Supplementary information on data formats and methodology

Biomass Energy Potential Mapping and Analysis Tool (BEPMAT)

Divyansh Singhal (Indian Institute of Technology, Delhi)
Dr. Vignesh Sridharan (Imperial College London)
Dr. Adam Hawkes (Imperial College London)

Contact information:
singhaldivyansh83@gmail.com

Data Sources and Types

GAEZv4 hosts all the following raster datasets on AWS, and they provide the link to these rasters. We use these links to extract the rasters and then conduct our analysis.

Raster datasets used (from GAEZv4) :

- The land is classified into 57 Agro-Economic Zones; CRS = WGS84; resolution at the equator = 30 arc-seconds (0.9 km x 0.9 km) (For more details on how GAEZv4 classifies land into different Agro-Economic Zones, please view their documentation [3].)
- Production values and Harvested Area for 2000 and 2010; CRS = WGS84; Resolution at equator = 5 arc-minutes (9 km x 9 km).
- Exclusion Areas and Tree Cover Share; CRS = WGS84; Resolution at equator = 30 arc-seconds (0.9 km x 0.9 km).
- Potential Yield for future years under different RCPs, water conditions and using different models; CRS = WGS84; Resolution at equator = 5 arc-minutes (9 km x 9 km).

Other raster datasets used :

- SEDAC pastureland 2000 dataset for identifying the percentage of each grid cell under pasture; CRS = WGS84; Resolution at equator = 5 arc-minutes (9 km x 9 km) [4], [5]

Besides raster datasets, this project extensively uses RPR, SAF and LHV values for different crops from different regions worldwide. The final RPR, SAF and LHV tables are as follows, along with their references and assumptions made :

Assumptions regarding downscaling of data

To preserve the crop data, data like classification and tree cover share with higher resolution have been converted into lower resolution. Since each pixel in the lower resolution will contain many pixels of the higher resolution, we have opted for the Mode Resampling method for the AEZ Classification and Exclusion Areas and Average Resampling for the share of Forests Pastureland.

File types and packages

Since most of our data is in the form of rasters, we will be working extensively with pixels and the values in them. In the code, you will be able to see that often, we have converted these rasters into NumPy arrays, have conducted operations on them and then converted them back to the rasters. All the modifications on the rasters have been done using Python libraries GeoPandas, Rasetrio and others. We have used the MemoryFile datatype in Rasterio extensively, which allows us to create rasters in the active memory without the need to download these to the computer.

Detailed Overview of Landcover types and how land is classified in BEPMAT

To explain the methodology on which the model is built, we will first classify the geographical land into various categories (Shapefiles for the regions are generated through a PyPI library GADM, which provides the GADM shapefiles for the selected regions): (1) Total Available Land, (2) Cropland, and (3) Marginal Land.

We will first identify the total available land suitable for growing crops by removing some areas (like protected areas, glaciers, deserts, extremely steep areas, etc.) from the total area of the selected geographic location. The GADM maps give us the shapefile for the selected region, and we need to be able to find the area of this selected region in terms of Km^2. Simple multiplication of the width and height of a pixel and the addition of all the pixels does not work since the width and height vary with longitude and latitude owing to the curvature of the Earth. To approximate the pixel areas as simply as possible, we use the following approximation: This work assumes a simple spherical shape for the earth. 111 km is the number of kilometers in one degree of latitude at the equator; as we go up towards the poles, this decreases and since it is 111 km at the equator which is 0 degrees and 0 km at the poles which is 90 degrees cosine is a decent approximation.

Pixel area = Pixel width * Pixel height * (111319.9) * 2 * cosine (Latitude)

Testing the approximation in places a decent distance above the equator:

- Spain: Area in hectares by the approximation: 50748752 hectares ~ 507,487 km^2;
  Area by the World Bank: 505,965 km^2 [7]
- Iceland: Area in hectares by the approximation: 10150445 hectares ~ 101,504 km^2;
  Area by the World Bank: 101 km^2 [7]

This shows that the error is manageable even for the higher latitudes. Therefore, we can use this approximation. Once we have the total area and verify that it is like the real documented area, we remove the areas where we cannot grow crops, like deserts, glaciers, and protected areas. This is done by selecting the pixels to be removed from different rasters, storing their locations, combining them into a single GeoDataFrame and then removing the masking pixels at these locations by assigning them NaN values. For e.g.: we select all the pixels classified as desert, glaciers, and very steep regions from the AEZ Classification57 raster and the pixels with forest cover over 50% from the forest cover raster and then store their locations into a single GeoDataFrame. Once we have successfully identified the total available land, this can further be split into two regions: one where crops have been growing according to the past data (called Cropland) and the second where we can look to grow crops in the future specifically with extracting energy in mind (called Marginal Land). We already have the dataset for the harvested area in 2010, and this is what we will use for Cropland Area calculations throughout this project. Next, we will subtract this from the total available land and calculate the marginal land area. We take our Total Available Land and subtract the Harvested Area raster from 2010, and this gives us the remaining area in each pixel, which can be considered Marginal Land. The harvested area removed from the net area is calculated by summing up the harvested area for all crops.

Once we obtain all the areas we need for the raw biomass potential from crop residues in a region, we will identify the crops in each region as described below. We have the data for crop yields already via GAEZv4; we have calculated the area, and their product gives us crop production. Then, we will use the Residue-to-Product Ratio (which tells us for 1kg yield of a crop how much residue is produced in kg) for each crop residue type to get the amount of each residue. We will then multiply this by the Surplus Availability Factor, which tells us how much of the available residue can be used for energy production (We cannot use the complete residue since some of it is needed for maintaining land fertility, feeding livestock and other purposes). Finally, we will multiply the obtained value with the Lower Heating Value (since we are not considering drying and removal of moisture in our calculations separately) to give us the net maximum energy that can be extracted.

$$\text{Energy} = \text{Crop Production} \times \text{RPR} \times \text{SAF} \times \text{LHV}$$

Total Available Land

This will be the total land available to us for growing crops. This involves the total geographical area of a region minus certain regions unfit for the growth of crops or the regions covered by pasturelands, water bodies and forests. According to the GAEZv4 documentation and research, the following lands are considered unsuitable for crop growth:

- Continuous or Discontinuous permafrost
- Very Cold Thermal Zone TZ6 (Very Cold)
- M1 moisture class (Desert or Arid)
- L1 and L2 Land Classification which include Water bodies and Built Area
- S1, S2 and S5 soil classification (very steep areas, waterlogged/hydromorphic soils and others like rocks, glaciers)

Land will be removed from the AEZ 57 Classification Zones raster, and the following classes will be removed from the GADM maps to give us the remaining area where crops can potentially be grown:

- Dominantly very steep terrains (AEZ 49)
- Land with severe soil/terrain limitations (AEZ 50)
- Dominantly hydromorphic soils (AEZ 52)
- Desert/Arid Climate (AEZ 53)
- Arctic/Very-Cold climate (AEZ 55)
- Dominantly built-up land (AEZ 56)
- Dominantly inland water (AEZ 57)

We have also used the tree cover raster from GAEZv4 and removed pixels containing more than 50% of the forest, keeping forest preservation in mind. Similarly, we have used a pastureland raster dataset from 2000 and have removed the pixels with more than 50% of the area under pastureland. Finally, we have also removed exclusion areas which include the following : (1) IUCN category in WDPA (World Database of Protected Area) ; (2) WDPA, not an IUCN category; (3) KBA(Key Biodiversity Areas), outside WDPA protected area; (4) GLWD-3 (Global Lakes and Wetlands Database) class 4–9, outside protected area and KBA polygon; and (5) Buffer zone around protected area. We suggest reading the documentation for further details on the exclusion area categories or the removed land use types, [3].

Cropland

In this study, we categorise the total available land to us into two parts: the cropland and the marginal land. The cropland is the land area over which crops are grown primarily for consumption as food. The distinction between cropland and marginal land is made to ensure food security while maximising the energy output from a region. For our studies, we took the land under cultivation in 2010 as the cropland. We also assume that the area under cropland remains the same in future. |For example: If in 2010, in the state of Punjab, India, the area used for the production of wheat was 10,000 hectares, we will in all our future calculations, assume that on these 10,000 hectares, only wheat will be grown to ensure food security to an extent. We make these assumptions since there is no way to predict whether the future cropland area will increase or decrease. The yields for the cropland will vary due to the changes in the weather, soil, moisture conditions, irrigation method and input levels. So, using the above assumption, we can find the area under cropland (2010 dataset used) and then multiply it by the future yields based on the selected conditions and models. This gives us the raw biomass potential energy from the cropland. We also assume that the RPR, LHV and SAF values will remain constant throughout the future since we cannot predict them for the future.

Marginal Land

The area of land remaining after removing the cropland from the total available land and where we can grow crops is classified by us as marginal land. Since the crops grown on the cropland are fixed, we have no control over what the biomass energy potential is; it will be fixed. But in the case of marginal land, our algorithm will check for each crop and then decide which crop should be grown in that pixel to maximise the energy output from that pixel. Our algorithm simply goes over all the crops and stores the crop and its corresponding data, giving the maximum energy for each pixel. We can simply obtain the raw biomass energy potential from the marginal land using this. Once we have the biomass energy potential from both the cropland and the marginal land, we will simply add them up for the selected region and obtain the area's net raw biomass energy potential.

Types of Outputs

Past Data (2010 and 2000): The raster data for crop production, and harvest area for the years 2000 and 2010 has been used to predict the biomass energy potential during those years. Since 2010 is considered the base cropland area, only cropland data has been available in the past years. These biomass energy potential values are primarily used as a comparison tool to show how the energy potential from the cropland varies in the future years under different pathways (This uses the water supply as Total). There are also options available to just get the biomass potential energy past data along with the choice of choosing from various water supply levels: Rainfed, Irrigated or Total.

Projections and Analysis (2011-2040, 2041-2070, 2071-2100): The future values for the cropland and marginal land are calculated as described in the above classification of lands. The future input level is always taken as high based on the assumption that in the future, technology and resources will have developed enough to ensure a high input. The cropland values use 2000 and 2010 values as a reference for comparison, whereas marginal values do not have anything of this type. Both the calculations as described above use yields in them and different rasters of yields available for each RCP and Model based on different water supply methods: Available water content of 200 mm/m (under irrigation conditions) and Available water content of 200 mm/m (under rain-fed conditions). This allows users to choose whether to evaluate under a condition where the land is being rainfed or irrigated. The function Total Raw Biomass Potential simply calculates Cropland and Marginal Land in the selected regions and adds them up to give the final target output.

Discussion and Scope for further research

We feel that there is immense scope for development in the area of extracting energy from biomass. In this work, we have simply calculated the raw biomass potential from crop residues, but in reality, all that energy cannot be extracted, and therefore there is further scope to incorporate energy conversion pathways into this tool. Another feature the tool currently lacks is accounting for groups like cereals, pulses, and vegetables in case of future cropland yield. This does not currently significantly impact the end results since cropland yields are significantly less than marginal land, and marginal land does accommodate everything. We will look to add this feature later. Also, we have used a very rough estimate for the area of each pixel; the main reason behind this is that we wanted to keep this as simple as possible. Involving different projections and UTM Zones would make everything much more complicated, but there is certainly a scope for improvement in this regard in this tool.

There is also immense scope for adding an economic layer over the existing methodology, which would help the users get an idea of how economical it is to grow a particular crop for energy. They will then be able to select the crop which provides a high amount of energy but at a relatively cheaper cost. This would further enhance the usability of this tool. The ability to research many of the methods described above depends on the availability of data. Currently, there is a lack of availability of the data for the cost of crop residue or the data required to calculate the energy efficiency of the conversion pathways for each crop. This leaves immense potential for further research in this area. Also, the used RPR and SAF values are taken from different regions across the world due to the lack of availability of data for various crops in the same regions. We believe that if the required data is generated or found, this tool's applicability could be greatly increased, and the results could be much more accurate and tailored to the selected region.

References

1. FAO and IIASA. Global Agro Ecological Zones version 4 (GAEZ v4). [10/12/23]]Accessed DATE. URL: http://www.fao.org/gaez/
2. FAO Global Agro Ecological Zones http://www.fao.org/gaez/ [10/12/23]
3. *Global agro-ecological zone V4 – Model documentation*. (2021). FAO. https://www.fao.org/3/cb4744en/cb4744en.pdf
4. Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley. 2010. Global Agricultural Lands: Pastures, 2000. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H47H1GGR. Accessed: 27/12/2023.
5. Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley. 2008. Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000. Global Biogeochem. Cycles 22 (1): GB1003. https://doi.org/10.1029/2007GB002952.
6. https://data.worldbank.org/indicator/AG.SRF.TOTL.K2?locations=ES
7. https://sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/RCPs.html
8. https://nbviewer.org/github/Divyansh8321/BioMass/blob/main/Sample%20Use.ipynb