Biodiesel Sim: Crowdsourcing Simulations for Complex Model Analysis

Derek Riley, Xiaowei Zhang,
University of Wisconsin- Parkside
Kenosha, WI
rileyd@uwp.edu, zhang010@rangers.uwp.edu

Xenofon Koutsoukos
Vanderbilt University
Nashville, TN
xenofon.koutsoukos@vanderbilt.edu

Keywords: Crowdsourcing, Human Computation, Biodiesel

Abstract

Biodiesel is an alternative fuel source that can be easily made by novices with an inexpensive home-made reactor using waste vegetable oil, but producing high quality fuel using a home-made reactor is difficult due to the complexity of the chemical interactions and the configurations of the processor. A biodiesel processor is a complex system that can be modeled and simulated using formal modeling methods, but accurate modeling can require prohibitively expensive analysis. In this work we present our model analysis approach designed to 'crowdsource' simulations of a formal Stochastic Hybrid System model of biodiesel production. The approach uses an iPad application called Biodiesel Sim, which allows users to play a simulation game that calculates trajectories of a formal stochastic model of biodiesel production. The results of simulations run in the game are collected in a centralized database, which can be searched by users or researchers via a website to encourage users to improve their simulations. The framework used in this application can be used as a template for analysis of other formal models.

1. INTRODUCTION

Biodiesel is an important biofuel that has the potential to significantly reduce reliance on fossil fuels and improve emissions from diesel vehicles. While the chemical reactions that produce biodiesel are well-understood, and it is possible to build an inexpensive reactor in a garage, it is still a challenge to design and run a biodiesel reactor that can produce high quality fuel from a variety of oil feed stocks inexpensively. Smaller, adaptable processors have the potential to be economically more viable than larger processors using a single virgin feedstock [7], so approaches that enhance their study are important. In this work, we present our Biodiesel Sim game designed to 'crowdsource' model simulation and analysis by encouraging iPad users to simulate a formal model of biodiesel production in various configurations to learn how to produce quality fuel inexpensively. As the users run simulations, the results are stored in a centralized, searchable database that is accessible to users to help them improve their efforts.

Crowdsourcing approaches put humans in the loop to help guide computers to improve the analysis of complex data or systems. Crowdsourcing (sometimes called 'human computation') is a relatively new term, so the definition is not established formally, but all crowdsourcing-type problems share similarities including orchestrated coordination of a large number of distributed human 'crowdworkers' and solving problems through small, convenient contributions. The tasks performed by humans are often accomplished through carefully-designed applications called 'games with a purpose,' and computers then coordinate and facilitate the results of the human computation [30].

Crowdsourcing approaches have been used to unlock insights into protein folding [8], geometric reasoning [13], genetic algorithms [19], and more. Crowdsourcing tasks vary widely, but they all share similarities including orchestrated coordination of a large number of distributed individuals and solving problems through small, convenient contributions. The potential applications of crowdsourcing approaches have been increasing as the number and sophistication of mobile devices in the general population increases.

Not all algorithms that involve humans and computer collaboration are considered crowdsourcing approaches. Online collaborations such as Wikipedia or forums are not generally considered as crowdsourcing because the activity is primarily directed by participant intuition and interest. Crowdsourcing problems in comparison are coordinated a priori to and contain a limited amount of creative content [23].

Several challenges exist when developing crowdsourcing approaches. First, splitting the large problems into smaller tasks requires a deep understanding of the underlying problem as well as human and computer abilities. Designing an engaging, attractive application that will attract many users and be intellectually stimulating enough to entice the users to continue using the tool is difficult. Further, combining large numbers of results to gain meaningful insights is often difficult due to the quantity and quality of results that are collected with crowdsourcing algorithms. Lastly, because crowdsourcing tools often target non-expert participants, the algorithms are vulnerable to malicious manipulation [9].

In this work we present a crowdsourcing approach to explore the state space of a Stochastic Hybrid System (SHS) model of biodiesel production. We have created an iPad application called Biodiesel Sim that allows users to execute a biodiesel model simulation using our SHS model of biodiesel production in the form of a game. The game is intended to

educate users about biodiesel production, and it allows them to run their own custom simulations of a batch biodiesel production system to attempt to minimize the cost. The game presents users with the results of their simulations (the chemical concentrations and production cost) and encourages them to try to simulate the system again to improve the results. The results of all simulation trajectories are sent to a central database, which can be searched by users and used for data mining including probabilistic verification or other analysis.

Biodiesel processors have been previously modeled and simulated to refine their design. Proprietary chemical engineering deterministic modeling tools such as ProSim [22], Aspen [3], and others provide a rich set of design and analysis tools. The focus of the studies that use these tools have been varied from analyzing catalysts and processing methods to reducing cost through feedstock optimization [16, 31, 12, 32]. However, these tools require in depth, technical knowledge of the chemical reactions and the modeling tools.

We model our batch biodiesel reactor with a SHS model for several reasons. SHS are a formal modeling paradigm that can be used to model discrete, continuous, and stochastic dynamics of any biochemical system with known kinetic reaction properties. It is important to capture all three types of dynamics (discrete, continuous, and stochastic) because modern biodiesel reactors are complex and exhibit a combination of all three types of dynamics. Our model of biodiesel production is based on a previous, validated SHS model [28, 24].

Well-established deterministic analysis techniques have been used to study the dynamics of the biodiesel reactions [2] and catalysts that can be used to drive the reactions [18]. These previous analysis methods for SHS models have used brute force approaches to analyzing the state space and improve performance using high-performance computing with statistical model checking techniques or other verification methods [24, 1]. For highly complex systems, brute force analysis is too computationally expensive. Our crowdsourcing approach distributes the computational cost of the brute force approach and uses human ingenuity to explore only 'interesting' areas of the state space.

2. CROWDSOURCING BIODIESEL ANALYSIS

The problem that Biodiesel Sim is attempting to solve is the exploration of the solution state space of a formal SHS model and identification of potentially optimal solutions. Specifically, we are interested in finding initial configurations of the model variables that lead to inexpensive, high quality fuel. Our solution to this problem is to encourage and collect the results of distributed simulations on iPads used by novice users and use the results to populate a central repository that can be searched and analyzed by users or computers. The architecture of the approach is demonstrated in Figure 1.

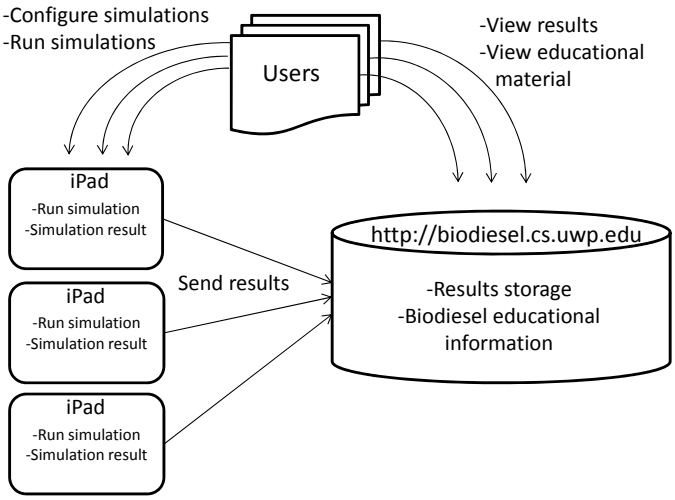

Figure 1. Biodiesel Sim application architecture

Our architecture (shown in Figure 1) allows any number of iPad users to simultaneously configure and run simulations on their individual iPads. Users are then presented with the results of their simulation (the cost of the fuel and the ending chemical concentrations), and the results are automatically sent to the central database. Results from all users are stored in the database (hosted on the companion website) where all users can see results from all other users. The website also presents biodiesel educational material to help inform users to understand the underlying process and model to improve their accuracy.

Previous analysis methods use a brute-force approach [20], but brute force methods are computationally expensive and intractable for large models. The brute force approach uses little to no knowledge of the underlying problem, so it is inherently inefficient because it explores regions of the state space that have a low probability of yielding an optimal solution. Instead, our crowdsourcing approach uses human users to identify patterns and select configurations to explore the state space.

Humans are inherently good at identifying patterns and applying knowledge of an underlying problem to find a solution, especially if the humans are informed about the underlying problem characteristics and motivation. Therefore, by educating users about the underlying processes involved and allowing users to use their unique pattern-identification skills allows for a more intelligent and effective analysis of the system. We encourage the unique strengths of human pattern matching by presenting the user with detailed, engaging data to encourage them to continue to search and test configurations. Educational materials are featured prominently on the companion website to encourage users to learn and become better contributors to the project.

The Biodiesel Sim application is designed to leverage

two aspects of crowdsourcing. First, inherent human pattern matching is encouraged by informing the user of the details of the simulation as the simulation progresses. Patterns of how the simulations progress are visually distinct in the numerical and pie chart output during a simulation. Patterns are also identifiable in the saved simulation data available in the companion website. The identification of these patterns is meant to lead the user to hypothesize about how untested configurations might proceed and encourage users to test their hypotheses by running more simulations. The more educated users are, the more likely they are to identify and test good hypotheses.

The second crowdsourcing aspect encourages users to access simulation results of other users via the companion website to find previous simulation results from all other users of the application. Searching previous simulations allows a user to learn from others' work and hopefully find new and improved contributions.

Users can search simulation results using several filtering criteria on the companion website. By finding 'successful' simulation configurations in the database, any user can load and try out similar configurations to see if they can find a better configuration. Further, the database of all simulations can be searched by an automated computer program using data mining techniques to find untested configurations that are likely to lead to good results. The use of an automated data mining program is left for future work in this project.

One of the challenges of crowdsourcing is splitting up the overall task into meaningful smaller tasks that humans and computers can accomplish individually. We solved this challenge by allowing the computer (iPad) to compute the numerical simulation component, which is challenging for a human to do. Humans then identify patterns and select configurations, which is difficult for a computer to do. Computer also store and sort the simulation data, which are difficult tasks for humans but are easily accomplished by a computer. By separating the work in this way the strengths of computers and humans are leveraged while the weaknesses of each are avoided.

Malicious manipulation and attracting crowdworkers are two other challenges of any crowdsourcing application. By designing the application around a well-known chemical system, it is intended to attract the attention of a large number of novice users. We have also designed the application to be visually attractive and intellectually stimulating to help attract crowdworkers. Designing the application for the iPad is also intended to attract a large number of crowdworkers due to the large number of people who own iPads. We do not feel that paying crowdworkers for this application is appropriate due to the nature of the task.

By carefully controlling the data that is sent and sorted by the server, we have made it difficult for malicious manipulation of our approach. Security measure that sanitize the data and filter results further assure that malicious manipulation is minimized. As more users create simulation trajectories, the website will aggregate and present the data to further enhance the feedback mechanism of crowdsourcing.

There are several characteristics of the formal model analysis problem that make it an appealing choice for a crowdsourcing approach. Splitting the analysis tasks between humans and computers is relatively straightforward because the concept of running a single simulation and choosing initial configuration parameters is well-understood among the general population of people. Storing and sorting data is a strength of computing systems, so it is natural to assign the computer to manage and store the data and provide access via a website.

Therefore, virtually any formal model could be analyzed using crowdsourcing approaches if the simulation methods execute in a reasonable amount of time and the simulation method can be implemented on a mobile device. One of the limitations of the crowdsourcing approach is the need to attract 'crowdworkers,' so the simulation must be engaging or have broad appeal. If the simulation doesn't have broad appeal, crowdworkers can be attracted by offering monetary incentives as part of the process. There is an active community of people across the world who produce biodiesel and are interested in biodiesel research (especially minimizing cost). This drives inherent interest in the biodiesel model, which is why we have not adopted the approach of paying 'crowdworkers.'

3. BIODIESEL SIM IPAD APPLICATION

The Biodiesel Sim application is designed to allow users to easily manipulate input configurations for a simulated biodiesel processor. Users choose an initial simulation configuration by dragging slider bars to select initial input chemical concentrations, temperature, and reaction times. The application then simulates the processing of biodiesel (using a SHS model of biodiesel production), and the application presents the results of the simulation on the screen of the application. Results are also sent to a central database for analysis.

We chose to develop the application for the iPad due to the large number of iPad users and the large amount of screen real estate available to show simulation details to users to help inform their simulation configurations. We expect that people currently producing biodiesel or interested in producing biodiesel as well as students and others will download and play the application to learn more about biodiesel production. Simulations of users are saved and combined, so the collective crowd intelligence can be leveraged via the companion website of the application.

The initial screen of the application provides a link to more information about the application as well as options to log

in or play the game. A screenshot of the application is seen in Figure 2. Users can create an account to log in for free, which allows them to search and track their simulations, see all simulations in the database, and communicate with other users via a message board. Users can also just play the game without logging in. These 'guest' users still have their results saved, but they are unable to search the online database (described in a later section of this paper).

Figure 2. Welcome screen of the Biodiesel Sim application

To configure the initial simulation settings in the application, users manipulate the initial oil amount, methanol amount, thermostat temperature settings, and catalyst amount used in the reaction. Users are also allowed to set the reaction length (η) and settling time (μ). The overall running time is the sum of the reaction time and the settling time. The users' input choices are all bounded by using slider bar inputs to ensure test configurations are realistic (for example: negative values for any chemical concentration are disallowed). The configuration settings can be seen in the top right of Figure 3.

The user is presented with a 'run' button that allows them to start the simulation once they have finished adjusting the input parameters. After pressing the run button, the users are shown an animation on the left side of the screen including liquid entering a beaker via a faucet and the beaker heating up via fire underneath. Both animations use particle systems to improve their realism. While these visuals do not affect the simulation outcome, they are intended to educate the use about the process of creating biodiesel so they are enticed to learn more about the process.

Pressing the run button starts the animation and a discrete time step simulation of the underlying SHS model using the *HMM* simulation algorithm (discussed in a later section). The input values selected by the user are used to initialize the simulation parameters. The simulation algorithm is run directly on the iPad, and users are shown details about how the simulation is progressing as it is running.

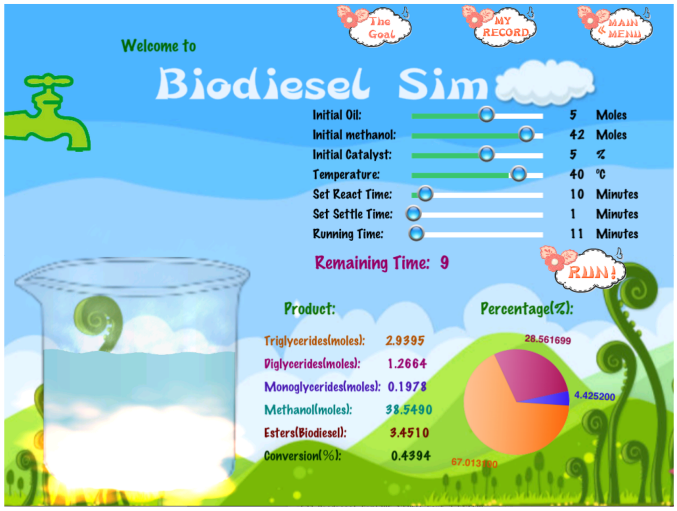

Figure 3. Screenshot of the running Biodiesel Sim application

Users are not allowed to select the simulation time step size since the step size has a significant impact on the quality and the run time of the simulation. The simulation time step selected for the game ($\Delta t = 0.02$) was chosen to allow for a relatively fast simulation with good accuracy (consistent with validated results).

As the simulation executes on the iPad, the amount of time left in the simulation is shown to the user (as seen in Figure 3). The user is also able to see the current percentages of each chemical in solution via a color-coded pie chart that updates with each simulation step. For more precision, the user is also shown the raw numerical chemical concentrations at each time step so they can see how the simulation is progressing (to help them learn how the reactant concentrations evolve over time). The screen shot shown in Figure 3 shows the pie chart and raw chemical concentrations for an example run in the lower right corner. It is intended that advanced users will use this presented information to help them find new ways to adjust the input parameters to find better solutions.

After completion of a simulation run, the application attempts to establish a connection to the database server, and if a successful connection is made, the simulation results (the initial configuration and resulting score) are sent to the database. The application passes the data using a URL and calls PHP code on the server to connect to the database. In this way, calls to the database can be sanitized and protected from security threats such as SQL injections.

After the simulation has completed, the user is shown a dialog box that summarizes their simulation results. It tells them whether fuel was successfully produced as well as other summary data. We consider fuel to be 'successfully' produced if 98% of the resulting product is biodiesel and only 2% un-

reacted oil. Too much unreacted oil in the final product can damage sensitive fuel injectors and emissions equipment in modern diesel engines. An example dialog box that summarizes the simulation is shown in Figure 4.

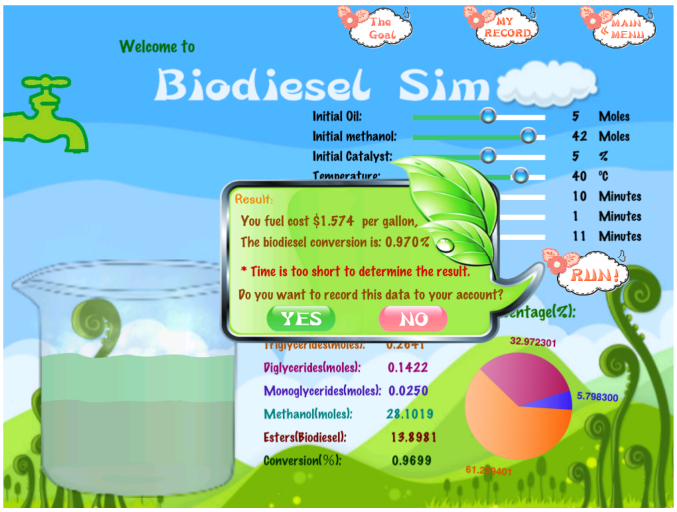

Figure 4. Screen shot at the end of a simulation

The 'score' for the game is the fuel cost since the goal of the game is to minimize the fuel cost (while still producing quality fuel). The fuel cost is calculated assuming the initial oil is free (as is the case with many waste oil reactors). The fuel cost takes into account the cost of methanol, catalyst, and the electricity to heat the tank. All of these variables change depending on how the user configures their simulation. Equation 1 below shows the equation used to calculate the cost where T is the temperature, M is the amount of methanol, TG is the concentration of triglycerides, DG is the concentration of diglycerides, and MG is the concentration of monoglycerides. The initial oil quantity is translated into TG, DG, and MG by dividing the initial oil amount into the percentages of TG, DG, and MG found in canola oil (a commonly used oil in biodiesel reactors).

$$cost = \frac{(T/1050 - 0.012) * \mu + M/5}{TG + DG + MG} \quad (1)$$

The biodiesel model is implemented in the application using the native iOS programming language, Objective C. A custom implementation of the *HMM* simulation algorithm was written in Objective C and adapted to update the user interface with results after each time step [29]. Since the *HMM* algorithm is a discrete, fixed step simulation method, the main calculations are done in a loop that executes until the simulation end time is reached.

Chemical concentrations are stored in the program as instance variables, and they are updated with each time step. As they change, the user interface is updated including the pie chart and other components. The differential equations used to calculate the chemical concentration changes are hard coded into the program, but adapting the simulation components to accommodate changes to the chemical dynamics or model other chemical systems is not difficult.

The simulation executes relatively quickly on an iPad, but the length of time is dependent on the configuration chosen by the user. The simulation performs many calculations, so it is processor-intensive, and on older iPads, it requires more time to execute. On new iPads, the simulation takes almost one second of 'real' time to execute one second of 'simulated' time. The simulation time step selected ($\Delta t = 0.02$) was chosen to allow for a relatively fast simulation with good accuracy. Larger step sizes were tested, and produced faster simulation results, but the accuracy was compromised, so we determined that $\Delta t = 0.02$ is the appropriate step size. As newer processors and hardware become available, the simulations will become much quicker.

4. MODELING AND SIMULATION

We chose SHS as our modeling paradigm because it is a flexible, efficient modeling paradigm that can capture discrete, continuous, and stochastic modeling components. In this section, we present details of the formal modeling and simulation approach and our model of biodiesel production.

4.1. SHS Modeling and Simulation

SHS have been used to model and analyze complex, interesting systems with discrete, continuous, and probabilistic dynamics [25, 26, 6, 1, 4, 5, 14]. SHS contain a set of discrete states $q \in Q$, invariants associated with the discrete states $X_q \subseteq \mathbb{R}^n$, and continuous dynamics associated with the discrete states $x(t) \in \mathbb{R}^n$. Discrete transitions between the states occur either because the continuous state x satisfies the transition guard $x(t) \in \partial X^q$ (guarded transition) or based on an exponential firing rate λ (probabilistic transition). A reset measure R is associated with any transition. The hybrid state at time t is $s(t) = (q(t), x(t))$.

To define the execution of the system, we denote $(\Omega, \mathcal{F}, P)$ the underlying probability space, and consider an $\mathbb{R}^p$-valued Wiener process $w(t)$ and a sequence of *stopping times* $\{t_0 = 0, t_1, t_2, \ldots\}$. Let the state at time t_i be $s(t_i) = (q(t_i), x(t_i))$ with $x(t_i) \in X^{q(t_i)}$. While the continuous state stays in $X^{q(t_i)}$, $x(t)$ evolves according to the stochastic differential equation (SDE)

$$dx = b(q,x)dt + \sigma(q,x)dw \quad (2)$$

where the discrete state $q(t) = q(t_i)$ remains constant and the solution of (2) is understood using the Itô stochastic integral [17]. A sample path of the stochastic process is denoted by $x_t(\omega), t > t_i, \omega \in \Omega$.

The next stopping time t_{i+1} represents the time when the system transitions to a new discrete state. The discrete transition occurs either because the continuous state x exits the

invariant $X_{q(t_i)}$ of the discrete state $q(t_i)$ (guarded transition) or based on an exponential distribution with transition rate function λ (probabilistic transition). Therefore, t_{i+1} can be defined as the minimum between two other stopping times: (i) The first hitting time of the boundary $\partial X^{q(t_i)}$ defined as $t^*_{i+1} = \inf\{t \geq t_i, x(t) \in \partial X^{q(t_i)}\}$ and (ii) a stopping time τ_{i+1} described by an exponential distribution with a firing rate λ. At time t_{i+1} the system will transition to a new discrete state and the continuous state may jump according to the reset measure R. The evolution of the system is then governed by the SDE (2) with $q(t) = q(t_{i+1})$ until the next stopping time.

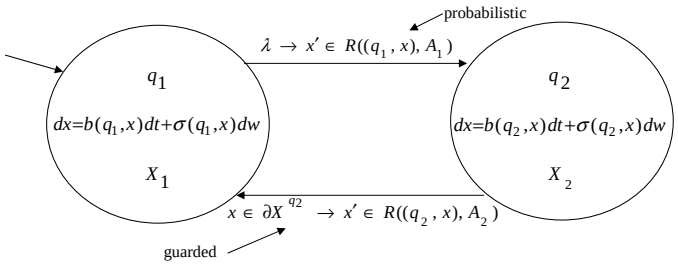

Figure 5. Stochastic hybrid system

Figure 5 shows a generic SHS model with two states and two transitions (one probabilistic and one guarded). The continuous dynamics of each state are defined by the associated stochastic differential equations. The probabilistic transition fires at the firing rate λ, and the guarded transition fires when x hits the boundary $x \in \partial X^{q_2}$. The logical condition $x \in \partial X^{q_2}$ is often referred to as the guard of the transition. Upon firing of a transition, the state resets according to the map $R((q,x),A)$. A full description of the formal model for SHS can be found in [27].

To simulate SHS, simulation of SDEs must be combined with simulation of the guarded and probabilistic discrete transitions in a way that captures the formal execution semantics. We developed a simulation method for SHS that performs numerical integration of SDEs with the Milstein Method (MM) and uses probabilistic discrete transition detection to improve accuracy. We introduced an improved simulation method, the Hybrid MM (or *HMM*), and described in detail in a previous work [29].

4.2. Biodiesel Production Model

Biodiesel is made from vegetable oil and other chemicals by a process called transesterification [21]. The process involves six chemical species and six highly-coupled reactions (Table 1). Vegetable oil in its purest form is made up of triglycerides TG; however, it breaks down into diglycerides DG and monoglycerides MG as it is heated. An alcohol, methanol M, is combined with the TGs, DGs, and MGs to convert them into biodiesel esters E and glycerine GL.

The kinetic rate equations (shown in Table 1) are used to calculate the kinetic coefficients of each reaction at various temperatures (T). Since temperature significantly effects the rates at which reactions occur, it is important to use accurate models of the kinetic coefficients. Our kinetic rate equations were derived using the Arrhenius equation and known dynamics of the reactions [21].

Reaction	Kinetic Rate
$TG+M \rightarrow DG+E$	$K_1(T) = 3.9 \times 10^7 e^{\frac{-13000}{2.0T}}$
$DG+E \rightarrow TG+M$	$K_2(T) = 5.6 \times 10^5 e^{\frac{-9900}{2.0T}}$
$DG+M \rightarrow MG+E$	$K_3(T) = 5.9 \times 10^{13} e^{\frac{-20000}{2.0T}}$
$MG+E \rightarrow DG+M$	$K_4(T) = 9.9 \times 10^9 e^{\frac{-14000}{2.0T}}$
$MG+M \rightarrow GL+E$	$K_5(T) = 5400 e^{\frac{-6400}{2.0T}}$
$GL+E \rightarrow MG+M$	$K_6(T) = 22000 e^{\frac{-9600}{2.0T}}$

Table 1. Biodiesel reactions and kinetic rate equations

It is critical to determine whether or not a biodiesel processor will be able to produce high quality biodiesel, which will pass the American Society for Testing and Materials (ASTM) biofuels tests. Studies have shown that only very small quantities of oil ($TG+DG+MG$) in the final biodiesel will allow the resulting fuel to meet ASTM specifications [10]. Challenges exist in designing biodiesel processors because different feed stocks have different concentrations of TGs, DGs, and MGs, and these concentrations have a direct impact on the catalyst used and the amount of M required to make high quality fuel. The ASTM requirements limit the amount of methanol dissolved in the final biodiesel; however, to meet this requirement most biodiesel production systems use post-processing washing techniques which clean the excess methanol from the biodiesel [32].

Our SHS model of biodiesel production has seven discrete modes with transitions and transition guards as shown in Figure 6. The discrete modes include a mode for heating the reacting mixture (q_1), cooling the mixture (q_2), letting the glycerine settle out from the mixture (q_3), testing the mixture (q_4), and states for success (q_6) and failure (q_5) of the reaction (to produce high quality fuel). Transitions between states use guards to define the dynamics that allow the system to transition between discrete states. The guards use Boolean logic to determine if they can fire or not, and some transitions also include a reset, which is executed if the transition fires. Once a guard is satisfied, we assume that it fires immediately. We use continuous dynamics in each discrete state to model a timer, the temperature, and fluctuations in chemical concentrations. The timer is used to trigger some transitions between specific states states. The continuous dynamics of the timer (x_8) are as follows: $dx_8 = 1.0dt$.

To accurately model the reactions, the rate at which the individual reactions fire must be calculated using accurate tem-

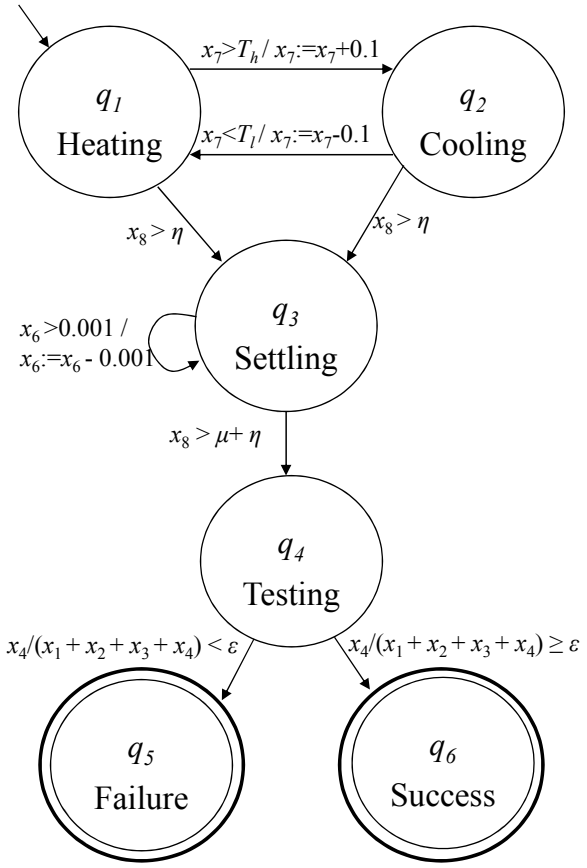

Figure 6. SHS model of biodiesel production system

perature and pressure dynamics. The rate a_i at which chemical reactions occur can be calculated using the stoichiometry defined by the type of reaction. We consider reactions of the form: $V + X \rightarrow Y + Z$ where chemical species V, X, Y, and Z have concentrations v, x, y, and z, and K_i is the kinetic coefficient for reaction i. For these types of reactions the reaction rate is $a_i = K_i vx$. For other types of reactions, the rate can be calculated similarly [15].

Since chemical dynamics are inherently stochastic, SDEs are an ideal modeling paradigm for the chemical concentrations. We use the method presented in [11] to model the concentrations of the chemicals in the system at each mode. The equations used to model the chemical reactions are presented in a previous work [24].

Biodiesel in small scale production systems is often made in a processor using a simple heater and thermostat. The chemical reactions involved are highly sensitive to temperature, so heating the reacting liquid is necessary to ensure quality biodiesel is successfully produced. However, using too much heat wastes time and energy, so processors typically utilize a simple thermostat that controls the temperature. The temperature is modeled with continuous dynamics where the heater switches on when the temperature drops below T_l and turns off when the temperature is above T_h. The resets (R) are used to ensure that the dynamics are sufficiently adjusted to avoid infinite switching behavior. The SDE for modeling the temperature is given by

$$dx_7 = \begin{cases} 0.02(300 - x_7)dt + 0.01dW & \text{for } q_1 \\ 0.01(-x_7)dt + 0.01dW & \text{for all others} \end{cases} \quad (3)$$

As the chemical reactions produce biodiesel, glycerine (GL) is formed as a byproduct of the reaction. Since the presence of glycerine inhibits the successful production of high quality biodiesel, separation of the glycerine from the reacting liquid is necessary. We model the glycerine separation in our biodiesel model with gravity settling using a discrete settling state (q_3). After a certain amount of time η of reacting in states q_1 and q_2, the mixture is allowed to settle. In the settling state, there is a self-transition that repeatedly fires until the glycerine level is below 0.001. This captures the behavior of an operator draining the glycerine from the bottom of the tank until its concentration is below a given threshold.

A simpler version of the biodiesel SHS model without the glycerine settling state was validated for correctness by comparing simulation results generated from a basic model with experimental results collected in an experimental reactor [28]. The model presented in this paper is not validated in its entirety, but it is assembled from components that are validated or follow existing modeling methodologies using the formal SHS framework.

The biodiesel model we have presented can be exhaustively analyzed with brute force simulation methods to achieve verification, but with a more complex model, exhaustive verification analysis becomes computationally too expensive due to the state space explosion problem [20]. Therefore, our approach of using simulations calculated by the 'crowd' to distribute SHS analysis is useful for large models and easily scales to virtually any size of crowd.

5. BIODIESEL SIM SERVER

The server for Biodiesel Sim hosts the companion website to Biodiesel Sim (http://biodiesel.cs.uwp.edu) and the database for storing simulation results. The website was designed to educate users to improve their pattern matching skills and to give users access to the database (and its searching capabilities). Several pages are dedicated to education of how biodiesel is made and how biodiesel can be modeled and simulated in a computer. The website is the 'front end' of the database that stores and tracks all simulation data and coordinates communication between humans and computers. Information about biodiesel production, the biodiesel model, and crowdsourcing on the website is intended to teach users about biodiesel production and encourage users to re-run simulations to improve their 'score.' A screenshot of the main page

of the website that shows the layout and basic information is shown in Figure 7.

Figure 7. Screenshot of the main web page

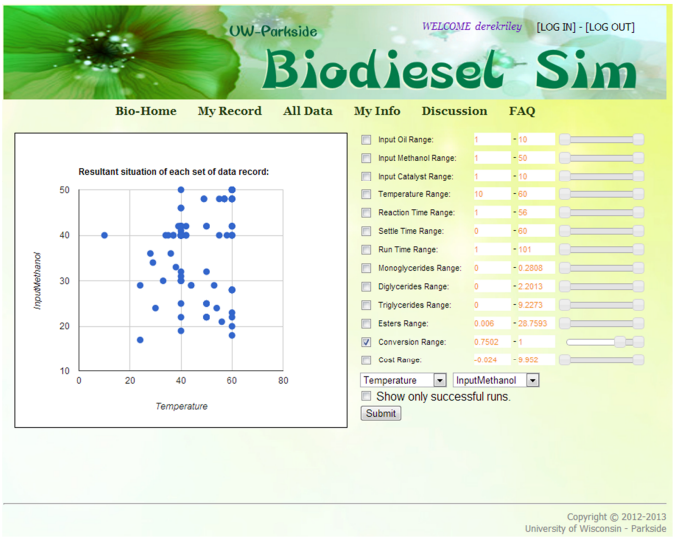

Figure 8. Screenshot of the data chart display in the web page

The website stores and presents all simulation results from every simulation run, so a registered user can explore the 'All Data' section to find the configurations from other users. It is important that the data be easily searchable so users will be encouraged to learn from other people's simulations and be inspired to try new configurations. Filters are provided to make the searching the data easy, and the data is presented in graphical form to aid pattern matching for users.

In the 'All Data' section we have included a customizeable 2D plot of data that can be changed by the user to analyze all simulation runs in the database (as seen in Figure 8). Users can choose the variable to put on the dependent and independent axes of the chart, and they can choose to limit the range of any variable to filter the displayed results (to highlight certain attributes that they might be interested in). In this way users are encouraged to datamine the information to find potential improvements to their simulations. Future versions of the application are in development that will automatically datamine the database to provide initial configuration suggestions for users based on other users' successful runs.

The Biodiesel Sim application allows users to play the simulation as a guest or as a registered user. Results of all simulations from any users are saved in the central database, but only registered users can search and see others' data.

6. RESULTS ANALYSIS

We have analyzed the data in the database to better understand how users are using the application, and it has revealed some interesting results. Users of the application have logged hundreds of simulation trajectories in the months the application has been available. Many of the simulation trajectories result in poor quality or expensive fuel, which does not help identify optimizations. However, many simulations have identified new local optimal solutions.

Most users do not register initially, so their simulation trajectories are stored in the database, but they cannot see other simulations to improve their score. Initial simulations by new users often do not lead to successful results. After users register, they are much more likely to identify local optimal solutions and run more simulations. Many users never register, however, which is an indication that improvements in the application are necessary to attract and retain more users.

It appears that most users initially use the application with the default settings, and many continue to play the game and log new simulation trajectories that ultimately identify local optimizations. Simulation trajectories that are generated after the user has become more familiar with the tool typically lead to more optimal solutions. It is left to future work to develop a mechanism to change the initial configurations for new users to improve their initial scores.

Some of the simulations have revealed new and interesting configurations to create high quality fuel at low costs. We leave it to future work to data mine the results and create a better feedback tool that will coordinate verification by suggesting configurations to users based on previous results.

7. CONCLUSIONS

In this work we present our crowdsourcing approach to simulate and analyze a complex SHS model of a biodiesel batch reactor. The model we use is a formal biochemical model, and our approach can be used to simulate and analyze

any SHS model. We leveraged the popularity of biodiesel to entice users ('crowdworkers') to download and use our application. Our results demonstrate that this type of approach can potentially be used to help find optimizations of complex models.

In the future we will be incorporating more advanced, automated data mining methods to enhance the crowdsourcing as well as other features to make the game more intuitive and more enticing to play. We also hope to release versions of the application for mobile phones to reach more potential users.

Biodiesel Sim can be downloaded via the Apple marketplace by searching for 'Biodiesel' or 'Biodiesel Sim' ('https://itunes.apple.com/us/app/biodiesel-simulation/id648569612?mt=8').

Acknowledgements This work is supported in part by the National Science Foundation (CNS-1238959) and AFOSR (FA9550-13-1-0227).

REFERENCES

[1] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry. Computational approaches to reachability analysis of stochastic hybrid systems. In *Hybrid Systems Computation and Control*, volume LNCS 4416, pages 4–17, 2007.

[2] M. Allen, G. Prateepchaikul, and . The Modelling of the Biodiesel Reaction. http://journeytoforever.org/biofuel_library/macromodx.html, 2012.

[3] Aspen. http://www.aspentech.com/, 2012.

[4] M. Bernadskiy, R. Sharykin, and R. Alur. Structured modeling of concurrent stochastic hybrid systems. *FORMATS LNCS*, 3253:309–324, 2004.

[5] H. Blom, G. Bakker, and J. Krystul. Probabilistic reachability analysis for large scale stochastic hybrid systems. *IEEE Conference on Decision and Control*, pages 3182–3189, 2007.

[6] M. Bujorianu and J. Lygeros. Theoretical foundations of general stochastic hybrid systems: Modeling and optimal control. In *IEEE Conf. on Dec. and Cont.*, 2004.

[7] M. Canakci. The potential for restaurant waste lipids as biodiesel feedstocks. *Bioresource Technology*, 98(1):183–190, 2005.

[8] S. Cooper and et al. Predicting protein structures with a multiplayer online game. *Nature*, 466:756–760, 2010.

[9] A. Doan, R. Ramakrishnan, and A. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4), 2011.

[10] S. Fernando, P. Karra, R. Hernandez, and S. Jha. Effect of incompletely converted soybean oil on biodiesel quality. *Energy*, 32:844–851, 2007.

[11] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.

[12] M. Haas, A. McAloon, W. Yee, and T. Foglia. A process model to estimate biodiesel production costs. *Bioresour. Technol.*, 97(671), 2006.

[13] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visual design. In *Human Factors in Computing Systems Conference (CHI)*, 2010.

[14] J. Hespanha. Stochastic hybrid systems: Application to communication networks. In *Hybrid Systems Computation and Control*, volume LNCS 2993, pages 387–401, 2004.

[15] J. Hespanha and A. Singh. Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems. *Int. J. on Robust Cont., Special Issue on Control at Small Scales*, 15:669–689, 2005.

[16] N. Jaya. Kinetic modeling of transesterification reaction for biodiesel production using hetergeneous catalyst. *IJEST*, 3(4):3463–3466, 2011.

[17] A. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[18] A. Kapil, K. Wilson, A. Lee, and J. Sadhukhan. Kinetic modeling studies of heterogeneously catalyzed biodiesel synthesis reactions. *Ind. Eng. Chem. Res.*, 50(9):4818–4830, 2011.

[19] A. Kosorukoff. Human based genetic algorithm. *Trans on Systems, Man, and Cybernetics*, 5, 2001.

[20] X. Koutsoukos and D. Riley. Computational methods for verification of stochastic hybrid systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 38(2):385–396, 2008.

[21] H. Noureddini and D. Zhu. Kinetics of transesterification of soybean oil. *JAOCS*, 74(11):1457–1463, 1997.

[22] ProSim. http://www.prosim.net/, 2012.

[23] A. Quinn and B. Bederson. Human computation: A survey and taxonomy of a growing field. In *Human Factors in Computing Systems Conference (CHI)*, 2011.

[24] D. Riley and X. Koutsoukos. Probablistic verification of a biodiesel production system using statistical model checking. *Mathematical and Computer Modeling of Dynamical Systems*, TBD(TBD):TBD–TBD, 2013.

[25] D. Riley, X. Koutsoukos, and K. Riley. Safety analysis of sugar cataract development using stochastic hybrid systems. In *Hybrid Systems Computation and Control*, volume LNCS 4416, pages 758–761, 2007.

[26] D. Riley, X. Koutsoukos, and K. Riley. Simulation of stochastic hybrid systems with switching and reflective boundaries. In *Winter Simulation Conference*, pages 804–812, 2008.

[27] D. Riley, X. Koutsoukos, and K. Riley. Modeling and analysis of the sugar cataract development process using stochastic hybrid systems. *IET: Systems Biology*, 3(3):137–154, 2009.

[28] D. Riley, X. Koutsoukos, and K. Riley. Reachability analysis of stochastic hybrid systems: A biodiesel production system. *European Journal of Control, Special Issue on Stochastic Hybrid Systems*, 16(6), 2010.

[29] D. Riley, X. Koutsoukos, and K. Riley. Simulation of stochastic hybrid systems using probabilistic boundary detection and adaptive time stepping. *Simulation Modeling Practice and Theory*, 18(9):1397–1411, 2010.

[30] L. von Ahn. Games with a purpose. *IEEE Computer Magazine*, 39(6), 2006.

[31] A. West, D. Posarac, and N. Ellis. Assessment of four biodiesel production processes using hysys.plant. *Bioresource Technology*, pages 6587–6601, 2007.

[32] Y. Zhang, M. Dube, D. McLean, and M. Kates. Biodiesel production from waste cooking oil: 1 process design and technological assessment. *Bioresource Technology*, 89:1–16, 2003.