**A data visualization tool for biomass valorization in Brazil**

Roger Sampaio Bif,a Larissa Thais Bruschi,a Moisés Teles dos Santos,a\*

*aDepartment of Chemical Engineering, Polytechnic School, University of São Paulo, Av. Prof. Luciano Gualberto, Trav. 3, 380, São Paulo, 05508-900, Brazil*

*moises.teles@usp.br*

**Abstract**

Brazil's diverse climate zones, vast territory, varied soils, and abundant water resources contribute to its rich biodiversity, making it ideal for biomass exploitation. This work deals with the creation of a user-friendly, interactive data visualization tool for biomass valorization. This involves collecting georeferenced data on biomass production and availability and developing an ontology to represent the Brazilian biomass supply chain. The Python programming language with libraries like Pandas, Matplotlib, and Streamlit, are used for data analysis, mapping, and creating web-based visualization applications. The proposed tool provides dynamic maps, showing biomass distribution and availability, highlighting regional resource disparities, assisting in region-specific planning and decision-making. This tool will be further integrated into optimization frameworks, helping to identify opportunities and to develop effective strategies for biomass valorization in Brazil, supporting the decision-making process.

**Keywords**: data visualization, biomass valorization, ontology.

* 1. **Introduction**

Brazil presents natural competitive advantages for biomass exploitation, such as the presence of various climatic zones, vast territory, a wide range of soils, and abundant water resources, reflecting in its biodiversity. These aspects demonstrate the Brazilian potential for the development of economic activities based on renewable resources. It also highlights the perspective to generate employment, income, and technological advancements in decentralized regions, where biomass can be harvested or cultivated. In the last ten years, Brazil has increased its participation in the global food market, rising from US$ 20.6 billion to US$ 100 billion (CEPEA/ESALC, 2020). Projections indicate that by 2030, approximately one-third of globally traded agricultural products will originate from Brazil (Forster-Carneiro et al., 2013). To produce biofuels, the biomass production, harvesting, and preprocessing can represent 44 % of the selling prices (Roni et al., 2019), presenting the dependence of the biorefinery success in the efficient supply chain management for biomass valorization projects. In this context, the valorization of the agricultural waste through science, technology, and innovation can contribute to the minimization of environmental impacts, promotion of the circular economy, enhancement of the economic performance of the sectors involved, and social development in associated rural communities.

The decision-making process on the entire supply chain requires a laborious data gathering procedure on collecting, transforming, and evaluating several data sources from different domains, such as engineering, accounting, agriculture, economics, and others. These complex datasets also need to be converted to standards that can be used by an optimization model, which increases the complexity of the data gathering process. To assist this process, the model input parameters and decisions variables can be ontologically represented. The ontology specifies the components in classes, that are characterized by attributes, only shared by the individuals of the class. Classes can interact with each other for relations, that define the properties (Marquardt et al., 2010). Ontology development can be found in different domains, as in the study of Okibe et al. (2023) for the valorization of sugarcane bagasse with different uses in the world largest’s sugarcane producers. In conceptual design, the ontology can also be used, for screening and selection of ingredients of cosmetics based on its properties, and to evaluate the formulation proposed (Gabriel et al., 2023). Hazard and operability studies can also be beneficiated by ontology, since these processes are laborious and depend on large quantities of data, algorithms based on ontologies can infer hazard and its propagation, not susceptible to human errors (Single et al., 2020).

In complex systems that depend on strategic and tactics decisions, data mining, visualization and ontology are crucial for computational decision-making tools, and can enhance efficient planning, particularly in systems involving renewable resources like biomass. Therefore, the objectives of this work are to gather georeferenced data about biomass production and availability, to create a data visualization tool for biomass, that is interactive and user friendly, and to formulate an ontology that represents the Brazilian supply chain of biomass uses, to support decision-making and to assist future optimization studies about agricultural wastes valorization.

* 1. **Methodology**

The data visualization tool was developed by data collection and user interface programming. Relevant and up-to-date biomass data from Brazil was collected and analyzed, creating a comprehensive database. The biomass types added to the database were: crops such as maize, sugarcane, soybean, palm, sunflower and agricultural residues from rice, potato, sugarcane, orange, maize, soybean and wheat. The data regarding crops were obtained from the Municipal Agricultural Production research (IBGE, 2021). To estimate the amount of residue produced for each of these crops, the production was multiplied by a residue coefficient, described by Ferreira-Leitao et al. (2010).

* + 1. *Visualization tool development*

The Brazilian territory is divided into federative units, mesoregions, microregions and municipalities (IBGE, 2021). The Instituto Brasileiro de Geografia e Estatística (IBGE) provides the files for the regions, in a shapefile format, widely used in Geographic Information Systems (GIS). After downloading the files, the PyShp library (Lawhead, 2022) was used to convert them from shapefile to json format, which is understandable to humans. However, the maps obtained can lead to a high level of detail, which means heavy files and low performance for the visualization software. For this reason, the Ramer–Douglas–Peucker algorithm (Douglas and Peucker, 1973) was used to reduce the map resolution and generate lighter files. Given that the boundaries of a region are represented by a polygon, the algorithm reduces the number of points of that polygon while keeping its general shape, this makes curves look more jagged and generates smaller files.

Afterwards, using Python’s object-oriented functionality, the *BiomassMap* class was created. Its function is to associate the biomass production in a region to its geographical position by joining the production dataset to the map dataset (json) using Pandas (Mckinney, 2015). A color scale was created based on the production values, and the *BiomassMap* class creates a Matplotlib figure (Hunter, 2007) that displays a biomass production map; the color of each region is related to the amount of biomass produced, defined by the color scale.

Similarly, the classes *DynamicUnits* and *StaticUnits* were created to display strategic points on the map, including storage units, regional capitals, and consumer centers. The difference between the classes is the fact that *DynamicUnits* has an associated magnitude (such as storage capacity) and the *StaticUnits* does not. Matplotlib (Hunter, 2007) was used to plot the points on the figure previously created by the *BiomassMap* class. Finally, the Streamlit library (Streamlit, 2023) was used to create an interactive graphical user interface (GUI) on the web, allowing the user to select the biomass type, the region to be displayed, the dynamic and static units of interest. Streamlit can identify the selections and interact with the classes described, updating the map based on the user’s choices.

* + 1. *Ontology development*

In the proposed ontology, classes were created to represent biomass, purchase sites, technologies, production and storage units, products, residues and consumer centers. The ontology created is presented in the class diagram of Figure 1.



Figure 1: Proposed ontology for biomass valorization

The core of the ontology is the *Technology* class; it represents the conversion technology used in a process, such as transesterification and fermentation, objects of the class. *Technology* has a composition relationship with three classes: *ProcessesUnit*, *Residue* and *Product*. For example, if transesterification is a *Technology* object, biodiesel could be a *Product* object and glycerol a *Residue* object, any biodiesel plant could be a *ProcessUnit* object. The class *Utility* is also associated with *Technology* and electrical energy could be one of its objects.

The classes *Chemical* and *Feedstock* are related to *Technology* and its objects could be methanol and soybean oil, respectively, considering the biodiesel production scenario.

Both *Chemical* and *Feedstock* classes are related to a *PurchaseSite*, which represents geographically where these goods can be purchased. Likewise, the class *ConsumerCenter* is related to *Product*. The *StorageUnit* class has a composition relationship with *Product* and, like *ProcessUnit*, inherits the attributes from *BaseUnit*.

* 1. **Visualization tool demonstration**

The application interface developed allows the user to select the feedstock, as well as to select a single region to get a detailed vision of the feedstock distribution. The biomass availability is presented in the map by a color scale, representing the quantity produced, shown in Figure 2. The use of color scales offers an intuitive and quick visual comprehension of the data. The flexibility to choose the biomass enables the user to assess the availability and potential of each feedstock type, making informed choices about which sources to prioritize based on factors like distance to storage and consumer centers.



Figure 2: Production of soybean waste in Brazil

The option to select a single region or visualize the whole Brazilian territory provides an understanding of biomass distribution and availability across different areas, which helps to identify regional disparities in resources and enables planning for each region based on its characteristics, as shown in Figure 3, for São Paulo region.



Figure 3: Peanut production in São Paulo region

The availability of peanuts in the state of São Paulo is more concentrated in the north and west of the state, far from the capital as shown in the Figure 3. To implement projects to valorize wastes from this production chain, knowing the location of the largest producers and the consumer market for a possible product is very important in the strategic decisions of these projects.

* 1. **Case study for ontology application**

An example for the developed ontology use is the production of orange essential oil through steam distillation (Ortiz-Sanchez, et al. 2024). As shown in Figure 4, steam distillation is a *Technology* object, the feedstock used is orange peel, steam is used as a utility and the product is orange essential oil. The exemplified purchase site is an industrial orange juice plant, located in Matão (region of São Paulo) – Brazil, which is one of the largest plants in the country (CitroSuco, 2023), and has orange peel as a residue of the orange juice production.



Figure 4: Ontology for orange essential oil production from orange peel

Process and Storage units locations and capacities should be determined by solving a supply chain optimization problem, using the aforementioned classes. Data such as transportation costs and the geographical position of the possible purchase sites, consumer centers, production costs and yields should be used as input in an optimization framework, to obtain deterministic results for a supply chain problem. However, it is important to note that the detailed resolution of this optimization problem is not part of this study.

* 1. **Conclusions**

The data visualization tool supports decision-making by providing a comprehensive view of biomass availability, operational capacity, and regional distribution. It assists decision-makers to identify opportunities, make informed choices, and develop effective strategies for biomass utilization, thereby contributing to sustainable and efficient resource management. In addition, the data visualization tool helps in interpreting the results of an optimization problem, as it allows the upload of these data into the database and visualization of the origins of the raw materials, which operational centers are used, and which consumer centers can buy the products. Furthermore, the presented ontology organizes and structures Brazil’s biorefinery sector and divides it into defined categories. The proposed data visualization tool and ontology will be further used integrated to optimization models, to assist decision-makers to identify opportunities, make informed choices, and develop effective strategies for biomass valorization in Brazil.

* 1. **Acknowledgments**

The authors thank the financial support from FAPESP agency (projects 2022/02809-0, 2023/05420-0 and 2023/12377-3).

**References**

A. Gabriel, J. Serna, V. Plantard–Wahl, A. Le Jemtel, V. Boly, V. Falk, 2023, Decision making software for cosmetic product design based on an ontology, Computer Aided Chemical Engineering, 52, 1987-1992.

CEPEA/ESALC, 2020, PIB Do Agronegócio, Série Histórica Do PIB Do Agronegócio Brasileiro.

CitroSuco, 2023, A CitroSuco. Available at: < https://www.citrosuco.com.br/a-citrosuco/#governanca> Accessed in 28/07/2023.

D. H. Douglas, T. K. Peucker, 1973, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Cartographica: the international journal for geographic information and geovisualization, University of Toronto Press, 10, 2, 112–122.

IBGE - Instituto Brasileiro de Geografia e Estatística, 2021, Malha Municipal. Available at: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=acesso-ao-produto>. Accessed in 09/02/2023.

J. D. Hunter, 2007, Matplotlib: A 2d graphics environment, Computing in Science & Engineering, 9, 3, 90–95.

J. I. Single, J. Schmidt, J. Denecke, 2020, Computer-Aided Hazop: Ontologies and Ai for Hazard Identification and Propagation, Computer Aided Chemical Engineering, 48, 1783-1788.

J. Lawhead, 2022, PyShp. Available at: <https://github.com/GeospatialPython/pyshp>. Accessed in 09/02/2023.

M. C. Okibe, M. Short, F. Cecelja, M. Bussemaker, 2023, Ontology Modelling for Valorisation of Sugarcane Bagasse, Computer Aided Chemical Engineering, 52, 3363-3368.

M. Ortiz-Sanchez, J. C. Solarte-Toro, P. J. Inocencio-García, C. A. C. Alzate, 2024, Sustainability analysis of orange peel biorefineries, Enzyme and Microbial Technology, 172, 110327.

M. S. Roni, D. N. Thompson, D. S. Hartley, 2019, Distributed biomass supply chain cost optimization to evaluate multiple feedstocks for a biorefinery, Applied Energy, 254, 113660.

Streamlit, 2023, Streamlit Documentation. Available at: <https://docs.streamlit.io/>. Accessed in 09/02/2023.

T. Forster-Carneiro, M. D. Berni, I.L. Dorileo, M. A. Rostagno, 2016, Biorefinery study of availability of agriculture residues and wastes for integrated biorefineries in Brazil. Resources, Conservation and Recycling,77, 78-88.

V. Ferreira-Leitão, L.M.F. Gottschalk, M. A. Ferrara, A. L. Nepomuceno, H. B. C. Molinari, E. P. S. Bon, 2010, Biomass Residues in Brazil: Availability and Potential Uses, Waste Biomass Valor, M. Ortiz-Sanchez, J. C. Solarte-Toro, P. J. Inocencio-García, C. A. C. Alzate1, 2024, Sustainability analysis of orange peel biorefineries, Enzyme and Microbial Technology, 172, 110327.

W. Marquardt, J. Morbach, A. Wiesner, A. Yang, 2010, OntoCAPE A Re-usable Ontology for Chemical Process Engineering, Springer, In RWTHedition.

W. Mckinney, 2015, Pandas, python data analysis library. Available at:< http://pandas.pydata.org>. Accessed in 09/02/2023.