Fossil-Free Frontier: Optimizing the Renewable Energy Mix

Andreas Greiler Basaldúa
Universität der Künste
MA Design and Computation
andreas.greiler.b@gmail.com

Abstract

The energy transition away from fossil fuels is a race against time in which the stakes are among the highest in human history. While many nations and international communities are making efforts to transition out of their coal-, oil- and gas-driven economies, it is worth asking how these transitional efforts compare to what could actually be done in a set time-frame. This project's main goal is a linear programming model that re-envisions ambitious yet feasible renewable energy goals for key regions by focusing on the trade-offs between different fossil-free energy sources. For this, the project focuses on the data of four common fossil-free energy sources: wind, solar, nuclear and hydro, as well as relevant variables such as Cost, Reliability, Existing Energy Mix and Public Approval.

I. Introduction

The energy transition away from fossil fuels is a race against time in which the stakes are among the highest in human history. While many nations and international communities are making efforts to transition out of their coal-, oil- and gas-driven economies - a development partly driven by sinking renewable energy costs - it is worth asking how these transitional efforts compare to what could actually be done in a set time-frame.[1]

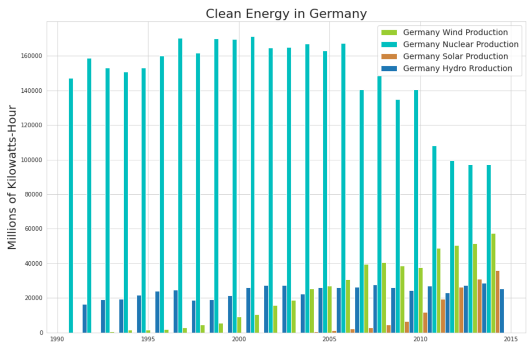


Figure 1: *Visible decline in nuclear energy production as wind and solar grow. Source: UN Data-set*

Despite the decline in the cost of key renewable energy sources, such as wind and solar, some are sceptical as to whether renewable energy can provide the reliability we have grown accustomed to thanks to nuclear energy and fossil fuels.

"Fossil-Free Frontier" uses a simple linear programming model to re-envision new and feasible fossil-free energy mixes based on the cost, reliability and public approval of key energy sources. We chose these three constraints because they play a key role in real-world energy policy and provide an intuitive idea to the trade-offs between different energy types (see trade-off structure in Figure 2).

At this point it is worth mentioning that this project is not meant as a standalone effort to be concluded upon the completion of this paper. Part of the work done behind the comparatively simple model presented in this paper was done to pave the way for future analysis by students considering to examine similar problems. Because of this, a relatively high amount of effort was devoted to the cleaning and organizing of the over 1 million observations contained in the United Nations Energy Statistics Data-Set.

One result of this effort is the Renewables_Selected_Countries data-set, which contains a simple overview of the energy balances of ten major economies covering a period of 24 years (1990-2014). This data-set as well as the Python code used for its aggregation are available in the attachment of this paper, together with the original United Nations Energy Statistics Data-Set.[2]

I. Finding the Right Problem

Often, finding and carefully defining the right problem to tackle can be a challenge in itself. The main optimization problem presented in this paper (Fossil-Free Frontier) was in fact the second idea explored in the context of this course. The first attempt at a problem to solve was a problem in realm of pedagogy and communication theory. Namely, how to optimize board game instructions for faster reading and improved comprehensibility. Developing general rules for how to improve such board game instructions might have theoretically shed some light onto how to generally improve the written communication of complex ideas when written out.

However, upon closer research, there were a few likely insurmountable challenges to this projects' quest. The first major challenge lay in the subjective nature of communication. Efficient communication often relies on knowing what the readers know so that the text can "pick the readers up where they are", and thus avoid boring the readers through redundancy or losing them due to the text being overly complicated. This subjectivity factor meant that it would likely be very challenging to arrive at general guidelines for what makes an efficiently written text.

Another challenge lay in defining objective variables for this line of inquiry. Some variables, such as "lexical density", which offers an approximation of the percentage of the informative words in a sentence, provided an interesting starting point, but were ultimately insufficient to motivate the further pursuit of this project.[3]

The quick brown fox jumped swiftly over the lazy dog.

Figure 2: *Example for lexical density: lexical words (nouns, adjectives, verbs, and adverbs) are colored green.*

With the guidance of Professor Hromada, I eventually settled on instead pursuing a project that had a more clearly defined scope and that capitalized on my background in the field of energy and climate change. I settled on "Fossil-Free Frontier" because this project would give me the opportunity to get hands-on experience with simple linear programming and deepen my understanding of data analysis without the added ambiguity of having to define a novel line of inquiry.

II. Methods

For this project, I used a combination of desk research, data cleaning (numpy, pandas), data visualization (matplotlib, pyplot), and linear programming in PuLP (inspired by Prof. Hromada's Diet Problem code).

I. Desk Research

The first step was deceptively straightforward: to gather the energy balances (e.g. main energy sources) for ten key countries, such as Germany, France, Italy and China. While theoretically straightforward, finding this sort of data proved a challenge in itself. The reason was that, while energy balance data is widely available at the country level, there were very few sources that provided all necessary data (different energy types), with the necessary granularity (useful units, such as Kilowatt-hours) and consistently covered multiple countries. Since the goal was to use the data for a comparative linear programming project, getting the energy balance data from a single source is strongly preferable to pooling the data multiple sources, some of which may have slight deviations from one another.

An additional challenge was to find a source that contained values for different years, which would add an interesting time component to the range of potential analyses. Even large data collectors such as the official sites for the United Nations and the World Bank do not provide comprehensive data-sets. Instead, users are required to submit precise data queries, which makes data retrieval highly impractical for our purposes.

Thanks to a classmate's recommendation, I was finally able to find a comprehensive data-set on the Kaggle.com: the United Nations Energy Statistics Data-set (henceforth "UN data-set"). Ironically, this data-set containing over one million observations with variable values as nebulous as "Anthracite – stock changes" or "Blast Furnace Gas – Own use by coke ovens" was too large and too granular for immediate analysis. Consequently, the first step towards meaningful analysis was to clean and organize the data.

```
RangeIndex: 1189482 entries, 0 to 1189481
Data columns (total 7 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   country_or_area         1189482 non-null  object
 1   commodity_transaction   1189482 non-null  object
 2   year                    1189482 non-null  int64
 3   unit                    1189482 non-null  object
 4   quantity                1189482 non-null  float64
 5   quantity_footnotes      163946 non-null   float64
 6   category                1189482 non-null  object
```

Figure 3: *Overview data for the raw UN data-set*

II. Data Cleaning and Visualization

The UN data-set covers all energy production, consumption and other variables for over 200 countries and over a period of 24 years (1990-2014), resulting in over one million observations. To gain some familiarity with the data-set and its interpretations, I implemented some data visualization code inspired by previous work by Rihad Variawa, who conducted some preliminary data visualization of the data-set.[4] However, his approach to the data-set proved relatively inefficient and immutable, resulting in messy code and time-consuming data manipulation.

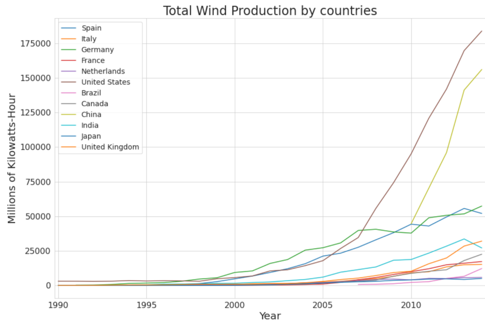


Figure 4: *Total wind energy production for key countries. Source: UN Data-set*

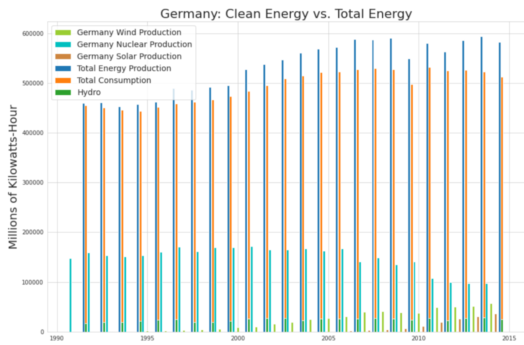


Figure 5: *Germany: Clean energy vs. total energy figures. Source: UN Data-set*

The two tallest pillars (blue and orange) represent total production and consumption respectively.

Thus, after an initial period familiarization and visualization, I focused on streamlining the relevant data by implementing code that enabled faster data extraction and manipulation. Most of this data cleaning work was new yet exciting territory for me.

Some of the steps included:

- Creating dictionaries to map country initials to country variable names
- New tables with key information by country, all years

- New tables for individual energy sources (Wind, Solar, Nuclear, Hydro), for all countries, all years

- New merged data-set with Wind, Solar, Nuclear and Hydro data for all countries as well as totalConsumption and totalProduction

- Creating a new Master data-set with all the above information focusing only on five European countries (Germany, France, Netherlands, Italy, UK) and five non-European countries (China, Canada, Brazil, India, Japan)

	country_or_are	year	unit	quantity_totalProduction	quantity_totalConsumption	quantity_wind
0	United States	1990	Kilowatt-hours, milli	2987971	2633575	3066
1	United States	1991	Kilowatt-hours, milli	3035748	2772927	3051
2	United States	1992	Kilowatt-hours, milli	2998601	2775452	2917
3	United States	1993	Kilowatt-hours, milli	3098678	2873029	3053
4	United States	1994	Kilowatt-hours, milli	3136954	2956258	3483
5	United States	1995	Kilowatt-hours, milli	3225827	3041978	3196
6	United States	1996	Kilowatt-hours, milli	3314012	3127976	3410
7	United States	1997	Kilowatt-hours, milli	3330572	3174192	3254
8	United States	1998	Kilowatt-hours, milli	3437248	3281328	3018
9	United States	1999	Kilowatt-hours, milli	3494337	3369885	4802
10	United States	2000	Kilowatt-hours, milli	3862116	3499463	5650
11	United States	2001	Kilowatt-hours, milli	3701978	3456063	6806
12	United States	2002	Kilowatt-hours, milli	3885542	3555705	10459
13	United States	2003	Kilowatt-hours, milli	3913885	3585012	11300
14	United States	2004	Kilowatt-hours, milli	4006486	3636065	14291
15	United States	2005	Kilowatt-hours, milli	4135191	3731531	17881
16	United States	2006	Kilowatt-hours, milli	4137671	3748623	26676
17	United States	2007	Kilowatt-hours, milli	4190757	3849177	34603

Figure 6: *New master data-set containing all relevant information for ten key countries.*

III. Adding New Constraint Variables

For simplicity's sake, I decided to model the linear programming problem after the Diet Problem, which is one linear programming application that uses the dietary values of certain food items to arrive at an "optimal" food combination given certain constraints, such as the required daily calorie intake for an average adult male.

To apply this model to the climate balance of any country, each energy source needs to map onto certain values that simulate some of the trade-offs these energy balances are subject to in the real world.

To this end, I collected the following data for the ten countries in the new data-set from different online sources:

- Levelized cost of energy (relative cost indicator), by energy source (e.g. 46 for Wind, 51 for Solar)[5]

- Approval rates in % by energy source, by country (e.g. Nuclear approval is 16% in France and 3 percent in Germany)[6]

- Capacity factor rates (relative reliability indicator) by energy source (e.g. 24% for Solar, 39% for Hydro, 93% for Nuclear)[7]

The above values can be subject to significant variation depending on many factors, but suffice in order to set up a simple model. Generally speaking, the trade-offs can be visualized in the following way:

Figure 7: *Germany: Public Approval of fossil-free energy sources in % vs. their Reliability (Capacity Factor) in %*

Illustration of a simple trade-off (assuming the goal is maximization)

IV. PuLP Linear Programming Model

The linear programming model uses the above variables to arrive at a new energy balance for any given country. For example, one can tune the program so that it solves a problem that either minimizes the cost or maximizes the reliability of the total energy mix. For the purpose of this paper we will focus on the former optimization problem: cost minimization.

V. Diversity and Bounds

However, since different energy sources tend to have significantly different values, optimizing the program for one factor, such as cost, can result in an unrealistically homogeneous energy mix. For example, minimizing the cost will result in an energy mix consisting only of wind energy, because it it the source with the the cheapest levelized cost of energy (LCOE).

A simple constraint to counteract this problem is to set a customizable range (the deviation variable) within which the optimal energy mix can deviate from the previous year's mix. Imagine setting the maximum deviation to 10 percent: this would mean that the new amount of wind energy can only be 10 percent higher or lower of the previous year's value. This constraint guarantees that all energy sources that make up over 10 percent of the energy mix will be represented in the new "optimal" solution.

III. Results

In this section, we will walk through one possible application of the model. The model is programmed dynamically, so that it can easily be tweaked to solve the energy mix problem for any country, any year and for any deviation range as long as the required values are available in the UN data-set.

Let's model an energy mix that minimizes the overall cost of all energy sources combined. We pick a country (Germany), a base year (2014), and choose a deviation range of 10% from the base year. Before we solve for the new optimal energy mix we need to have an idea for where to set the constraints variables of Reliability and Approval.

To obtain these numbers, we can model the most cost-effective scenario by taking away the constraint variables. This way, our model will output the lower bound (-10% from base year) for all energy sources. This output will provide a useful ballpark for our constraint variables of Reliability and Approval.

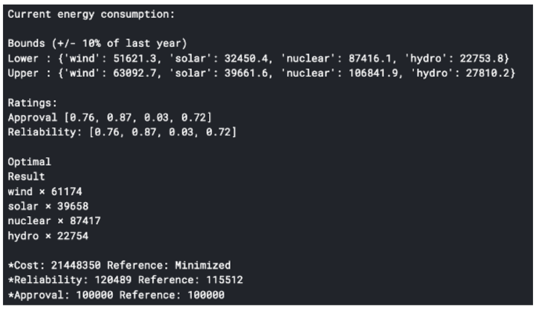
```
Optimal
Result
wind × 51622
solar × 32451
nuclear × 87417
hydro × 22754

USE VALUES BELOW AS BENCHMARK REFERENCES FOR THE NEXT SECTION:

*Cost: 20641401 Reference: Minimized
*Reliability: 115512 Reference: None
*Approval: 86470 Reference: None
```

Figure 8: *Result output from strict cost minimization towards lower bound.*
We can use its constraint values (Reliability and Approval) as a ballpark for the next model run.

Now, if we increase the required Approval variable from 86,470 to 100,000 (meaning: we want an energy mix that scores higher with public opinion), we get the expected results: we see significant increases in the two most popular energy sources, wind and solar, but no increase in the two less popular energy sources, nuclear and hydro.



```
Current energy consumption:

Bounds (+/- 10% of last year)
Lower : {'wind': 51621.3, 'solar': 32450.4, 'nuclear': 87416.1, 'hydro': 22753.8}
Upper : {'wind': 63092.7, 'solar': 39661.6, 'nuclear': 106841.9, 'hydro': 27810.2}

Ratings:
Approval [0.76, 0.87, 0.03, 0.72]
Reliability: [0.76, 0.87, 0.03, 0.72]

Optimal
Result
wind × 61174
solar × 39658
nuclear × 87417
hydro × 22754

*Cost: 21448350 Reference: Minimized
*Reliability: 120489 Reference: 115512
*Approval: 100000 Reference: 100000
```

Figure 9: *Result output from cost minimization problem with higher Approval constraint.*
We see an increase in wind and solar.

Conversely, if we instead increase the required Reliability variable from 115,512 to 130,000 (meaning, we want to get more energy from a more reliable source), we see a steep increase in the most reliable energy source, nuclear energy, alongside an increase in the cheapest source: wind energy.

```
Current energy consumption:

Bounds (+/- 10% of last year)
Lower : {'wind': 51621.3, 'solar': 32450.4, 'nuclear': 87416.1, 'hydro': 22753.8}
Upper : {'wind': 63092.7, 'solar': 39661.6, 'nuclear': 106841.9, 'hydro': 27810.2}

Ratings:
Approval [0.76, 0.87, 0.03, 0.72]
Reliability: [0.76, 0.87, 0.03, 0.72]

Optimal
Result
wind × 63091
solar × 32451
nuclear × 98803
hydro × 22754

*Cost: 23036279 Reference: Minimized
*Reliability: 130000 Reference: 130000
*Approval: 95528 Reference: 86470
```

Figure 10: *Result output from cost minimization problem with higher Reliability constraint.*
We see an increase in nuclear and wind.

The most immediate next step would be to make the use of the model as well as the interpretation of its results more user friendly. Some approaches to achieve this will be outlined in the next section.

IV. Where do we go from here?

As mentioned in the introductory section of this paper, this project should ideally serve as a starting point for further analyses and applications. The lightweight Renewables_Selected_Countries master data-set and the new energy tables can provide easy access points for such projects.

That aside, the project at hand, the Fossil-Free Frontier model, can use various improvements. One such improvement is to make the use of the model more accessible and user-friendly. Currently, the process by which the user sets the values for the constraint variables (Cost, Reliability, Approval) remains counter-intuitive at best.

Similarly, the changes in the "optimal" energy quantities are displayed in absolute terms, which makes it difficult to notice which quantities have changed and to what extent. Using a relative change indicator (e.g. wind energy registered a +/- 8 10% change from the previous year) could help users quickly read and interpret the model's results.

While the current values used in the model allow for the model to work, the model would benefit from more accurate and reliable data across all value groups as well as some interdependencies between variables.

Finally, one powerful aspect was left entirely unexplored in the current project: the data-set covers a period of 24 years! Over two decades' worth of highly granular energy data would allow for interesting analyses of trends over time, such as the speeds for energy transitions across different geographies.

V. Acknowledgments

Thanks to Professor Daniel Hromada for advising this project from start to finish, and to my classmate Freder for his pointer to Kaggle.com, the place where I found the data-set that made this project possible.

Notes

1. An overview of recent changes in energy prices:
https://www.forbes.com/sites/energyinnovation/2020/01/21/renewable-energy-prices-hit-record-lows-how-can-utilities-benefit-from-unstoppable-solar-and-wind/?sh=4d3d90e42c84

2. Data Attachment (Google Drive):
https://drive.google.com/drive/folders/1QcwwcnulknM3tPFJTUTn14yNtuWvmquE?usp=sharing

3. Informal overview of lexical density:
https://www.analyzemywriting.com/lexical$_{d}ensity.html$

4. Rihad Variawa's Data Visualization:
https://drive.google.com/file/d/14x1HiFpIQQk8GfNLwFj2qCXeJTWcV5Ji/view?usp=sharing

5. Overview for levelized energy costs:
https://en.wikipedia.org/wiki/Cost$_{o}f_{e}lectricity_{b}y_{s}ource$

6. Comprehensive document with different public opinion metrics:
https://drive.google.com/file/d/1POcN0rpKAeHdwQ9ruvuhLIffq3aN0N7T/view?usp=sharing

7. Overview of capacity factors for different energy sources:
https://www.energy.gov/ne/articles/nuclear-power-most-reliable-energy-source-and-its-not-even-close