use profiles, so the ESs provided are also unique. As a result, the two sites would have different opportunities for ESs to provide value. We compare identical decision contexts to elucidate the different models’ abilities to provide information pertinent to HESs in these decision contexts. We note these assessments are hypothetical because we are not currently working closely with user groups interested in water quantity at either site. However, we believe this work provides significant value to users of these models and the scientific community in organizing an understanding of these different tools’ applications to HESs in different decision contexts, which should lead to more responsible land management.

3.2 MODEL COMPARISON

We compare the results from two conceptually distinct hydrologic models: the InVEST water model [Tallis and Polasky, 2009] and SWAT (Soil Water Assessment Tool) [Gassman et al., 2007]. We briefly describe each model here.

3.2.1 INVEST
The InVEST suite of models was developed by the Natural Capital Project \( \text{[www.naturalcapital.org]} \) to assess the benefits and tradeoffs of ESs under a variety of scenarios, and to help include those ESs in the decision-making process. As such, the InVEST models are simple and have very low data requirements so that they can be used throughout the developing world and by non-experts. In addition, numerous simple InVEST models can be applied simultaneously to consider tradeoffs across ESs and avoid undesirable effects in which increasing one ES diminishes another. However, these simple models have just begun being rigorously tested to assure that their results are sufficiently accurate to lead to appropriately-informed decisions [Hamel et al., 2015]; reaching such decisions remains an ongoing goal of the Natural Capital Project.

The InVEST water model is one of the most commonly used InVEST model components [Sharp, personal communication]. The InVEST water model uses a Budyko curve approach [Budyko, 1974; Zhang et al., 2004] with a Fu-type equation [Yang et al., 2008] to calculate the average annual water yield for each subbasin based upon the vegetation and aridity of the watershed, and assuming no change in storage. The InVEST water model includes two main components: (1) a dryness index, which includes reference evapotranspiration, annualized crop coefficient, and precipitation; and (2) a storage component which includes soil depth, plant root depth, soil water storage, and a seasonality term. The seasonality term lacks clear guidance, but one option we take here is to use it as a calibration parameter to assure the correct average annual water yield across the whole watershed. This approach results in a perfect calibration since we calibrate to a single average annual water yield for the watershed.

3.2.2 SWAT

The SWAT is a semi-distributed, process-based hydrological model that has been widely used to investigate water resource questions and assess the effects of land use and land management changes on hydrologic processes [Arnold et al., 1998; Gassman et al., 2007]. SWAT uses specific physical and semi-empirical approaches to model distinct processes in the hydrological cycle, and connects those processes to create a hydrologic simulator. SWAT models infiltration/runoff using a curve number approach; soil moisture, a shallow aquifer that can provide baseflow, and a deep aquifer that can be recharged; plant growth and evapotranspiration that are dependent on both water and nutrient availability; in-stream flow; and human management operations such as irrigation and fertilization. SWAT simulates processes in hydrologic response units—spatially discontiguous areas with similar land uses, soils, and slopes. SWAT simulates other processes at the subbasin level, especially those which route water through channels. SWAT’s large suite of processes allows for numerous feedbacks in the socio-eco-hydrologic system that might be missed when modeling only a subset of the components. The many-component hydrologic simulator approach used by SWAT contrasts with
the InVEST water model, which treats the whole watershed as an integrated system and looks for the system-wide response without
detailed simulation of the many related processes.

SWAT requires hydrologic flow data and significant calibration effort to ensure that the hydrologic processes simulated by the model
are accurate at various spatial and temporal scales, though uncalibrated models have been shown to perform satisfactorily at the
annual scale [Srinivasan et al., 2010]. User-friendly tools have been developed to make this calibration easier [Abbaspour et al.,
2007; Vrugt and Robinson, 2007]. In addition, user-friendly tools simplify the process of building a model by spatially combining the
necessary datasets such as topography, weather, soil [Sheshukov et al., 2012], and land use and management [Winchell et al., 2007].

SWAT has been used to estimate the response of watersheds to many different scenarios, and the effects of those scenarios on ESs
[Gassman et al., 2007; Van Liew et al., 2007; Francesconi et al., 2016]. Logsdon and Chaubey [2013] developed their suite of ES
indicators based partly upon a carefully developed SWAT model of the Wildcat Creek Watershed in Indiana. The details of the SWAT
model calibration and validation for the Upper Upatoi Creek Watershed are included as Supplement 3 to this paper.

3.2.3 ENSURING EQUIVALENCE IN MODEL INPUTS

We compare the results of InVEST with those from SWAT, advancing several pieces of work which suggested this opportunity
[Lautenbach et al., 2010; Vigerstol and Aukema, 2011]. We isolate the differences in model structure to determine both the models’
accuracy and the value of the model results for making ES decisions, rather than allowing variations in model inputs to create
differences. However, we note that the models might often be used with distinct input data, leading to different results based upon
input data. To isolate the differences in model structures we use inputs to the InVEST models which are equivalent to those in the
SWAT models. In particular, we use the SWAT subbasin delineations, precipitation (grouped by subbasin), PET (also grouped by
subbasin, using SWAT to calculate PET via the Penman-Monteith approach in Wildcat Creek Watershed and the Hargreaves
approach in Upper Upatoi Creek Watershed), and SWAT crop characteristics as inputs to InVEST. We take particular care to set the
annual crop coefficient in InVEST to be equivalent to that in SWAT. To do so we:

i. determine the daily crop coefficient in SWAT by taking \( k_{c,daily} = \frac{ET}{PET} \);

ii. eliminate crop coefficients for water-stressed days, since crop coefficients are generally defined for a well-watered surface

iii. average the crop coefficients by month to avoid overweighting the wet periods relative to the dry periods \( k_{c,month} = \frac{\sum k_{c,daily,wet}}{n} \) where \( k_{c,daily,wet} \) are the members of \( k_{c,daily} \) with no water stress in a given month, and \( n \) is the number

of days with no water stress in a month
3.3 COMPARISONS

We compare both (1) the hydrologic accuracy of the InVEST model vs. the SWAT model, and (2) the usefulness of the model results in three specific decision contexts (Figure 1). First we compare the hydrologic results of the InVEST water model and SWAT simulations at the subbasin scale. We compare both the average annual water yield and the ranked order of subbasins by water yield. We choose the subbasin scale because it is the natural scale for outputs of both InVEST and SWAT. The InVEST water model is run on a pixel scale, but because it is based on the Budyko curve which treats catchments as an integrated system, the model guidance clearly states that results should be considered only at the subbasin scale. SWAT’s HRU structure aggregates flow outputs in stream sections at the subbasin scale.

We compare these subbasin-scale water yields and ranks in both watersheds in several ways (Figure 2 and Supplement 1). Scatterplots compare the correlation between the two models’ estimates of water yield per subbasin, and build maps display their spatial distribution. We also make both scatterplots and maps of the water yield ranks for further analyses. We calculate uncertainties for the SWAT average annual water yields based on the standard error of the mean in the annual water yield over the year

iv. then average $k_{c_{month}}$ to obtain an average annual $k_c$.

Figure 1: This paper compares both the hydrologic outputs of an ES-specific model (the InVEST water model) and a more typical ecohydrologic model (SWAT) both in terms of their hydrologic outputs and the decision implications from their use. The left side of this figure shows the process of an ES-specific model, which starts with assessing a hydrologic response and then provides an estimate of the ES (the typical process runs vertically in this figure). The right side shows the process of a typical hydrologic model, which is designed to provide a hydrologic estimate and then additional tools are used to estimate ESs to help inform decisions about policies or investments around land use. We compare across these two types of models (comparison in the horizontal) both at the hydrologic step and at the decision implication level, as indicated by the bold, numbered arrows. [1.5 column image]
period of simulation. The InVEST water model provides only an average annual estimate since it assumes no change in storage, so its uncertainty cannot be assessed in the same manner.

We use SWAT as a comparison for InVEST, but note that we do not explicitly know the true water yield for each subbasin because we do not have measurements of discharge for each subbasin. Rather we are using these hydrologic models to downscale our understanding of the hydrologic response based upon local topography, pedology, and ecology. We use SWAT output as a baseline for two primary reasons: (1) the detailed daily-scale split-sample calibration performed; and (2) the broad use and verification of SWAT around the world. For (1), the SWAT model estimates match both the high peaks and the long-term baseflows seen in both watersheds, so we believe that the model is representing the flow from distinct land uses in the watershed reasonably. Since we know which land uses we expect to have relatively quick responses and which would be slower, we believe that our calibrations should result in a reasonable spatial distribution of water yield. For (2), SWAT has 2200 citations in their literature database, showing its wide use for land use planning and management around the world. Many of those studies have used SWAT to perform spatial downscaling and examine smaller spatial-scale hydrologic responses [e.g., Bekele and Nicklow, 2005; Rouholahnejad et al., 2012], and we do the same here.

Second we compare the usefulness of InVEST’s annual water yield estimates to the SWAT model’s daily water yield estimates and the Logsdon ES indices (applied both with and without a SWAT model to downscale its estimates) for specific land management decisions. We choose three land management decisions where HESs may either meet regulations or provide a return on investment, and consider the usefulness of these tools for making decisions in those contexts. This comparison is necessarily less quantitative than the comparison of SWAT and InVEST’s water yields, but the choice of three specific decision contexts makes the comparison more concrete.

We consider the following three decision contexts and assess the ability of the models to answer the associated questions:

1. Environmental flows

   **Context:** In many locations, regulations require that local rivers meet minimum environmental flows for each day of the year or for a certain portion of the year to protect in-stream and riparian species or otherwise support the local environment [e.g., Wilson and Carpenter, 1999]. Models must provide estimates of the flows on the short timescales common for environmental flows as well as spatially explicit estimates of the origin of these flows to support this decision context.
Questions: Are environmental flow regulations being met for this watershed? What subbasins should be prioritized for land use management to enhance or ensure environmental flow regulations are met?

2. Land use investments for municipal water supply

Context: ES approaches have been driving investment in conservation and restoration to secure municipal water supplies.

This approach has been most widely applied in Latin America, where The Nature Conservancy’s Water Fund approach [Goldman, 2009] has 17 operating water funds and 26 others in various stages of development. In these water funds, utilities, interested businesses, and environmental groups invest money in restoring or conserving natural areas to secure water resources in terms of water quantity, timing, and quality.

Questions: What subbasins should be prioritized towards HESs to improve the water supply quantity assuming that there is enough storage available to avoid challenges with seasonality?

3. Land use investments for rainfed irrigation

Context: At smaller scales, farmers and ranchers have begun to look at using ES approaches to improve water supplies with rainfed irrigation. Though not addressed here, the farmers may also be looking to reduce nutrient loads or draw down groundwater to avoid salinization. These investments need to be extremely local, often at the field scale, to assure a correct spatial distribution of water and nutrients across and between fields. The estimates of hydrologic response to land management need to be at very high spatial and temporal resolution to assure that the responses are correct in space and that water is available at the correct time for growing.

Question: Where would we target field-scale land use investments towards HESs to improve rainfed irrigation?

4 RESULTS

4.1 HYDROLOGIC RESULTS

Subbasin-level average annual water yields from the InVEST water model match the estimates from SWAT quite well in Wildcat Creek Watershed but not well in the Upper Upatoi Watershed. Figure 2 shows scatterplots of estimates of water yield and ranks of the subbasins by water yield for both models at both sites. Maps of the water yields are included in the Supplement 1.

\[\text{For the purpose of this paper we refer to high-ranked as being a low-number, which also corresponds to a relatively high water yield; thus the sub-basin ranked 1 has the highest water yield and the highest rank.}\]
In Wildcat Creek Watershed the InVEST average annual water yield estimates have similar values and spatial distribution to the SWAT annual water yield estimates as shown in Figure 2. The largest water yields in both models come from the portion of the watershed with the largest human impact, the low-density urban zone in the center north. Similarly, the top 5 ranks match perfectly (Figure 2b); all of these are in this more developed zone. The InVEST estimate of average annual water yield is always within 1 standard deviation of the average annual water yield estimate from SWAT.

The InVEST estimates of average annual water yield vary in a few locations relative to the SWAT estimates in the Wildcat Creek Watershed. In general, InVEST overpredicts the average annual water yield in the low-density urban areas, as seen by the 1:1 line being above trend for these more developed subbasins (Figure 2a). However, the InVEST overestimates are typically within 10%, a
reasonable error given the lower effort to run InVEST (Figure 2a). In addition, while the rankings match quite well, there are some local variations (Figure 2b). SWAT estimates a set of subbasins along the northern edge of the watershed to have much higher ranks than does InVEST. In addition, SWAT ranks a couple of subbasins in the middle of the watershed much lower than InVEST, and SWAT ranks two small subbasins just west of the low-intensity urban zone higher than InVEST. These rankings differ by as much as 17 places, almost 1/3 of the total range of 54 (see S1.1).

In contrast, the SWAT and InVEST estimates diverge significantly in the Upper Upatoi Creek Watershed (Figure 2c,d). The SWAT estimates of water yield vary relatively little across the Upper Upatoi Creek Watershed (Figure 2c), probably due to its largely baseflow-supported hydrograph and its status as a major recharge zone for the Cretaceous aquifer system [Peck et al., 2011]. The InVEST estimates have a significantly larger range, although most of the estimates are within 20% of the SWAT estimates (Figures 2c). The large range of InVEST water yield estimates results in a near-random relationship between the ranks of the two watersheds (Figure 2c).

The average annual water yield estimates from the two models also have strong spatial differences. Both models estimate that a single subbasin in the south-central portion of the watershed has the largest water yield (see S1.2). This subbasin may be an anomaly, as it covers much of a large artillery field where data is sparse due to the presence of unexploded ordnance. However, the water yield estimates vary significantly elsewhere. The InVEST water model estimates that the northwestern part of the watershed provides a large water yield, while SWAT estimates the water yield from that area will be average. The northwestern part of the watershed consists of some low-density urban zones interspersed with significant areas of forest and wetlands. While the urban zones could provide a higher water yield than the forests in much of the rest of the watershed, the SWAT calibration suggests that these low-density urban zones supply quick runoff after a storm but that the well-drained sandy soils elsewhere in the watershed support a great deal of long-term baseflow that provide similar total water yields on a long-term basis. In addition, SWAT suggests that the wetlands in the southern part of the watershed should have an above-average water yield, while InVEST estimates them as only about average. The two models also differ spatially, with the ranking errors in the same areas as the water yield measures mentioned above. The ranking errors are up to 34 places out of 37 total subbasins, leading to significant uncertainty in the value of InVEST’s ranked estimates.

4.2 Decision Application

In addition to comparing the hydrologic results, we also consider the how each models’ results could inform ES-based land management in the decision contexts discussed in the methods. Note that we do not specify whether these models provide
estimates that are accurate enough to make a good decision; that depends strongly on a number of climatic, biophysical, and other variations as well as the social and ecological context of the decision-making process [Ostrom, 2009]. Rather we discuss whether the model results can be reasonably applied in this decision context.

4.2.1 ENVIRONMENTAL FLOWS

The InVEST annual water yield model does not provide estimates of daily flows, and so it cannot be applied to land management decisions about environmental flows. Recently the Natural Capital Project has focused on building a monthly water yield model; however, even at this timescale a model may not provide actionable information on environmental flows as environmental flows may be defined on the daily or weekly temporal scale.

The SWAT model provides estimates of flow at a daily time period, sufficient temporal resolution to assess environmental flows. If an environmental flow policy is clearly stated in terms of flow rate and number of days, the SWAT results can be used to estimate whether regulatory requirements are likely to be met and thus to make land management decisions about conserving or improving ESs to meet such a requirement.

The Logsdon ES indices clearly address whether environmental flows are being provided as ESs. Their fresh-water provisioning index (FWPi) compares the mean flow to a required environmental flow (and provides a suggestion in the absence of regulatory requirements) as well as how often the environmental flow requirement is violated. This index is calculated from any hydrograph, and a model like SWAT can downscale the estimates of discharge to smaller spatial scales so that these indices can be applied at smaller spatial scales than those of stream gauge measurements (though with some error in the downscaling process).

4.2.2 LAND USE INVESTMENTS FOR MUNICIPAL WATER SUPPLY

There are two primary cases for land use investments in municipal water supply: (1) those where significant storage is available and (2) those where there is not significant storage available.

The InVEST water model can provide useful guidance on land use investments for municipal water supply in case (1), because it provides long-term estimates of water yield, and as long as sufficient storage is available this timing does not matter. The InVEST water model provides estimates of water yield at the subbasin scale, so it can provide information on prioritizing land use investments across subbasins. This application depends upon InVEST’s accuracy. The InVEST water model may not provide accurate estimates in areas with significant baseflow or with significant snow or shallow soils, as discussed previously.

Seasonal variations in discharge will play a significant role when there is not significant storage in place (2). The InVEST annual water model does not estimate the seasonal variations, and so cannot be used in this case without externally applying some considerations...
of the seasonal variation. The under-development InVEST monthly water model could address (2) by accounting for seasonal variations in discharge.

The SWAT results provide sufficient temporal and spatial resolution to assess land investments for municipal water supply HESs. SWAT is often calibrated at the watershed scale, and it is commonly used to assess the hydrologic response of the landscape at the subbasin scale [e.g., Bekele and Nicklow, 2005; Rouholahnejad et al., 2012], so it could be used to prioritize investments in municipal water supply HESs.

The Logsdon ES indices cannot, by themselves provide information about the value of land use investments for municipal water supply because they provide estimates of the whole watershed provision of ESs. Instead, these indices need to be coupled with a tool like SWAT to downscale the discharge estimates. When this is done, the Logsdon FWPI can provide information about how changes in land use might affect water provisioning ESs.

4.2.3 LAND USE INVESTMENTS FOR RAINFED IRRIGATION

The InVEST water model cannot provide useful guidance for rainfed irrigation ES land use investments because it does not provide predictions at the appropriate spatial or temporal scale. As an average annual model, it is not possible for this model to assure that water will be available at the correct time for irrigation unless significant storage is present, which is less common for smaller rainfed irrigation systems. More importantly, the InVEST water model predicts the integrated behavior of the catchment so its behavior for individual pixels should not be trusted; rather only its estimates of a significant area of hydrologic response should be trusted. As noted above, the InVEST water model guidance suggests that results should only be used at the subbasin scale, far larger than the field-scale resolution necessary to assess the availability of water for rainfed irrigation.

The basic SWAT model has a temporal resolution that is likely sufficient for planning land use investments in ESs for rainfed irrigation, but lacks the spatial scale necessary to predict hydrologic response at the field scale. SWAT is a semi-distributed model, without the full spatial connectivity of a fully distributed model. Therefore SWAT’s estimates of discharge should not be assessed at the pixel (or field) level. Newer developments with SWAT include using common land-units as the base unit [Kalcic et al., 2014], running SWAT on a grid [Rathjens and Oppelt, 2012], and running SWAT using landscape units [Volk et al., 2007] could be applied to target field-scale investments in ESs. However, these newer SWAT applications are not yet common.

The Logsdon ES indices do not apply at small enough spatial or temporal scales to assess land use investments in ESs for rainfed irrigation. These indices generally assess watershed-scale responses where local gauges can provide hydrographs to estimate ES provision. As stated previously, these indices can use a model like SWAT to downscale to subbasin responses. However, these
indices have not been designed to assess pixel-scale hydrologic responses. In addition, these indices may not reasonably assess water supply ESs on the very short temporal scale necessary for rainfed irrigation due to the time-averaging component of the indices. However, the temporal scale of these indices is less explored than the spatial, and could potentially be modified for this assessment.

5 DISCUSSION

5.1 HYDROLOGIC ACCURACY

In the first part of this work we investigated the hydrologic estimates provided by two different models used for estimating the hydrologic component of ESs: SWAT and InVEST. The two models provided similar estimates of the spatial distribution of annual water yield in the largely agricultural region of the Wildcat Creek Watershed in Indiana. The models provided much more variable estimates of the spatial distribution of annual water yield in the largely forested Upper Upatoi Creek Watershed in Georgia.

Hydrologic models for ESs do not need to be perfectly accurate because the decision-making process occurs in a complex social-ecological system that includes numerous factors that play significant roles beyond the accuracy of the hydrologic model estimates. The hydrologic model only needs to be accurate enough to guide investments in the correct directions given all the other sources of uncertainty. While we do not analyze the necessary model accuracy in such social-ecological systems here, we do suggest that future work should investigate how accurate models need to be to drive correct decisions in such complex systems.

The mismatch of water yield estimates between SWAT and InVEST in the Upper Upatoi Creek watershed raises significant questions. As discussed earlier, we generally trust the SWAT model estimates more, but we do not have subbasin-level data which would allow us to check this conclusively. We performed significant investigation to understand the source of the mismatch, which we present here. We were unable to determine a clear cause of mismatch, but suggest it is likely due to the difference in storage dynamics between the two models.

We regard the SWAT results as more accurate than the InVEST results because of the detailed daily-scale split-scale calibration we performed (see Supplement 3). The SWAT model estimates match both the high peaks and the long-term baseflows seen in this watershed, so we believe that the model is representing the flow from distinct land uses in the watershed well. For example, we expect that the low-density urban subbasins in the northern part of the watershed should provide quick runoff, and therefore provide much of the discharge to match the high peaks. The long-term baseflow must be provided by other subbasins according to a water-balance argument, and SWAT’s simple shallow aquifer model simulates that.
We examined the mismatch between water yield estimates in the Upper Upatoi Creek watershed in relationship to the good match in the Wildcat Creek watershed in several ways, but were unable to find a clear reason for the difference. We considered the watershed area, climate, ecosystem, flow components, and local storage. The two watersheds differ in area by a factor of about two (2083 km$^2$ for Wildcat Creek Watershed versus 870 km$^2$ for Upper Upatoi Creek Watershed), but do not expect that this difference in area will cause a significant difference in the accuracy of the models as both have been used for both smaller and larger sites. We note, however, that InVEST’s integrated catchment approach is likely to work better on larger sites where heterogeneity is averaged over larger scales and local deviations from the no-storage change assumption are less likely to occur. Both the Upper Upatoi Creek and Wildcat Creek watersheds have fully humid climates without strong seasonality in precipitation, so that should not cause a significant difference between the two. Wildcat Creek does receive a significant portion of its precipitation as snow, whereas Upper Upatoi Creek does not; however, since InVEST does not have a snow model of any kind this seems unlikely to be the source of the error. InVEST and SWAT match better in the few strongly human-influenced subbasins of Wildcat Creek, which is surprising since the Budyko Curve approaches that underlie the InVEST annual water yield model were developed for natural, forested watersheds more like that in Upper Upatoi Creek. We also searched for relationships between the water yield ratio ($\text{WY}_{\text{InVEST}}/\text{WY}_{\text{SWAT}}$) and both land use types and components of flow (overland flow, soil interflow, and groundwater flow), grouping land uses in similar categories (see S2). We did not find any clear relationship to explain the mismatch in the Upper Upatoi Creek Watershed.

We believe that much of the mismatch between SWAT and InVEST estimates of average annual water yield in the Upper Upatoi Creek Watershed comes from the significant baseflow seen in that watershed. The strong baseflow component suggests that much of the watershed is able to store large amounts of groundwater, as would be expected in the sandy soils of the area and in the local aquifer recharge zone [Peck et al., 2011]. The discharge never drops much below 3 m$^3$/s, even when it has not rained appreciably for many months. In the SWAT model, much of the Upper Upatoi Creek Watershed has reduced surface flow responses to storms, instead having significant percolation into the aquifer which provides a source of baseflow during dry periods. In addition, changes in the groundwater and soil storage may violate the InVEST model assumption of no storage change, which may account for the difference in these models. We do observe a reasonably strong correlation between the subbasin-level storage change and water yield ratio in the Upper Upatoi Creek watershed ($r^2=0.44$). However, we see a similar range of subbasin-level storage change in the Wildcat Creek Watershed with virtually no relationship to water yield ratio, so this does not uniquely describe the difference between the SWAT and InVEST models.

Further work is needed to understand the value of InVEST for assessing water yield in watersheds like Upper Upatoi Creek, and to assure its estimates of ESs don’t lead to misinformed decisions about ES-based land management. Despite the mismatch, InVEST’s
estimates of average annual water yield are in the correct range after calibration. Figure 2 shows that only one subbasin has more
than 20% error. However, InVEST’s estimate of the water yield spatial distribution varies significantly from the SWAT estimate. The
InVEST model estimates that the northern subbasins contribute much more water than the southern, and so might lead to
investments in the southern subbasins for increased water supply. In contrast, the SWAT model estimates a relatively consistent
water yield across different subbasins, so that investments should be much more evenly distributed and driven primarily by other
factors.

We suggest further work is needed on two fronts: 1) to assess where InVEST’s average annual water yield estimates are reasonably
accurate and where they vary significantly from reality or the results of more sophisticated models (these considerations should be
focused on the areas where model can be reasonably applied to the decision context); and 2) to assess how accurate model
estimates need to be to make well-informed decisions.

5.2 Decision Application

In addition to the tests of hydrologic accuracy, this work considers the ability of different model results to inform particular
decisions. The InVEST annual water yield model, SWAT, and the Logsdon ES indices can provide valuable information for some
decisions, but none of them serve all purposes.

We selected three distinct decision contexts for considering ESs in land management decisions, as discussed earlier. Each decision
context has distinct temporal and spatial scales associated with them (Figure 3). As shown, the spatial and temporal scales of these
decision contexts overlap the spatial and temporal scales at which models can estimate the hydrologic response of watersheds. A
perfect decision-making process would include information at all spatial and temporal resolutions to consider all the possible HESs
affected by land management. However, this large amount of information would likely overwhelm decision-makers who often ask
for simple indices to assist in making decisions. We also note that hydrologists debate the appropriate scales of representation for
hydrologic processes [e.g., Sivapalan et al., 2003], so it may not be realistic to represent all scales well within one model.
Figure 3 illustrates one suggested way forward; it couples the temporal and spatial scales of distinct ES decision contexts with the ability of models to provide information at these scales. We believe that this format helps visualize the process of matching the scale of models to different decisions (a process that hydrologists have been doing for as long as there have been hydrologic models) and will provide useful guidance as groups push to include HESs in decision-making processes.

Spatial Scale

- Global
- Regional
- Basin
- Sub-basin
- Reach
- Field

Temporal Scale

- Daily
- Monthly
- Yearly
- Decadal

Figure 3: The different models investigated here can be used to address distinct decisions around HESs due to changes in land use management. These differences can be shown in a figure displaying the temporal and spatial scale at which each model operates (colored boxes, see key in the lower-left corner) as well as the HES decision context (noted in the words, with a pale box around each). This figure does not show where each model may give accurate results, which requires other considerations such as land use type, soil depth, and amount of baseflow. [1.5 column image]

We chose three decision contexts focused around water quantity to show this approach. These decision contexts do not represent all of the potential ways in which HESs may meet investment or regulatory goals through land management. In particular, InVEST, SWAT, and the Logsdon ES indices all include models/components/indices focused on water quality. Many investments are being made to reduce sediment and nutrient loading in both the US and abroad [Goldman et al., 2010; Keeler et al., 2012]; therefore this work could be expanded upon by performing a similar investigation of water quality.
SWAT, InVEST, and the Logsdon ES indices each have their own strengths but none serve all purposes. The InVEST water model is very simple to run and may provide accurate results for some sites at large spatial and temporal scales; systematic assessment of its accuracy is ongoing. SWAT estimates the hydrologic response of a watershed at higher temporal resolution, but does not put its results in an ES context. In addition, SWAT’s heavier computational load means that it is resource-intensive at very large spatial scales, and its additional complexity means that building high-quality SWAT models requires expertise that may not be available in many decision contexts. The Logsdon ES indices target ESs specifically, which has great value, but can only estimate these services where either hydrologic measurements or available or when linked to a hydrologic model. Ideally models for HESs would couple the high spatial and temporal resolution of a model like SWAT with the conceptual information of the Logsdon ES indices.

6 Conclusion

We assessed an ES-specific hydrologic model (the InVEST annual water yield model) against a standard ecohydrologic model (SWAT) in two distinct locations. We investigated both the hydrologic accuracy of the simple model and the value of each models’ results for considering the HESs that could be obtained through land management and their use to support investments or meet regulations.

We compared hydrologic results from the InVEST water model and SWAT for the first time in the peer-reviewed literature; this has significant value towards validating the InVEST water model which is being widely used to assess ESs. We found that the InVEST water model’s estimates matched those for SWAT very well in the Wildcat Creek Watershed in Indiana, but did not match the SWAT estimates in Upper Upatoi Creek, GA well, and suggested a few reasons why InVEST might perform better in the former. We also considered three distinct decision contexts to determine whether each of these models might be useful in them, and considered how useful the Logsdon ES indices might be in those same decision contexts. We showed a mapping of HES decision contexts against model spatial and temporal scales as one way to organize our assessments of ESs and to evaluate when particular approaches might be useful. While this work is not a comprehensive assessment of methods to evaluate HESs, we hope that it sets a precedent for useful assessment of models and the ES decisions that they can duly support.

7 Acknowledgements

The authors would like to thank Stanford University’s Department of Earth System Science and the NSF Graduate Research Fellowship Program for financial support. In addition, the authors would like to thank Steve Gorelick and Yonas Ghile for their thoughtful comments through the early stages of this work, and helpful comments from the editor and reviewer.


8 SUPPLEMENTS

SUPPLEMENT 1

Supplement 1 contains maps which display the detailed spatial patterns of both average annual water yield values and average annual water yield rank, for both Wildcat Creek Watershed and Upper Upatoi Watershed.

SUPPLEMENT 2

Supplement 2 contains a table which describes the groups of land uses analyzed to investigate systematic differences between InVEST and SWAT in the Upper Upatoi Watershed.

SUPPLEMENT 3

Supplement 3 describes the SWAT model built for the Upper Upatoi Watershed, including data sources, parameters selected through calibration, and the final calibration statistics.
Click here to download e-component: 20160125_Supplement1_Maps.docx
Click here to download e-component: 20160125_Supplement2_Tables.docx
Click here to download e-component: 20160211_Supplement3_UpatoiSWAT.docx