Biomass Assessment Tools
Visualizing Trends in Biofuel Production Cycles

Lyndsey Franklin
Department of Computer Science
University of Maryland, College Park
lyndsey@cs.umd.edu

Dung Quang (Bryan) Ta
Department of Computer Science
University of Maryland, College Park
bryanta@cs.umd.edu

Michael Szaszy
Department of Computer Science
University of Maryland, College Park
mszaszy@cs.umd.edu

Scott Abromowitz
College of Information Studies
University of Maryland, College Park
abromo@umd.edu

ABSTRACT
Microalgae are receiving increased global attention as a potential sustainable energy crop for biofuel production. Several strains of microalgae have shown great potential to provide adequate biomass to be converted into oil alternatives for mass consumption. Experts working in biofuels production have compiled many simulations and models that predict the yield of many strains of microalgae under a number of environmental conditions while others have made contributions towards understanding the economics of harvesting, refining, and releasing output to markets. Individual models have related and interconnected inputs and outputs such that a simple, single visualization is incapable of supporting their analysis. A lack of support tools has placed researchers at a disadvantage in this respect. We present a prototype of our Biomass Assessment Tools website to fulfill the need for sophisticated information visualization support in the domain of biofuels research. Through expert review (n = 4), we demonstrate our tool’s potential to allow new trend finding analysis to occur over data generated by predictive models.

Keywords
Information Visualization, Geographic Information Systems
INTRODUCTION

Rising fuel costs and environmental concerns have put increasing pressures on scientists and policy makers to explore alternative sources of fuel to prepare for the demands of the future. Often, the focus of alternative energy sources lies in a small number of popular potential methods including solar cells, wind farms, and hydropower. Each potential alternative comes with its own benefits and complications. However, each alternative is typically thought of as providing alternative power for electric consumption. These alternatives are not typically considered as substitute fuel sources for other power consumers such as automobiles, planes, and other forms of transportation that are not run on electricity. Additional options, such as biomass, may prove more feasible as alternative sources of fuel for oil and fossil-fuel dependent systems [6].

Microalgae are receiving increased global attention as a potential sustainable energy crop for biofuel production. Several strains of microalgae have shown great potential as providing adequate biomass to be converted into oil alternatives for large-scale consumption. One study has found that microalgae crops have the potential to produce fuel in an amount equivalent to 48% of US petroleum imports, but at the cost of 5.5% of the land area in the United States and nearly three times the water currently used for irrigated agriculture [14]. Important steps to realizing the potential of these strains of algae at a full “energy scale” include: 1) Quantifying the demands and trade-offs that large-scale algal biofuel production will place on available resources; and 2) Analyzing whether or not the economics of this new industry can be positively realized. The economics at the site-scale and at the enterprise scale, where economies of scale may be attained, must be considered. Impacted resources are similar to other forms of alternative energy but also include some additional concerns more specific to the growth of biomass including water, land, nutrients/fertilizers, environment, and infrastructure. Biofuel production at a large scale could alleviate the reliance on traditional oil resources, but at the cost of increasing reliance on agricultural resources and competition with other agricultural endeavors. Determining scenarios where biofuel production provides an economically feasible alternative to fuel demands is important for encouraging its adoption as well as its continued research [9].

Data analysis is critical for the success of bringing fundamental research results to the research community and private industry. Experts working in biofuels production have compiled many simulations and models predicting the yield of many strains of microalgae under a number of environmental conditions while others have made contributions towards understanding the economics of harvesting, refining, and releasing output to markets. Individual models have related and interconnecting inputs and outputs in which a single, simple visualization is not capable of supporting the analysis. A lack of support tools has placed researchers at a disadvantage in this respect. Analysis of models is often conducted in isolation and the inclusion of more than a single source of model data requires a great deal of manual effort. One from Geographic Information Systems (GIS) literature is that of Fernandes and Costa (2010) where biomass residues are found to be economical alternative energy sources after extensive manual statistical analysis. Entire teams of researchers are sometimes required to integrate the findings of related models resulting in time wasted in the coordination of data, rather than productively spending the time in data analysis. This is despite the fact that model outputs are often related to common site locations represented by similar GIS data and can be associated with moderate effort. Technology and techniques to explore interactions in multivariate data are available but underutilized in the biofuels domain. Ideally, interconnections between models could be visualized through the use of interactive multiple coordinated views; nonetheless, current practice involves manually retrieving data from multiple sources and examining results in tabular format. Advanced tools capable of pulling information from multiple model sources typically represent data from only one source at a time on a single, often static, map-based visualization which impedes on multi-model analysis. This places a high demand on short-term memory because there is neither interaction with data once it has been retrieved nor a process for overview, zoom and filter, or details on demand.

To address trade-off analysis within the domain of biofuel production, we introduce the Biomass Assessment Tools. Our prototype website simplifies the process of exploring biofuel production cycles in the context of economics of land values by providing a single tool that reflects the constraints of user determined scenarios. With data from multiple model sources, the prototype represents an improvement over the current practice of analyzing model outputs in isolation. Multiple, configurable visualizations improve on current practice by providing greater flexibility for representing multivariate data. The coordination between these multiple visualizations allows for greater depth of analysis by assisting users in following data across
the available representations. Despite its analytic power, the Biomass Assessment Tools website allows users to understand their data without burdening them with the requirement of learning a general purpose data analysis tool. Several widely used visualizations have been included in its design and improved upon to increase their value in answering analytic questions of model output data.

The rest of this paper is organized as follows. The next section reviews several related works which are typical examples of GIS based efforts to support analysis of multivariate biodiversity or biofuel related data or policy analysis. These related works are generally limited in their ability to support a great number of scenarios and have been meticulously designed to answer a single analytic question. The following section introduces the multivariate data expected by the Biomass Assessment Tools. After this we introduce the Biomass Assessment Tools by discussing our client-driven design requirements and our resulting website prototype in detail. An evaluation by domain experts actively working in the domain of biofuels production and policy follows this and we conclude with a discussion of future directions.

RELATED WORKS

Much research is available in the field of GIS data and biomass analytics. Fernandes and Costa (2010) approach the analysis of biomass residues for energy production using statistical analysis and GIS databases and is representative of current practice with respect to the process of biomass analytics. This work focuses on a unique region of Portugal where small-scale biomass plants have been proposed as options for heating purposes and inputs to the analytic process include maps of forest and agricultural residues, administrative boundaries, rivers and roads, and statistical data for cultivated areas and types of cultivation. Conclusions from their analysis indicate that biomass plants have large potential for providing economic and environmental advantages over fossil fuel-based heating systems. Despite the fact that research of this kind demonstrates the importance of continued analysis of alternative energy sources for encouraging their adoption, the research process is under-supported and dependent on extensive manual effort.

Bottero et al. (2013) describe an analytic system with GIS capabilities to support decision makers in understanding human pressures on biodiversity. In particular, they take advantage of maps as an easily understandable output of various models and produce sustainability maps and identify regions in need of intervention based on factors such as natural protected areas, ecosystem quality, ecological value, land cover, water bodies, and human pressure. The work presented here follows in this line of research but extends the available output of such a model based system with multiple interactive visualizations.

The need for policy decisions to satisfy multiple criteria is not limited to the biomass or biologic systems fields of research. In a similar process to [2], Foster and McDonald (2000) combine raster and vector based data with a GIS framework to produce maps indicating areas of potential hazard to water quality. They utilize predictive models to quantify risk frequency and impact. More unique to their work, however, is the use of a medium specification PC and their intention that their analysis process be usable across a larger intranet of utility companies. We also expose models to a larger audience of users who are not necessarily domain experts that have had direct input to model development though we extend the work by not limiting our analytic capabilities to GIS based maps alone.

Policy decisions and their effects on water resources are the domain of mDSS [5] which was developed for assessing nitrogen pressure from agriculture on water resources at the European level. A statistical model is used to calculate a simplified nitrogen balance based on sources including organic and mineral fertilizers. However, while the mDSS provides multiple views into a specific scenario, it requires extensive domain knowledge to configure scenario parameters making it an impractical tool for a wide variety of users who could derive utility from analyzing a given agricultural scenario. And, while users are ultimately able to complete a trade-off analysis between a set of predetermined options, there is not a simple way to interactively adjust parameters and observe their effects. Despite its intent as a decision support system, Policy makers would need pre-generated reports from subject experts in order to benefit from the system. Our work differs from this highly detailed process by providing a system where both researchers and non-domain experts may benefit from data exploration.

Uganda and Thailand are the settings of two case studies in providing decision makers with GIS based systems for multiple objective decision-making for two very different purposes: disease control and livestock production development [8]. Both cases center on competing economic, health, and structural demands with very real impact to the regions involved. Their results include increased stakeholder
involvement but also caution future endeavors on the importance of having high quality data and the complications of working with uncertainty with respect to the decision rules. The work presented here has similar encouraging results with regards to stakeholder involvement and eagerness to adopt decision supporting analytic tools. However, as revealed through our work with expert reviewers discussed in later sections, the need for high quality data will always be a concern to users. Our current work also differs in that our analytic process is iterative and interactive as opposed to the static analysis of current conditions performed in [8].

Other risk management attempts have focused on combining geologic models with artificial intelligence techniques to develop spatial decision support systems for allocating resources during earthquakes or landslides [13]. Soil distribution and elevation maps are combined with vegetation conditions to classify fragile areas as likely landslide locations. GIS modules are used for combining maps and providing distance and slope calculations. Our prototype extends this combination of GIS and algorithmic site classification by supplying map based visualizations of sites of interest but also providing a mechanism to directly explore the classification of sites into clusters with our Site Cluster visualization.

Sustainability planning in industrial settings has also been addressed by multi-criteria decision support systems. Ruiz et al. (2012) analyze influential location factors and propose a multi-criteria evaluation model intended to support decisions about zoning policy using a semi-structured decision process. A GIS system and compatible tools are utilized to produce maps differentiating zones according to sustainability criteria. While the production of maps to present results has been a dominant theme of many decision support systems, we have chosen to view maps as one of many viable options for data communication and exploration. In particular, our map visualization makes use of multiple synchronized maps to provide trending information otherwise un-supported by standard report maps. Ruiz et al. also employ a zone evaluation method which involves ranking model variables in accordance to their impact on the application objectives. While we do not currently score biofuel sites in our prototype, we do make use of several ranking options based on parameter values in representing sites in our Top Ranked Sites visualization.

Finally, we draw some inspiration from efforts such as [10] which demonstrate the possibility of online environments for coordination of multiple data sources for the purpose of analysis. Sridhar et al. (2011) also present information about subsurface water resources through HTML/JavaScript based visualizations. A GIS system is used for integrating information across data sources.

**DATA DESCRIPTION**

The Biomass Assessment Tools website is designed to allow users to explore multivariate data pertaining to the production of oil from freshwater algal sites across the United States. For this work, the term ‘site’ refers to an individual location which may be chosen as a candidate for beginning biomass energy production. Currently, spatial and numerical models capture resource requirements, production potential, and sustainability metrics for bioenergy feedstock. Climate, water, land, and infrastructure data, along with environmental constraints, biomass growth rates, and resource requirements all interact to produce a model result that needs to be analyzed for feasibility in the context of regional goals for environment and economy.

To accomplish this analysis, model outputs are collected on a per site basis resulting in a dataset where sites are represented by aggregate data of varying granularity. Algal output, oil production, and water use per site are aggregated on monthly, seasonal, and annual totals providing a view into the production cycle of each site. In addition to these aggregated data, sites are also linked with data regarding their location, the cost of purchasing the land for the site, the area associated with the site, and the cost of leveling and readying the site for biofuel production. These data address the economic factors that complicate the adoption of widespread biofuel production.

A single run of these data generating models for a single strain of freshwater algae produces approximately 11,000 sites from regions across the country. Aggregates represent the average values for the year, season, or month based on the previous 30 years of data. In future iterations of our prototype, data per site will be expanded to include the energy production per site, the energy cost of maintaining each site, and the energy cost to bring produced oil to markets. This expanded dataset will allow further investigation into the tradeoffs between economic realities and production potentials.

**DESIGN AND IMPLEMENTATION**
For ease of integration with existing models and data sources, our prototype is built as a HTML5, CSS3, SVG, and JavaScript based web-portal. This foundation allows rapid integration with current customer services and enables easier prototype transfer at the end of the project. To leverage some pre-developed visualization packages, we utilized the D3.js (Data-Driven Document) library [1]. The Twitter Bootstrap front-end development framework and jQuery UI (for sliders) provide the foundation for the front-end interface to interact with the data filters. Interactive filtering was accomplished through use of the Crossfilter JavaScript library [4]. For the sake of rapid development, data was converted from CSV data hosted by a remote web service to raw GeoJSON that is downloaded by the client browser immediately upon entering the website. In a production ready system these files would be replaced by direct connection to the web services to enable on-demand runs of models analysis of specific scenarios. The current prototype has been developed using Firefox and is not guaranteed to behave correctly in other browsers.

Working with our clients, the following features were established as minimum requirements for the system:

- **Web-based environment**: To speed up adoption by the user base, the project should produce a web portal so that minimal software installation is required to support analysis
- **Site Clustering Capabilities**: User needs the ability to cluster pond sites based on
  - Water Demand
  - Oil Production
  - Land Value
  - Water Source Availability
- **Drill-Down Capabilities** for assessing trends: Users must be able to see nationwide trends in their data as well as pond-site specific details
- **Site by Site Comparisons**: Users must be able to compare the yearly oil production between multiple selected pond sites
- **Seasonal Aggregates**: Users must be able to view seasonal trends for pond sites
- **Data Filtering**: Users must be able to reduce the size of their dataset to sites matching predefined criteria

Elaborating from these requirements, the following usage scenarios were drafted to guide our design process:

- **Policy Maker**: Several state-level policy makers are in the process of determining which sites in their state would make good candidates for biofuel production. They have a fixed budget that is restricted at the state and federal level, which determines how much they are able to spend per site. They must also limit the amount of freshwater required at each site to avoid conflicting with competing needs for water such as the state’s agricultural irrigation system. They are interested in maximizing the amount of oil produced to demonstrate the efficacy of the state’s biofuels program.
- **Researcher**: A research scientist is searching for candidate sites nationwide to market a new production technique. The technique is most applicable to freshwater sites with a certain expected production level because of restricted water availability. The scientist needs to identify several sites across the nation with similar expected yearly production levels that have similar water requirements to set up experimental studies to test their new technique.
- **Industrial Planning Committee**: A planning committee is looking to market their local area as suitable for biofuels production. The committee is competing with other locations across the nation for federal development funds. They are interested in putting together a report that demonstrates not only their low development costs and consistently high production potential, but also their greater than expected total oil output compared against other key competitors to strengthen the local area’s argument.

**VISUALIZATIONS**

With such a large dataset, assessing trends is a complex analysis process that requires a variety of views into the data. The multiple aggregates within the data represent the same site data in a variety of time periods to allow the data to answer questions about production trends across multiple timescales. Visualizations were chosen to allow site data to be viewed at each of these timescales: monthly, seasonal, and yearly. Coarse filtering of data at the yearly level allows for rapid reduction of dataset size to sites that meet users’ ultimate per-site requirements while seasonal and monthly based visualizations allow users to understand how yearly production levels are reached over the course of the year.

Figure 1 shows the Biomass Assessment Tools website with a small dataset of around 430 potential biofuel production sites. Several of the sites have been selected within this dataset to show how selection is highlighted and coordinated across the
Multiple visualizations. Each visualization can be hidden and reshown by clicking its name in the title bar. This allows users to compare results of otherwise distant visualizations as needed. The available filters are located in a control region along the left side of the website and are applied by manipulating range sliders, selecting from a dropdown list, or checking and unchecking items from a list. In the following sections we discuss the features of each visualization in detail.

Maps

The Map visualization provides a series of individually configurable maps to allow users to see trends across the dataset geographically. Each site is represented as a circle and may be color encoded to represent any property of the data. By default, there are four maps per row that allow users to take advantage of small multiples and dedicate one map to the data of each of the four seasons (winter, spring, summer, and fall). The number of rows and the number of maps per row is configurable though this is not currently available at the front-end UI. Legends describe the data domain and are available...
for each map. Tooltips indicate the ID number of each site. Zooming and panning are enabled to allow users to explore any region in detail and the maps are synchronized to display the same regions (see Figure 3).

Selection is possible by clicking on the “S” button in the control bar. This allows users to indicate a region of interest and select the points in that region. Selected sites are drawn as solid bronze circles with a solid black outline. Both incoming and outgoing selection is represented with the same encoding (see figure 4).

**Site Clusters**

The Site Cluster visualization is based on the D3 icicle-partition visualization [1] and provides a hierarchical clustering [3] of sites in the dataset as determined by the seasonal data associated with each site. Once more, four clustering options are currently available: on seasonal oil production, on seasonal algal production, on seasonal water use, and on land values. Each site and cluster in the hierarchical clustering algorithm is represented as a 4D vector (1 dimension for each season) with values populated based on the user selected clustering option. The Euclidian distance between these vectors is used to determine similarity. When sites or clusters are merged to create new clusters, then the new cluster vector is defined as the centroid of all its children. The hierarchical method uses a greedy, top-down
approach that successively pairs the two closest items together as a cluster in the next level of the hierarchy until the root level (e.g., level containing 1 node) is reached. Cluster labels are determined by the first site of the cluster to provide consistency between cluster-processing. Clusters of size one representing an individual site are labeled with the site ID number. Higher-level clusters are labeled with the size of the cluster and if space is available the location of the first site of the cluster. Active clustering occurs as sites are filtered from the dataset through usage of the filtering controls as well as when users select a different property of the data to cluster on. A click-to-zoom navigation allows users to drill into each cluster to view sub-clusters and the individual sites represented by each cluster. Mouseover tooltips provide the detailed seasonal averages of each cluster on demand.

Selection is possible from the Site Clusters visualization by clicking on the “S” button in the visualization’s tool bar. In selection mode, zooming is disabled and a mouse click adds or removes entire clusters from selection. To alert users to the active selection, the “S” button becomes bronze while selection mode is enabled and zooming is disabled. Selections are represented with bronze borders around the rectangle indicating the cluster as shown in Figure 5.

Figure 5 Selected clusters are outlined in bronze. Clusters are labeled with the location of a site represented by the cluster and the cluster’s size, or the cluster/s size only if space is not available.

Figure 6 Selections from other visualizations are traced to their representative clusters. A cluster which has all its represented sites selected is itself selected and outlined in bronze.

Figure 7 Monthly averages for water use, algal output, and oil output for all sites in the dataset are represented with orange, blue, and green lines. When selections have been made in other visualizations, the monthly averages for water, algal, and oil are all represented as additional bronze lines.
Incoming selections from other visualizations are also represented with bronze borders but include all clusters in the hierarchy from the leaf clusters of the individual sites to the largest cluster with all its children in the selection as pictured in Figure 6.

**Monthly Averages**

The monthly averages visualization represents the oil production, algal output, and water use of all the sites in the filtered dataset throughout the year. Each property is depicted as a single line in the graph and all sites are represented in each line. Values are log-scaled to accommodate the greatly differing domains of oil, algae, and water and allow the trends of each to be visible on a single chart. Vertical bars are extended from March, June, and September to visually delimit seasonal boundaries for rapid comparison. Mouseover tooltips provide details of maximum and minimum values for each line on demand. Averages for each line are recomputed as filtering narrows the sites of the dataset.

Incoming selection is represented as an additional set of three lines that show the oil production, algal output, and water use of the currently selected set of sites. These additional lines are displayed along with the baseline average of the filtered dataset and are indicated by their bronze color and labeling (see Figure 7). Mouseover tooltips also provide details of maximum and minimum values on demand for the lines of the selected sites.

Because the representation of each site is aggregated with every other site of the filtered dataset, the ability to distinguish a single site from the other sites is not possible. As such, selection is unattainable within the monthly averages visualization.

**Top Ranked Sites List**

The top ranked sites list shows the individual sites of the filtered dataset in detail and includes the annual oil production, algal output, water use, area, purchase cost, and leveling cost of each site. Sites are ranked according to one of these data properties and the exact property is selectable by the user. For convenience, the list ranks oil production, algal output, and area in descending order and water use, purchase cost, and leveling cost in ascending. The size of the visualization is also configurable and users may choose to view up to 30, 50, 100, or all of the filtered dataset sites in the list.

Selection is available as a toggle by clicking on individual sites. Unselected sites become selected sites with a single mouse click and selected sites become unselected with a single mouse click. When selected, the backgrounds of individual sites become bronze and lettering converts to white to distinguish them from unselected sites (see Figure 8).

**EVALUATION**

To improve the usability of our prototype, four experts currently working in the domain of biofuels research and policy in three phases evaluated our prototype. After the first phase, we implemented several of the requested features that were suggested during the review and returned to three additional experts for additional reviews. One expert provided guidance early in our design process, but did not work with us again until a functional prototype was built. The other experts worked with our functional prototype only. The focus of the reviews was to elicit qualitative feedback about additional features to include in later iterations of our prototype.
evaluate the effectiveness of the prototype’s current iteration.

**Review Phase 1**

Our first review session was scheduled as a two hour meeting with our project team and our expert reviewer. We began with a brief description of the site and its primary features and then followed with three example usability tasks. Each task consisted of a usage scenario and four subtasks to accomplish the goal of the scenario (see Appendix A). Our site was shown on a laptop connected to a projector so that all meeting attendees could easily see the actions taken to complete each scenario and subtask. Our expert followed a think aloud protocol during task completion and each scenario was paused multiple times for detailed feedback about a recently completed step from our expert and a follow-up discussion.

Overall our expert responded positively and enthusiastically to our initial prototype. At the end of the third scenario, our expert examined the results and said “Look at this- this is cool! Look what you found!” Several discussions centered on task completion resulted in the request for additional features to take the supported analysis to a deeper level. This session concluded with the following list of features to be implemented immediately in the prototype:

- Selection within filtered data to understand where individual sites were represented across visualizations
- Relocation of the controls to show/hide visualizations to the top navigation bar so that the controls for individual visualizations could be hidden when the visualization was hidden.
- Legends to indicate the color scaling of the map visualization
- Vertical bar additions to the line visualization to delimit where seasons begin during the year.

Following this first phase of expert review, the project team implemented the requested features, which have been discussed in detail in previous sections. Several additional features were requested during the review but were determined to be less important; as such, the functions were not immediately implemented in our prototype:

- Site Locations added to available filters to allow finer-grained location control
- Convert current State filtering to a check-list so that multiple states can be included as a filter at a time
- Expanded data capabilities: include additional model sources for more comprehensive analysis

The inclusion of additional model data is discussed in the Results and Discussion section. Ultimately, the inclusion of site locations and the conversion of the state filter to a list were abandoned due to a limitation of the CrossFilter JavaScript library. In its current implementation (version 1.1 at the time of this writing), CrossFilter does not support multiple category filtering on categorical data. Due to time constraints, our project team could not extend the CrossFilter framework to support more complex actions on the state and location attributes due to their categorical nature. Conversion to a numeric scale proved to be a poor alternative because CrossFilter interpreted the numeric encodings of states and locations to be continuous numbers for which min and max operations were applicable.

**Review Phase 2**

Our second review was held as a one hour meeting with a member of our project team and an expert working in the field of Environmental applications who was not affiliated with our client organization. We again began with a description of the site and its primary features as well as an additional briefing on the nature of the dataset and the particular problem space. We then proceeded to complete the same usage scenarios and tasks that we used in our first review phase. A laptop was used to display our website which the expert manipulated. Again, our expert followed a think aloud protocol and scenario completion was paused for discussion of recently completed tasks.

Our second expert regarded our prototype as an “encouraging start.” He immediately observed that the level of granularity of data was not yet adequately complete to support using our prototype to answer specific scientific questions. He felt that domain experts would probably desire to see more information complexity presented and more clear indications of the exact underlying data values than were currently available. He suggested that we
extend our filtering system to allow fine-grained control over specific parameters after coarse-grained filtering was applied to limit the dataset. A second suggestion was to make use of the map visualization’s background to display more information. Domain relevant clustering algorithms were also suggested as a means to extend the current cluster visualization.

Our expert focused primarily on the map and cluster visualizations and did not make much use of the monthly averages or top ranked sites until prompted. Because of their more limited interactions, he felt “there was less do to there” and that “[they were] for seeing what was left after the rest was done” rather than using them as part of his iterative investigation. He did, however, indicate some relevant literature from the domain of scientific visualization as sources of refinement for the monthly averages graph with respect to alternative Y-axis scales.

At the conclusion of our review, our second expert was eager to share the prototype with other experts within his organization. We agreed upon visiting again for a full demo in the near future as a follow-up task.

**Review Phase 3**

Our third expert review was conducted as a one hour phone meeting with remote experts. Our experts were asked to watch a 10 minute video demonstration of our prototype and then use our working prototype hosted on a server internal to their workplace network. Early feedback was provided by email and our experts were asked to complete the same usage scenarios and tasks as our first and second reviewers.

Our experts of phase 3 described our overall prototype as “phenomenal” and were eager to discuss ways in which to extend the work further.

Our remote experts were able to complete all three usage scenarios after watching the video. One expert did not immediately understand the “TOP” or “S” control buttons of the Site Cluster visualization but also admitted did not watch our demo video in its entirety. We still interpret this feedback as indication that our controls are not yet completely intuitive to users without early guidance and will work on improving this in our next prototype. Our experts also requested the ability to directly input values for each of the filters as a short cut. The sliding, range filters were useful for reducing the data quickly for exploration purposes. However, they felt that given a set of specific tasks such as those in our usage scenarios with necessarily explicit values to filter on direct input would have led to faster results.

From this feedback we can interpret that our goals of supporting exploration are well met, however additional features and controls will need to be added for future users who are familiar with the website and have become more expert in its use.

Our use case scenarios were considered very realistic by our experts who used them as starting points of discussion on including additional model data in our prototype. Scenario 1 prompted them to discuss providing data to compute a cost-per-gallon metric reflecting the water and energy inputs needed to
produce oil per site. Scenario 2 needed more refinement. After some debate between experts, it was decided that the scenario should be altered in the future to have users look for high algal mass producing sites with high water use and that the goal of the scenario should be to reduce water demand. Our experts strongly approved of the temporal aspect of Scenario 3 indicating that “spotty output is a real problem, so is getting all your water at once” and that we “hit the nail on the head with this one”.

In reporting the results of their usage scenarios, our experts provided screen captures with explanations of what each screen shot meant to their analysis (see Appendix C). From these screen captures we could see that our experts use the Maps and Site Cluster visualizations along with the dynamic filters to perform the bulk of their analysis and confirmed what they had found using the Monthly Averages and Top Ranked Sites (see Figure 9). Our experts were also able to answer the questions of usage scenario #3 based on the displays of the Site Cluster visualization. They did, however, indicate that better labeling of the clusters might be appropriate as location is not included in the cluster vectors at this time. Instead, clusters could be labeled simply with their size as a starting point.

RESULTS AND DISCUSSION

The most frequent request from our expert reviewers and from our internal users was to extend the complexity of visualized data by including the output of additional models. In particular, models providing data on annual operating costs and capital costs were suggested as the next logical inclusions. As one expert explained via email, “we don’t [want to] build biomass plants that require more energy than they produce (let’s call it the Toyota Prius Effect 😊).”

Desired operating costs would include model output, which provides a detailed per site breakdown of energy requirements at various stages of production such as mixing, carbonating, centrifuge, and pumping. Model output would provide additional filtering options as well as trend data for net power consumed and the net cost for power to provide perspective on the generated energy per site and its relative levels compared with energy consumed during production. Capital cost data would address additional economic factors impacting the start-up of sites. Additional considerations would include the costs of production systems and equipment, nutrient costs, and total capital costs per gallon of output. This would provide data to analyze trends in rates of payoff per site to estimate which sites are more profitable and recover start-up costs faster.

Other suggested models include those producing data on water delivering costs. Data from this model could complement the current water-use data represented in our website by addressing the comparative costs and energy required to divert water of different types (ground vs. fresh vs. saltwater) to sites. A similar model would provide data for nutrient supply costs per site to indicate sites that may be more natively fertile and distinguish such sites from others that have prohibitive nutrient needs.

Along with the expansion of data and model sources, additional visualizations were suggested as mechanisms for supporting specific types of analysis. One extension suggested is to include calculated daily means for oil, algae, and water data in our visualizations and to aggregate these daily means over months and seasons. These daily means would provide another specific baseline that individual sites could be compared against. A suggested improvement to the top ranked sites list was to develop a singular scoring metric to account for oil production, water use, and land costs in a weighted measure to provide an additional summary ranking. This compound ranking metric may prove even more useful as the extension of data sources and inclusion of additional models makes ranking by single data properties less effective and may take the form of a cost-per-gallon metric.

CONCLUSION AND FUTURE WORK

The Biomass Assessment Tools website has been demonstrated to be a successful application of information visualization techniques to a domain burdened by data and without support for deep cross-model analysis. By associating individual sites with production values (oil output, algal output, and water use) and economic data (purchasing costs, leveling costs), we provide a central location to understand biofuel production trends and the economic factors that limit its widespread adoption. Multiple coordinated views supply a mechanism of overview, zoom and filter, and details on demand to support analysis in ways novel to the domain of biofuels production.

Future work will include the expansion of represented data to drive more complex analysis of biofuel production trends and encourage their greater adoption as an alternative source of fuel for oil-based energy consumers.
ACKNOWLEDGEMENT
We would like to thank Professor Ben Shneiderman of the University of Maryland for his mentorship and Tim Seiple and Andre Coleman, our client advisors, of the Pacific Northwest National Laboratory for their assistance and support that they provided us throughout our project. We also thank Dr. Joseph Jaja and Nino Zuljevic for their expertise and encouraging feedback.

CREDITS
Lyndsey Franklin
Team Manager
Programming (Clusters, Monthly Averages, Top Ranked Sites, Filtering, Main Index)
Report
Video (Script, Screen Recording)
User Scenarios
PowerPoint Presentation (Editing)

Michael Szaszy
Data Preparation
Programming (UI Design, Maps, Main Index, Filtering)
Report (Editing)
PowerPoint Presentation (Editing)

Dung Quang (Bryan) Ta
Programming (UI Design, Monthly Averages, Filtering)
Report (Editing)
PowerPoint Presentation (Editing)

Scott Abromowitz
CSS (Bootstraps)
Report (Original Outline, Editing)
Video (Voice, Editing)
User Survey
Git Repository Hosting
PowerPoint Presentation

REFERENCES


[3] Cluster-fck
http://harthur.github.com/clusterfck


APPENDIX A

Scenario 1: Informing Policy
Assume you are a state-level policy maker in the process of determining which freshwater sites in your state would make good candidates for biofuel production. You have a fixed budget that is restricted at the state and federal level that determines how much money you can spend per site. In addition, the amount of freshwater you can use per site is limited to avoid conflicting with your state’s agricultural irrigation system.

Your goal is to maximize the potential amount of oil produced to demonstrate your state’s efficacy for biofuel production. To meet this goal, perform the following tasks:

1. Restrict the visualized sites to those in the state of Louisiana.
2. Limit the purchase cost of sites to $300,000 or less.
3. Limit the annual water use of sites to 1,400,000 L/Ha or less.
4. List the top 5 oil producing sites.

Scenario 2: Research Application
Assume you are a research scientist looking for freshwater sites nationwide to market a new oil production technique. The technique is designed to increase oil production in freshwater sites with low algal output and high water usage to make a site more productive.

To meet this goal, perform the following tasks:

1. Restrict the annual algal output to between 30,000 kg/Ha or fewer.
2. Restrict the annual water use to 10,000,000 L/Ha or greater.
3. List the top 10 water-using sites and the states where these sites are located.
4. List the top 10 algal output sites and the states where these sites are located.

Scenario 3: Stable Algal Production
Assume you are a researcher looking for a high-level understanding of algal production and water use through the entire year across the country.

Your goal is to understand the general trend of biofuel output, either increasing or decreasing, and water usage, either increasing or decreasing, through the year for the United States. To meet this goal, perform the following tasks:

1. Set up the Maps visualization to show seasonal biofuel totals and seasonal water use totals.
2. Identify the region of the country with the highest biofuels production year-round.
3. Identify the region of the country with the highest water use year-round.
4. Identify four groups of regions with similar algal output and name one site that is typical of each group.

APPENDIX B

User Survey
Users who made use of the website on the client’s internal network were asked to complete a short survey to gain additional feedback from these early experiences. Eight questions appeared on the survey:

Biomass Assessment Tools Survey
Please respond to this short 8 question survey. Your responses will help to improve our application greatly. Thank you! Remember all responses are optional!

- How would you rate the application's ability to filter specific criteria? Can you find sites that fit specific criteria? (location, cost, area, etc.) Scale 1-10, 1 = Difficult, 10 = Easy
- Font size and legibility? Issues with visibility? Scale 1-10, 1 = Barely Legible, 10 = Very Legible
- Which feature did you find most useful and why? The ability to zoom? Select water sites?
- Which features did you find the most difficult to use and why?
- Are there any features that you wish to see added to improve the program (Design, accessibility, readability, etc.)? If so what? Icons? Colors?
- Would you use a tool such as this in your daily work routine? Yes or no and why?
- Did you find any bugs/flaws in the program? If so what? Browser issues? Rendering issues? Too slow to use the program? Visualizations not appearing?
APPENDIX C

Expert Screen Captures
For our third phase of evaluations, our experts were given access to our working prototype and asked to complete the tasks of our usage scenarios in Appendix A. As part of his feedback process, one of our experts provided screen captures for his results and explained what each screen capture answered.

Figure 10 Screen Capture 1 indicates the experts answer to Scenario 1. The top 5 oil producing sites in the state of Louisiana that can be purchased for $300,000 or less and use less than 1,400,000 L/Ha water annually are in Houma and Lake Charles LA.

Figure 11 Screen Capture 2 shows the top ten water using sites requiring more than 10 million L/Ha water annually while producing less than 30,000 kg/Ha algal output annually. These are all found in Washington state and answer the questions in Scenario 2.
Figure 12 Screen Capture 3 shows the unfiltered dataset clustered on seasonal water use. Expert 3 indicated that the sites in the state of Florida were the most stable regions based on map coloring and clustering results. Vero Beach Florida is the location of several sites with high year-round biofuels production while Lavon Dam Texas is a site in the cluster with high water use year-round.