Development of a watershed modelling approach to assess the impacts of biofuel feedstock on water quality in the Upper Mississippi River Basin

D. J. Lampert & M. M. Wu
Argonne National Laboratory, Lemont, USA

Abstract

The United States Energy Independence and Security Act of 2007 mandates production of 36 billion gallons of renewable fuel per year by 2022 in the United States, most of which will be met by biofuels. An essential component in meeting this challenge will be to produce biofuel feedstock in a sustainable manner, including quantifying the effects of biofuel production on water quality. Herein, the integration of several recently developed geographic information systems (GIS) datasets and the Hydrologic Simulation Program in Fortran (HSPF) using the Python Programming Language are presented. The approach allows for rapid development and calibration of an HSPF model for assessing agricultural runoff.
Keywords: watershed modelling, geographic information systems, pesticides, Python programming.

1 Introduction

Because of changes in technology, land use, and rapid population growth, both the quantity and quality of water in engineered and natural systems are increasingly stressed. As a result it is increasingly important to assess the sustainability of government policies that impact water availability and water quality. The production of energy is intimately linked to water, and thus sustainable energy production requires assessments of water quantity and quality. The United States Energy Independence and Security Act of 2007 mandates production of 36 billion gallons of renewable fuel per year by 2022 in

the United States. Most of this demand will be met by the production of biofuels. Meeting this demand will impact both water quantity (through increased irrigation and different farming practices) and water quality (associated with production of different crops).

Biofuels can be produced from edible biomass sources such as sugars, starches, plant oils and fats (first generation biofuels). However, these first generation biofuels utilize only a fraction of the biomass created from their cultivation is utilized for energy production [1]. As a result, first-generation biofuels have limited carbon sequestration benefits and may impact regional food supplies. As an alternative to first generation sources, biofuel can be produced from second generation sources composed of lignocellulosic biomass such as wood chips, grasses, crop residue, and municipal and industrial wastes. Second generation biofuels have different soil erosion and chemical runoff characteristics than first generation sources [2]. They offer potentially greater carbon sequestration benefits, lower energy inputs, and lower water usage [3]. However, these sources also suffer from high capital and production costs and require large transportation-related energy expenditures. Other higher generation biomass fuel sources exist including algae and photosynthetic bacteria. These sources offer potentially high carbon sequestration and biomass density benefits albeit with very high monetary costs and water usage requirements.

Making accurate assessments of the potential costs and benefits of each of the different biofuel production pathways is a critical component in the sustainable development of biofuels. Currently, biofuel sustainability is assessed using life cycle analysis (LCA) to determine net energy use, net carbon emissions, and net water usage. In addition, other factors must be considered to determine the long-term feasibility of various biofuel-production practices including soil erosion and water quality.

The cultivation of different biomass sources will affect water quality through the production and application of fertilizers, pesticides, and herbicides, wastewater discharges from biofuel production facilities, and from spills during fuel transport [4]. Different biomass production techniques will have different impacts on water quality. For example, the removal of corn stover from agricultural could significantly alter pesticide loadings due to the increased water runoff and soil erosion. The United States Environmental Protection Agency (EPA) recently released a recommendation to determine appropriate best management practices (BMPs) to avoid or mitigate negative environmental effects from biofuel production, including water quality [4]. Predicting many of these effects requires a model capable of linking the changes in land use with water quantity and quality and soil erosion.

2   Methods

2.1 Watershed model selection

The literature on large-scale hydrologic watershed modeling from non-point sources (i.e., from agricultural runoff) was reviewed and two models were

selected for final analysis, the Hydrologic Simulation Program Fortran (HSPF) and the Soil and Water Assessment Tool (SWAT). Several studies have been performed comparing these two models, and the general consensus was that SWAT is easier to use, more frequently applied, and scales to large watersheds; HSPF has a more robust hydrologic and water quality component, but is normally limited to smaller watersheds and has a more difficult learning curve. Because sophistication and accuracy were deemed more important than ease-of-use, HSPF was selected as the best model for water quality assessment of biofuels.

2.2 Hydrologic Simulation Program in Fortran (HSPF)

HSPF is based on the Stanford Watershed model originally developed in the 1960s at Stanford University [5]. The model divides a watershed into a series of pervious and impervious land segments, each of which feed into a reach. The reaches are connected together to form a watershed. Typically HSPF land segments are based on land use categories that can be thought of us homogeneous; e.g. forests, agriculture, developed, etc. Flow rates in reaches can be used to compare model simulations to gauge station values for model calibration. A calibrated model can then be used to assess how changes in land use or climate will affect water quantity and quality.

The model has changed very little from the version that was released over 30 years ago, which is a testament to how well the original program was written. However, the format of the input files is cumbersome to develop and the input data sources available today are very different than those available at that time. The original source code for HSPF is available through the United States Geological Survey (USGS) website. The source was studied and compiled using the open source GNU Fortran compiler [6] on both Microsoft Windows 7 [7] and on the open-source Linux-based operating system Ubuntu [8]. It may be useful to run simulations on either operating system. The HSPF source code contains 14 test programs, including the input files and the expected output. Each of the test files was run and compared to the output file. The results were consistent which indicated a successful compilation.

HSPF utilizes two input files to project water quality parameters. One is the user control input file (UCI), which is a text file that contains all the commands to call the relevant modules for each part of the watershed. The other is a watershed data management (WDM) file, which is a direct-access, unformatted binary file of time series of data inputs (e.g., a list of the hourly rainfall volume at a weather station over the period of the simulation). HSPF utilizes the commands from the UCI file and the data in the WDM file to generate time series predictions of model output (e.g., the concentration of a pesticide at a particular point in the watershed).

2.3 Python Interface to HSPF

To streamline the model development process, an interface to HSPF was developed using the Python Programming Language [9], including the

NumPy [10], Osgeo [11], and matplotlib [12] modules. NumPy provides support for large, multi-dimensional arrays and matrices, and a library of high-level mathematical functions to operate on these arrays. NumPy also contains an interface, f2py, that allows Python programs to call Fortran subroutines. Osgeo is used for geospatial data calculations, and matplotlib is used for plotting.

A series of Python and Fortran scripts were developed to provide straight-forward access to the needed subroutines in the complied HSPF library, including the main routine HSPFBAT. Additional subroutines were developed to provide access to the subroutines used by HSPF can be used to create the WDM file. These scripts allow a user to create and manipulate datasets in the WDM files including creating a WDM file; adding time series datasets to it; adding time series attributes (time step, units, locations, etc.) to a dataset; adding data to the datasets, and reading the timeseries data and attributes into conventional Python formats.

2.4 Geospatial data

Development of a watershed model with HSPF requires geospatial data from geographic information systems (GIS) describing the subbasin catchment areas, overland flow parameters, subbasin reach slope and geometry, and subbasin land use types. HSPF makes no assumptions about a watershed; thus it is incumbent on the user to properly supply the data for a watershed model.

The USGS divides the surface water in the United States into a series of hydrologic unit codes (HUCs) of varying length (corresponding to the size of the catchment area). The National Hydrography Dataset Plus is the most detailed map of the surface water in the United States. The NHDPlus dataset divides streams into cataloguing units with a unique 14-digit code. The original source data for various 2-digit HUCs for V2 of NHDPlus are available at (http://www.horizon-systems.com/NHDPlus/NHDPlusV2_home.php). The source data include the 30-m National Elevation Dataset (NED), flowlines and the catchment area shapefiles, and several databases of shape attributes. The raw data for HUC 07 (Upper Mississippi River Basin, UMRB) were downloaded and used to generate input parameters for HSPF.

The United States Department of Agriculture's National Agricultural Statistics Service (NASS) has released a GIS dataset containing 255 land use categories on a 30-m grid called the Cropland Data Layer (CDL). The land use categories from the CDL were filtered into 10 relevant categories for developing watershed models for the UMRB. Table 1 summarizes the categories.

2.5 Watershed delineation

A methodology was developed to automatically delineate a USGS 8-digit HUC based on the criteria of establishing a subbasin outlet co-located with USGS gauge stations and ensuring a maximum area of 400 km^2. Each subbasin was then divided into land segments based on the land use categories listed in Table 1. The input UCI file was then generated using the GIS data for the watershed.

Table 1:   Land use categories for watershed modelling.

Corn
Soybeans
Other grain
Developed
Water/wetland
Forest
Hay/alfalfa
Fallow land
Pasture/grass
Other

2.6 Climate data

Climate data for the HSPF model, including hourly precipitation, daily minimum and maximum temperature, and daily pan evaporation data were obtained from the National Climate Data Center (NCDC). For a given watershed, the nearest available pan evaporation and temperature gage values were used to construct a single hourly time series of values for the watershed. For each subbasin within a watershed, a specific hourly time series was developed based on the nearest available gages. A methodology was developed to automatically extract data for a USGS 8-digit watershed and create the input WDM file for HSPF.

3 Results

The raw data for HUC 07100008 (Lake Red Rock, IA) are shown in Figure 1. There are approximately 3400 NHDPlus flowlines and catchments and 14 USGS
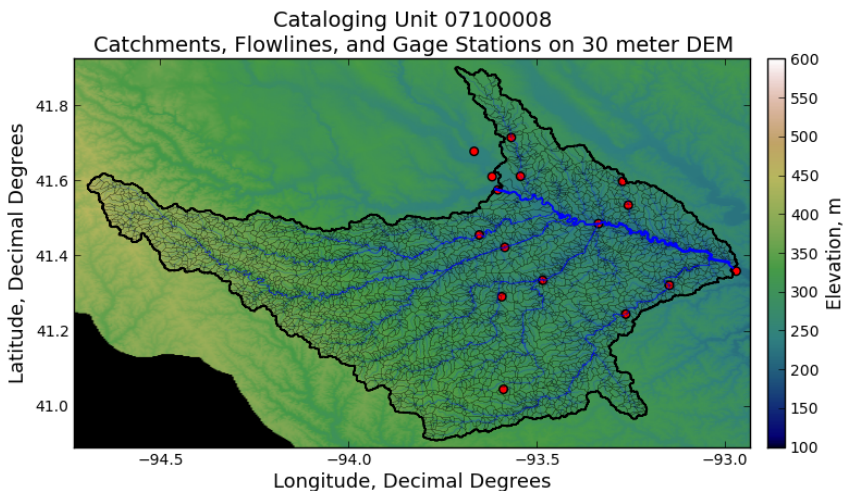

Figure 1:   Raw GIS data for Lake Red Rock, IA.

flow gauge stations at the watershed. Figure 2 shows the data after processing. Figure 3 shows the land use data after filtering into the 10 land use categories for the watershed. The fraction of each land use type was calculated in each subbasin.

The raw and computed GIS data and the NCDC climate data were used to build an HSPF model for the watershed. The model hydrology is currently being calibrated. Following the completion of the hydrology calibration, the model will be calibrated for sediment transport and then pesticides. The calibrated model can be used to assess future scenarios where different land use will be
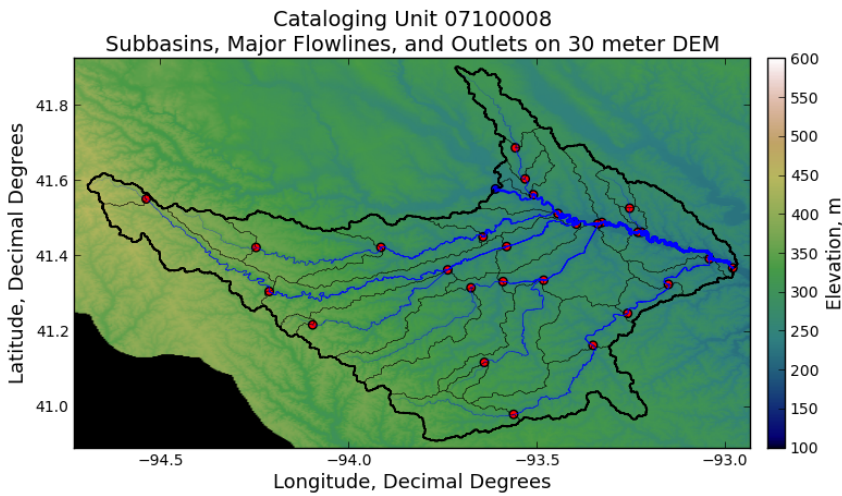

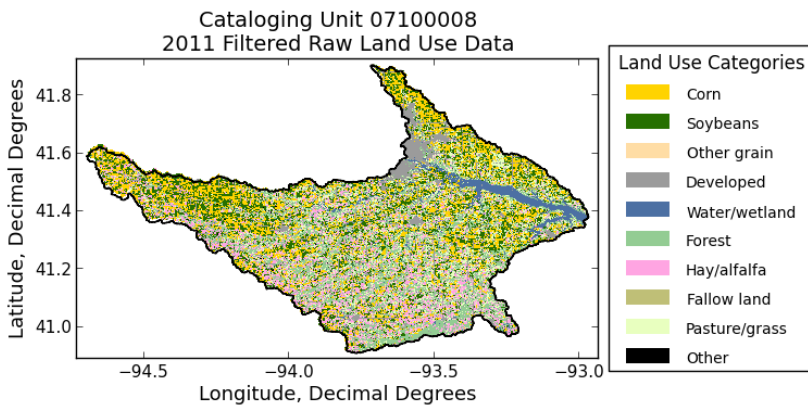Figure 2:   HSPF processed data for Lake Red Rock, IA.

Figure 3:   Raw land use data for 2011 for Lake Red Rock, IA.

allocated, or for assessing climate change effects. The modelling approach can be utilized at other 8-digit watersheds throughout the UMRB to answer questions related to groundwater recharge, evapotranspiration (for life cycle assessments), water quality effects, and sediment transport loadings downstream.
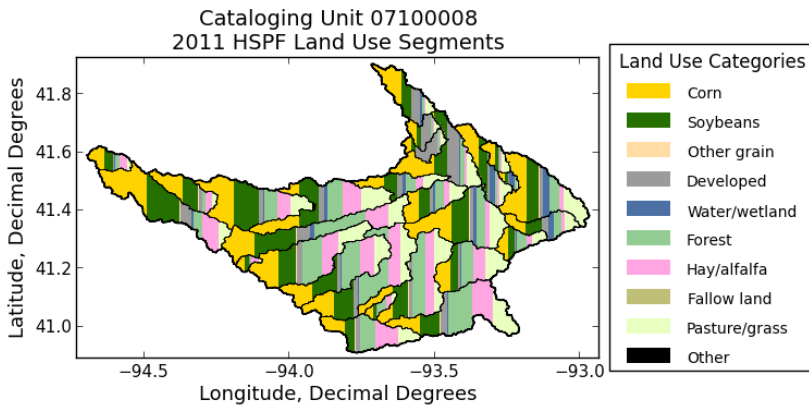

Figure 4:   Land segments for HSPF modelling of Lake Red Rock, IA watershed for 2011.

4   Conclusions

The results of this study will be useful to policy-makers for weighing the water quality impacts of biofuel production against potential benefits. The study will also help farmers and other stakeholders understand best management practices to utilize to minimize impacts on water quality. Energy and water are closely related, and as biofuel production increases the associated effects on water will be more important. Therefore developing scientifically sound and computationally efficient methods for analysing the energy-water relationship now will be helpful in answering questions as they arise in the future. The information provided by these models will be useful in assessing the sustainability of biofuels, which are a growing source of energy production throughout the world.

References

[1]   van der Laak, W.W.M., Raven, R.P.J.M., & G.P.J. Verbong. Strategic Niche Management for Biofuels: Analysing Past Experiments for Developing New Biofuel Policies. *Energy Policy*, 35, pp. 3213-3225, 2007.

[2]   McLaughlin, S. & Walsh, M. Evaluating Environmental Consequences of Producing Herbaceous Crops for Bioenergy. *Biomass and Bioenergy*, 14, pp. 317-324, 1998.

[3]   Semere, T. & Slater, F.M. Ground Flora, Small Mammal and Bird Species Diversity in Miscanthus (Miscanthus x giganteus) and Reed Canary-grass (Phalaris Arundinacea) Fields. *Biomass and Bioenergy*, 31, pp. 20-29, 2007.

[4]   U.S. EPA. *Biofuels and the Environment: the First Triennial Report to Congress (External Review Draft). United States Environmental Protection Agency, Washington, DC, EPA/600/R-10/183A,* 2011.

[5]   Crawford, N. H., & Linsley, R. K. Digital Simulation in Hydrology's Stanford Watershed Model 4. Transportation Research Board of the National Academies: Technical Report Number 39, 1966.

[6]   GFortran, GNU compiler collection (gcc), Version 4.6.2, GNU General Public License, 2012.

[7]   Windows 7, Redmond, Washington, Microsoft Corporation, 2009.

[8]   Ubuntu Version 11, GNU General Public License, 2012.

[9]   Python Programming Language, Version 3.2.3, Python Software Foundation, 2012.

[10]   NumPy Version 1.6.2, Python Software Foundation, 2012.

[11]   OSGeo Version 2.0.8, Open Source Geospatial Foundation, 2012.

[12]   Matplotlib Version 1.2.0, Python Software Foundation, 2012.